COMPUTER PROGRAMMING

4 BOOKS IN 1

Python for Data Science
Hacking with Kali Linux
Computer Networking
Python Programming

Jason Callaway

COMPUTER PROGRAMMING

4 Books In 1:
Data Science, Hacking with Kali Linux, Computer Networking for Beginners, Python Programming

Coding Language for Machine Learning and Artificial Intelligence.

Table of Contents

CONCLUSION

BOOK 1: Python for Data Science

Master Data Analysis from Scratch, with Business Analytics Tools and Step-by-Step techniques for Beginners. The Future of Machine Learning & Applied Artificial Intelligence

Download the Audio Book Version of This Book for FREE

If you love listening to audio books on-the-go, I have great news for you. You can download the audio book version of this book for **FREE** just by signing up for a **FREE** 30-day audible trial! See below for more details!

Audible Trial Benefits

As an audible customer, you will receive the below benefits with your 30-day free trial:

- FREE audible book copy of this book

- After the trial, you will get 1 credit each month to use on any audiobook

- Your credits automatically roll over to the next month if you don't use them

- Choose from Audible's 200.000 + titles

- Listen anywhere with the Audible app across multiple devices

- Make easy, no-hassle exchanges of any audiobook you don't love

- Keep your audiobooks forever, even if you cancel your membership

- And much more

Click the links below to get started!

[>> For Audible US](#)

[>> For Audible UK](#)

[>> For Audible FR](#)

[>> For Audible DE](#)

Introduction

Python and Data Science are two most popular technical terms that you will hear everywhere. The combination of these two terms will provide you an added advantage in current tech era. As it has been proven every year, the application and significance of Python is expanding daily, especially with the data science community and data analytics society.

In this book, we will guide you on how to get started:

Why read on? You will learn how to apply data analysis, which is cooler and advanced than using Microsoft Excel. Secondly, you will also learn how to implement Python data analysis and develop a mindset of an expert data analyst.

The most important, you will learn Machine learning and Python, and how it works in real-world problems. You will see several examples of how the modern methods of a data analyst suit in solving modern problems.

This is vital because a lot of data provides us with numerous opportunities to gain insights and make a useful impact in any field. This phenomenon also delivers new challenges that demand new technologies and methods. Besides, this demands new skills and mindsets to successfully navigate through the problems and successfully tap the biggest potential of the opportunities available.

This book will provide you a great foundation to advance into a complex data analysis with Python.

Python is one of the most common tools used in data analysis. In a certain survey conducted by Analytics India Magazine, it was discovered that 44% of data scientists prefer to use Python than SQL, SAS, and R. Not only that,

Python is the only best general-purpose programming language that you can rely on.

As you will be learning the important technical skills needed in Python to complete data science projects, you should focus on practice. Start to search for data from various sources and begin to play around with them. It is highly advised that you apply a lot of operations as possible so that you become familiar with different approaches.

The Python programming language is a great tool for any data scientist because provides effective libraries that carry the tools of the algorithms and models that are necessary for analysis.

Chapter 1: Introduction to Data Analysis
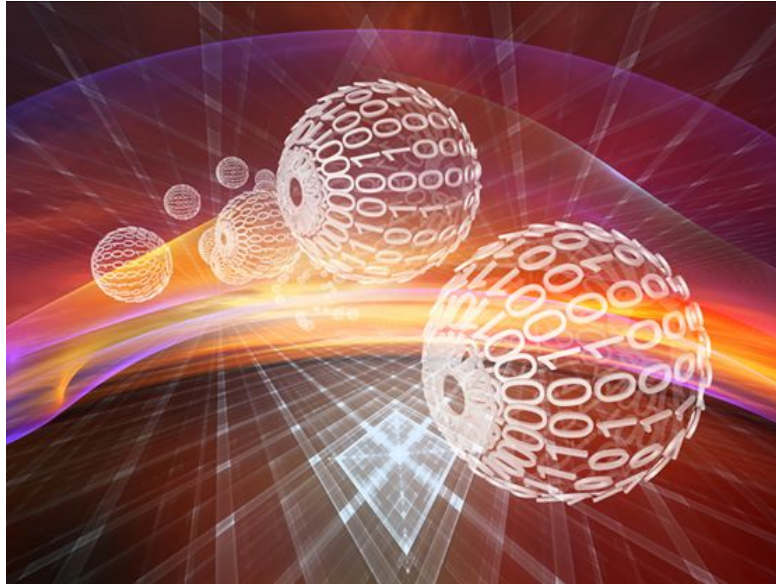
The act of inspecting, cleaning, transforming, and modeling data with the intention to find relevant information and reaching a conclusion to boost the decision-making process is referred to as Data Analysis.

There are different facets and techniques for data analysis. The data analysis in statistics is categorized into exploratory data analysis, confirmatory data analysis, and descriptive statistics. Data requires to be cleaned. Data cleaning is the procedure of fixing the outliers and other inaccurate and unwanted information. There are different types of data cleaning processes to use based on the type of data to be cleaned.

Business intelligence explores the data analysis that extends heavily on aggregation, slicing, disaggregation, dicing, and concentrating on the business information. Predictive analytics refers to the use of statistical models for predictive forecasting. Text analytics describes the application of statistical, structural models, and linguistic models to derive and categorize the information from texts. All these describe different varieties of data analysis.

Python for Data Science

Why Select Python

Python is a simple, clear, and elaborate programming language. That is the reason many scientists and engineers go for Python to implement numerous applications. Probably they prefer handling the main task quickly rather than spend hundreds of hours mastering the nuances of a "complex" programming language.

This lets scientists, engineers, researchers, and analysts to dive into the project more quickly. As a result, they gain important insights into the minimum amount of time and resources. This does not imply that Python is fit and perfect programming language on where to complete data analysis and machine learning. Languages such as R may have advantages and properties Python doesn't have. However, Python is a good starting point, and you may attain the right knowledge of data analysis if you use it in future projects.

Python vs. R

You may have already come across this in Reddit, Stack Overflow, and other forums and websites. You may have also searched for other programming languages because even learning Python or R requires a few weeks and months. It is a huge time investment, and you don't want to make any mistake.

To avoid any confusion, just begin with Python because the usual skills and concepts are easy to transfer to other languages. In some cases, you may need to adopt a whole new approach to thinking. But all in all, understanding how to apply Python in data analysis will provide you a channel to solve many complex problems.

Some may say that R is designed for statisticians, especially when it comes to easy and better data visualization properties. It's also easy to learn,

especially if you plan to use it for data analysis. On the flipside, Python is a bit flexible because it goes past data analysis. Most data scientists and machine learning practitioners may have selected Python because the code they write can be integrated into a dynamic and live web application.

While it's open for debate, Python is still a great choice for beginners or anyone who wants to get started with data analysis and machine learning. It is quite easy to learn, and you can dive into full-time programming if you see that this suits you well.

Widespread Application of Data Analysis

There are a lot of packages and tools that enable the usage of Python in Data analysis and machine learning. Some of these packages consist of Numpy, scikit-learn, and Pandas. These tools simplify the data analysis process.

Additionally, graduates from the university can dive into data science because most universities nowadays provide an introductory course in computer science with Python as the main programming language. The change from computer programming and software development can happen so fast because many people already have the correct foundation to begin learning and using programming to tackle real-world data problems.

Another reason for Python's popularity use is that there are numerous resources that will show you how to complete anything. If you've any questions, it is because someone else has requested that, and another solved it for you. This makes Python more popular because of the presence of resources online.

Clarity

Because of the ease of learning and Python's syntax clarity, experts can concentrate on the crucial features of their projects and problems. For

instance, they can apply Numpy, TensorFlow, and scikit-learn to gain insights rather than build everything from scratch.

This generates another level of clarity because professionals can concentrate on the nature of the problem and its effects. They can also come up with effective approaches to handling the problem rather than being bombarded with the many challenges a given programming language presents.

The main aim should always be on the challenge and the opportunities it may bring. It only requires a single breakthrough to change the entire thinking about a given problem, and Python can achieve that because of its ease and clarity.

Types of Data Analysis

Descriptive Analysis

Data science is related to information retrieval and data collection approaches with the aim of reconstituting past events to determine patterns and identify insights that help determine what happened and what made it happen. For instance, looking at sales figures by region to identify customer preferences. This section requires that you remain familiar with statistics and data visualization approaches.

Predictive Analysis

Data science is a means to determine the chances that some events are currently taking place or will occur in the future. In the following case, the data scientist considers past data to determine explanatory variables and create statistical models that can be used in other data points. For instance, trying to predict the probability that a certain credit card transaction is fraudulent in real-time. This section is usually related to the machine learning field.

Prescriptive Analysis

In this case, data analysis is considered as a way to make informed decisions, or probably data-driven decisions. The focus should be to search for multiple options and applying simulation approaches and optimize the results. For instance, maximizing the supply chain by aiming to reduce the operating costs.

Typically, descriptive data science aims to address the question of what does the data tell me. On the other hand, the predictive analysis answers the question of why is the data behaving in this way, and prescriptive analysis respond to the question of how you optimize the data toward a given goal.

Why Data Analysis Is on the Rise?

There are several factors that lead to the rise of data science.

First, the amount of data gathered increases at an exponential rate. According to a certain IBM Marketing Cloud research, around 2.5 quintillion bytes are generated daily. This is around 2.5 billion bytes, but only a small fraction of this data is ever analyzed, leaving a lot of missed opportunities on the table.

Secondly, we are experiencing a cognitive revolution that began some years ago. Nearly every industry is going for AI technology, which comprises of natural language processing (NLP) and machine learning. Although these sections existed for some time, they have recently achieved renewed attention to the point that they are now popular courses in colleges. It is now clear that, if they are to survive, companies require to be more agile, move faster, and shift into digital businesses, and as the time available for decision-making reduces, they must be fully data-driven.

If you also include the fact that AI algorithms require high-quality data, you can start to understand the vital role played by data scientists.

Another reason is that with the growth of cloud technologies and the rise of platforms as a Service (PaaS), access to computing engines and storage has never been easier. Running big data workloads is now available to smaller organizations or any person with a credit card. In turn, this is driving the growth of innovation across the board.

Because of the following reasons, there is no question that data science is here with us and that its development will continue for a long time. Still, we cannot ignore the fact that it hasn't achieved its full potential and generated the right results, especially in enabling companies to transform into data-driven organizations. In most cases, the problem lies in achieving that next step, which is to convert data science and analytics into a key business activity that provides clear-sighted and intelligent business decisions.

Summary of the Data Science Process

Sticking to a structured method in data science helps you to increase your chances of success in a data science project at the lowest cost. It also makes it possible to handle a project as a team, with every team member concentrating on what they do best. However, this procedure might not be helpful for every project type or be the only means to implement the right data science.

The typical data science procedure comprises of six steps through which you will iterate.

Here is a short introduction to each of the steps.

1. The first step is to set a research goal. The main focus here is to ensure all the stakeholders know the what, how, and why of the project.

2. The next step is data retrieval. You need to have data ready for analysis, so the following step involves searching for the best data and getting access to the data from the data owner. The result is that raw data probably demands polishing and transformation before it is usable.

3. Once you have the raw data, the next thing is to prepare it. This involves transforming the data from a raw type into data that's directly usable in your models. To achieve this, you will identify and correct various types of errors within the data, integrate data from various sources, and change it. If you complete this step successfully, you can move to data visualization and modeling.

4. The fourth step is data exploration. The purpose of this step is to attain a deep understanding of the data. You will search for patterns, deviations, and correlations depending on visual and descriptive techniques. The insights you derive from this step will allow you to begin modeling.

5. Lastly, you will get to the most interesting part. And that is data modeling. This is the time when you try to make the insights or predictions outlined in your project charter. Now is the time to generate heavy guns, but remember that research has taught us that always a mix of simple models tends to overcome a complex model. If you have completed this phase, you're almost done.

6. The last step of the data science model is the results presentation and automating the analysis. One focus of a project is to change a process or make better decisions. You might still need to convince the business that your findings will, in fact, change the business process as required. This is the

point where you can shine in your influence role. The importance of the following step is obvious on a strategic and tactical level. Some projects require you to complete the business process over and over again. Hence, automating the project saves time.

The reality is that this is not going to be a straightforward process. In most cases, you will regress and iterate between the various phases.

Following the above six steps pays off based on a higher project success ratio and an increased effect of research results. This process helps you attain a well-defined research plan, a great understanding of the business question, and clear deliverables before you begin looking at the data. The first steps allow you to concentrate on achieving high-quality data as input for your models. This way, your model will do well in the later stages.

Another advantage of sticking to a structured approach is that you work a lot more in a prototype model while you look for the best model. When creating a prototype, you will probably attempt multiple models and won't concentrate heavily on problems such as program speed or writing code standards. This enables you to concentrate on generating business value.

Not every project is created by the business itself. Insights derived at the time of analysis or the arrival of new data can generate new projects. When the data science team comes up with an idea, work has already been conducted to present a preposition and identify a business sponsor.

Dividing a project into smaller stages will enable employees to work together as a team. It is impossible to be an expert in everything. You don't need to learn how to upload all the data to all the various databases, determine an optimal data scheme that works not only for your application but also for other projects within your company, and then monitor all the statistical and data mining techniques.

Prerequisite and Reminders

By now, you need to be familiar with the Python syntax plus other basic things in Python. Some of the basic things you need to be conversant include functions, variables, loops, and lists. You don't need to be a specialist, but it's important to have the right knowledge, so the rest become smoother.

You don't need to make it hard because programming revolves around telling the computer what is supposed to be done. Then, the computer should be able to understand and successfully implement your instructions. You may only need to write several lines of code to suit your application.

Plus, a lot of things that you will implement in Python for data analysis are already pre-built for you. So, you may only need to copy and run the code. But remember that mastering Python programming is still important. This will also provide you with confidence because you understand how something operates.

Do You Need Some Expertise in Mathematics?

Data analysis implies working with numbers and mining useful insights from them. But do you really need to be an expert on numbers and mathematics?
Successful data analysis using Python always demands great skills and knowledge in math, programming, and the field you are working on. This implies that you don't have to be an expert in any of them.

Don't allow a lot of experts to fool you because many of them are fake or just inexperienced. What you really need to know is what's the next thing to do so you can successfully complete your projects. You won't be an expert in anything once you read all the chapters here. But this is enough to provide you a better knowledge about Python and data analysis.

When we consider the mathematical expertise, it is probably right that you're familiar with mean, standard deviation, and other common terms in statistics. While extending further into data analysis, you may come across calculus and linear algebra. If you have the time and interest to explore them, you can always do anytime, or later. This may or may not provide you with the advantage of the specific data analysis project you're working on.

Again, remember that it is about solving problems. The main focus should be on how to handle the challenge and successfully overcome it. This should apply to all fields in business and science. Don't allow the myths to distract you. Concentrate on the key concepts, and you will be fine.

Chapter 2: Python Review

In this chapter, we shall look at the Python programming language basic stuff. This chapter tries to teach you Python in brief. It's so much as a cheatsheet, so it will only remind you of some basic features to start you off. Typically, if you really want to master a language, you need to commit time and program it for a while. This chapter assumes that you're already familiar with Python programming and will, therefore, skip the majority of the non-language–specific stuff. The important keywords will be pointed out so you can easily spot them.

We shall concentrate on Python 3 because that is the version you need to use. All the examples we shall provide to you are written in Python 3.

Properties

First, Python is a strongly typed language.

Getting help

Help in Python is usually available within the interpreter. If you want to learn how an object operates, all you need to do is call **help (<0bject>)!** Additionally, the most important are dir (), which displays all the object's methods, and **<object>._doc_,** which displays all the documentation string.

Let's set some expectations here, so you know where you're going. This is also to introduce the limitations of Python, data analysis, data science, and machine learning (and also the key differences). Let's start.

Data Analysis vs. Data Science vs. Machine Learning

Data Analysis and Data Science are almost the same because they share the same goal, which is to derive insights from data and use it for better decision-making.

Often, data analysis is associated with using Microsoft Excel and other tools for summarizing data and finding patterns. On the other hand, data science is often associated with using programming to deal with massive data sets. In fact, data science became popular as a result of the generation of gigabytes of data coming from online sources and activities (search engines, social media). Being a data scientist sounds way cooler than being a data analyst. Although the job functions might be similar and overlapping, it all deals with discovering patterns and generating insights from data. It's also about asking intelligent questions about the nature of the data (e.g., Are data points form organic clusters? Is there really a connection between age and cancer?). What about machine learning? Often, the terms data science and machine learning are used interchangeably. That's because the latter is about "learning from data." When applying machine learning algorithms, the computer detects patterns and uses "what it learned" on new data. For instance, we want to know if a person will pay his debts. Luckily, we have a sizable dataset about different people who either paid his debt or not. We also have collected other data (creating customer profiles) such as age, income range, location, and occupation. When we apply the appropriate machine learning algorithm, the computer will learn from the data. We can then input new data (new info from a new applicant) and what the computer learned will be applied to that new data. We might then create a simple program that immediately evaluates whether a person will pay his debts or not based on his information (age, income range, location, and occupation). This is an example of using data to predict someone's likely behavior.

Possibilities

Learning from data opens a lot of possibilities, especially in predictions and optimizations. This has become a reality, thanks to the availability of massive datasets and superior computer processing power. We can now process data in gigabytes within a day using computers or cloud capabilities. Although data science and machine learning algorithms are

still far from perfect, these are already useful in many applications such as image recognition, product recommendations, search engine rankings, and medical diagnosis. And to this moment, scientists and engineers around the globe continue to improve the accuracy and performance of their tools, models, and analysis.

Drawbacks of Data Analysis & Machine Learning

You might have read from news and online articles that machine learning and advanced data analysis can change the fabric of society (automation, loss of jobs, universal basic income, artificial intelligence takeover). In fact, society is being changed right now. Behind the scenes, machine learning and continuous data analysis are at work especially in search engines, social media and e-commerce. Machine learning now makes it easier and faster to do the following:

- Are there human faces in the picture?
- Will a user click an ad? (Is it personalized and appealing to him/her?)
- How to create accurate captions on YouTube videos? (recognize speech and translate into text)
- Will an engine or component fail? (preventive maintenance in manufacturing)
- Is a transaction fraudulent?
- Is an email spam or not?

These are made possible by the availability of massive datasets and great processing power. However, advanced data analysis using Python (and machine learning) is not magic. It's not the solution to all problems. That's because the accuracy and performance of our tools and models heavily depend on the integrity of data and our own skill and judgment. Yes, computers and algorithms are great at providing answers. But it's also about asking the right questions. Those intelligent questions will come from us

humans. It also depends on us if we'll use the answers being provided by our computers.

Accuracy & Performance

The most common use of data analysis is in successful predictions (forecasting) and optimization. Will the demand for our product increase in the next five years? What are the optimal routes for deliveries that lead to the lowest operational costs? That's why an accuracy improvement of even just 1% can translate into millions of dollars of additional revenues. For instance, big stores can stock up certain products in advance if the results of the analysis predict an increasing demand. Shipping and logistics can also better plan the routes and schedules for lower fuel usage and faster deliveries. Aside from improving accuracy, another priority is on ensuring reliable performance. How can our analysis perform on new data sets? Should we consider other factors when analyzing the data and making predictions? Our work should always produce consistently accurate results. Otherwise, it's not scientific at all because the results are not reproducible. We might as well shoot in the dark instead of making ourselves exhausted in sophisticated data analysis. Apart from successful forecasting and optimization, proper data analysis can also help us uncover opportunities. Later, we can realize that what we did is also applicable to other projects and fields. We can also detect outliers and interesting patterns if we dig deep enough. For example, perhaps, customers congregate in clusters that are big enough for us to explore and tap into. Maybe there are unusually higher concentrations of customers that fall into a certain income range or spending level. Those are just typical examples of the applications of proper data analysis. In the next chapter, let's discuss one of the most used examples in illustrating the promising potential of data analysis and machine learning. We'll also discuss its implications and the opportunities it presents.

Chapter 3: Important Python Libraries

You can get a mind map describing software that can be used to analyze data at xmind.net. However, you cannot install all of this software in this chapter. This chapter will explore how to install SciPy, matplotlib, Numpy, and IPython on different platforms. You will also see some sample code that used NumPy.

NumPy is a powerful Python library that features numerical arrays and functions.

SciPy is a powerful Python library, which slightly overlaps NumPy. NumPy and SciPy historically shared their code but were later separated.

Matplotlib provides a plotting library featuring NumPy. You will learn more about matplotlib during data visualization.

IPython contains an architecture for interactive computing. The most critical part of the following project is the IPython shell. You will learn more about the IPython shell in the following chapter.

The installation instructions for the rest of the software will be provided in the rest of the book. When you're done reading this chapter, you will identify pointers on how to determine extra information online if you stumble or are uncertain about the right way to solve problems.

This book uses applications based on Python, so that is why you need to have Python installed on your computer. Some operating systems come with Python already installed, but you will need to check whether the Python version is compatible with the software version you want to install. There are different Python implementations. In the following book, we shall

stick to the standard CPython implementation, which is considered compatible with NumPy.

The software application that we will install in this chapter features binary installers for Windows, Mac OS X, and Linux distributions. There are still source distributions in case you prefer that.

Install the Software and Setting Up

You will learn how to install and set up NumPy, matplotlib, and IPython on Linux, Windows, and Mac OS X. Let's explore the process in detail.

For Windows

When you want to install the above Python libraries on a Windows operating system, it's a straightforward process that we dive into detail. You only require to download an installer, and a wizard will direct you through the installation steps. We will provide you with the steps to install Numpy. The steps to install the remaining libraries are the same. The instructions to use are as follows:

1. Get a Windows installer from the SourceForge website or any other legit website. The current versions may change, so just select the one that suits your setup best.
2. Select the right version from the available list.
3. Next, open the EXE installer by double-clicking it.
4. Next, you will see a description of Numpy and its properties. Now, click on the Next button.

Numpy Arrays

Once you are through with the installation of Numpy, now is the time to explore Numpy arrays. Numpy arrays are more efficient compared to Python lists when it comes to numerical operations. Numpy arrays are

specialized objects that have broad optimization. Numpy code demands less explicit loops than equivalent Python code.

If we recall some of the mathematics topics you did learn in high school, I am sure you can remember scalars and vectors. For example, the number 2 is a scalar. When you add 3 to 2, you will be doing a scalar addition. We can create a vector out of a group of scalars. In Python programming, you will have a one-dimensional array. This concept can still be extended to a higher dimension. Conducting an operation on two arrays, like addition, can be reduced to a group of scalar computations.

In summary, Numpy is faster than the pure Python code. One thing that is clear, you can acquire the same results whether you use NumPy or not. But the result displayed differs in presentation. Remember that the result from the Numpysum () function doesn't have a lot of commas. This is because you are not working with a Python list but with a NumPy array.

Using IPython as a Shell

Data analysts, engineers, and scientists like to experiment. IPython was built by scientists with the goal of experimentation. The interactive environment provided by IPython is considered by many as a direct solution to MATLAB.

Below is a list of properties of the IPython shell:

- History mechanism
- The pylab switch
- Tab completion, which helps you identify a command
- Access to Python debugger and profiler

Web Scraping Using Python Libraries

In this section, you will scrape the "Rate My Professor" website. A brief description of Rate My Professor website, it's a website that has ratings of schools, professors, and universities. In this website, you can search for any professor or school and see their ratings before registering their courses. It is a great feature that helps you to learn more about your professor or the university that you wish to join. In this section, you will learn how to scrape and extract crucial professor's tag. While this is not an illegal process, mass scraping of data from this website may result in your IP address being blacklisted. So, you should only attempt it once or twice, but don't do it foolishly.

Web Scraping

Also known as scraping, data harvesting, or data extraction is a mechanism used to mine data from websites. In some cases, web scraping can be vital when you get the data that you are searching for straight from the web, but sometimes, it is a bad way to handle it because it's like stealing the expensive data from the website without their permission. However, you should reduce your scraping process to only once or twice so that you don't fall in trouble.

The most important libraries for web scraping include:

1. Beautiful soup
2. Requests

Here are the steps that we would be following.

1. First, we import the relevant libraries
2. Finding the URL and keeping it in a variable
3. Sending a request to the website using the requests library
4. Applying the Beautiful Soup library to find the HTML data from the website.

5. Using soup to identify all methods to get the necessary tag that we are searching for.
6. Removing all the HTML tags and converting it to a plain text format.

You may be wondering the type of tags to extract, well in the Rate My Professor website, every professor has his or her respected tags. We will try to extract these tags.

Before we start, ensure that you scrape the data at a low pace, and you can use a VPN service to change your IP address. This prevents your IP address from being blocked. Hope you will follow the instructions.

One thing that you should note is that there is no need to explain each and every line of code because Python code is self-explanatory. But you will not be confused because everything will be made clear in an easy format. So, this guide is written in such a way that everybody can understand regardless of their programming level.

You may find a lot of online tutorials, but this guide is easy to understand because the codes are explained. However, some parts are a mechanical process, wherein you need to follow them.

Let's dive in!

1. First, import the necessary libraries

Let us import several libraries such as Beautiful Soup and Requests.

import requests
from bs4 import BeautifulSoup

2. Find the URL and store it in a variable

Let store the professor's URL within a variable called "url". The website URL is [Rate My Professor.](#)

3. Send a request to the website using the requests library

In the following case, we apply the requests library by passing "url" as the parameter. Be careful that you don't run this numerous times. If you receive like Response 200, then it's a success. If you find something different, then there is something wrong with maybe the code or your browser.

4. Apply the Beautiful Soup library to get the raw HTML data from the website.

Here, we apply the BeautifulSoup by passing the page.text as a parameter and applying the HTML parser. You can attempt to output the soup, but displaying the soup doesn't provide you the response. Instead, it contains a huge percentage of HTML data.

Here is a code snippet.

soup = BeautifulSoup(page.text, "html.parser")

5. Applying the soup.findAll method to find the respected tag that you are searching for.

Here is where you will be adding the tags that you're searching for. To find the tag name, all you need to do is to right-click on the relevant tag or click Ctrl-Shift-I on the tag within the webpage. Next, a webpage with the required tag will open for you to your right-hand side.

Next, you can copy the HTML tag and class if available, and then place it within the soup.findAll method. In the following case, the HTML tag is "span" and the class is "tag-box-choosetags".

proftags = soup.findAll("span", {"class": "tag-box-choosetags" })
proftags

6. Eliminating all the HTML tags and changing it to a plain text format

In the following step, you need to remove all the HTML tags and change it to a text format. This can be done using the get_text method aligned within the for loop. This changes the HTML into the text format.

for mytag in proftags:
*print(mytag.get_text())***Hilarious (11)**
Caring (9)
Accessible outside class (8)
Amazing lectures (5)
Clear grading criteria (4)
Inspirational (4)
Group projects (3)
Respected (3)
Gives good feedback (2)
EXTRA CREDIT (2)
Participation matters (1)
Lecture heavy (1)
Test heavy (1)
So many papers (1)

Therefore, you get the above information that you are looking for. You get all the tags of the professor. This is how you scrape the data from the internet using Requests and Beautiful Soup libraries.

In summary, Python has three core data science libraries which many others have been built:

- Numpy
- SciPy
- Matplotlib

For simplicity, you can consider Numpy as your one-stop for arrays. Numpy arrays are different from the typical Python lists in many different ways, but a few to recall is that they are faster, consume less space, and have more functionality. It is important to remember, though, that these arrays are of a fixed size and type, which you define at creation. No infinitely appending new values like you want with a list.

SciPy is defined on top of Numpy and provides a lot of the optimization, linear algebra, and statistics functions you will need. Although sometimes, Numpy has similar functionality, SciPy's functionality is better. If you want to compute a correlation coefficient or generate some normally distributed data, then SciPy is the library for you.

Matplotlib

This is probably not grabbing any beauty awards, but it is the major library for plotting in Python. It has a lot of functionality and enables you to have huge control when needed.

2nd generation

The main libraries are great, and you will find yourself using them a lot. There are three 2nd generation libraries, which have been highly developed on top of the main one to give you additional functionality with less code.

Pandas were designed to make data analysis easier in Python. Pandas make it easy to load structured data, compute statistics on it, and slice and dice the data in whichever way you want. It is one of the most indispensable tools during the data exploration and analysis stage. However, it's not good to use it in production because it doesn't scale very well to large datasets. You can gain a huge boost in production by converting your Pandas code to raw Numpy.

Although Matplotlib isn't the best out of the box, Seaborn makes it easy to define beautiful visualizations. It is centered upon Matplotlib, so you can still apply Matplotlib functionality to augment Seaborn charts. It also makes it easier to define complex chart types.

Chapter 4: Data Manipulation

The Python Pandas library is an open-source project that has easy to use tools for data manipulation and analysis. A significant amount of time in any machine learning project will be required to prepare the data and analyze the basic trends and patterns before creating any models.

The Pandas library has risen into a powerhouse of data manipulation tasks in Python because it was built in 2008. With its elaborate syntax and flexible data structure, it's easy to learn and support faster data computation. The development of Numpy and Pandas library has stretched Python's multi-purpose nature to compute machine learning challenges. The acceptance of Python language in machine learning has been critical since then.
This is one of the reasons highlighting the need for you to master these libraries.

In the following chapter, you will learn about how to use Numpy and Pandas libraries for data manipulation from scratch. We shall mix both theory and practical aspects.

First, we will recap the syntax and commonly used functions of the respective libraries. Then later work on a real-life data set.

So, you need to be good at the basics of Python. No further knowledge is required. Also, ensure you have Python installed in your machine.

6 Key Things You Need to Know About Numpy and Pandas

- The data properties of Pandas are designed on top of the Numpy library. In one way, Numpy is a dependency of the Pandas library.
- Pandas is perfect at handling tabular data sets made of unique variables. Besides that, the Pandas library can still be used to perform even the most naïve of tasks such as loading data or performing feature engineering on time series data.
- Numpy is suitable for performing basic numerical computations like median, range, mean, etc. Besides that, it supports the development of multidimensional arrays.
- The Numpy library can also be used to mix C/C++ and Fortran code.
- Keep in mind, Python is a zero-indexing language, unlike R language, where indexing begins at one.
- The best thing about learning Pandas and Numpy is the strong, active community support you gain from the world.

Just to give you a grasp of the Numpy library, we'll quickly revise its syntax structures and some critical commands such as indexing, slicing, concatenation, etc. All these commands will be useful when using Pandas. Let's get started.

Getting Started with Numpy

First, load the library and confirm its version, just to be sure that we are not using an older version.

```
import numpy as np
np.__version__
'1.12.1'
```

Next, build a list of numbers running from 0 to 9.

L = list(range(10))

Then convert integers to a string. This method of working with lists is known as a list comprehension.

List comprehension provides a versatile means to deal with list manipulation tasks easily.

Array Indexing

The important thing to remember is that indexing in Python starts at zero.
x1 = np.array([4, 3, 4, 4, 8, 4])
x1
array([4, 3, 4, 4, 8, 4])

#assess value to index zero
x1[0]
4

#assess fifth value
x1[4]
8

#get the last value
x1[-1]
4

#get the second last value
x1[-2]
8

#in a multidimensional array, we need to specify row and column index
x2

```
array([[3, 7, 5, 5],
 [0, 1, 5, 9],
 [3, 0, 5, 0]])
```

#1st row and 2nd column value
x2[2,3]
0

#3rd row and last value from the 3rd column
x2[2,-1]
0

#replace value at 0,0 index
x2[0,0] = 12
x2
```
array([[12, 7, 5, 5],
 [ 0, 1, 5, 9],
 [ 3, 0, 5, 0]])
```

Array Slicing

Now, we'll learn to access multiple or a range of elements from an array.

y = np.arange(10)
y
array([0, 1, 2, 3, 4, 5, 6, 7, 8, 9])

#from start to 4th position
y[:5]
array([0, 1, 2, 3, 4])

#from 4th position to end
y[4:]
array([4, 5, 6, 7, 8, 9])

#from 4th to 6th position
y[4:7]
array([4, 5, 6])

#return elements at even place
y[: : 2]
array([0, 2, 4, 6, 8])

#return elements from first position step by two
y[1::2]
array([1, 3, 5, 7, 9])

#reverse the array
y[::-1]
array([9, 8, 7, 6, 5, 4, 3, 2, 1, 0])

Array Concatenation

Many a time, we are required to combine different arrays. So, instead of typing each of their elements manually, you can use array concatenation to handle such tasks easily.

#Using its axis parameter, you can define row-wise or column-wise matrix

np.concatenate([grid,grid],axis=1)
array([[1, 2, 3, 1, 2, 3],
 [4, 5, 6, 4, 5, 6]])

Until now, we used the concatenation function of arrays of equal dimension. But, what if you are required to combine a 2D array with a 1D array? In such situations, np.concatenate might not be the best option to use. Instead, you can use np.vstack or np.hstack to do the task. Let's see how!

x = np.array([3,4,5])
grid = np.array([[1,2,3],[17,18,19]])
np.vstack([x,grid])
array([[3, 4, 5],
 [1, 2, 3],
 [17, 18, 19]])

#Similarly, you can add an array using np.hstack
z = np.array([[9],[9]])
np.hstack([grid,z])
array([[1, 2, 3, 9],
 [17, 18, 19, 9]])

Besides the functions we have learned above, there are other mathematical functions available in the Numpy library, such as divide, multiple, mod, sin, cos, var, min, max, abs, etc. which you can apply to complete basic arithmetic calculations. Be free to go back to Numpy documentation for additional information on such functions.

Let's proceed to Pandas now. Ensure you follow every line below because it will enable you to perform the computation using Pandas.

Getting Started with Pandas

The load library - pd is just an alias. The pd is used because it's short and literally abbreviates Pandas. Still, you can use any name as an alias.

import Pandas as pd

Next, create a data frame. Dictionary is used here where keys get converted to column names and values to row values.

data = pd.DataFrame({'Country': ['Russia','Colombia','Chile','Equador','Nigeria'],
 'Rank':[121,40,100,130,11]})
data

	Country	Rank
0	Russia	121
1	Columbia	40
2	Chile	100
3	Equador	130
4	Nigeria	11

We can perform a quick analysis of any data set using:
data.describe()

	Rank
count	5.000000
mean	80.400000
std	52.300096
min	11.00000

25%	40.000000
50%	100.000000
75%	121.000000
max	130.000000

Remember, describe () method performs summary statistics of integer/double variables. To find complete information about the data set, we can apply the info () function.

Among other things, it shows the data set has 5 rows and 2 columns with their respective names.

#Let's create another data frame.
data = pd.DataFrame({'group':['a', 'a', 'a', 'b','b', 'b', 'c', 'c','c'],'ounces':[4, 3, 12, 6, 7.5, 8, 3, 5, 6]})
data

#Let's sort the data frame by ounces - inplace = True will make changes to the data
data.sort_values(by=['ounces'],ascending=True,inplace=False)

	group	ounces
1	a	3.0
6	c	3.0
0	a	4.0
7	c	5.0
3	b	6.0
8	c	6.0
4	b	7.5
5	b	8.0

2	a	12.0

Still, you can sort the data by not just one column but numerous columns as well.

data.sort_values(by=['group','ounces'],ascending= [True,False],inplace=False)

	group	Ounces
2	a	12.0
0	a	4.0
1	a	3.0
5	b	8.0
4	b	7.5
3	b	6.0
8	c	6.0
7	c	5.0
6	c	3.0

Typically, we get data sets with duplicate rows, which are just noise. As a result, before we train the model, we need to ensure we eliminate such inconsistencies within the data set. Here is how we can remove duplicate rows.

#sort values
data.sort_values(by='k2')

	K1	K2
2	one	3

1	one	2
0	one	1
3	two	3
4	two	3
5	two	4
6	two	4

#remove duplicates - ta da!
data.drop_duplicates()

	K1	K2
2	one	3
1	one	2
0	one	1
3	two	3
4	two	3
5	two	4

Here, we removed duplicates based on matching row values across all columns. Alternatively, we can also remove duplicates based on a particular column. Let's remove duplicate values from the k1 column.

data.drop_duplicates(subset='k1')

	K1	K2
o	one	3
3	two	3

Now, we will learn to categorize rows based on a predefined criterion. It happens a lot while data processing where you need to categorize a

variable. For example, say we have got a column with country names, and we want to create a new variable 'continent' based on these country names.

In this case, we will follow these steps:

```
data = pd.DataFrame({'food': ['bacon', 'pulled pork', 'bacon',
 'Pastrami','corned beef', 'Bacon', 'pastrami', 'honey ham','nova lox'],
                'ounces': [4, 3, 12, 6, 7.5, 8, 3, 5, 6]})
data
```

Now, we want to create a new variable that indicates the type of animal which acts as the source of the food. To do that, first, we'll create a dictionary to map the food to the animals. Then, we'll use map function to map the dictionary's values to the keys. Let's see how is it done.

```
meat_to_animal = {
'bacon': 'pig',
'pulled pork': 'pig',
'pastrami': 'cow',
'corned beef': 'cow',
'honey ham': 'pig',
'nova lox': 'salmon'
}
```

```
def meat_2_animal(series):
 if series['food'] == 'bacon':
 return 'pig'
 elif series['food'] == 'pulled pork':
 return 'pig'
 elif series['food'] == 'pastrami':
 return 'cow'
 elif series['food'] == 'corned beef':
```

```
return 'cow'
elif series['food'] == 'honey ham':
return 'pig'
else:
return 'salmon'
```

Another method to create a new variable is by using the assign function. With the following tutorial, you can continue to discover the new functions, and you will learn how powerful Pandas are.

data.assign(new_variable = data ['ounces']*10

We constantly find missing values within the data set. A quick approach for inputting missing values is by completing the missing value using any random number. Not only missing values, but you may find a lot of outliers within your data set, which demands replacement.

Let's proceed and learn about grouping data and creating pivots in Pandas. It's an immensely important data analysis method which you'd probably have to use on every data set you work with.

Importing Data

Pandas has tools to help read data from an extensive source. Since the dataset being used is a csv file, you will require to access read_csv function. This function has numerous options for parsing data. In a lot of files, the default choice works well.

```
import pandas as pd
train_values = pd.read_csv('train_values.csv')
train_labels = pd.read_csv('train_labels.csv')
```

For data to be analyzed, it is important to train the values and labels into a single dataframe. Pandas deliver a merge function that will integrate dataframes on either indexes or columns.

Missing Data

Pandas has multiple functions to handle missing data. To begin with, the isna() function can be important to learn the number of missing values present in the data.

The basic function of this appears at each value within a row and column, and it will display True if it is absent and false in case it is not. Therefore, it is possible to create a function that displays the fraction of missing values in every column.

In our dataset, there are no missing values available. But in case it was present, then we could apply a function to replace another value, or a function to remove the rows featuring missing values.

Anytime you apply fillna(), you have several options. You can decide to replace it with a static value which can be a string or a number. Also, you can substitute it with computations. It is true that you will need to apply different techniques for various columns based on the types of data and missing values.

The following code illustrates how you can apply pandas functions to fill the numerical values using mean.

```
train|train.select dtypes(include |'int64', 'float64'|).columns|
train|train.select dtypes(include_|'int64', 'float64'|).columns|.apply(lambda x:x.fillna(x.mean()))
```

Visualize the Data

When you plot values in Pandas, it doesn't appear attractive. However, if you want to highlight trends from data, it can always be the most effective means to achieve this.

The basic function for plotting that is normally used is the plt.plot(). Whenever you plot in Pandas, you will need to refer to the matplotlib API. Therefore, it is critical that you use matplotlib first.

This function supports different visualization types, including histograms, boxplots, and scatter plots. Where the plotting function in Pandas becomes critical is when you use it with other data aggregation functions.

To integrate the value_counts() using bar plot choice provides a rapid visualization for specific features. In the following code, we are exploring the distribution using this method.

```python
import matplotlib.pyplot as plt
% matplotlib inlinetrain['thal'].value_counts().plot.bar()
```

When you apply the groupby function, you require to plot mean.

```python
train.groupby("slope_of_peak_exercise_st_segment")['resting_blood_pressure'].mean().plot(kind='bar')
```

The Pandas pivot tables can still be used to create visualizations of compiled data. In this example, a comparison is made, and the relationship to heart disease available.
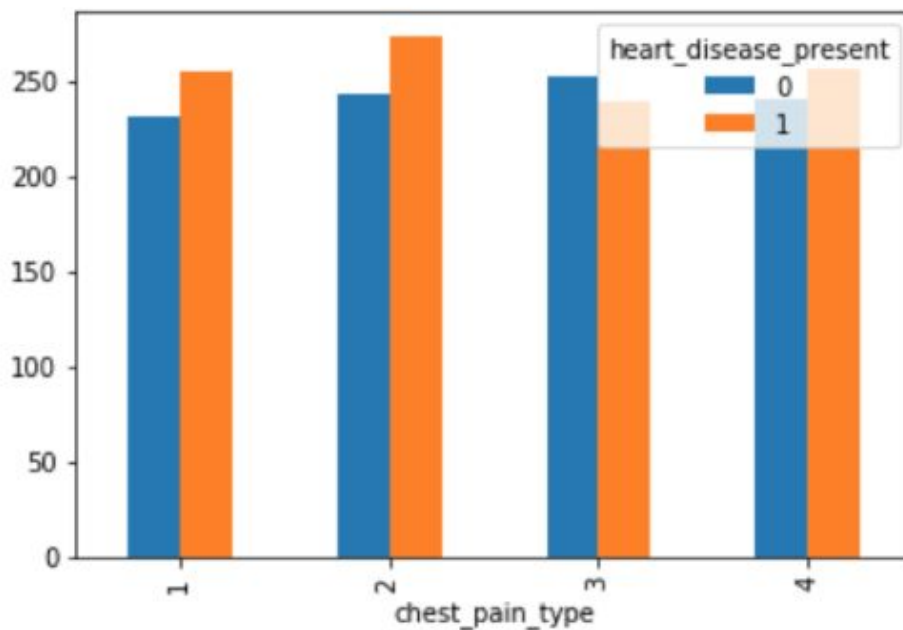
Transformation of Feature

Pandas does have different functions that can be applied in feature transformations.

For instance, the most popular machine learning libraries expect data to be in numerical form. As a result, it is important to change any non-numeric properties. The right way to achieve this is by hot encoding. Pandas have a function for this known as the get_dummies. Once this method is applied to a data column, it changes every unique value into a new binary column.

Another style which a feature may require to be converted for machine learning is called binning. A great example of this data set is the age feature. It can be significant to gather the ages into ranges for the model to learn. Pandas do have a function pd.cut that can be applied.

```
bins = [0, 30, 40, 50, 60, 70, 100]
train['age_group'] = pd.cut(train['age'], bins)
train['age_group'].value_counts().plot(kind='bar')
```
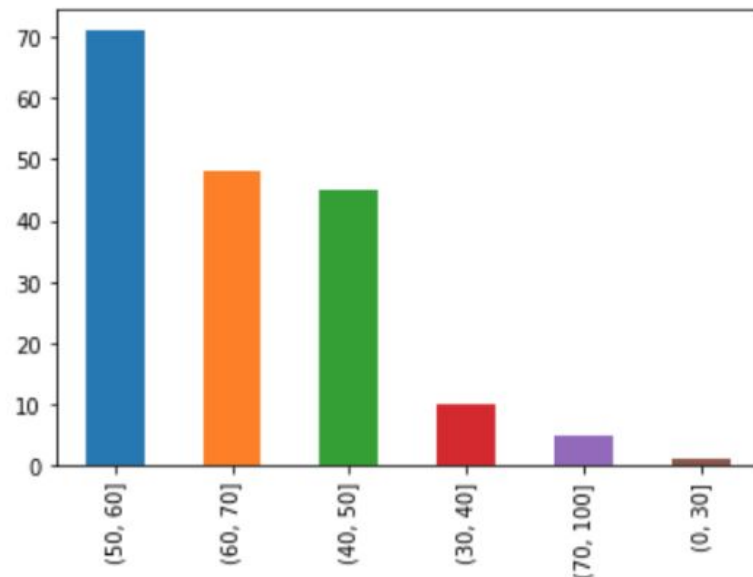

This is only a small introduction to different properties in Pandas for application in the early parts of the machine learning project. There are a lot of features to data analysis, manipulation, and the Pandas library itself. This can be a time-consuming stage, and Pandas deliver different tools and functions that can make the process efficient.

Chapter 5: Data Aggregation

This represents the first part of aggregation and clustering using Pharo DataFrame. This will only handle the basic functionality like clustering a data series using values of a separate series of corresponding size and using aggregation functions to the grouped data structures.

The next iterations will deal with functionality extended based on the targeted scenarios. The implementation is likely to change into something optimized.

Definition of Data Frame

This represent spreadsheet such as data structures that deliver an API for cleaning, slicing, and analyzing data.

In case you want to read more about the DataFrame project, you need to consider the documentation.

Split-Apply-Combine

The split-apply-combine is a technique where you categorize a certain task into manageable parts and then integrate all the parts together.

The data aggregation and grouping facilitates the production of summaries for analysis and display. For example, when you calculate the average values or creating a table of counts. This is a step that adheres the split-apply-combine procedure.

1. Separate the data into sections based on a given procedure.
2. Use the function to every cluster independently.

3. Combine the results using a data structure.

Implementation

In this part, you will discover how the grouping and aggregation function is being implemented. In case you don't want these details, you can skip to the next part.

Take, for instance, this message that has been sent to firstSeries object:

```
firstSeries groupBy: secondSeries.
```

Once this message is sent, firstSeries will define an object of DataSeriesGrouped, which divides firstSeries into various subseries depending on the values of secondSeries.

The collection of subseries is later kept as an object of DataSeries whose keys are equivalent to the special values of the secondSeries and values store the subseries of firstSeries. That will match each of those unique values.
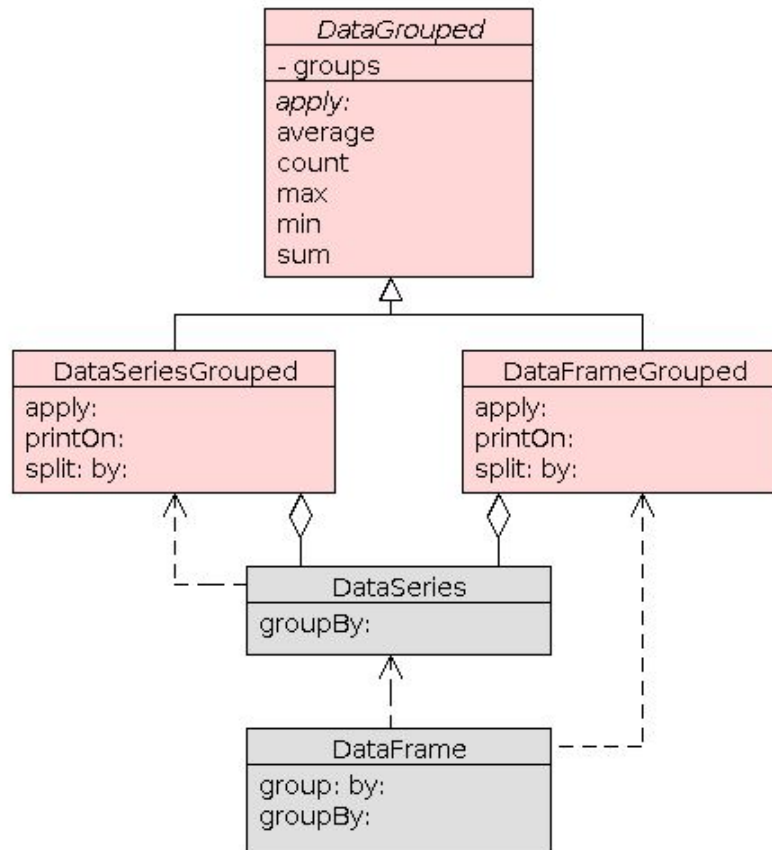
When a receiver of a groupBy: message is a DataFrame, it creates an instance of DataFrameGrouped, which splits the data similarly to the way DataSeriesGroup does it, except the values of groups series are sub-data frames, not subseries.

This means groups represent DataSeries that hold keys that match the unique values contained in a series which the data frame is identified. Where the data frame is categorized by a single column, this column is removed from data frame before the grouping. Therefore, this eliminates data duplication because the same values will be kept as keys.

For the case of DataSeriesGrouped, every subseries will be attached to a scalar, and all the following scalars will be joined into a DataSeries. When it comes to the DataFrameGrouped, it will include the block to every

column of each sub-data frame and display the eventual matrix of scalars as a new DataFrame.

Aggregation happens with the use of messages. It requires a block as an argument, and uses it on every value of the groups series, and then integrates into a new data structure.

The most common aggregation functions, such as average, min, and max deliver shorter messages. In the following iteration, these messages are useful and work as shortcuts.

```
average
  ^ self apply: [ :each | each average ].
```

However, these messages will carry the optimized implementations of the likened aggregations because it is necessary that these functions are time and memory efficient.

Let's examine the grouping series.

The easiest example of using this groupBy operator is to classify the values of a series using values of the same size.

```
bill := tips column: #total_bill.
sex := tips column: #sex.bill groupBy: sex.
```

The result of the above query will be an object. This object will separate the bill into two series.

Because a lot of time, you need to classify the group series that resemble columns of a single data frame. There is a useful shortcut.

How to Group Data Frames?

Besides the shortcut for classifying columns. The DataFrame has a method for classifying one of its columns.

The response of the above query will be an object of DataFrameGrouped, keeping two different data frames for smokers and non-smokers.

The smoker column will be removed from the above data frames because its values will be kept as keys within a DataFrameGrouped object. Additionally, the different groups of smokers and non-smokers will enable the complete reconstruction of the smoker column when needed.

The aggregation functions represent the ones that accept different input and display a scalar value that sums up the values of that particular series. These refer to statistical functions such as min, max, stdev, and many more.

Once the data has been combined, next you can use aggregation function to get the integrated data structure that sums up the original data.

```
grouped := tips group: #total_bill by: #day.
grouped apply: [ :each | each average round: 2].
```

Since the grouping is being done to a column of DataFrame by a separate column, the result will be a DataSeries object.

As said before, the DataGrouped presents shortcuts for popularly applied aggregation functions such as count, sum, min, and max. At the moment, these are shortcuts, but in future, they will execute the optimized aggregations that will be used faster.

Once the data frame was grouped into an object of DataFrameGrouped, we can also apply an aggregation function to this object. DataFrameGrouped implements the apply: message in such a way that the function is applied to each column of each sub-data frame, producing the scalar value. These scalars are then combined into a new data frame.

The result of this query will be a data frame containing the number of non-empty cells for each column, corresponding to 'Male' and 'Female' rows

```
         | total_bill  tip  smoker  day  time  size
---------+---------------------------------------------
Female   | 87          87   87      87   87    87
Male     | 157         157  157     157  157   157
```

Chapter 6: Data Visualization

Data Visualization is an important element for every data scientist. During the early periods of a project, you will need to perform an exploratory data analysis to identify insights into your data. Creating visualizations allows you to simplify things, particularly with a wide dimensional dataset. Towards the end of your project, you need to deliver the final result in a transparent and compelling manner that your audience can understand.

Data Visualization to the End-User

Usually, the data scientist has a role in submitting their insights to the final user. The results can be conveyed in different ways:

- *A single presentation*. In **the** following case, the research questions consist of one-shot deals because the business decision extracted from them will direct the organization to a given course for several years to come. For instance, company investment decisions.
- *Do you distribute the goods from two distribution centers or just one*? Where are they supposed to be located for the best efficiency? When the decision is made, the exercise might not be repeated until you retire. In the following case, the results are generated as a report with a presentation as the icing on the cake.
- *A new viewport on data*. The most common example of this is customer segmentation.

For sure, the segments themselves will be send using reports and presentations, but in essence, they comprise of tools, but not the final result itself.

Once a clear and important customer segmentation is identified, it can be supplied back to the database as a new channel on the data from which it was extracted.

From this point, people can create their own reports. For instance, how many products were sold to every customer segment?

- *For a real-time, Dashboard-Your functions as a data scientist doesn't complete once you have the new information. You can send your information back to the database and get done with it. However, when other people start to create reports on the discovered gold nugget, they can interpret it incorrectly and generate reports that don't make sense.*

Since you are the data scientist that found this new information, you need to set the example. In the following case, you need to create the first refreshable report so that the rest can learn from it and use your footsteps. Creating the first dashboard is still a means to reduce the delivery time of your insights to the final user who wants to make use of it daily. By doing this, they already have something to build upon until the reporting department discovers the time to establish a permanent report regarding the company's reporting software.

You may have discovered that some important elements are at play:

- First, what type of decision are you supporting? Is it strategic or operational? Strategic decisions need you to conduct an analysis and generate a report. But still, operational decisions require the report to be updated often.

- *What is the size of your organization?* For smaller organizations, you will deal with the general cycle. This one ranges from collection to reporting. For bigger teams, reporters

could be available to create the dashboards for you. Still, in the last part, creating a prototype dashboard can be relevant because it provides an example and reduces the delivery time.

Matplotlib is a great Python library that can be used to build your Data Visualizations. However, designing the data, parameters, and plots can get messy and tiresome to do regularly.

This section will guide you through data visualizations and create some rapid and easy functions with the help of Python's matplotlib. You will learn how to create basic plots using Seaborn, Matplotlib, and Pandas visualization.

Python offers many graphing libraries that are packed with a lot of features. No matter whether you want to describe an interactive, or complex Python plots, it delivers a powerful library.

To provide some overview, the popular plotting libraries consist of:

- Seaborn. This has an advanced interface and important default styles.
- Plotly. This is important in the development of significant plots.
- Visualization using Pandas. It is an easy to apply interface and has been built on Matplotlib.
- Ggplot. This one relies on R's ggplot2 and applies Grammar of Graphics.

Matplotlib

This is the most common Python library. It is a low library type that has Matlab interface which has more freedom.

Matplotlib is important for creating basic graphs such as bar charts, line charts, and histograms. You can import the following library by using the following line of code:

```
import matplotlib.pyplot as plt
```

Line Chart

The Matplotlib library allows the development of a line chart by applying the plot method. Still, it is possible to create multiple columns using a single graph by plotting and looping every column on the same axis.

Histogram

By using Matplotlib, you can design a histogram using the hist method. In case you relay categorical data like column points, it will help determine the likelihood of each class happening.

Bar Chart

In case you want to show data in a bar chart, then the bar function is useful. The bar chart is not automatically created using the frequency of a category, so you will require to use Pandas to achieve this. The bar chart is useful for grouping data that doesn't spread categories because it can become messy.

Visualization Using Pandas

Pandas represent an advanced level of open-source library. It is a simple library that represents data structures and data analysis tools.

Visualizations, with the help of pandas, makes it easy to create data frame and series. Still, it has an advanced level of API than the Matplotlib. As result, minimum code is required for similar results.

If you want to install Pandas, then you require to run the pip command.

The Objective of Visualization

Data communication and exploration are the major focus of data visualization. Once data is visualized, the patterns become visible. You will immediately tell whether there is an increasing trend or the relative magnitude of something in connection to other factors. Rather than tell people the long list of numbers, why not display to them for better clarity? For instance, let's consider the worldwide trend search on the word 'bitcoin'.
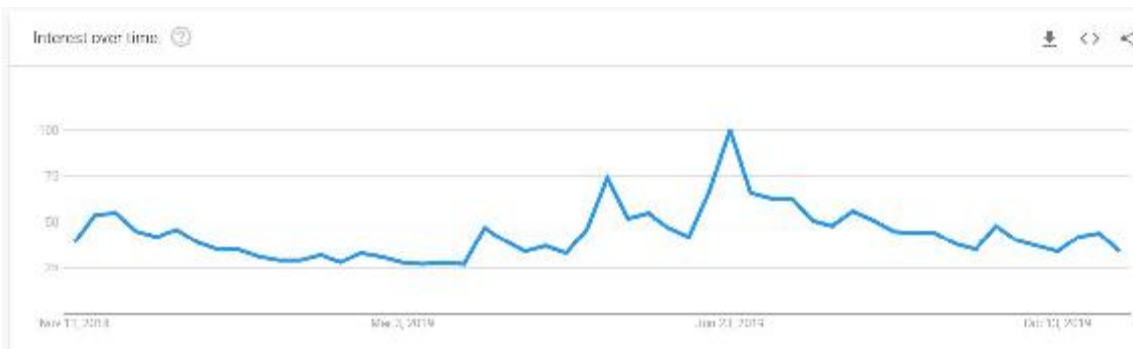
You should see that there is a temporary rise in the bitcoin interest, but it starts to decrease after the peak. Overall, during the peak interval, there's a huge hype connected to the technological and social effects of bitcoin. Again, the following hype decreases because people understand it, or it's a common thing related to hypes.

No matter the situation, data visualization helps us to determine the patterns in a very clear style. Keep in mind the importance of data visualization is to explore data. In the following case, you can quickly choose the patterns as well as the data send to us.

This is critical when you submit it to the public audience. Others may decide to go for a quick brief of the data without rushing into detail. You

don't really need to disturb them with texts and numbers. What presents a wide effect is the way you build them using numbers and texts. What sets a big difference is how you define the data so that individuals can quickly recall its importance.

This is where data visualization becomes helpful to allow people to mine data and communicate whatever you are trying to speak.
There are numerous methods of visualizing data. Some of these methods have already been discussed already.

The Simplest Method to Complex Visualization of Data

To create excellent data, visualizations are a powerful skill that every data scientist needs to be aware.

It is more than just creating beautiful charts, representing dataset's information in a way that it is easy for individuals to learn. When you have the right visualization, an individual can quickly learn the patterns and information that is found beneath the data.

In the early stages of a project, you will conduct an exploratory data analysis to generate insights into your data. Creating visualizations will increase your analysis.
At the end of your project, it is vital to submit your final results in a brief and compelling manner such that any audience can be able to read.

There's no doubt your visualizations to the next stage will let you defeat your next presentation.

This section will explore ways in which you can define an attractive, complex data visualization. You will apply the plotly python library that is excellent in creating interactive visualizations.

Overview of Plotly

Plotly represent an interactive, browser-depended graphic python library. It is a library that allows you to improve the visualization capabilities compared to the standard Matplotlib.

There are two benefits of applying Plotly instead of other Python libraries such as Matplotlib, Pandas, and Seaborn. That is:

1. The ease of application. This will define an interactive plot and other complex graphics. Performing the same operation using other libraries takes a lot of work.
2. It provides additional functionalities.

Since Plotly is designed from D3.js, the plotting capability is powerful than other plotting libraries.

The Sunburst charts and many more are possible using plotly.

Building Attractive Plots Using Plotly

Plotly is useful in building fancy plots.

To start, first, let's import plotly and its internal graph objects component. You will also import Pandas to load the dataset.

```
import plotly
import plotly.graph_objs as goimport Pandas as pd
```

To read the dataset, you basically write a one-liner in Pandas.

Scatter Plots

For this particular section, we are going to plot a scatter plot for sales price against year built. To achieve that, you will need to define a scatter graph object and store it in a trace.

```python
trace = go.Scatter(
  x = data['YearBuilt'],
  y = data['SalePrice'],
  mode = 'markers',
  showlegend = True
)
plot_data = [trace]
```

Then, to plot, you only write a single line.

```python
plotly.offline.plot(plot_data, filename='basic-scatter')
```

The following command will create a new tab within your browser with the plot.

Graph interactivity comes automatically built-in with Plotly.

Box Plots

This time, we will look at the box plots.
The process is quite similar. For that reason, we are going to define a graph object, store it into a trace, and then represent it in a browser.

```
import plotly
import plotly.graph_objs as go

import Pandas as pd

data = pd.read_csv('train.csv')

trace = go.Box(
 x = data['YearBuilt'],
 y = data['SalePrice'],
 marker = {'color': 'green'},
 showlegend = True,
)
plot_data = [trace]

plotly.offline.plot(plot_data, filename='basic-box')
```

The box plot will feature attractive properties with box plots. By default, we attain the same zooming, panning, and point of selection. Now that the box plot exists, if you hover around each box plot, it will reveal the following:

- Median
- 1st and 3rd quartiles
- Min and Max values of the data range
- The upper and/or lower fences if there are outliers

Heat Maps

Heat maps are a critical tool for a data scientist. They are effective for displaying the association between multiple feature variables in a single graph plus the relative significance of each relationship.

To demonstrate the way your Heat Maps can be improved with Plotly. We are going to create a correlation matrix of the House Prices dataset as a heat map.

```
import plotly
import plotly.graph_objs as go

import Pandas as pd

data = pd.read_csv('train.csv')

corrmat = data.corr()

trace = go.Heatmap(z=corrmat, x=corrmat.columns.tolist(), y=corrmat.columns.tolist())

plot_data = [trace]

plotly.offline.plot(plot_data, filename='basic-heatmap')
```

Heatmaps in Matplotlib can be somehow difficult because you cannot identify the correct value of each cell_ you can only tell from the color. You can write the code to make it interactive, but that's probably the hassle in Matplotlib.

Plotly provides interactivity beyond the box, so when you plot a heat map, you get an attractive overview and an option to confirm exact values when needed.

Both the pan-and-zoom functionality of Plotly is super clean, providing an easy means to perform a comprehensive exploration from a visual point of view.

These are just to indicate the significance and possibilities of applying plotly. Keep in mind that you can create a publication using quality data visualizations. Additionally, you can change the example codes to your objective. There's no need to invent something unique. You can copy the right sections and apply them to your data.

Probably, in the future, there will be easier and effective methods to build data visualizations, especially when dealing with huge datasets. You can still build animated presentations that can change with time. Whichever

way, the main goal of data visualization is to communicate data. You can select other methods, but the goal normally remains the same.

In the following chapter, you have learned general aspects of data visualization. You have learned that data visualization is the practice of understanding data by representing it in a graphical style so that trends may not be seen exposed.

Python provides many different graphic libraries that are packed with lots of different attributes.

In this section, we have looked at Matplotlib, Pandas visualization, and Seaborn. In the coming chapters, we will take a look at the complex topics that are specific to machine learning and complex data analysis. The original goal is to make sure that you know the common features and terms applied in data science.

Chapter 7: Machine Learning

Machine learning is a division of AI that delivers technologies with the ability to learn from past scenarios without any programming. Machine learning is abbreviated as ML. This field is concerned with the creation of computer applications that process data and learn from it.

The learning process starts with observations, direct experience, or examples to build patterns from the data and use these patterns to predict the future. The main goal of machine learning is to enable computers to automatically learn without intervention by humans

With ML, it is possible to analyze massive quantities of data. ML generates useful results, but still we may require different resources to reach this state. Extra time may be required to train the machine learning models.

Machine Learning Algorithms Classifications

Let's start with the supervised learning

Supervised Learning

For the case of supervised learning, the human is required to provide both the inputs and the outputs which are required and furnish the feedback based on the accuracy of the predictions during the time of training. After completion of the training, the algorithm will have to use what was applied to the next data.

The idea of supervised learning can be said to resemble learning under a teacher's supervision. The teacher provides examples to the student, and the student comes with new rules and knowledge based on these examples so that to apply it somewhere else.

It is also perfect for you to know the distinction between the regression problems and classification problems. When it comes to regression issues, the target is a numeric value, while in classification, the target is a class. A regression task can help compute the average cost of all houses in London. On the other hand, a classification task will allow you to determine the types of flowers based on the length of the petals and sepals.

Unsupervised Learning

When it comes to unsupervised learning, the algorithms don't expect to be supplied with the output data. A technique called deep learning, which is an iterative approach, is employed to revise the data and reach a conclusion. This makes them perfect for processing tasks that are complex compared to the supervised learning algorithms. In other words, unsupervised learning algorithms learn from examples without responses to these. The algorithm finds patterns from the examples on its own.

The supervised learning algorithms work the same way humans determine any similarities between two or more objects. Most of the recommender systems you come across when buying items online work based on unsupervised learning algorithms. In the following case, the algorithms find what to recommend to you what you have bought before. The algorithm has to approximate the type of customers whom you resemble, and a suggestion is derived from that.

This seems to be the essence of true Artificial Intelligence wherein the computer can learn without human intervention. It's about learning from the data itself and trying to find the relationship between different inputs (notice there's no expected output here in contrast to Regression and Classification discussed earlier). The focus is on inputs and trying to find the patterns and relationships among them. Perhaps there are natural clusters, or there are clear associations among the inputs. It's also possible that there's no useful relationship at all.

Reinforcement Learning

This type of learning happens when the algorithm is presented with examples that don't have labels, as it is the case with unsupervised learning. But the example can be achieved with the help of a positive or negative feedback based on the solution suggested by the algorithm. It is hooked with applications in which the algorithm has to make decisions, and these decisions are linked with a consequence. It is same as trial and error in human learning.

Errors become necessary in learning when they are connected with a penalty such as cost, loss of time, and pain. When it comes to reinforced learning, some actions are likely to succeed than others.

ML processes resemble those of data mining and predictive modeling. In both cases, searching through the data is a must so as to draw patterns then adjust the actions of the program accordingly. A great example of ML is the recommender systems. If you buy an item online, you will see an ad that is associated to that item, and that is a great example of ML.

How to Approach a Problem

Many data scientists approach a problem in a binary way. Does the task fall under Supervised or Unsupervised Learning?

The quickest way to figure it out is by determining the expected output. Are we trying to predict y values based on new x values (Supervised Learning, regression)? Is a new input under category A or category B based on previously labeled data (Supervised Learning, Classification)? Are we trying to discover and reveal how data points aggregate and if there are natural clusters (Unsupervised Learning, Clustering)? Do inputs have an interesting relationship with one another (do they have a high probability of co-occurrence)?

Many advanced data analysis problems fall under those general questions. After all, the objective is always to predict something (based on previous examples) or explore the data (find out if there are patterns).

What is Deep Learning

Deep learning refers to a sub-branch of ML that involves algorithms that are motivated by the function and structure of the brain known as the artificial neural networks. This makes machines to do what is natural to humans, and that is, learn from the past. This is the system behind the idea of driverless cars.

It is with the help of deep learning that a computer can perform classification tasks. Deep learning models can attain accuracy, which sometimes exceeds human-level performance.

Neural Networks with Scikit-Learn

Neural networks refer to ML technology that attempts to copy the natural biological neural networks operate. Humans have the capacity to identify patterns with a very high degree of accuracy. Anytime you see a cow, you can immediately recognize that it is a cow. This also applies to when you see a goat. The reason is that you have learned over a period of time how a cow or a goat looks like and what differentiates between the two.

Artificial neural networks refer to systems that try to imitate the capabilities of human learning via a complex architecture that resembles the nervous system of a human being.

The idea behind artificial neural networks is actually old. But recently it has undergone massive reemergence that many people (whether they understand it or not) talk about it.

Why did it become popular again? It's because of data availability and technological developments (especially massive increase in computational

power). Back then, creating and implementing an ANN might be impractical in terms of time and other resources. But it all changed because of more data and increased computational power. It's very likely that you can implement an artificial neural network right in your desktop or laptop computer. And also, behind the scenes, ANNs are already working to give you the most relevant search results, most likely products you'll purchase, or the most probable ads you'll click. ANNs are also being used to recognize the content of audio, image, and video.

Many experts say that we're only scratching the surface, and artificial neural networks still have a lot of potential. It's like when an experiment about electricity (done by Michael Faraday) was performed, and no one had no idea what use would come from it. As the story goes, Faraday told that the UK Prime Minister would soon be able to tax it. Today, almost every aspect of our lives directly or indirectly depends on electricity.

This might also be the case with artificial neural networks and the exciting field of Deep Learning (a subfield of machine learning that is more focused on ANNs).

The Structure of Neuron

A neutron is made up of the cell body, having a number of extensions from it.

Majority of these are in the form of branches commonly known as "dendrites".

A long process or a branching exists, and this is referred to as the "axon". The transmission of signals begins at a region in this axon, and this region is known as the "hillock".

The neuron has a boundary, which is known as the "cell membrane". A potential difference exists between the inside and the outside of the cell membrane. This is known as the "membrane potential".

If the input becomes large enough, some action potential will be generated. This action potential then travels will then travel down the axon and away from the cell body.

A neuron is connected to another neuron by synapses. The information leaves the neuron via an axon and is then passed to the synapses and to the neuron, which is expected to receive it. Note that a neuron will only fire once the threshold exceeds a certain amount. The signals are very important as they are received by the other neurons. The neurons use the signals or the spikes for communication. The spikes are also responsible for encoding the information which is being sent.

Synapses can either be inhibitory or excitatory. When spikes arrive at the excitatory synapse, the receiving neuron will be caused to fire. If the signals are received at an inhibitory synapse, then the receiving neuron is inhibited from firing.

The synapses and the cell body usually calculate the difference in the incoming inhibitory and excitatory inputs. If this difference is found to be too large, the neuron will be made to fire.

Back Propagation

To ensure that a neural network completes a problem, the units must be changed so that we can limit the error between the original output and the required output. In other words, the difference of the weights must be calculated by the network. To simplify everything, the network has to follow up with the changes in error. The backpropagation algorithm is popularly used in the error computation.

In case you have the network units as linear, then this particular algorithm will be easy to master. For the algorithm to compute the error difference of the weights, it requires to first understand the rate at which the error changes because the activity level is being changed.

For the case of the output units, the error difference is found by computing the difference between the real and the target output.

To compute the error change rate for the unknown measurement, make sure all the weights between the hidden unit and output units are calculated.

We can proceed and multiply the weights using the error derivatives in the weights, and then the product is compiled. The sum you find will be equivalent to the error change rate for the hidden unit. Once you get the error change rate in the weights of the hidden layer which is just before the output layer, we will manage to calculate these error changes for the other layers.

The computation for this will happen from one layer to the next, and in the direction which is different from the one transmitted through network.

This indicates where the name "back propagation" originates from. After the error change rate has been determined for some unit, the error difference for the weights for all the connections of the weight can be computed easily. The error difference for the weights can used by multiplying the rate of error change with activity through incoming connection.

Scikit-Learn

Scikit-learn offers users with various algorithms with the help of the Python interface. The library itself was designed in Python, and some of its algorithms were built in Cython to provide better performance. Scikit-learn is a better library for creating machine learning models. This library is open source and is supported by BSD license.

The Neural Networks Using TensorFlow

TensorFlow refers to a Google framework for the development of a deep learning model. Tensor flow depends on data-flow graphs for numerical calculation. TensorFlow has allowed machine learning to become easy. It

simplifies the process of getting data, creating predictions, and changing future results.

TensorFlow

The artificial neural network features a simple structure that requires the multiplication of matrices. The inputs are first multiplied with random weights and then relayed through an activation function. Next, the output values are extracted for making predictions.

This demonstrates that this network is far from the truth. The loss metric is used. A higher loss function implies a dumper model. To enhance network knowledge, some optimization has to be carried out by regulating the weights. Stochastic gradient is applied to change the values of the weights in the correct direction. Once the weights have been modified, the network can apply a new batch of data to test new knowledge.

The error may be lower than what you had initially, but not small enough. The optimization procedure should be implemented iteratively until the error is reduced, such that no additional knowledge can be extracted.

Despite that, the challenge with this particular model is that it has no memory. This implies that the input and output are independent.

What this shows is that the model doesn't care about what took place. This sparks several questions as when you need to predict time series because the network will require to have information about past words or historical data. To solve this issue, a new architecture known as a recurrent neural network was created.

Preparing the Environment

Before you get into the practical side of machine learning and deep learning, you need to have two libraries running in your computer. The

libraries include:

- Scikit-Learn
- TensorFlow

Installing Scikit-Learn

Scikit-learn works in Python 2.7 and above. Before you can install Scikitlearn, make sure that you have Numpy and SciPy libraries already installed.

Additionally, make sure that you have the latest versions of the above libraries. Once you install the above libraries, you can proceed to install Scikit-learn on your computer. The process of installation can be achieved with the help of pip. This is a tool that has Python. If you want to install scikit-learn, first run the following command on your terminal.

```
pip install scikit-learn
```

This installation should proceed until it reaches completion. Again, you can use conda to install the scikit-learn. You can do this by running the following command:

```
conda install scikit-learn
```

Once you are through with the installation of scikit-learn, you require to import it into your Python program to apply algorithms. This can be achieved with the help of import statements, as shown below:

Import Scikitlearn

In case the command runs without any error, understand that the installation of scikitlearn went successful. However, if the command sends an error, then the installation was not successful. Now, you can use the scikit-learn to build your own machine learning models.

Installing TensorFlow

TensorFlow has APIs for programming languages such as C++. Java, and it comes packed with a third-party package for R language. You will learn how to install TensorFlow on Windows. For Windows, TensorFlow can be installed using pip.

The native pip will install the TensorFlow on your system without the need to go through a virtual environment. Keep in mind that the installation of TensorFlow using pip interface with other Python installations on your system. But, the best thing is that you only need to run a single command and TensorFlow will be installed on your computer. Additionally, once TensorFlow is installed through pip, users will be enabled to run the TensorFlow programs from the directory they want.

If you want to install TensorFlow using Anaconda, you may need to establish a virtual environment. But within the Anaconda itself, it is advised that you install TensorFlow through pip install command instead of the conda install command.
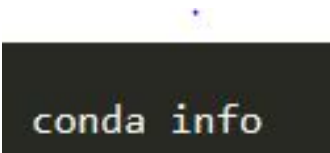
Make sure that you have Python3.5 + installed. Python3 comes with a pip3 program which can be used in the installation of TensorFlow. Therefore, we need to apply the pip3 install command to facilitate the installation.
This same command will install TensorFlow on your Windows system. Still, it is possible to install TensorFlow using the Anaconda package. The pip comes installed with Python, but Anaconda does not. In other words, for you to install TensorFlow using Anaconda, you will need to have the

Anaconda installed. So, navigate to the Anaconda website, and you will find all the instructions to guide you on the installation.

Once you have the Anaconda installed, then you can search for a package called conda, which is better for the management of virtual environments and installation of packages. To use the above package, you need to start the Anaconda.
Go to Windows, select Start, and click "All Programs", expand the "Anaconda…" folder. This should start the anaconda on your system. If you want to see the details of the conda package. Run the following command on the terminal.

This should return more details regarding the package manager.

```
(tensorenviron) C:\Users\admin\Documents>conda info
Current conda install:

               platform : win-64
          conda version : 4.3.27
       conda is private : False
      conda-env version : 4.3.27
    conda-build version : 3.0.22
         python version : 3.6.2.final.0
       requests version : 2.18.4
       root environment : C:\Users\admin\Anaconda3  (writable)
    default environment : C:\Users\admin\Anaconda3\envs\tensorenviron
       envs directories : C:\Users\admin\Anaconda3\envs
                          C:\Users\admin\AppData\Local\conda\conda\envs
                          C:\Users\admin\.conda\envs
          package cache : C:\Users\admin\Anaconda3\pkgs
                          C:\Users\admin\AppData\Local\conda\conda\pkgs
           channel URLs : https://repo.continuum.io/pkgs/main/win-64
                          https://repo.continuum.io/pkgs/main/noarch
                          https://repo.continuum.io/pkgs/free/win-64
                          https://repo.continuum.io/pkgs/free/noarch
                          https://repo.continuum.io/pkgs/r/win-64
                          https://repo.continuum.io/pkgs/r/noarch
                          https://repo.continuum.io/pkgs/pro/win-64
                          https://repo.continuum.io/pkgs/pro/noarch
                          https://repo.continuum.io/pkgs/msys2/win-64
                          https://repo.continuum.io/pkgs/msys2/noarch
            config file : None
             netrc file : None
           offline mode : False
             user-agent : conda/4.3.27 requests/2.18.4 CPython/3.6.2 Windows/7 W
indows/6.1.7601
          administrator : False

(tensorenviron) C:\Users\admin\Documents>
```

There is something special about the Anaconda. It allows us to develop a virtual Python environment with the help of the conda package. This particular virtual environment is a single copy of Python with the ability to maintain its own files, directories, and paths so that a person can work with particular versions of Python or other libraries affecting Python projects.

Virtual environments offer a means of separating projects and eliminate problems that may come up because of version requirements and different dependencies across multiple components. Keep in mind that the above virtual environment will remain different from your normal Python environment, implying that the packages running in the virtual environment will not impact the ones you have within your python usual environment.
You need to build a virtual environment for the TensorFlow package. This can be conducted with the help of conda create command. The command uses the following syntax:

```
conda create -n [environment-name]
```

In the above example, you need to assign the above environment the name tensoreviron. You can create it by running this command:

```
conda create -n tensorenviron
```

You will be requested to let the process of building the environment to proceed or not. Enter "y" to accept and hit the enter key on the keyboard. The installation will proceed successfully.

Once you create an environment, you need to activate it so that it is possible to use it. The activation can happen with the help of the activate command, followed by the environment name.Now that you have the TensorFlow environment activated, you can proceed to install the TensorFlow package in it. You can achieve that by running the command below.

```
conda install tensorflow
```

You will get a complete list of packages that will be installed with the TensorFlow package. You will be asked to allow the installation of this package. Just type "y" and press the enter key on your computer keyboard. The installation of the above packages will start immediately. Remember that the installation procedure may take longer, so you need to be patient. Despite that, the speed of your internet connection will determine how long this process is going to take.

The steps of the installation process will also be displayed in the prompt window. After a few minutes, the installation process will finish, and it will be time for you to confirm whether the installation was successful. You can do this by navigating to the Python's import statement. The statement should be executed from the Python terminal. While still on the anaconda

prompt, enter the phrase Python and press the enter key. This should direct you to the Python terminal. Now execute the following command.

```
import tensorflow as tf
```

If the package was not installed successfully, you would get an error, otherwise, the installation of the package was successful.

Chapter 8: Artificial Neural Networks

For us, humans, it's easy for us to identify objects and digits. It's also easy for us to understand the meaning of a sentence or piece of text.

However, it is a completely different case with computers. What's trivial and automatic for us might be an enormous task for algorithms and computers.

In contrast, computers can achieve long and complex mathematical computations while we humans are terrible at it. It's interesting that the abilities of humans and computers are complementary. However, the natural step is to replace humans at what they are expert at.

In the future, we might not be able to tell the difference whether whom we are talking to is human or not.

How the Brain Works

The most popular and promising style is the application of artificial neural networks; these are loosely motivated by how our brains and neurons work. The prevailing model concerning how the brain functions is by neurons processing and sending signals.

While it's not a 100% accurate method concerning the brain and neurons, this model is important for most applications.

This is the issue with artificial neural networks where there are neurons in one of few layers receiving and sending signals. Below is a basic demonstration from TensorFlow playground:
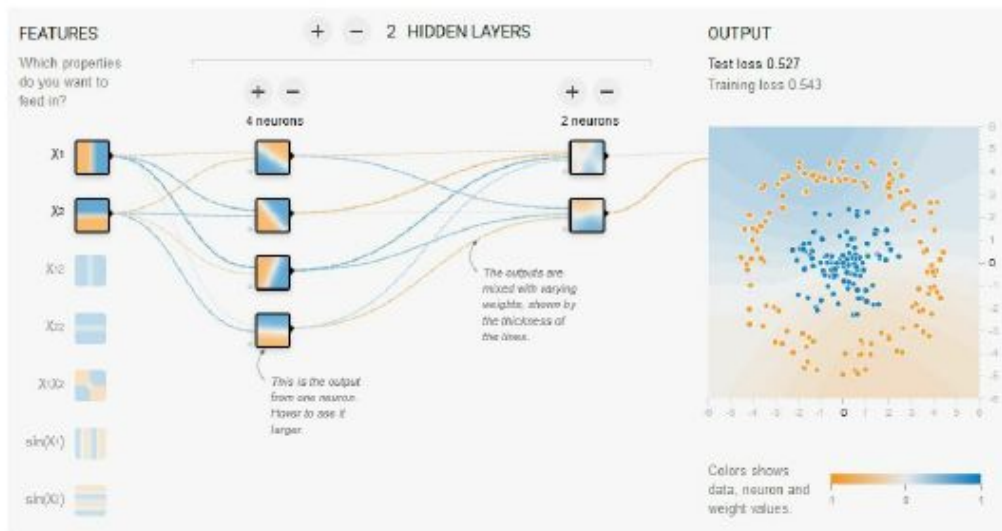
Remember that it began with the inputs, and then they are linked with 2 "hidden layers" of neurons.

Finally, there's a result wherein the data was already computed iteratively to define important models or generalization. In most cases, the way artificial neural networks function is similar to how supervised learning functions. When it comes to ANNs, we like to accept a large number of training examples and then create a system that supports learning from the above examples.

At the time of learning, the ANN automatically infers rules for recognizing an image, audio, text, or any other kind of data.
.
As you may have already noticed, the accuracy of recognition heavily relies on the quantity and quality of our data. At the end of the day, it's Garbage

In Garbage Out. Artificial neural networks tend to learn from the data fed into it.

We may still boost the accuracy and performance via means other than enhancing the quality and quantity of data.

Constraints and Opportunities

The notion behind ANN dates back long ago. However, in recent times it has undergone huge reemergence that most people discuss about it.

Why did it increase in popularity? The reason is the availability of data and technological development. Long ago, building and executing an ANN was impractical based on resources and time.

However, it all changed because of the increased speed of computation and data. It is likely that you can execute an artificial neural network right in your laptop computer.

Additionally, ANNs is already working behind the scenes to present the most important search results. ANNs are also being used to identify the content of image, video, and audio.

Most professionals believe that we are only lightly scratching ANN. It's like when an experiment about electricity was executed, and no one had an idea what function would emerge from it.

Let's See an Example

With the presence of TensorFlow playground, it's possible to identify a quick idea of how it operates. You can visit their website and record separate words such as activation, learning rate , and regularization.

Select the "Play" button and wait to see a cool animation. Be careful to the output at the far right. After some time, it will appear this:

The connections increase in clarity as well as other features. Keep in mind that the output has a clear Blue region.

This might be a classification problem wherein the blue dots are part of the Class A while the orange ones are part of the Class B.

As the ANN runs, remember that the division between Class A and Class B becomes clearer. And that is because the technology keeps learning from

the training examples. As learning becomes strong, the classification becomes accurate.

By learning more about TensorFlow playground, it provides the easiest way to understand how neural networks work. It's a rapid visualization even it is not 100% accurate representation. Therefore, we can understand the hidden layers, features, and output.

We can still introduce some changes in the learning rate, the ratio of training to test data, and the number of hidden layers. For example, we can set the number of hidden layers to 3 and adjust the learning rate to 1. You will see something like this:

If you press the Play button and allow it to run for some time, the image will remain like this:

Keep an eye on the output. You should see that the Classification appears worse. However, surrounding most of the yellow points under the Yellow region, there seems to be a lot of misses. This happens because of the parameter change.

For example, the Learning Rate has a big effect on the level of accuracy and striking the correct convergence. If you make the learning rate to be low, convergence may take a lot of time. However, if the learning rate is very high, we may not attain the convergence because we have overshot it and missed it.

There are different methods to attain convergence within an appropriate time. Learning Rate is simply right, a lot of hidden layers, perhaps fewer or more features to apply. However, over-optimizing everything may not make economic sense. It's right to set a clear goal at the beginning and stick to it.

In case there are great opportunities that arise, you may want to enhance the parameters and boost the performance of the model.
However, if you want to get some idea on the appearance of ANN in python, you can take a look at the following code:

```python
X = np.array([ [0,0,1],[0,1,1],[1,0,1],[1,1,1] ])
y = np.array([[0,1,1,0]]).T
syn0 = 2*np.random.random((3,4)) - 1
syn1 = 2*np.random.random((4,1)) - 1
for j in xrange(60000):
    l1 = 1/(1+np.exp(-(np.dot(X,syn0))))
    l2 = 1/(1+np.exp(-(np.dot(l1,syn1))))
    l2_delta = (y - l2)*(l2*(1-l2))
    l1_delta = l2_delta.dot(syn1.T) * (l1 * (1-l1))
    syn1 += l1.T.dot(l2_delta)
    syn0 += X.T.dot(l1_delta)
```

This is a simple example. When it comes to the real world, ANN would appear long and complex when designed from scratch. Luckily, the way you work with them is becoming more "democratized," which implies people with limited technical backgrounds would manage to take advantage of them.

Chapter 9: How to use Scikit-Learn

Now that you are done with the installations, you can begin to use the libraries. We will begin with the Scikit-Learn library.

To be able to use scikit-learn in your code, you should first import it by running this statement:

import sklearn

Loading Datasets

Machine learning is all about analyzing sets of data. Before this, we should first load the dataset into our workspace. The library comes loaded with a number of datasets that we can load and work with. We will demonstrate this by using a dataset known as *Iris*. This is a dataset of flowers. The following code shows how we can use scikit-learn to load the dataset:

```python
# Import scikit-learn library
from sklearn import datasets
# Load iris dataset
iris= datasets.load_iris()
# Confirm by printing the shape of the data
print(iris.data.shape)
```

Simple Linear Regression

We need to use our previous example, which is, predicting the number of marks a student will score in a test depending on the number of hours they have studied for the test. It is a simple linear regression task since we only have two variables.

Import Libraries

Run the following Python statements tom import all the necessary libraries:

```
import numpy as np
import pandas as pd

import matplotlib.pyplot as plt
```

This file is saved in an MS Excel file named *student_marks.csv*. It's in the directory where my Python scripts are found, so no need to give the path leading to the file. The .csv extension in the filename shows that it is a comma-separated values file.

The following statement will help us to import the dataset into the workspace.
We are using the Pandas library (we imported it as *pd*) for this:

dataset = pd.read_csv('student_marks.csv')

We can now explore the dataset to know more about it and see what it has.
Go directly to the Python terminal and type this:

dataset.shape

Which means that the dataset has 25 rows and 2 columns.
To see the first five rows of the data, call *head()* function.

However, you may get an error when you attempt to print the data. The cause of the error could be that Pandas is looking for the amount of information to display, so you should provide *sys* output information.
The error can be solved by modifying your code to the following:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sys
sys.__stdout__ = sys.stdout
dataset = pd.read_csv('student_marks.csv')
print(dataset.head())
```

We have simply provided the information to the *sys* library.
To see the statistical details of the dataset, we call the *describe()* function as follows:

dataset.describe()

We can now plot the data points on a 2-D graph and see how they are distributed. You no appreciate why we imported the *Matplotlib* library. The following code will help you to plot the data points:

```
dataset.plot(x='Hours', y='Marks', style='o')
plt.title('Hours vs Marks(%)')
plt.xlabel('Hours')
plt.ylabel('Marks (%)')
plt.show()
```

The code returns the following plot:
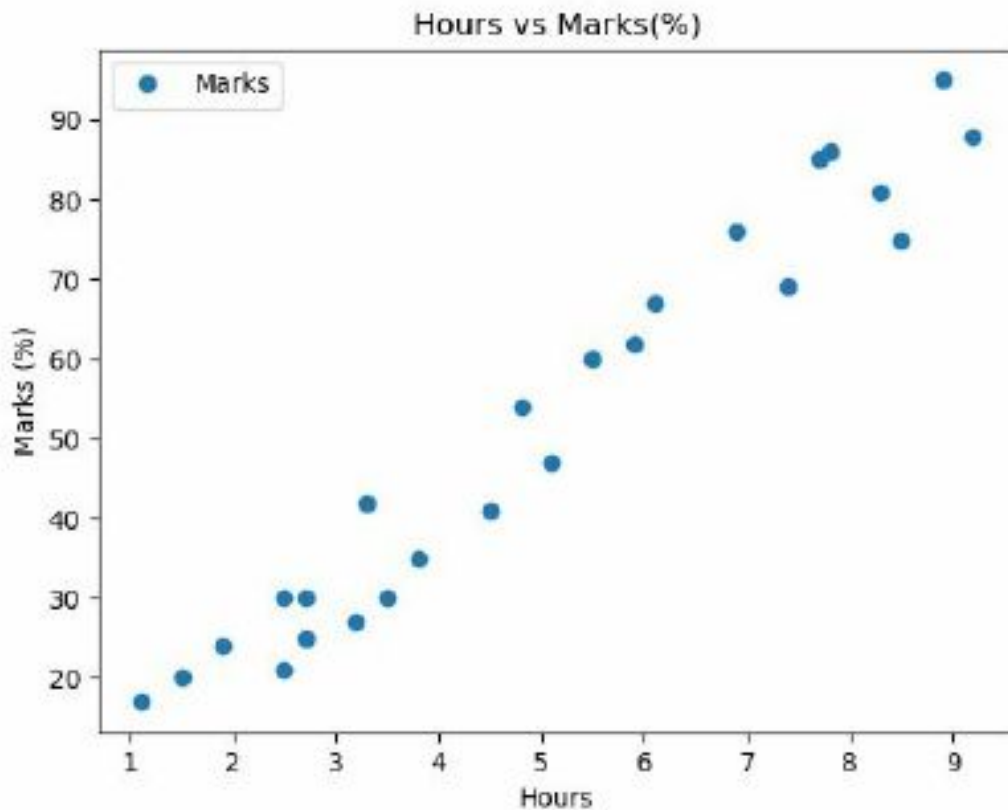
We have called the *plot()* function provided by the Pandas library. We passed the column names to this function, and it was able to create and display the plot. The *show()* function helped us to display the plot.

Data Preparation

The preparation of the data should involve subdividing it into labels and *attributes*. Attributes should create independent variables, and labels create dependent variables.

Our dataset has only two columns. We are predicting Marks based on Hours. This means Hours will form the attribute while Marks will form the label. The attributes and labels can be extracted by running the following code:

```
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, 1].values
```

The X variable will store the attributes. Note that we have used -1 because we need all columns to be assigned to attributes except the last one, that is, Marks. The y variable will store the labels. Here, we have used 1 since the column for Marks is at index 1. Remember that column indexes begin at index 0. At this point, we have the attributes and labels for the dataset. We need to divide our data into two sets, namely the *training* and *test* sets. The Scikit- Learn library provides us with a method named "train_test_split()" which can be used for this.

Training the Algorithm

We will be training the algorithm using the *LinearRegression* class, which must be imported from Scikit-Learn. The import can be done as follows:

from sklearn.linear_model import LinearRegression

Now that we have imported the class, we need to instantiate it and give the instance the name *linear_regressor*. This is demonstrated below:

linear_regressor = LinearRegression()

Let us now call the *fit()* method and pass the training data to it:

linear_regressor.fit(X_train, y_train)

As we had stated earlier, Linear Regression works by finding the best values for the slope and the intercept. This is what we have done above. These two have been calculated, so we only have to view their values.
To see the intercept, run the following command:

print(linear_regressor.intercept_)

Predicting

In the training done above, we have created a linear regression model, which is the equation. The values for the slope and the intercept are known. We can make predictions based on the data we preserved as the training set. The following statement helps us make predictions from the test data:

pred_y = linear_regressor.predict(X_test)

We have simply created a numpy array named *predict_y*. This will have all the predicted values for y from the input values contained in the *X_test* series. We now have the actual values for the X_test as well as the predicted values. We need to compare these two and see the amount of similarity or difference between the two. Just run the following code:

df = pd.DataFrame({'Actual': y_test, 'Predicted': pred_y})
print(df)

The model is not accurate, but the values are close to each other.

Evaluating the Accuracy

We now need to evaluate the accuracy of our algorithm. We need to determine how well the algorithm performed on the dataset. When it comes to regression algorithms, three evaluation metrics are used. These include the following:

1. Mean Absolute Error.
2. Mean Square Error.
3. Root Mean Squared Error.

Multiple Linear Regression

You now know how to do a Linear Regression when you have two variables.

However, this is not a true reflection of what we have in the real world. Most of the problems the world is facing involve more than two variables. This explains why you need to learn Multiple Linear Regression. The steps between the two are almost the same. However, the difference comes when it comes to evaluation. When evaluating the multiple linear regression model, we need to know the factor with the highest impact on the output variable. We also need to determine the relationship between the various variables.

We need to demonstrate this by prediction the consumption of fuel in US states. We will consider factors like per capita income, gas taxes, paved highways, and the proportion of persons who have a driver's license.

Let us first import the libraries that we need to use:

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Again, you may get an error for trying to print the contents of the dataset. Use the method we used previously to solve the problem. You should have the following code:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import sys
sys.__stdout__ = sys.stdout
dataset = pd.read_csv('fuel_consumption.csv')
print(dataset.head())
```

Data Preparation

In the example of simple linear regression, we subdivided the data into attributes and labels. In this case, we will be directly using the column names for this purpose. Run the code given below:

```python
X = dataset[['Tax', 'Income', 'Highways',
'Licence']]
y = dataset['Consumption']
```

We have four columns for the attributes (the independent variables) and one column for the label (the dependent variable).

Let us now subdivide the data into training and test sets. 80% of the data will be used as the training set while the remaining 20% will be used as the test set:
First, let us import the *train_test_split()* method from Scikit-Learn:

from sklearn.model_selection import train_test_split

Run the following command to do the division of the data:

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=0)

Training the Algorithm

We now need to train the algorithm. This can be done by calling the *fit()* method, as we did previously. Let us first import the *LinearRegression* class from the Scikit-Learn library:

from sklearn.linear_model import LinearRegression

Next, create an instance of the above class then use it to call the *fit()* method:

linear_regressor = LinearRegression()
linear_regressor.fit(X_train, y_train)

Note that this is a multiple linear regression and we have many variables. The linear regression model has to find the optimal coefficients for each attribute. You can see the chosen coefficients by running the following command:

coeff = pd.DataFrame(linear_regressor.coef_, X.columns, columns= ['Coefficient'])
print(coeff)

```
          Coefficient
Tax        -40.016660
Income      -0.065413
Highways    -0.004741
Licence   1341.862121
```

This means that any unit increase in fuel tax will lead to a decrease of 40.02 million gallons in gas Consumption. Also, a unit increase in the proportion of the population with Driver's license will lead to an increase of 1.342

billion gallons in gas Consumption. The results also show that average Income and Highways have a very small impact on gas Consumption.

Predicting

We will make the predictions using the test data. The following script can help us do this:

pred_y = linear_regressor.predict(X_test)

At this point, you have the actual X_text values as well as the predicted values. We need to perform a comparison between these two to determine their similarities and differences. This can be done by running the following script:

df = pd.DataFrame({'Actual': y_test, 'Predicted': pred_y})
print(df)

Evaluating the Accuracy

We now need to evaluate the algorithm in terms of performance. This can be done by determining the values for the various types of errors. These include the MAE, RMSE, and MSE. This requires us to firm import the *metrics* class from Scikit-Learn:

from sklearn import metrics

The calculation of the error values can then be done using the following script:

```
print('MAE:', metrics.mean_absolute_error(y_test, pred_y))
print('MSE:', metrics.mean_squared_error(y_test, pred_y))
print('RMSE:', np.sqrt(metrics.mean_squared_error(y_test, pred_y)))
```

The value for the root means the square error is 68.31, as shown above. This is slightly greater than 10% of the mean value for gas consumption in all the states. It is true that our algorithm was not very much accurate, but we can still use it to make predictions. The big error could be brought by a number of factors. Maybe there was a need for more data. We have used data for only one year. Having data collected over multiple years could have helped us improve the accuracy of the model. Also, we had assumed that our data has a linear relationship. This could not be the case, hence the big error. I recommend that you visualize the data to see whether this is true. Also, the features we have used could not be correlated.

Chapter 10: K-Nearest Neighbors Algorithm

The KNN algorithm is highly used for building more complex classifiers. It is a simple algorithm, but it has outperformed many powerful classifiers. That is why it is used in numerous applications data compression, economic forecasting, and genetics. KNN is a supervised learning algorithm, which means that we are given a labeled dataset made up of training observations (x, y) and our goal is to determine the relationship between x and y. This means that we should find a function that x to y such that when we are given an input value for x, we are able to predict the corresponding value for y. The concept behind the KNN algorithm is very simple. We will use a dataset named Iris. We had explored it previously. We will be using this to demonstrate how to implement the KNN algorithm.

First, import all the libraries that are needed:

```python
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

Splitting the Dataset

We need to be able to tell how well our algorithm performed. This will be done during the testing phase. This means that we should have training and testing data. The data set should be divided into two parts. We need to split the data into two parts. 80% of the data will be used as the training set, while 20% will be used as the test set. Let us first import the *train_test_split* method from Scikit-Learn.

Feature Scaling

Before we can make real predictions, it is a good idea for us to scale the features. After that, all the features will be evaluated uniformly. Scikit-Learn comes with a class named *StandardScaler,* which can help us perform the feature scaling. Let us first import this class.

We then instantiate the class then use it to fit a model based on it:

```
feature_scaler = StandardScaler()
feature_scaler.fit(X_train)
X_train = feature_scaler.transform(X_train)
X_test = feature_scaler.transform(X_test)
```

The instance was given the name *feature_scaler*.

Training the Algorithm

With the Scikit-Learn library, it is easy for us to train the KNN algorithm. Let us first import the *KNeighborsClassifier* from the Scikit-Learn library:
from sklearn.neighbors import KNeighborsClassifier

The following code will help us train the algorithm:

```
knn_classifier = KNeighborsClassifier(n_neighbors=5)
knn_classifier.fit(X_train, y_train)
```

Note that we have created an instance of the class we have created and named the instance *knn_classifier*. We have used one parameter in the instantiation, that is, *n_neighbors*. We have used 5 as the value of this parameter, and this basically, denotes the value of K. Note that there is no specific value for K, and it is chosen after testing and evaluation. However, for a start, 5 is used as the most popular value in most KNN applications.

We can then use the test data to make predictions. This can be done by running the script given below:

pred_y = knn_classifier.predict(X_test)

Evaluating the Accuracy

Evaluation of the KNN algorithm is not done in the same way as evaluating the accuracy of the linear regression algorithm. We were using metrics like RMSE, MAE, etc. In this case, we will use metrics like confusion matrix, precision, recall, and f1 score. We can use the *classification_report* and *confusion_matrix* methods to calculate these metrics. Let us first import these from the Scikit-Learn library: from sklearn.metrics import confusion_matrix, classification_report
Run the following script:

```
print(confusion_matrix(y_test, pred_y))
print(classification_report(y_test, pred_y))
```

The results given above show that the KNN algorithm did a good job of classifying the 30 records that we have in the test dataset. The results show that the average accuracy of the algorithm on the dataset was about 90%. This is not a bad percentage.

K Means Clustering

Let us manually demonstrate how this algorithm works before implementing it on Scikit-Learn:

Suppose we have two-dimensional data instances given below and by the name D:

D = { (5,3), (10,15), (15,12), (24,10), (30,45), (85,70), (71,80), (60,78), (55,52), (80,91) }

Our objective is to classify the data based on the similarity between the data points.

We should first initialize the values for the centroids of both clusters, and this should be done randomly. The centroids will be named c1 and c2 for clusters C1 and C2 respectively, and we will initialize them with the values for the first two data points, that is, (5,3) and (10,15) . It is after this that you should begin the iterations. Anytime that you calculate the Euclidean distance, the data point should be assigned to the cluster with the shortest Euclidean distance. Let us take the example of the data point (5,3):

```
Euclidean Distance from the Cluster Centroid c1 = (5,3) = 0
Euclidean Distance from the Cluster Centroid c2 = (10,15) = 13
```

The Euclidean distance for the data point from point centroid c1 is shorter compared to the distance of the same data point from centroid c2. This means that this data point will be assigned to the cluster C1   The distance from the data point to the centroid c2 is shorter; hence, it will be assigned to the cluster C2. Now that the data points have been assigned to the right clusters, the next step should involve the calculation of the new centroid values. The values should be calculated by determining the means of the coordinates for the data points belonging to a certain cluster. If for example for C1 we had allocated the following two data points to the cluster:

(5, 3) and (24, 10). The new value for x coordinate will be the mean of the two:
x = (5 + 24) / 2
x = 14.5
The new value for y will be:
y = (3 + 10) / 2

y = 13/2
y = 6.5

The new centroid value for the c1 will be (14.5, 6.5).

This should be done for c2, and the entire process is repeated. The iterations should be repeated until when the centroid values do not update anymore. This means if, for example, you do three iterations, you may find that the updated values for centroids c1 and c2 in the fourth iterations are equal to what we had in iteration 3. This means that your data cannot be clustered any further. You are now familiar with how the K-Means algorithm works. Let us discuss how you can implement it in the Scikit-Learn library. Let us first import all the libraries that we need to use:

```python
import matplotlib.pyplot as plt
import numpy as np
from sklearn.cluster import KMeans
```

Data Preparation

We should now prepare the data that is to be used. We will be creating a numpy array with a total of 10 rows and 2 columns. So, why have we chosen to work with a numpy array? It is because the Scikit-Learn library can work with the numpy array data inputs without the need for preprocessing.

Visualizing the Data

Now that we have the data, we can create a plot and see how the data points are distributed. We will then be able to tell whether there are any clusters at the moment:

```
plt.scatter(X[:,0],X[:,1], label='True Position')
plt.show()
```

The code gives the following plot:
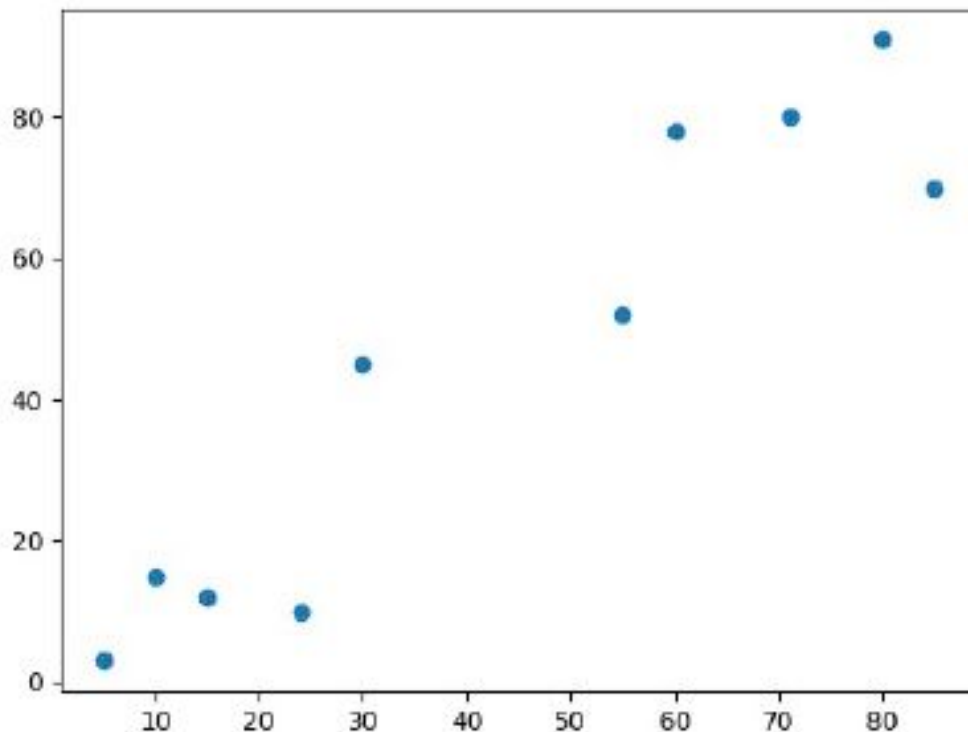
If we use our eyes, we will probably make two clusters from the above data, one at the bottom with five points and another one at the top with five points. We now need to investigate whether this is what the K-Means clustering algorithm will do.

Creating Clusters

We have seen that we can form two clusters from the data points, hence the value of K is now 2. These two clusters can be created by running the following code:

```
kmeans_clusters = KMeans(n_clusters=2)
kmeans_clusters.fit(X)
```

We have created an object named *kmeans_clusters,* and 2 have been used as the value for the parameter *n_clusters*. We have then called the *fit()* method on this object and passed the data we have in our numpy array as the parameter to the method. We can now have a look at the centroid values that the algorithm has created for the final clusters: print (kmeans_clusters.cluster_centers_) This returns the following: The first row above gives us the coordinates for the first centroid, which is, (16.8, 17). The second row gives us the coordinates of the second centroid, which is, (70.2, 74.2). If you followed the manual process of calculating the values of these, they should be the same. This will be an indication that the K-Means algorithm worked well.

The following script will help us see the data point labels:

print(kmeans_clusters.labels_)

This returns the following:

The above output shows a one-dimensional array of 10 elements that correspond to the clusters that are assigned to the 10 data points. You clearly see that we first have a sequence of zeroes, which shows that the first 5 points have been clustered together while the last five points have been clustered together. Note that the 0 and 1 have no mathematical significance, but they have simply been used to represent the cluster IDs. If we had three clusters, then the last one would have been represented using 2's.

We can now plot the data points and see how they have been clustered. We need to plot the data points alongside their assigned labels to be able to distinguish the clusters. Just execute the script given below:

plt.scatter(X[:,0],X[:,1], c=kmeans_clusters.labels_, cmap='rainbow')
plt.show()

The script returns the following plot:
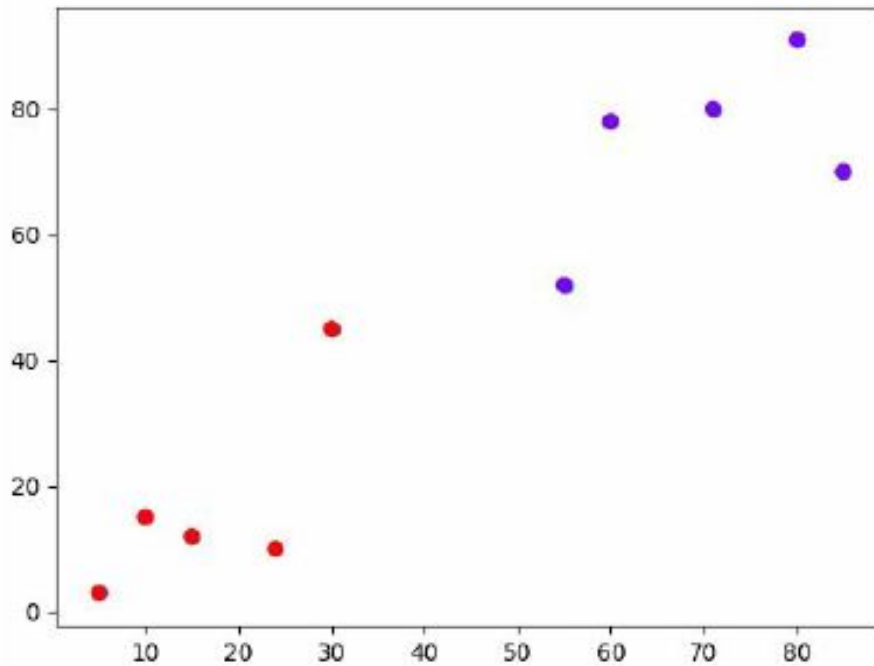
We have simply plotted the first column of the array named X against the second column. At the same time, we have passed *kmeans_labels_* as the value for parameter c, which corresponds to the labels. Note the use of the parameter *cmap='rainbow'*. This parameter helps us to choose the color type for the different data points.

As you expected, the first five points have been clustered together at the bottom left and assigned a similar color. The remaining five points have been clustered together at the top right and assigned one unique color. We can choose to plot the points together with the centroid coordinates for every cluster to see how the positioning of the centroid affects clustering. Let us use three clusters to see how they affect the centroids. The following script will help you to create the plot:

```
plt.scatter(X[:,0], X[:,1], c=kmeans_clusters.labels_, cmap='rainbow')
plt.scatter(kmeans_clusters.cluster_centers_[:,0]
,kmeans_clusters.cluster_centers_[:,1], color='black')
plt.show()
```

The script returns the following plot:
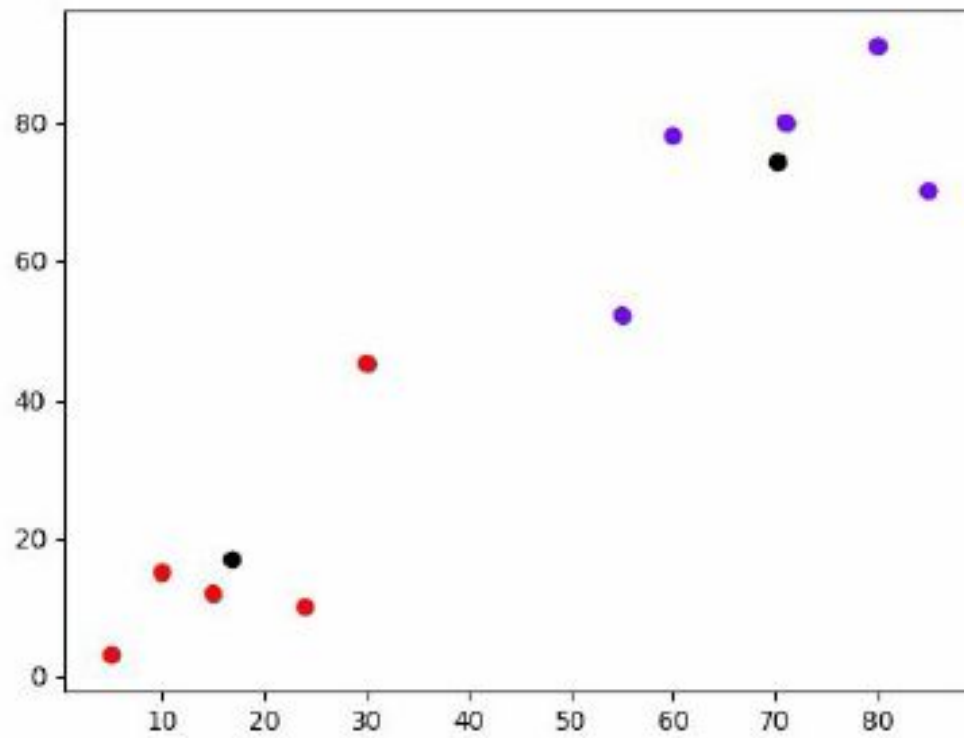
Chapter 11: Classification

Is it spam or not spam? This is one of the popular applications and examples of Classification. Similar to regression, Classification is also categorized under Supervised Learning. The model learns from labeled data. Then the system uses that learning to a new dataset.

For instance, there's a dataset with various email messages, and each dataset is classified as spam or not spam. The model may then identify features among email messages that are highlighted as spam.
When conducting a prediction, the model may attempt to identify those features in new email messages.

There are different ways of accomplishing a successful classification. Let's look a few of them:

Logistics Regression

In most classification problems, the objective is to confirm whether it's 0 or 1 by using two independent variables. For instance, provided that the Age and Estimated Salary determine a result such as when the person bought, how can we successfully build a model that represents their relationships, and use that for prediction.

This might sound confusing, and that is why you need to get an example.

Logistic Regression (Training set)
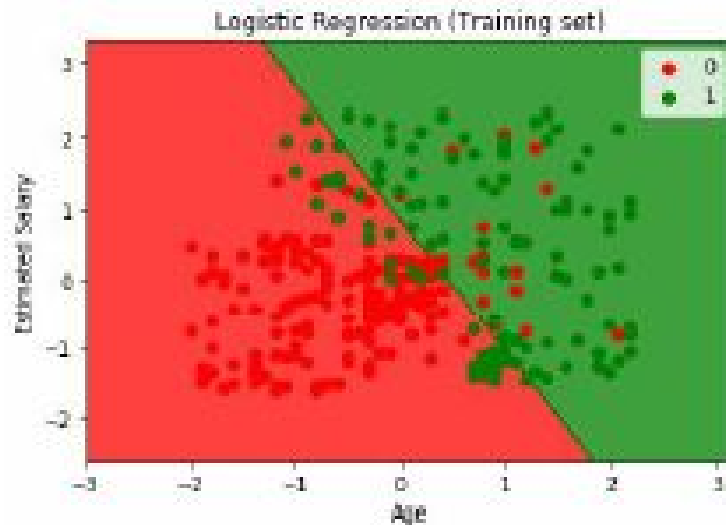
In the above examples, there are two variables Age and Estimated Salary. Every data point is then categorized either as 0 or 1. There's a line that separates the two. This method is based on probability.

As with Regression, wherein there's the black box, and the behind the scenes of Logistics regression can be difficult. The good thing is that its execution is straightforward when you apply scikit-learn and Python.

It is a usual step to first learn from the Training Set and then use that learning to test set. At the end of the day, this is the purpose of Supervised Learning. First, you will find there's training and supervision. Next, the lesson will be used in new situations.

As you discover in the visualization for the test set, the majority of the green dots fall under the green region. Therefore, the model can be best for predicting whether a person with specific age and estimated salary can buy or not.

```
# Feature Scaling
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

First, we convert the data into the same range or scale to avoid skewing or over-dependency on a given variable. In the dataset, the approximated salary is represented in thousands, while age is represented on a smaller scale. We need to set them in the same range so that we can expect a reasonable model.

Besides Logistic Regression, there are other methods of grouping tasks. Let's discuss them:

K-Nearest Neighbors

This algorithm is popular for creating advanced classifiers. It is an easy algorithm, but it has done better than other classifiers. That is the reason it's popular in genetics, forecasting, and data compression.

The KNN belongs to a supervised learning algorithm. In other words, we are presented with a dataset made up of training observations (x, y), and the main goal is to identify the association between x and y. So, we need to define a function x-y such that when we are provided with an input value for x, we can predict the corresponding value for y.

The idea behind the KNN algorithm is quite simple. It computes the distance of the new data point to all the other training data points. The distance can be of different types, including Euclidean, Manhattan, etc. The K-nearest data points are selected, where K can be an integer. Lastly, the

data point is assigned to the class, which most of the K data points belong to.

Remember that Logistic Regression appears to feature a linear boundary between 0s and 1s. Therefore, it lacks a few of the data points that require to be on the other end.
Luckily, there are non-linear models that feature a lot of data points in a more accurate style. One of them is the application of K-Nearest Neighbors. It operates by letting a new data point and then counting the number of neighbors that belong to either category. In case a lot of neighbors belong to category A than category B, then the new point should be part of category A.

Therefore, the classification of a given point depends on the majority of its nearest neighbors. This can be achieved the following line of code:

```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 5, metric = 'minkowski', p = 2)
classifier.fit(X_train, y_train)
```

Also, rather than starting from square one, we will import the prebuilt code that simplifies our task. The behind the scenes can be learned and studied. But for many reasons, the prebuilt ones are better to make reasonably helpful models.

Don't forget that the boundary is non-linear because of the various approach by the K-NN algorithm. Also, remember that the misses still exist. For you to capture all the misses, you need to use a larger dataset or a different method.

The Decision Tree Classification

Many data scientists also apply decision trees in classification. Building a decision tree involves dividing a dataset into smaller subsets while

separating them out. Take a look at the example below:

Are you able to see that the branches and leaves originate from the dataset division? The same can apply in classification.

When you run the entire code, you will see the following:

Decision Tree Classification (Training set)

Decision Tree Classification (Test set)

See the big difference when it is compared to Logistic regression and K-Nearest Neighbors (K-NN). In these two examples, there are only two boundaries. In the above decision tree classification, you can identify two points outside the main red region that are within the "mini red regions." In the following case, the model was able to retain the data points that would be impossible otherwise.

Random Forest Classification

Remember from the previous section on regression, random forest was described as a collection of decision trees. This applies to classification

wherein a lot of decision trees are applied, and the results are averaged.

Random Forest Classification (Training set)

Random Forest Classification (Test set)
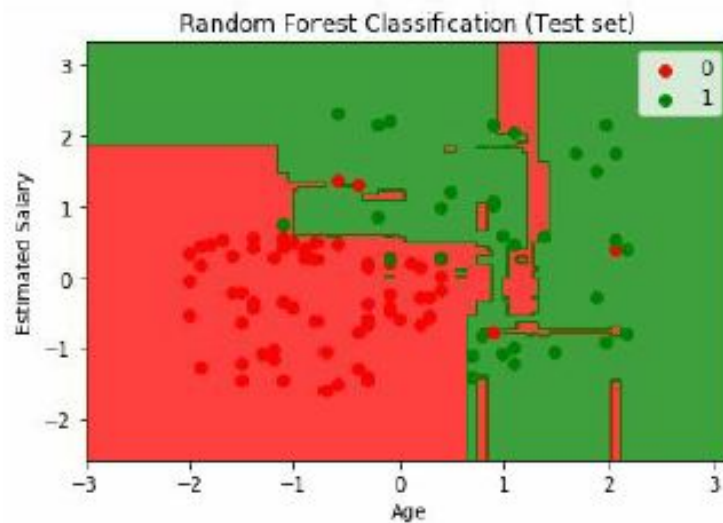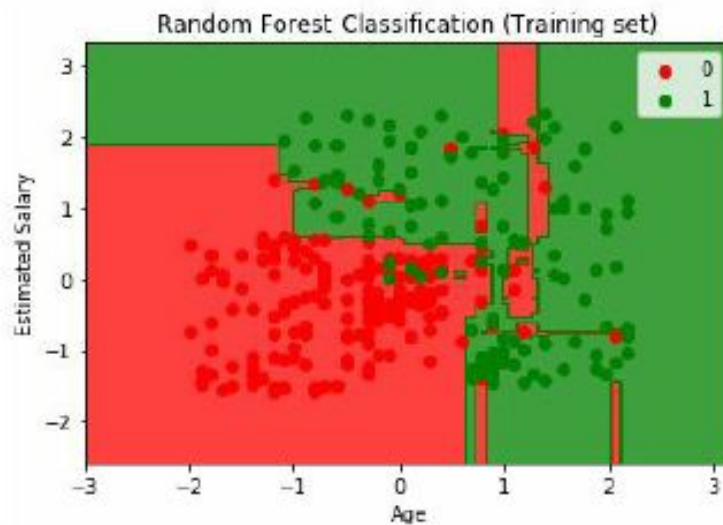
Take a look at the similarities between random forest and decision tree. At the end of the day, they pick a similar approach of dividing down a dataset into smaller subsets. The difference is that the Random Forest applies randomness and averages the decision trees to develop an accurate model.

Clustering

The preceding chapters explored supervised learning. We have mastered learning from "labeled" data. Already, there were correct responses and our role was to learn how to get those answers and use them in learning new data.

In the following chapter, it will be different. The reason is that it will be beginning with unsupervised learning, wherein there were no exact labels provided.

This means there's only input data, but there's no output data. Additionally, no supervision happens when learning from data.

In fact, unsupervised learning is considered to engulf the essence of artificial intelligence. That's because there is no much human supervision or intervention. As a result, the algorithms are left to discover things from data. This is the case in Clustering wherein the eventual goal is to reveal organic aggregates or "clusters" in data.

Objectives and Function of Clustering

This is a form of Unsupervised Learning where there are no labels, or in many cases, there are no truly correct answers. That's because there were no correct answers in the first place. We just have a dataset, and our goal is to see the groupings that have organically formed. We're not trying to predict an outcome here. The goal is to look for structures in the data. In other words, we're "dividing" the dataset into groups wherein members have some similarities or proximities. For example, each e-commerce customer might belong to a particular group (e.g., given their income and spending level). If we have gathered enough data points, it's likely there are aggregates.

At first, the data points will seem scattered (no pattern at all). But once we apply a clustering algorithm. The data will somehow make sense because

we'll be able to easily visualize the groups or clusters. Aside from discovering the natural groupings, Clustering algorithms may also reveal outliers for Anomaly Detection (we'll also discuss this later). Clustering is being applied regularly in the fields of marketing, biology, earthquake studies, manufacturing, sensor outputs, product categorization, and other scientific and business areas. However, there are no rules set in stone when it comes to determining the number of clusters and which data point should belong to a certain cluster. It's up to our objective (or if the results are useful enough). This is also where our expertise in a particular domain comes in. As with other data analysis and machine learning algorithms and tools, it's still about our domain knowledge. This way, we can look at and analyze the data in the proper context. Even with the most advanced tools and techniques, the context and objective are still crucial in making sense of data.

K-Means Clustering

One way to make sense of data through Clustering is by K-Means. It's one of the most popular Clustering algorithms because of its simplicity. It works by partitioning objects into k clusters (number of clusters we specified) based on feature similarity. Notice that the number of clusters is arbitrary. We can set it into any number we like. However, it's good to make the number of clusters just enough to make our work meaningful and useful. Let's discuss an example to illustrate this. Here we have data about Mall Customers ('Mall_Customers.csv') where info about their Gender, Age, Annual Income, and Spending Score are indicated. The higher the Spending Score (out of 100), the more they spend at the Mall.
To start, we import the necessary libraries:

```
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
%matplotlib inline
```

Then we import the data and take a peek:

	CustomerID	Genre	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
5	6	Female	22	17	76
6	7	Female	35	18	6
7	8	Female	23	18	94
8	9	Male	64	19	3
9	10	Female	30	19	72

In this example, we're more interested in grouping the Customers according to their Annual Income and Spending Score.

X = dataset.iloc[:, [3, 4]].values

Our goal here is to reveal the clusters and help the marketing department formulate their strategies. For instance, we might subdivide the Customers in 5 distinct groups:

1. Medium Annual Income, Medium Spending Score
2. High Annual Income, Low Spending Score
3. Low Annual Income, Low Spending Score
4. Low Annual Income, High Spending Score
5. High Annual Income, High Spending Score

It's worthwhile to pay attention to the #2 Group (High Annual Income, Low Spending Score). If there's a sizable number of customers that fall under

this group, it could mean a huge opportunity for the mall. These customers have high Annual Income, and yet they're spending or using most of their money elsewhere (not in the Mall). If we could know that they're in sufficient numbers, the marketing department could formulate specific strategies to entice Cluster #2 to buy more from the Mall.

Although the number of clusters is often arbitrary, there are ways to find that optimal number. One such way is through the Elbow Method and WCSS (within-cluster sums of squares).
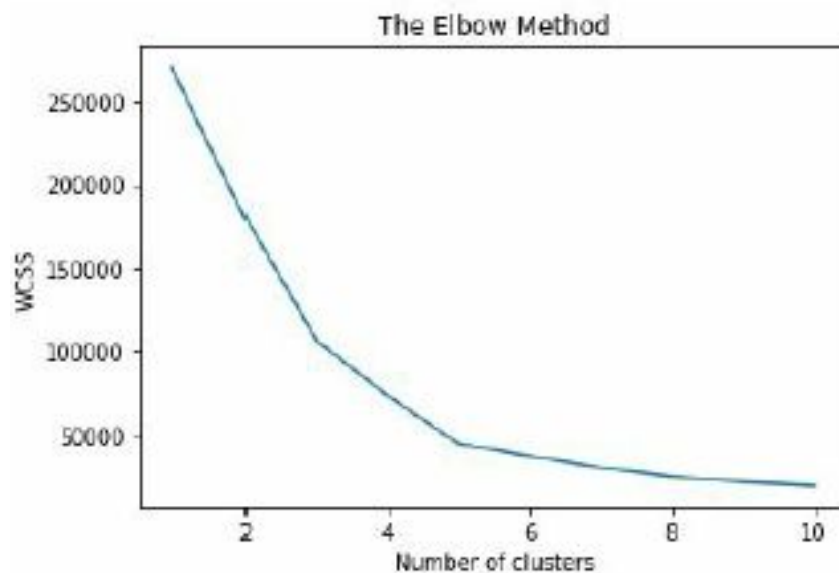
Notice that the "elbow" points at 5 (number of clusters). Coincidentally, this number was also the "desired" number of groups that will subdivide the dataset according to their Annual Income and Spending Score. After determining the optimal number of clusters, we can then proceed with applying K-Means to the dataset and then performing data visualization:

Clusters of customers
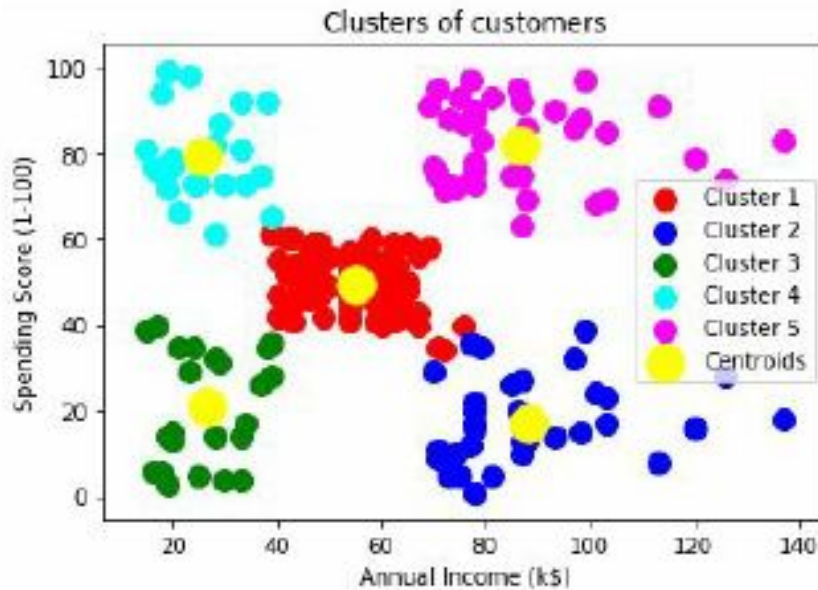
There we have it. We have 5 clusters and Cluster #2 (blue points, High Annual Income and Low Spending Score) is significant enough. It might be worthwhile for the marketing department to focus on that group. Also, notice the Centroids (the yellow points). This is a part of how K-Means clustering works. It's an iterative approach where random points are placed initially until they converge to a minimum (e.g., sum of distances is minimized). As mentioned earlier, it can all be arbitrary, and it may depend heavily on our judgment and possible application. We can set n_clusters into anything other

than 5. We only used the Elbow Method so we can have a sounder and consistent basis for the number of clusters. But it's still up to our judgment what should we use and if the results are good enough for our application.

Anomaly Detection

Aside from revealing the natural clusters, it's also a common case to see if there are obvious points that don't belong to those clusters. This is the heart of detecting anomalies or outliers in data. This is a crucial task because any large deviation from the normal can cause a catastrophe. Is a credit card transaction fraudulent? Is a login activity suspicious (you might be logging in from a totally different location or device)? Are the temperature and

pressure levels in a tank being maintained consistently (any outlier might cause explosions and operational halt)? Is a certain data point caused by wrong entry or measurement (e.g., perhaps inches were used instead of centimeters)? With straightforward data visualization, we can immediately see the outliers. We can then evaluate if these outliers present a major threat. We can also see and assess those outliers by referring to the mean and standard deviation. If a data point deviates by a standard deviation from the mean, it could be an anomaly. This is also where our domain expertise comes in. If there's an anomaly, how serious are the consequences? For instance, there might be thousands of purchase transactions happening in an online store every day. If we're too tight with our anomaly detection, many of those transactions will be rejected (which results in loss of sales and profits). On the other hand, if we're allowing much freedom in our anomaly detection, our system would approve more transactions. However, this might lead to complaints later and possibly loss of customers in the long term. Notice here that it's not all about algorithms, especially when we're dealing with business cases. Each field might require a different sensitivity level. There's always a tradeoff, and either of the options could be costly. It's a matter of testing and knowing if our system of detecting anomalies is sufficient for our application.

Chapter 12: Association Rule Learning

This is a continuation of Unsupervised Learning. In the previous chapters, we've discovered natural patterns and aggregates. There was not much supervision and guidance on how the "correct answers" should look like. We've allowed the algorithms to discover and study the data. As a result, we're able to gain insights from the data that we can use. In this chapter, we'll focus on Association Rule Learning. The goal here is to discover how items are "related" or associated with one another. This can be very useful in determining which products should be placed together in grocery stores. For instance, many customers might always be buying bread and milk together. We can then rearrange some shelves and products so the bread and milk will be near to each other. This can also be a good way to recommend related products to customers. For example, many customers might be buying diapers online and then purchasing books about parenting later. These two products have strong associations because they mark the customer's life transition (having a baby). Also, if we notice a demand surge in diapers, we might also get ready with parenting books. This is a good way to somehow forecast and prepare for future demands by buying supplies in advance. In grocery shopping or any business involved in retail and wholesale transactions, Association Rule Learning, can be very useful in optimization (encouraging customers to buy more products) and matching supply with demand (e.g., sales improvement in one product also signals the same thing to another related product).

Explanation

So how do we determine the "level of relatedness" of items to one another and create useful groups out of it.? One straightforward approach is by counting the transactions that involve a particular set.

Transactions	Purchases

1	Egg, ham, hotdog
2	Egg, ham, milk
3	Egg, apple, onion
4	Beer, milk, juice

Our target set is {Egg, ham}. Notice that this combination of purchases occurred in 2 transactions (Transactions 1 and 2). In other words, this combination happened 50% of the time. It's a simple example, but if we're studying 10,000 transactions and 50% is still the case, of course, there's a strong association between egg and ham. We might then realize that it's worthwhile to put eggs and hams together (or offer them in a bundle) to make our customers' lives easier (while we also make more sales). The higher the percentage of our target set in the total transactions, the better. Or, if the percentage still falls under our arbitrary threshold (e.g., 30%, 20%), we could still pay attention to a particular set and make adjustments to our products and offers. Aside from calculating the actual percentage, another way to know how "popular" an itemset is by working on probabilities. For example, how likely is product X to appear with product Y? If there's a high probability, we can somehow say that the two products are closely related. Those are ways of estimating the "relatedness" or level of association between two products. One or a combination of approaches might be already enough for certain applications. Perhaps working on probabilities yields better results. Or, prioritizing a very popular itemset (high percentage of occurrence) results to more transactions. In the end, it might be about testing different approaches (and combinations of products) and then seeing which one yields the optimal results. It might be even the case that a combination of two products with very low relatedness allows for more purchases to happen.

Apriori

Whichever is the case, let's explore how it all applies to the real world. Let's call the problem "Market Basket Optimization." Our goal here is to generate a list of sets (product sets) and their corresponding level of relatedness or support to one another. Here's a peek of the dataset to give you a better idea:

shrimp,almonds,avocado,vegetables mix,green grapes,whole wheat flour,yams,cottage
cheese,energy drink,tomato juice,low fat yogurt,green tea,honey,salad,mineral
water,salmon,antioxidant juice,frozen smoothie,spinach,olive oil
burgers,meatballs,eggs
chutney
turkey,avocado
mineral water,milk,energy bar,whole wheat rice,green tea
low-fat yogurt
whole-wheat pasta,french fries
soup,light cream,shallot
frozen vegetables,spaghetti,green tea
french fries

Those are listed according to the transactions where they appear. For example, in the first transaction, the customer bought different things (from shrimp to olive oil). In the second transaction, the customer bought burgers, meatballs, and eggs.

As before, let's import the necessary library/libraries so that we can work on the data:

import Pandas as pd
dataset = pd.read_csv('Market_Basket_Optimisation.csv', header = None)

Next is we add the items in a list so that we can work on them much easier. We can accomplish this by initializing an empty list and then running a for loop (still remember how to do all these?):

transactions = []
for i in range(0, 7501):
transactions.append([str(dataset.values[i,j]) for j in range(0, 20)])

After we've done that, we should then generate a list of "related products" with their corresponding level of support or relatedness. One way to accomplish this is by the implementation of the Apriori algorithm (for association rule learning). Thankfully, we don't have to write anything from scratch.

We can use Apyori, which is a simple implementation of the Apriori algorithm.

It's prebuilt for us and almost ready for our own usage. It's similar to how we use scikit-learn, Pandas, and numpy. Instead of starting from scratch, we already have blocks of code we can simply implement. Take note that coding everything from scratch is time-consuming and technically challenging. To implement Apyori, we can import it similarly as how we import other libraries:

from apyori import apriori

Next is we set up the rules (the levels of minimum relatedness) so we can somehow generate a useful list of related items. That's because almost any two items might have some level of relatedness. The objective here is to include only the list that could be useful for us.

rules = apriori(transactions, min_support = 0.003, min_confidence = 0.2,

min_lift = 3, min_length = 2)

Well that's the implementation of Apriori using Apyori. The next step is to generate and view the results. We can accomplish this using the following block of code:

```
results = list(rules)
results_list = []
for i in range(0, len(results)):
results_list.append('RULE:\t' + str(results[i][0]) + '\nSUPPORT:\t' + str(results[i][1]))
print (results_list)
```

When you run all the code in Jupyter Notebook. It will be messy and almost incomprehensible. But if you run it in Spyder (another useful data science package included in Anaconda installation), the result will look a bit neater: Notice that there are different itemsets with their corresponding "Support." The higher the Support, we can somehow say that the higher the relatedness. For instance, light cream and chicken often go together because people might be using the two to cook something. Another example is in the itemset with an index of 5 (tomato sauce and ground beef). These two items might always go together in the grocery bag because they're also used to prepare a meal or a recipe.

This is only an introduction to Association Rule Learning. The goal here was to explore the potential applications of it to real-world scenarios such as market basket optimization. There are other, more sophisticated ways to do this. But in general, it's about determining the level of relatedness among the items and then evaluating that if it's useful or good enough.

Chapter 13: Reinforcement Learning

Notice that in the previous chapters, the focus is on working on past information and then deriving insights from it. In other words, we're much focused on the past than on the present and future. But for data science and machine learning to become truly useful, the algorithms and systems should work on real-time situations. For instance, we require systems that learn real-time and adjust accordingly to maximize the rewards.

What is Reinforcement Learning?

This is where Reinforcement Learning (RL) comes in. In a nutshell, RL is about reinforcing the correct or desired behaviors as time passes. A reward for every correct behavior and a punishment otherwise. Recently RL was implemented to beat world champions at the game of Go and successfully play various Atari video games (although Reinforcement Learning there was more sophisticated and incorporated deep learning). As the system learns from reinforcement, it was able to achieve a goal or maximize the reward.

One simple example is in the optimization of click-through rates (CTR) of online ads. Perhaps you have 10 ads that essentially say the same thing (maybe the words and designs are slightly different from one another). At first, you want to know which ad performs best and yields the highest CTR. After all, more clicks could mean more prospects and customers for your business. But if you want to maximize the CTR, why not perform the adjustments as the ads are being run? In other words, don't wait for your entire ad budget to run out before knowing which one performed best. Instead, find out which ads are performing best while they're being run. Make adjustments early on, so later, only the highest-performing ads will be shown to the prospects.

It's very similar to a famous problem in probability theory about the multiarmed bandit problem. Let's say you have a limited resource (e.g., advertising budget) and some choices (10 ad variants). How will you allocate your resource among those choices so you can maximize your gain (e.g., optimal CTR)?

First, you have to "explore" and try the ads one by one. Of course, if you're seeing that Ad 1 performs unusually well, you'll "exploit" it and run it for the rest of the campaign. You don't need to waste your money on underperforming ads. Stick to the winner and continuously exploit its performance. There's one catch, though. Early on, Ad 1 might be performing well, so we're tempted to use it again and again. But what if Ad 2 catches up and if we let things unfold Ad 2 will produce higher gains?

We'll never know because the performance of Ad 1 was already exploited. There will always be tradeoffs in many data analysis and machine learning projects. That's why it's always recommended to set performance targets beforehand instead of wondering about the what-ifs later. Even in the most sophisticated techniques and algorithms, tradeoffs and constraints are always
there.

Comparison with Supervised & Unsupervised Learning

Notice that the definition of Reinforcement Learning doesn't exactly fit under either Supervised or Unsupervised Learning. Remember that Supervised Learning is about learning through supervision and training. On the other hand, Unsupervised Learning is actually revealing or discovering insights from unstructured data (no supervision, no labels).

One key difference compared to RL is in maximizing the set reward, learning from user interaction, and the ability to update itself in real-time. Remember that RL is first about exploring and exploiting. In contrast, both Supervised and Unsupervised Learning can be more about passively learning from historical data (not real-time). There's a fine boundary among the 3 because all of them are still concerned about optimization in one way or another. Whichever is the case, all 3 have useful applications in both scientific and business settings.

Applying Reinforcement Learning

RL is particularly useful in many business scenarios, such as optimizing click through rates. How can we maximize the number of clicks for a headline? Take note that news stories often have limited lifespans in terms of their relevance and popularity. Given that limited resource (time), how can we immediately show the best performing headline?

This is also the case in maximizing the CTR of online ads. We have a limited ad budget, and we want to get the most out of it. Let's explore an example (using the data from Ads_CTR_Optimisation.csv) to better illustrate the idea: As usual, we first import the necessary libraries so that we can work on our data (and also for data visualization).

```
import matplotlib.pyplot as plt
import Pandas as pd
%matplotlib inline #so plots can show in our Jupyter Notebook
We then import the dataset and take a peek
dataset = pd.read_csv('Ads_CTR_Optimisation.csv')
dataset.head(10)
```

In each round, the ads are displayed, and it's indicated which one/ones were clicked (0 if not clicked, 1 if clicked). As discussed earlier, the goal is to explore first, pick the winner, and then exploit it. One popular way to

achieve this is by Thompson Sampling. Simply, it addresses the exploration-exploitation dilemma (trying to achieve a balance) by sampling or trying promising actions while ignoring or discarding actions that are likely to underperform. The algorithm works on probabilities and this can be expressed in code.

When we run the code and visualize:
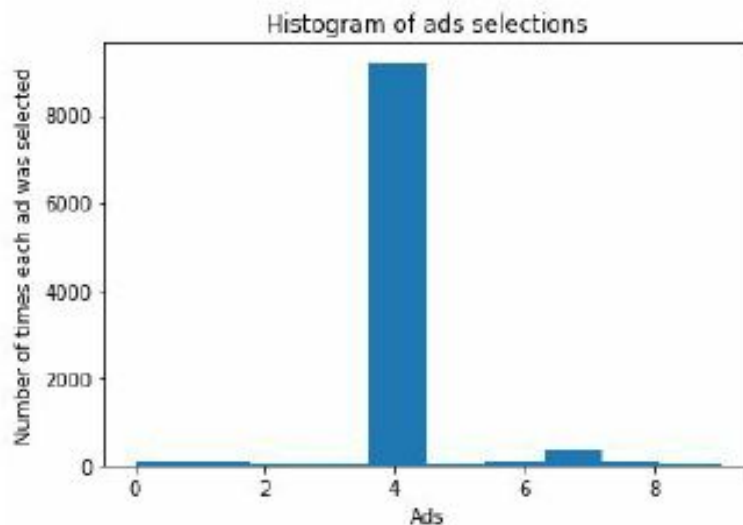
Notice that the implementation of Thompson sampling can be very complex.

It's an interesting algorithm which is widely popular in online ad optimization, news article recommendation, product assortment, and other business applications.

There are other interesting algorithms and heuristics such as Upper Confidence Bound. The goal is to earn while learning. Instead of later analysis, our algorithm can perform and adjust in real-time. We're hoping to maximize the reward by trying to balance the tradeoff between exploration and exploitation (maximize immediate performance or "learn more" to improve future performance). It's an interesting topic itself, and if you want to dig deeper, you can read more online resources.

Chapter 14: Ensemble Methods

This chapter will deal with ensemble methods. Anytime you experience uncertainty while you make real-time decisions, do you look for a friend to help you? Well, the fact is that we make decisions depending on the collective knowledge that we find from friends. Ensemble uses the same concept in ML.

The idea behind is to develop more data models and generate the final prediction depending on the prediction from all the individual models. For a regression problem, the end result can be the mean prediction value from all the models. In a classification environment, a lot of the vote is selected to determine the output class.

The popular idea is to have numerous models, each model presenting slightly different outputs on the training dataset. Some models learn the properties of the data better compared to others. The belief that the final result from all these models should be better than the result generated by anyone of them.

As already said, the notion behind the ensemble is to integrate different models together. These models can be of the same kind or different. For example, you can merge a neural network model output by applying a Bayesian model. This chapter will be restricted only to ensemble of the same type of models.

 Merging the same type of models is normally common in Data science society through bagging and boosting techniques.

The bootstrap aggregation is a perfect approach for creating a lot of models and combining their output to generate the last prediction. Every model within a bagging ensemble uses only a single section of the training data. The idea behind bagging is to cut down data overfitting. As previously said, you want every model to be unique from the rest. So, you need to sample the data using replacement for training every model and thus introduce variability. Another approach used to introduce variation within the model

is to sample the properties. You don't submit all the features to the model, but different models attain specific properties.

Bagging can easily be parallelized. According to the parallel processing framework available, models can be generated in parallel using different samples of the training dataset. The bagging method doesn't work with linear predictors like linear regression.

Boosting is part of an ensemble technique that produces various complex models. It works sequentially by training the newer models based on the errors in the original models. Every model trained is associated with a weight, which is calculated based on how better the model has done on a given data.

When the final prediction is performed, the above weights define the percentage of influence that a certain model has over the final result. Boosting doesn't result in parallelism like bagging. And this is because the models are built in a sequence. Thus, it is impossible to parallelize.

The errors presented by classifiers in the sequence are considered as hard instances to classify. The framework is created in a manner that the models come later in the sequence, or erroneous predictions are generated by the first predictor and attempt to improve them.

In general, the weak classifiers are used to boost, for example, a decision stump, which describes a decision tree using a single splitting node and two leaves.

Mastering the Bagging Method

Ensemble methods belong to the committee-based learning family. If we can succeed in introducing variation in the dataset, it will lead to models with slight differences.

You take advantage of bootstrapping to fed to these models variation within the dataset. Bootstrapping is the means of randomly sampling a certain dataset for a specific number of cases, with or without replacement. For bagging, you need to leverage bootstrapping to generate. For instance, let us assume m has different datasets, and we create a model for each of them.

Finally, you take advantage of the output of all the models to produce the last prediction in case of regression issues.

For example, you bootstrap the data m times; you will get m models, that is, y, m values and the final prediction would include:

$$Y_{final(x)} = \frac{1}{m} \sum_{i=1}^{m} y_m(x)$$

For the case of classification problems, the final result is determined depending on voting. Therefore, you have one hundred models in your ensemble, and you have a two-class classification problem using class labels as {+1, -}.

Suppose more than 50 models predict the output as +1, you need to declare the prediction as +1.

Randomization is a unique approach by which variability can be applied in the model creation exercise. An example is to define randomly a collection of properties for each model within the ensemble.

An example is to choose a random subset of properties for each model in the ensemble. That way, different models will consist of different sets of properties. This approach is known as random subspace method.

When you apply stable models, bagging may not attain huge results. Bagging is important if the underlying classifier is highly sensitive to even small changes to the data. For example, the decision trees, which are unstable. The unpruned decision trees are the best selection for bagging. However, you need to consider a Nearest Neighbor Classifier. K is a stable model. But you can still leverage the random subspaces, and bring some instability into the nearest neighbor methods.

In the following recipe, you will realize how to leverage bagging and random subspaces on a K-Nearest Neighbor algorithm. You will acquire a classification problem, and the final prediction will rely on the majority of the voting.

How to Do It

Import the relevant libraries, and write a function get_data () to generate a dataset to work.

Then proceed to define three functions:
- Function build_single_model to build a simple KNearest neighbor model with the provided data.
- Function build_bagging_model, -this function implements the Bagging routine.
- Function view_model to inspect the model you have developed.

Finally, you will write the main function, which will call the remaining functions.

Conclusion

Well, you don't need to be proficient in Python to conduct data analysis in Python. All you need to do is to master five Python libraries to effectively find a solution to a wide array of data analysis problems. So, you need to start learning these libraries one by one. Remember that you don't have to be a pro at building great software in Python to productively conduct data analysis.

Before you proceed further, make sure that you set up your programming environment and learn the way to work with Ipython notebook.

Then you can get started with learning Pandas profiling. This tool will help you achieve different animation on your project.
Using this tool is quite simple, just like installing and importing the Pandas profiling package.

Next, you need to learn how to plot Pandas data with the help of Plotly and Cufflinks. Experienced data scientists will be familiar with the use of matplotlib and pandas. So, it is important that you become familiar with the above tools.

Besides that, you need to also master the IPython magic commands. These are improvements that IPython provides on top of the standard Python syntax. The magic commands appear in two ways: line magics, which are defined using a single % prefix, and work on a single line of input, which is represented by a double %% prefix and work on numerous lines of input.
Additionally, you will need to learn how to create fancy features using Jupyter. This is a great tool that will support some HTML / CSS formatting in the markdown cells.

Don't forget, Numpy. This is another powerful Python package useful for scientific calculation. Having the correct understanding of Numpy will allow you to use tools such as Pandas. Remember, Pandas feature advanced data structures and manipulation tools to simplify data analysis.

That said, one thing that you should avoid doing is trying to learn every tool and library in Python at the same time. When you attempt to learn everything at once, you will spend a lot of time-shifting between different concepts, properties, and getting frustrated, and switching to something else.

So, you should try to concentrate on the following basic steps:

- Master python basics
- Master Numpy
- Learn Pandas
- Understand Matplotlib

From here, you can continue to expand on more topics.

BOOK 2: Hacking With Kali Linux

Practical Guide to Computer Network Hacking, Encryption, Cybersecurity, Penetration Testing for Beginners. The Secrets of VPN Services, Firewalls and the Linux Command Line

Download the Audio Book Version of This Book for FREE

If you love listening to audio books on-the-go, I have great news for you. You can download the audio book version of this book for **FREE** just by signing up for a **FREE** 30-day audible trial! See below for more details!

Audible Trial Benefits

As an audible customer, you will receive the below benefits with your 30-day free trial:

- FREE audible book copy of this book

- After the trial, you will get 1 credit each month to use on any audiobook

- Your credits automatically roll over to the next month if you don't use them

- Choose from Audible's 200.000 + titles

- Listen anywhere with the Audible app across multiple devices

- Make easy, no-hassle exchanges of any audiobook you don't love

- Keep your audiobooks forever, even if you cancel your membership

- And much more

Click the links below to get started!

[>> For Audible US](#)

[>> For Audible UK](#)

[>> For Audible FR](#)

[>> For Audible DE](#)

Introduction

I am very much delighted to see that you all have shown so much interest in learning about the basics and usefulness of Kali Linux. Kali Linux is one of the most effective software of today. It can also be regarded as a boon for all computing and networking people.

Kali Linux does the job of a security auditing software and it also helps in various aspects of networking and hacking. Kali Linux comes along with various information and security-related tasks such as reverse engineering, penetration testing and security research. Computer forensics is also a part of Kali Linux. Each and every service which is provided by Kali Linux is certified and comes along with all-over control along with wider aspects of accreditations.

Kali Linux belongs to the family of Linux distribution. Cyber security is the prime concern of this Linux distribution. Many of the companies today take help of Kali Linux for checking and tracing out their vulnerabilities for ensuring 100% security of their infrastructure. It is an open-source program and is thus totally free. Not only that but it is completely legal and can be used for various scenarios in an enterprise or organization.

Chapter 1: Basics of hacking

Hacking is nothing but unauthorized intrusion within a network or computer which is executed by attackers known as hackers. The attackers try to attack those systems which are vulnerable to threats. They keep their prying eyes open all the time, searching around for vulnerabilities. They can act as an individual or even work in a group. Not only might that but the hackers also function as a part of an organization which works with the motive of disrupting the functionalities of other organizations. Most of the time they try to alter the system of an organization and target the security infrastructure for breaching of information and gaining access. However, hackers not only work as attackers but also use their skills for finding out the weak spots along with the various vulnerabilities within a system. This is also carried out for finding and mending the weaknesses for preventing all forms of malicious attacks from entering the system.

Different Types of Hackers

There are various types of hackers in the world of hacking which perform different types of functions. The types of hackers help in defining the relationship between the systems and hackers which are trying to attack. The most common types of hackers are:

- **Black Hat Hackers:** The term black hat had its origin from the old Western movies in which the villains used to wear black hats. The black hat hackers act as individuals who try to have unauthorized access into the system of an organization or network for the purpose of exploiting the security infrastructure for various malicious reasons. The hackers of this type do not come with any sort of authority or permission for compromising the targets. They attempt to do damage by compromising the infrastructure of the security systems, shutting down the systems or also by altering the primary functions of a website or network. The primary intention of the black hat hackers is to gain all-over access or steal the information regarding finances, access various passwords or gain insights into other forms of personal data.
- **White Hat Hackers:** The white hat hackers are the second type of hackers but they act as the good guys. The white hat hackers work with various organizations for the purpose of strengthening the security of any system. The white hat hackers come with all sorts of permissions for engaging the targets and also compromise the same within the provided boundary of rules. The white hat hackers are also known as ethical hackers. The ethical hackers specialize in this field with various forms of ethical tools and techniques meant for hacking. They use special methodologies for securing up the information system of an organization. Contrary to the black hat hackers, the ethical hackers exploit the security system of a network and then check out for the backdoors after being legally permitted to perform

so. The ethical hackers always point out all forms of vulnerabilities that they dig out from the systems of the organizations to make sure that the gaps are mended for preventing exploitation by the malicious attackers.

- **Grey Hat Hackers:** The grey hat hackers gain access to the security systems of the organizations and networks in the same way just like black hat hackers do. But the grey hat hackers perform such actions without any form of malicious intent and disclose the vulnerabilities along with the loopholes to the agencies of law enforcement or various intelligence agencies. The grey hat hackers generally surf the internet and hack the computer systems for notifying the owners or the administrator of the network or system which contains various vulnerabilities which need to be mended immediately. The grey hat hackers might also extort the hacked systems by offering to inform about the defects for some fees too.

Common Tools of Hacking

For accomplishing the act of hacking, the hackers implement various types of techniques. Let's have a look at some of them.

- **Rootkits:** Rootkit acts like a program or a huge set of software which allows the attackers to gain complete access or control of a system or network which directly connects or interacts with the system of the internet. Rootkit was first introduced as a system of backdoor process for fixing various issues in regards to software. However, today this software is widely being used by the hackers for disrupting the functionality and control of a system from its actual owner or administrators. There are various ways in which rootkits can be installed in the system of the victim. The most common way of installing rootkit is by implementing phishing attacks along with social engineering. Once the rootkits have been installed in the system of the victim, the attacker gains access to the system secretly and

controls the overall functioning with which they can easily steal confidential data and information and can also shut down a system completely.

- **Keyloggers:** This is a very special type of tool which has been designed for recording and logging each and every key pressed on the victim system. The keyloggers record the stroke of the keys by staying attached to the Application Programming Interface or API. It tracks the key strokes when anything is being typed by using the keyboard in a system. The files which are recorded are then saved which contains various forms of information such as details regarding website visit, usernames, the record of opened applications, screenshots, bank details and many more. The keyloggers are also capable of capturing the personal messages, credit card details, passwords, mobile numbers and various other details which are generally typed in a system. The keyloggers generally arrive as a malware which allows the cybercriminals to breach all forms of sensitive data.

- **Vulnerability scanner:** A vulnerability scanner is used for the purpose of classifying and then detecting various forms of weaknesses in a system, network, communication system, computers etc. This is one of the most common forms of tool which is being used by the ethical hackers for finding out the potential vulnerabilities and loopholes and then fixes them up on urgent basis. However, a vulnerability scanner can also be used by the black hat hackers for checking the vulnerabilities and weak spots within a system and then finding out the proper tool for exploiting the same.

Techniques of Hacking

There are various techniques which are being used by the hackers for exploiting a system.

- **SQL Injection:** SQL or structured query language has been designed for the purpose of exploiting various forms of data in the database of the victim. This form of attack falls under the cyber attack which targets the databases via the statements of SQL for tricking the systems. This form of attack is generally carried out by the use of website interface which attempts in issuing the commands of SQL through a database for hacking the passwords, usernames and other related information related to the database. All those websites along with web applications which are coded poorly are very much prone to the SQL injection attacks. This is because the applications which are based on the web contains various user input fields like login pages, search pages, request forms related to support and products, comments section and many others which are very much susceptible to the attacks and can be very easily hacked by simple manipulation of the codes.

- **DDoS or Distributed Denial of Service:** It is a form of hacking attack in which the normal traffic of a server is distorted from entering the server and floods the traffic of the network. This ultimately results in denial of service as it serves just like a traffic jam which clogs the roads and also prevents the regular form of traffic from reaching the destination. All the devices of today such as IoT devices, computers, mobile phones etc. which connects with the network are very much prone to the attacks of DDoS.

- **MAC Spoofing:** Each and every form of device which are used by the people today come with network interface controller or NIC. It helps the users to connect with the network such as with the internet directly. The NIC of each device is accompanied with a MAC address which is assigned after various processes of hard coding. The MAC spoofing attack is a very deadly form of attack in which the hackers hide themselves and their system behind a customized and false MAC address. This reduces the risks on the part of the hackers from getting caught. So, you might give access to a new system thinking of it to be absolutely legitimate but it might happen that a hacker will hide himself behind a false MAC address which you cannot even realize.

By using this technique, the hackers can easily hack internet connection via Wi-Fi and can also gain access to all those devices which are connected to each other via LAN. The technique of MAC spoofing also leads to several forms of other serious crimes in which the hackers steal the identity of someone else and carries on with some serious form of data breaching in which someone will be held as guilty without even knowing about the actual hacker. However, there are various OS in the market today such as MAC and Windows which can easily connect with the LAN without using the MAC address.

Chapter 2: What is Ethical Hacking?

Ethical hacking is also called as intrusion testing, penetration testing and also red teaming. In simple words, it is the controversial technique of finding out vulnerabilities and weaknesses in a system simply by imitating the actions and intent of the malicious hackers. An ethical hacker is a person or security professional who uses his skills for the purpose of various defensive measures on part of the administrators of an information system. An ethical hacker is also known as a white hat or white hat hacker. By conducting various tests, an ethical hacker tries to find out the answers to the following questions:

- What are the locations, systems or information can the attacker gain access?
- What will the attacker see before setting his target?

- What will the attacker do with the information which is available in the system?
- Is anyone able to notice the various attempts made by the attacker to gain access?

The ethical hacker who has been given the job of penetration testing operates on the permission along with the knowledge of that organization for which he has been assigned the job of defense. There are various cases in which an organization will not be informing the security information team about all the activities which is going to be carried out by the ethical hacker just for testing the effectiveness and concise of the security information team. This whole thing is also known as double blind environment. For the purpose of effective and legal operation, the organization needs to inform an ethical hacker about all those assets and information which are meant to be protected, the potential sources of threats and the limit to which the organization will be supporting the efforts of the ethical hacker.

Process of ethical hacking
All the ethical hackers follow a strict process in order to get the best usable and to the point legal results. Let's have a look at the processes which are followed by the ethical hackers.

Planning
No matter what kind of project it is, for every successful project planning is of utmost importance. It provides the ethical hackers with the opportunity of thinking about what are the things that need to be done, set the goals which are to be reached and also for the assessments of risks for evaluating how to carry out a complete project. There are various factors which are considered by the ethical hackers before carrying out a project of ethical hacking. The list of factors includes culture, policies of security, laws, regulations, requirements of the industry and best practices. All of these

factors play an important role in the process of decision making when it comes to the initiation of ethical hacking.

The phase of planning in ethical hacking will be having an overall influence on how the process of hacking is being performed, the information which is collected and shared and will also be directly influencing the integration and delivery of the results into the program of security. The planning phase is the very first step and will be describing most of the details about the controlled attack of hacking. It will also be answering all forms of questions regarding hacking such as how the process of ethical hacking is going to be controlled and supported, what are the basic actions which needs to be performed and for how long will the process go on.

Reconnaissance

It is the process of searching for all those information which are freely available for assisting in the process of attack. This whole process can be as easy and simple as just using a ping or browsing the various newsgroups which are available on the internet for searching that information which is leaked by the employees or as tough and messy as digging through a huge trash of letter or receipts. This process can also include several other processes such as phone tapping, social engineering, network tapping and also data theft. The process of information searching will be limited only to the extent to which the organization and the ethical hacker will want to go for the purpose of recovering all the required information which they are looking out for.

The phase of reconnaissance introduces the deep relationship in between the tasks which needs to be completed and all those methods which will be needed for protecting the information and assets of the organization.

Enumeration

It is also known as vulnerability or network discovery. Enumeration is the process of obtaining all those information which is available readily from the system of the target, networks and application which are used by the target. It is also to be noted that the phase of enumeration is the actual point

where the thin line between malicious attacks and ethical hacking gets blurred very often as it is very easy and simple to go outside the dedicated boundaries which have been outlined in the original plan of attack. For the purpose of creating a clear picture of the environment of an organization, various techniques and tools are being used which are readily available. These available tools include NMap and port scanning. However, it is very easy to collect all the required information, it is very difficult to make sure of the value of information which is available in the hands of the hacker.

At the very first glance, the process of enumeration seems to be very simple in which data is collected then evaluated collectively for establishing a proper plan for more searching or building up a detailed matrix for the analysis or vulnerability phase. However, this phase is the actual phase in which the ability of ethical hacker in taking logical decisions plays a very important role.

Analysis of vulnerability

For the purpose of effectively analyzing all the data, an ethical hacker needs to employ a pragmatic approach which is logical in nature as well. In the phase of vulnerability analysis, all the information which has been collected is compared with all the known forms of vulnerabilities in the practical process. Any form of information is useful in the process, no matter from where it originates or what the source is. A small pinch of information can also help in finding out some new sort of vulnerability in the system and might also lead to several other discoveries of vulnerabilities which have not been found yet. The known form of vulnerabilities, service packs, incidents, updates along with various hacker tools helps in properly identifying the point of attack. The internet provides the ethical hackers with a huge amount of information which can be associated very easily with the system architecture along with weak and strong points in a system.

Exploitation

A considerable amount of time is spent for the purpose of evaluating and planning an ethical hack. It is very obvious that all of these planning will lead to some sort of attack. The level of exploitation of a system can be as

simple as running a very small tool in the system or as tough as a collection of many complex steps which needs to be executed in a proper way for gaining access to the system. The process of exploitation can be broken into a collection of subtasks which can be either one single step or a collection of various steps. As each and every step is performed, a process of evaluation takes place which ensures that the outcome which has been expected is met. Any form of divergence from the plan of attack can be graded into two parts:

- **Expectation:** Are the expected results of exploitation met or the results are conflicting with the assumptions of the organization?

- **Technical:** Is the target system behaving in a manner which is not at all expected, which is actually having an impact on the system exploitation and the system engagement in total?

Final analysis
Although the phase of exploitation comes with a huge number of validations and checks for ensuring the success of the hack, one last final analysis is needed for categorizing the system vulnerabilities in accordance to the exposure level and also for assisting in the drawing up of a plan for mitigation. The phase of final analysis links up the exploitation phase and the deliverable creation. A comprehensive image of the complete attack is needed for the construction of a bigger size picture of the current posture of the security environment of an organization and also for expressing the vulnerabilities clearly.

Deliverables
Deliverable communicates with the test results in a variety of ways. Some of the deliverables are concise and short in nature which only provides the vulnerabilities list along with the ways in which it can be mended whereas, the other form of deliverables can be detailed and long which will provide a

list of the probable vulnerabilities in a system which comes with the description regarding how the vulnerabilities were found, how they can be exploited, the results of having such vulnerabilities within the system and how to fix the situation. This phase is actually used by an ethical hacker in conveying his hack results to the organization. It can also be the case if the deliverables do not actually frighten the administrators, the test is considered as a fail.

Chapter 3: Cyber Security

In this world of today where technological innovations are taking place every day, the potential threats of cyber attacks are also increasing in equal pace. Cyber security plays a deep role in securing the information and data of the systems and networks in today's world of vulnerability. Cyber security is nothing but the employment of various tools and technologies for the purpose of securing the networks, programs, system data and network from the potential attacks, damages and various forms of unauthorized access. Cyber security is also known as security of information technology.

Cyber security and its importance

Most of the organizations and institutions such as military, government, medical along with financial bodies stores up an accountable amount of data on the systems of computers along with databases which can be found online. In most of the cases, the information which is being stored up in the

servers and databases are highly sensitive in nature, leakage of which can result in serious troubles for the concerned organization. Unauthorized access to the systems of the organizations along with the database can lead to data breaching along with the exploitation of the security infrastructure of an organization.

The organizations which are targeted might lose up all forms of sensitive data along with complete loss of access to the systems. As the volume of cyber attacks is increasing day by day, the organizations especially those which are concerned with national health and security are required to take some serious steps for safeguarding all forms of sensitive data. Cyber security is the ultimate option which can help an organization in protecting all its data and servers.

Cyber Security & Encryption

Encryption is the process of encoding communication in such a way so that only the authorized parties can encode the message of communication. It is done by using SSL/TLS and PKI protocols. The very reason why is it important so much stems from the process in which the internet was built up by using the protocol of HTTP. Hypertext Transfer Protocol or HTTP is of the same age that of the internet. HTTP is the protocol of communication which allows the servers in the web and the web browsers for communicating and displaying the information in a proper intended way. When a user visits a website, it is not actually the way it looks in the browser. Websites are built up of a bunch of codes which are sent to the web browsers which are then visually arranged by the browser in the way the web designer intended to do.

The main problem of HTTP is that it is not at all secure. So, any person who knows the process can easily spy on the connections of HTTP on the internet. In simple words, a third party can easily read along with manipulate a communication over HTTP between the clients and the servers. Encryption is the technique that actually comes into play in taking

care of the communication by serving the websites over the protocol of HTTPS. HTTPS is the secured version of HTTP. All the connections which are built over HTTPS are encrypted in nature. In simple terms, any form of communication over the protocol of HTTPS is highly secure. Encryption prevents spying on communication by the third parties. In case you are related with online business and you need to take the financial as well as personal details of the customers, make sure that your website is encrypted so that your customers are not at risk at the time of details exchange.

How does the process of encryption work?

The process of encryption begins when the web browser reaches one website which comes with an SSL certificate. The web server and the browser proceeds with what is known as SSL handshake. At the preliminary stages, the web browser verifies that the SSL certificate which is installed in the website is legitimate in nature and has been issued by a trustworthy authority of certification. After the web browser makes sure that the certificate is legitimate in nature, it starts to negotiate with the terms of the encrypted connection with the server.

When it comes to encryption, there are mainly two key pairs. The first is the asymmetric key pair which consists of the private and public keys. These keys have no function with the encryption bulk but they are used for authentication. When a web browser tests the authenticity of SSL certificate of a website, it makes sure that the certificate of SSL which is being questioned is actually the owner of the public form of key. It performs this by using up the public key for encrypting a small packet of data. If the web server is able to decrypt the data packet by using the respective private key and then send the packet back, it is proved that the server is the owner of the public key and everything is stated as verified. In case the web server fails to decrypt the data packet, the certificate of the server is taken as "not trusted".

The other key pair is the session keys. This form of keys is generated after the authenticity of the SSL certificate has been verified and all the terms regarding encryption have also been negotiated. While a public key can be used only for encrypting and a private key for decrypting, the session keys can be used for both the functions of encryption and decryption. The session keys are smaller in size and also less secure in nature when compared with the asymmetric form of counterparts. However, the session keys are strong enough for performing both the functions. The server and the web browser use the session keys for the rest of the communication. After leaving the site, the session keys which are being used are discarded and brand-new session keys are generated for the new visit.

Common Types of Cyber Attacks
Cyber attacks are increasing day by day with the innovations in the world of technology. There are various types of cyber attacks that can be found today where some are used most commonly such as phishing, malware, XSS and many more. Let's have a look at some of the most common types of cyber attacks.

Malware

Malware is a form of harmful software which is used for gaining access to the systems of the victims. The malware can also be called as viruses. Once a malware enters the victim system, it can lead to havoc starting from gaining overall control of the system to the monitoring of all sorts of actions, stealing sensitive data silently and also can lead to a complete shutdown of the system. The attackers use various ways for inserting malware in the target system. But there are also various cases in which the system users are being tricked into installing a malware in the system.

Phishing

Receiving emails with various unwanted links and attachments is a very common thing today. Such action of sending out harmful links and attachments via email is known as phishing. In phishing attacks, the attackers send out emails to the targets which seem like a trustable email. Most of the emails come with links and attachments which when clicked leads to the installation of malware in the system without even the user of the system knowing nothing. Some of the phishing links can also lead the users to a new website which might ask for confidential data such as bank and credit card details. Such websites are actually a trap which is used by the attackers for installing the malware in the target systems.

XSS

Cross-site scripting or XSS attack is used for targeting the users of a website directly. It is somewhat similar to the SQL injection attack and also involves injecting harmful codes in a website. But, in the case of XSS attacks, the websites are not attacked. In an XSS attack, the malicious code which has been injected in the website runs only in the browser of the user and can be used for stealing sensitive data such as username, password, bank details and many more.

Malware and Its Types

Malware is a form of malicious software which is being used for gaining access to the system of the victim. The cyber criminals design malware in a way which can be used for stealing data, compromising the functions of the computer, bypassing the access controls and many more.

Types of malware

There are various types of malware that can be found today. Let's have a look at them.

- **Adware:** Adware are those programs which are used for displaying advertisements on the websites which when clicked redirects to the website which is being advertised and also collects all forms of market data about the user. There are also various forms of pop-up adware that generally contains malicious links which can lead to harm of the system.
- **Spyware:** It is a software which is used for spying the target users. It has been designed for capturing and monitoring the activities of the users on the websites. Adware is also a form of spyware which sends out the activities of browsing of the users to the advertisers.
- **Worm:** Worm is a form of virus which is being used by the cybercriminals for the purpose of replicating themselves. Worms use computer networks for spreading and can lead to stealing or deletion of data. Many of the worms are also being designed for spreading only through the systems and do not lead to any form of harm to the systems.

Chapter 4: Linux Architecture
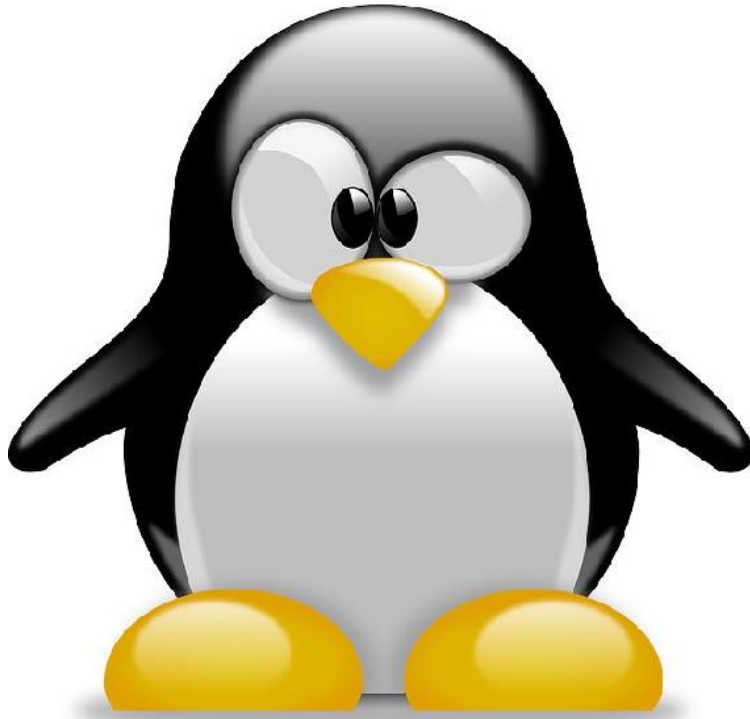
Linux is one of the finest operating systems which can be found today. It is open source in nature and is based on UNIX. It is just a simple OS like the commercial ones like Windows XP, Windows 10 and MAC OS. An OS is nothing but the graphical form of interface between the system of a computer and the user of the system. It comes with the responsibility of managing all the resources related to hardware that the system of a computer has and also helps in establishing communication in between the hardware and the software.

Open Source Software

An open source software is a software which has its source code available with the license with which the holder of copyright has the right to study

the software, change the settings and also distribute the same software with anyone he wants for any form of purpose.

Linux OS and its components

The Linux OS is composed of three different components.

- The Kernel
- The System Library
- The System Utility

The Kernel

The kernel functions as the core part of any form of OS. It is responsible for handling the tasks along with the hardware of the system of a computer. The CPU time and memory are the two examples of the entities which are being managed by the kernel. The kernel of an OS is of two types:

- **Microkernel:** The microkernel is a type of OS kernel. As its name goes by, it comes with a very basic form of functionality. It is the least amount of software which can provide with the environment which is required for the functioning of an operating system. This environment of kernel covers management of threads, low level management of address space and inter-process form of communication.

- **Monolithic kernel:** Monolithic kernel is the form of kernel which comes with various drivers along with it. It is an architecture of the operating system in which the operating system of a system works in the space of kernel. This form of the kernel is able to load or unload dynamically all the modules which are executable at the time of running. The monolithic form of kernel stays in the supervisor mode. The major point of difference between the micro kernel and the monolithic kernel

is that the monolithic form of kernel can alone define a very high level of interface over the hardware of the system of a computer.

Supervisor mode

The supervisor mode of the monolithic kernel is a flag which mediates from the hardware of a system. It can be easily modified by running the codes in the software system level. All form of system level tasks comes with this flag while they are operating or running. However, the applications of user space do not come with this flag set. The flag makes sure that whether the execution of machine code operations is possible or not such as performing various operations like disabling the interruptions or modifying the registers for various forms of descriptor table. The main idea behind having two different types of operation comes from the idea "with more amount of control come more responsibilities".

Any program in the supervisor mode is trusted so much that it will never fail as any form of failure will lead to crashing of the computer system. In simple words, the kernel is the component which is responsible for all form of activities of the OS. It is composed of various types of modules and also directly interacts with the base hardware. The kernel comes with all the necessary abstraction for the purpose of hiding all the low-level details of hardware to system or programs of application.

The System Library

The system library is composed of a collection of resources which are non-volatile in nature and are used up by the resources of the computer system and is mainly used for developing software. This comes with data configuration, help data, documentation, templates for messaging and many more. Generally, the term library is being used for describing a huge collection of implementations regarding behavior which is written down in terms of computer language. It comes with a perfectly defined form of interface which helps in invoking the behavior. So, this means that anyone

who wants to create a program of high level can easily use up the system library for the purpose of making system calls continuously.

The system library can be requested at a time by various individual forms of programs simultaneously, in order to make sure that the library has been coded in a way so that several programs can use up the library even when the concerned programs are not at all linked nor have a connection with each other. In simple terms, the system libraries are unique programs or form of functions built up of the system utilities or application programs which have access to all the features of the kernel. This form of library implements a majority of the functions related to the operating system of a computer and they are not required to have the rights of code access for the module of the kernel.

The System Utility

The programs of system utility are responsible for performing all forms of individual and specialized level tasks. The utility software is a form of system software. It has been designed for running the programs of application and hardware for a system of computer. The system software can also be considered as the interface between the applications of the users and the hardware. In simple words, the system utility software is the software of a system which has been designed for the purpose of configuring, analyzing, optimizing and maintaining a system of computer. The utility software works hand in hand with the operating system for supporting the infrastructure of a system, differentiating it from the software of application which is aimed for performing the various tasks directly which will be benefiting the normal users.

Characteristics of Linux architecture

Linux comes with various features that can help the regular users a lot.

Multiuser capability

This is the most unique characteristic of Linux OS in which the resources of a computer such as memory, hard disk etc. can be accessed by various users at a time. However, the users access the resources not from a single terminal. Each of the users is given an individual terminal for accessing the resources and operating them. A terminal consists of at least one VDU, mouse and keyboard as the devices for input. All the terminals are linked or connected with the primary server or Linux or with the host machine the resources of which and other peripheral devices like printer can be used by the users.

Multitasking
Linux OS comes with the capability of easily handling various jobs at a time. For example, a user can execute a command for the purpose of execution of a huge list and type in a notepad at the same time. This is intelligently managed by dividing the time of CPU by implementing the policies of scheduling along with the concept of switching of contexts.

Portability
Portability is the feature that made Linux OS so famous among the users. Portability does not mean at all that it can be carried around in CDs, pen drive or memory cars nor the size of the file is small. By portability, it means that the OS of Linux along with all its application can function on various types of hardware in the exact same way. The kernel of Linux and the application programs of the OS support the installation of the same on even those systems which comes with the least configuration of hardware.

Security
Security is considered as the most essential part of any operating system. It is really important for all those users and organizations who are using the system for various forms of confidential tasks. Linux OS comes with

various concepts of security for the purpose of protecting the users from any form of unauthorized access of the system and their data.

Main concepts of Linux security

Linux provides 3 main types of security concepts.

- **Authentication:** This helps in authenticating the user with the system by providing login names and password for the individual users so that their work cannot be accessed by any third party.
- **Authorization:** At the file level of Linux OS, it comes with limits of authorization for the users. There are write, read and execution permissions for every file which determines who all can access the files, who can modify the same and who all can execute the files.

- **Encryption:** This feature of Linux OS helps in encoding the user files into a format which is unreadable in format and is called cyphertext. This makes sure that even if someone becomes successful in opening up the system, the files will be safe.

Communication

Linux OS comes with a great feature for the purpose of communicating with the users. It can be either within the network of one single computer or in between two or more than two networks of a computer. The users of such systems can seamlessly exchange data, mail and programs through the networks.

Chapter 5: Basics of Linux Operating System

Linux is a simple operating system just like other operating systems such as Windows. As an OS, Linux helps in managing the hardware of a system and also provides services that the other software needs for running. It is regarded as a hands-on operating system. For example, if running an OS like Windows is like an automatic car, running Linux OS is like driving a stick. It might need some more work to do, but once the user gets a nice grip of the functioning of Linux, using the line of commands and also installing the packages will become super easy.

History of Linux

Linux is similar to the MAC OS X, which is also based on Unix. Unix was developed in the early 1970s with a primary goal of creating an OS which will turn out to be accessible and also secure at the same time for various users. In 1991, Linux was developed with the goal of distributing the features of Unix. It was launched as open-source software and till date, it is the same. Open source software is a software whose code is visible completely by the user and can also be modified according to need and can be redistributed. Linux is just the kernel and not a complete OS. The kernel provides for an interface between the hardware and requests from the user applications. The other part of the OS consists of utilities, GNU libraries and various other software. The OS as one complete unit is called as GNU/Linux.

A bit of servers

The Linode that the users have is a type of server. A server is nothing but a type of master computer which helps in providing various forms of service all over the network or across a connected network of computers. The servers are generally:

- Stays on all the time.
- It is generally connected with the internet or a group of computer networks.
- Consists of files and programs for the purpose of website hosting or for other content of the Internet.

As the server acts just like a computer, there are various similarities in between the Linode and the home computer. Some of the similarities are:

- The Linode is generally hosted on a physical form of machine. It sits on the available pool of data centers.
- Linodes uses up OS like Linux. It is another type of OS similar to Mac or Windows.
- Just like a user can easily install various applications in their PC, applications can be installed on Linode as well. All these applications which are installed on a Linode help in performing various tasks like hosting a website.
- A user can easily create, edit and delete files just like it can be done on a PC. The user can navigate through the directories as well just like PC.
- Just like a PC, Linodes are connected with the internet.

Things to consider before installing Linux

Before installing Linux, you need to make sure which distribution of Linux you want to install. Linux OS comes in various versions which are known as distributions. The distributions are similar to that of the versions of OS like Windows 7 or Windows XP. The new versions of operating systems like Windows are the upgraded versions. But, in case of Linux, the distributions are not upgraded but are of various flavors. Several distributions of Linux install various different software bundles.

Linux Distributions

The major difference between the distributions of Linux tends to be from the aspect of aims and goals of the distribution and which software bundles are installed rather than any form of difference in the Linux kernel code. RedHat Linux which consists of CentOS and Fedora and Debian Linux which consists of Ubuntu shares a huge amount of codes with one another. The kernels are more or less the same and the applications along with user utilities from the project of GNU are also similar. Some of the distributions of Linux have been designed to be as minimalistic and simple as possible whereas some has been designed having the current and the best software of the era. All the distributions of Linux aim at providing the best stability and reliability to the users.

In addition to the individual personality of distributions, you will also need to consider various factors which will help you at the time of choosing your desired distribution.

- **Release cycle:** The various distributions of Linux release the updates of their OS at different schedules. The distributions like Arch Linux and Gentoo uses a model of rolling release in which each individual package is released when they are declared as complete or ready by the developers. Distributions like Slackware, Debian and CentOS targets in providing the users with the most stable form of operating system which will be attainable as well and also releases the newer versions very frequently. Linux distributions such as Ubuntu and Fedora release its new versions after every six months. Selecting the release cycle which will be perfect for you also depends on various factors. The factors include the software that you require to run, the amount of reliability and stability that you require and the comfort level you are looking out for.

- **Organizational structure:** Although it might not directly affect the distribution performance, it is still one of the most distinguishing factors in between the Linux distributions. Some

of the Linux distributions like Gentoo, Debian, Slackware and Arch are all developed by the communities of independent developers while some of the other distributions such as Ubuntu, Fedora and OpenSUSE are developed by those communities which are being sponsored by different corporations. Distribution like CentOS is derived from the distributions which are produced commercially.

- **Common set of tools:** The various distributions of Linux uses different types of tools for performing various common tasks such as configuration of system or management of packages. Distributions like Ubuntu and Debian uses APT for managing the .deb packages, OpenSUSE uses .rpm package and CentOS along with Fedora also uses .rpm packages but manages all of them by using a tool known as yast. In most of the cases the distribution you choose will end up to that one distribution which comes with all the tools which you require and you are comfortable with.

The distributions are designed for performing in different situations. You are required to start with experimenting the distributions for finding out the one that fits you the best according to your need.

Linux security

When you start using a system based on Linux OS, you become the owner of your system security. The internet is filled up with people who are waiting to use the computing power of your system for satisfying their own goals. Linux offers the users with various security options that help the users in securing their system and tuning the same according to their need.

Finding your folders and files

Everything on a Linux system is in the form of a directory. In Linux, a folder is termed as a directory. Linux OS uses a well-balanced tree of various nested directories for keeping all the files in an organized manner. The directory of the highest level is known as the root directory. It comes designated with only one single slash. In Windows OS, you will come across various drives and disks. But this is not the case in Linux OS. There are several other sub-directories which lie under the root directory. Most of the systems based on Linux come with directories which are called as var and lib along with many others under the tree of the root directory.

The directory of lib consists of the system libraries whereas the directory of var consists of all sorts of files which are available in the system which are most likely to change like the mail messages and logs. The directories of Linux OS can also go inside the other directories.

Users and permissions

Linux OS uses a very powerful system for the users and its permissions for making sure that only the right people get access to the system files. As the owner of your Linux system, you can set the users and permissions for every directory. The file access system in Linux comprises of three categories.

- **Users:** A user account is assigned generally to a person or also to an application which requires access to the files in the system. You can provide user access to the system as many numbers you want.

- **Groups:** A group is the collection of one or more than one user. Groups are a great way of granting the same kind of access to various users at one time without the need for setting permissions for each individually. When you create an account of user, it gets assigned to a default group which comes with the same name as that of the name of the user. A user can be a part of as many groups as the user wants. Users who belong to a

group get all the permissions which are granted for that specific group.

- **Everyone:** This category is for everyone other than the groups and users. When someone accesses the system files without even logging in the system as one specific user, they fall into the category of everyone.

The next important thing that comes right after users is permissions. Each and every directory and file in a Linux system comes with three probable levels of access.

- **Read:** All the files that come with the permission of read can be viewed.
- **Write:** All the files that come with the permission of write can be edited.
- **Execute:** All the files that come with the permission of execute can be executed or run just like an application. When you start a new script or program, you start executing it.

Software installation in Linux

Like all the other things in the Linux system, software installation is also done by typing and then executing one specific form of text command. Most of the distributions in Linux come along with managers of package which makes it easier for installing or uninstalling any software in the system. Distributions such as Ubuntu and Debian use APT or the Advanced Packaging Tool package manager whereas CentOS and Fedora use YUM or Yellowdog Updater Modified manager of packages.

Chapter 6: Basic Linux Commands

Linux is one of the most famous operating systems that can be found today. However, Linux is not one complete OS, it is the kernel of an OS. Linux is also regarded as a clone of UNIX. Some of the most common distributions of Linux are Linux Mint, Ubuntu Linux, Red Hat Enterprise Linux, Fedora and Debian. Linux is primarily used in the servers. It can also be regarded that almost 90% of the internet is being powered by the servers of Linux. This is mainly because Linux is secure, fast and free as a kernel. Windows servers can also be used for the internet but the main problem that comes with Windows is its costing. This problem of costing can be easily solved by the servers of Linux. In fact, the operating system Android which runs in a majority of the smartphones today has also been made from the Linux kernel.

Linux shell

Linux shell is a form of program which receives the commands form the users of a system and transfers it to the operating system for the purpose of processing and then shows the result as well. The Linux shell is the main part of Linux OS. Its distributions come in graphical user interface or GUI but Linux basically comes with command line interface or CLI. For opening up the terminal of Linux shell, you need to press Ctrl+Alt+T in the Ubuntu distribution or you can also press Alt+F2, type in the gnome terminal and then hit enter.

Linux Commands

Let's start with some of the most basic commands of Linux.

pwd

When you open up the terminal first, you will be in the home directory of the user. For knowing exactly in which directory you are in, you can use the command pwd. It helps in giving out the exact path, the path which starts exactly from the root. The root is nothing but the base of the file system in Linux. It is generally denoted by using a forward slash (/). The directory of user generally looks like /home/username.

ls

By using the command ls, you can easily know what are the files within the directory in which you are in. You can also see each and every file which is hidden by using command ls –a.

cd

You can use the command cd for going to a directory. For example, if you are in the folder of home and you wish to go into the folder of downloads, you need to type cd Downloads and you will be in the downloads directory. You need to note that this command is very case sensitive. You are also required to type in the folder name exactly in the way it is in. However, this type of command comes with certain problems. For example, you are having a folder named as Raspberry Pi. In such case, when you enter the command as cd Raspberry Pi, the Linux shell will assume the second argument that comes with the command as a completely different entity and so what you will get in return is only an error message that will say that there is no such directory.

In such cases, you can use the backward slash which means use the command as cd Raspberry\Pi. The spaces are taken as : in Linux. If you type the command cd only and hit enter, you will get into the home directory again. In case you want to go back from a specific folder to a

folder just before that, you need to use "cd..". The two dots in the command represent the request of going back.

mkdir & rmdir

The mkdir command is being used for the purpose of creating a new folder or directory. For example, when you need to create a new directory such as DIY you need to enter command like mkdir DIY. Always remember that in case you want a directory named as DIY Hacking, you need to type it in as mkdir DIY\Hacking. You can use rmdir command for deleting the directory which you no longer need. However, always keep in mind that rmdir can only be used at the time of deleting a directory which is empty in nature. If you want to delete one directory which contains files, you need to use rm command.

rm

You can use the command rm for the purpose of deleting the directories and files. If you want to delete the directory only, you need to use rm –r command. When you use the rm command, it will delete the folder along with all the files in it.

touch

This command is used for creating new files. It can be anything, starting from a txt file which is empty to an empty form of a zip file. You can use the command like touch new.txt.

man & - - help

If you want to know in details about a command and how you can use it, you can use the command man. It helps by showing all forms of manual pages of all the commands. For example, if you enter man cd, it will show all the manual pages of the command cd. When you type in the name of the command along with the argument - - help, it will show in which way you can use the command.

cp

The cp command is used for copying files from the command line. It takes in two arguments, the first argument is the file location which is to be copied and the second is where to copy the file.

mv

The command mv is used for moving the files through the line of the command. You can also use this command for renaming a file. For example, if you need to rename a file "text" to "old" you can type in mv text old. It also takes in two arguments just like the command cp.

locate

The command locate is used for locating any file in the system of Linux. It is similar to the command of search in the system of Windows. This command might turn out to be very useful when you have no idea where a specific file is located or saved or what is the actual file name. When you use the argument –i with this command, it helps in ignoring the cases. So, for example, if you need to find a file which has the word "bye" in it, it will give out a complete list of all the Linux system files which contains the word "bye" when you use locate –i bye. In case you remember two words from the file name, you can easily separate the two by inserting an asterisk (*). For instance, for locating a file name with words "bye" and "this", you need to use locate –i *bye*this.

Intermediate commands

echo

This command helps in moving some part of data and most of the times text into a file. For instance, if you need to create a brand new text file or add up

to the already existing text file, you need to use the command as echo hello, my name is sunny>>new.txt. In this case, you are not required to separate the spaces in a sentence by using \ as in this you will need to put two triangular forms of brackets as you finish with the writing.

cat

You can use the command cat for displaying all the contents in a file. It is generally used for viewing programs easily.

nano & vi

nano and vid are the text editors which are installed already in the command line if Linux. The command nano is a form of good text editor which helps by denoting the keywords in colors and can also easily recognize most of the languages. The command vi is much simpler in form than nano. By using the command vi, you can create any new file or even modify files by using this form of editor. For instance, you need to create a new file with the name check.txt. You can easily create the same by the use of the command nano check.txt. You can also save the files after you are done with editing by using Ctrl+X and then Y for yes or N for no.

sudo

It is a very widely used command in the system of Linux. The command sudo stands for SuperUser Do. In case you want any of the command to be carried on with the privileges of root or administration, you can use the command sudo. For example, if you need to edit a file such as viz. alba-base.conf, which requires root permissions, you can type in sudo nano alba-base.conf. You can enter the command line of root by using sudo bash and then type the password of the user. You can also su command for doing the same but you are required to set in one root password before doing that. For setting the password, you need to type sudo passwd and then type in the new password of root.

df

You can use the df command for seeing the disk space which is available in every partition of the system. You just need to type df in the command line and then you can easily view each of the mounted partition along with the available and used space indicated in % along with in KBs. If you want to view the same in megabytes, type in df –m.

du
This command is used for knowing the usage of the disk by a file in the system. In case you are required to know the disk usage for one specific file or folder in the system of Linux, type in du followed by the folder or file name. For example, if you need to know the disk usage which is being used by the folder documents in the system of Linux, type in du Documents. You can also use ls –lah command for viewing the size of the files within a folder.

zip & unzip
You can use the command zip for compressing a file into an archive of zip. For the purpose of extracting files from zip archive use the command unzip.

uname
You can use this command for showing all the information about that system in which your Linux distribution is running. You can type in uname –a for printing the majority of the information about a system.

Chapter 7: Characteristics of Kali Linux and Why It Is So Important In The Hacking World

Kali Linux is a distribution of Linux which is based on Debian. It has been designed very significantly for the purpose of catering to the needs of the network analysts along with the penetration testers. The wide range of tools that come along with Kali Linux makes it the prime weapon of all the ethical hackers. Kali Linux was previously called Backtrack. Kali Linux is the successor of Backtrack with a more polished version of tools than Backtrack which used to serve the same purpose with a wide range of tools and making the OS jam-packed with several utilities which were not at all necessary. That is why the ethical hackers turned towards Kali Linux which provides tools required for penetration testing in a more simplified form for the ease of functioning.

Why this OS?

Kali Linux comes with a plethora of features. There are also various reasons that justify why one start using Kali Linux should.

- **Free of cost:** Linux is a free software and so all the distributions of Linux are also free of cost. Kali Linux has been and will also be free of cost always.
- **A wide array of tools:** Kali Linux can offer you with more than 600 different types of tools for penetration testing and also various tools related to security analysis.
- **Open-source software:** Linux is an open-source software. So, Kali Linux being a part of the Linux family also follows the much-appreciated model of being open-source. The tree of development of the OS can be viewed publicly on Git and all

the codes which are available with Kali Linux are also available for the purpose of tweaking.

- **Support for multi-language:** Although of the fact that the penetration tools are written in English, it is evident that Kali Linux supports multilingual use as well. It has been done to make sure that a greater number of users can operate the OS in their native language and can also locate the tools which they need for their job.
- **Totally customizable:** The developers of the tools for offensive security know that every user will not be agreeing with the model design. So, Kali Linux has been developed in a way so that it can be fully customized according to the need and liking of the user.

System requirements

Installing Kali Linux for the purpose of penetration testing is very easy. You just need to make sure that you have the required set of hardware. Kali Linux is supported on amd64, i386 and ARM. You all require:

- Minimum 20 GB of disk space for the installation of the software
- Minimum 1 GB of RAM
- One CD/DVD drive or virtual box

List of tools

Kali Linux comes with a wide range of tools pre-installed. Let's have a look at some of the most commonly used tools.

- **Aircrack-ng:** It is a tools suite which is used for the purpose of assessing Wi-Fi network security. It aims at some of the prime areas of security related to Wi-Fi.

1. **Monitoring:** It helps in capturing packet and also exports data to the text files for processing in the later stages by the third-party tools.
2. **Attacking:** It helps in replay attacks, fake access points, de-authentication and various others by the process of packet injection.
3. **Testing:** It helps in checking the Wi-Fi cards and other capabilities of the drivers.
4. **Cracking:** It helps in cracking WPA PSK and WEP.

- **Nmap:** Nmap, also known as Network Mapper, it is an open source and free form of utility for the purpose of network discovery along with auditing of security. Nmap uses up the raw packets of IP for determining which hosts are available on the desired network, what are the services are being offered by those hosts, what are the operating systems that they are using, which type of firewall or packet filters are being used and various other characteristics. Many of the administrators of network and systems also use it for:
  1. Inventory of network
  2. Managing the schedules of service upgrade
  3. Monitoring the service or host uptime

- **THC Hydra:** When you are required to crack one remote authentication service, THC Hydra can be used. It is capable of performing super fast dictionary attacks in opposition to 50 or more protocols which includes HTTP, FTP, SMB and HTTPS. It can be used easily for the purpose of cracking into wireless networks, web scanners, packet crafters and many more.

- **Nessus:** It is a form of remote scanning tool which is used for checking the security vulnerabilities of computers. It is not capable of blocking any form of vulnerabilities that the system

of a computer has but it can easily sniff all of them out by running more than 1200 checks for vulnerability and also sends out alerts when it is required to make the security patches.

- **WireShark:** It is an open-source analyzer of packet which anyone can use and that too free of charge. With the help of this tool, the user can easily see the network activities provided along with customizable reports, alerts, triggers and many more.

Features of Kali Linux

Kali Linux is a form of Linux distribution that comes along with a wide range of tools which are pre-installed in the distribution. It has been designed for the targeted users for ease of functioning. Kali Linux is more or less like the other distributions of Linux but it comes along with some added features too that help in differentiating it from the others. Let's check out some of the most unique features of Kali Linux.

- **Live system:** Unlike the other distributions of Linux, the primary ISO image that you are going to download will not only help in installing the OS but it can also be used just like a bootable form of live system. In simple words, Kali Linux can be used without even installing it in the system by just using the ISO image by booting the same. The live system of the distribution contains all the tools which are required by the penetration testers. So, in case your present system is not running on Kali Linux OS, you can easily use it by inserting the USB device and then reboot the same for running Kali Linux on your system.

- **Forensics mode:** While performing any kind of forensic related work on the system, generally the users want to avoid any form of activity which might result in data alteration on the system which is being analyzed. Unfortunately, most of the modern-day environments of desktop tend to interfere with this form of objective and tries to auto-mount any form of the disk which it detects. In order to avoid this form of behavior, Kali Linux comes with the forensics mode which can be enabled from the menu of reboot and it will result in disabling all such features. The live system of Kali Linux turns out to be so useful only for the purpose of forensics as it is readily possible to reboot any system of computer into the system of Kali Linux without even accessing or doing any kind of modification in the hard disks.

- **Customized Kernel of Linux:** Kali Linux is well-known for providing customized version of the recent kernel of Linux which is based on the latest version of Debian Unstable. This helps in ensuring solid support for hardware, precisely for the wide collection of wireless devices. The kernel of Linux gets patched with the support for wireless injection as some of the assessment tools regarding wireless security tends to rely on this form of feature. As most of the hardware devices need updated files of firmware, Kali Linux comes with the feature of installing the files by default along with all the firmware updates which are available in the non-free section of Debian.

- **Trustable OS:** The users of this security distribution wants to know that whether or not it can be trusted and as it has been developed plain sight, it allows anyone to easily inspect the codes of the source. Kali Linux has been developed by a very small team of developers who always follow the required practices of security. The developers also upload the source packages in signed format.

- **Customizable:** Each and every penetration tester has their own way of working and might not agree with the default configuration of the OS. Kali Linux is fully customizable which allows the users to customize the same according to their need. There are also various forms of live-build techniques that can be found online that helps in modifying the OS, install several other supplementary forms of files, run the arbitrary commands, install any other required packages and many more. The users can also customize the way in which the distribution functions.

Chapter 8: Installation of Kali Linux

If you are thinking about pursuing information security for your career, the primary thing that you require is to have an operating system which is focused only on system security. With the help of a proper operating system, you can easily perform various forms of tedious and time-consuming jobs very easily and efficiently. In the present situation, there are various OS available which are based on Linux. Out of the several distributions that can be found today, Kali Linux is regarded as the best choice for the purpose of information security and penetration testing. It is being widely used by the professional penetration testers and the ethical hackers for performing various activities related to their field along with the assessment of network security.

Kali Linux is regarded as the leading distribution from the house of Linux which is also being used for auditing of security. Kali Linux is the only OS related to ethical hacking and network security that comes pre-packaged with several different types of tools related to the hacking of command line which is required for various tasks related to information security. The tasks in which Kali Linux is most commonly used are application security, penetration testing, forensics related to computer system and security of network. In simple terms, Kali Linux is the one and only and the ultimate operating system which has been designed for the ethical hackers.

People who are connected with the world of ethical hacking and penetration testing use Kali Linux for some specific reasons. Kali Linux comes with more than 600 tools for penetration testing. The best part is Kali Linux is 100% customizable. So, in case you are not liking the present configuration of Kali Linux, you can easily customize it in the way you want. Another interesting thing about Kali Linux is that it comes with multilingual support. Although the tools are written in English, this allows people from

all provinces to use this OS using their own native language. It comes with the support of a wide collection of wireless devices. Kali Linux is such an OS which is developed in a secure form of environment. What makes Kali Linux so popular is the feature of being an open source nature of software which is free as well. It also comes with custom kernel which can also be patched for the purpose of injections.

How can you install Kali Linux?

The process of installing Kali Linux in your system is quite easy and simple. The users can also enjoy several options for installing the software. The most preferable options for installation are:

- Installation of Kali Linux by using hard disk
- Installation of Kali Linux by creating bootable Kali Linux USB Drive
- Installing Kali Linux by using software for virtualization like VirtualBox and VMware
- Installing Kali Linux by the process of dual booting along with the operating system

The most widely used options for installing Kali Linux are by using USB drive and installation by using VirtualBox or VMware. You need minimum 20 GB of free space in the hard disk of your system along with at least 4 GB of RAM if you are using VirtualBox or VMware. You will also require USB along with CD/DVD support.

Installing Kali Linux with the help of VMware

- Before you want to run Kali Linux in your system, you will require a virtualization software at the very first place. There are various options available today when it comes to choosing a virtualization software. You can start by installing VMware or

VirtualBox from the house of Oracle. After you have installed the virtualization software, you need to launch the same from the folder of applications.

- Now you are required to download the installation file for Kali Linux which you can easily find from the download page in the official website of Kali Linux. You can choose the one which you think will be meeting your needs. Along with the download file in the download page, you will also come across a wide variety of hexadecimal numbers which are used for the security-related jobs. You are required to check the image integrity which you are going to download. You need to check fingerprint SHA-256 for the file and then compare the same which has been provided on the download page of Kali Linux.

- After you have downloaded the installation file for Kali Linux, you are required to launch the virtual machine now. For this, you need to open the homepage of VMware Workstation Pro and then select create a new virtual machine. After you have created a new virtual machine, you need to select the iso file of Kali Linux followed by the selection of the guest OS. You will also need to configure all the details of the virtual machine which is Kali Linux in this case. Now you can start the Kali Linux virtual machine simply by selecting the VM for Kali Linux and then selecting the power on button which is green in color.

- After the virtual machine has powered up, a pop-up menu will be prompted in which you need to select the preferable mode of installation in the GRUB menu. You need to select the option graphical install. Click on continue.

- The next few screens will be asking you to choose your locale information like the preferred language in which you want Kali Linux to install, the location of your country along with the layout of your keyboard.

- Once you are done with all the required locale information, the installer will automatically start to install some required additional components for the software and then will also configure the settings related to network. After the components have been installed, the installer will ask you to enter the hostname along with the domain name for the purpose of installation. You are required to provide each and every appropriate information for proper installation of the software and for continuing with the installation.
- After you are done with all the above mentioned steps, you will need to set up a password for your machine of Kali Linux and then hit the continue button. Make sure that you do not forget to set a password for your Kali Linux machine.
- As you set up the password for your Kali Linux machine, the installer will then prompt you for setting up the time zone and will then pause the setup at the time of defining the disk partitions. The installer of the machine will give you four different choices regarding the disk partitions for the machine disk. In case you are not sure about partitioning your disk, the easiest option which is available for you is to select the option of Guided – Use Entire Disk which will be using up the entire disk space and will omit the process of disk partitioning. If you are an experienced user, you can select the option of manual partitioning for more granular options for configuration.
- You will now require to select the partitioning disk. However, the most recommended option is to select the option for all files in one partition for all the new users. After you gave selected the partitioning disk, select continue.
- Now you will need to confirm all the changes that you have made to the disk on the machine of the host. Make sure that you do not continue with the process as it will be erasing all the data which is available on the disk. Once you confirm all the changes in the partition, the Kali Linux installer will start

running the process of file installation. It might take a while and do not interrupt the process as the system will install everything automatically.

- Once all the required files have been installed, the system will be asking you in case you want to set up any network for the purpose of obtaining the future updates and pieces of software. Make sure that you enable this function if you are going to use the repositories of Kali Linux in the future. The system will then configure the manager of package related files.
- After this step, the system will be asking you to install the boot loader of GRUB. Click on yes and then select the device for writing up the required information of boot loader to the hard disk which is needed for booting Kali Linux.
- Once the installer has finished installing the boot loader of GRUB into the disk, select continue for finishing up the process of installation. It will then install some of the files at the final stage.

After you are done with all these steps, Kali Linux will be installed in your system and you can start using the same for the purpose of penetration testing and network security. You can also use Kali Linux in your system by simply creating a USB bootable drive without even installing the software in the system.

Chapter 9: Applications and Use of Kali Linux

Kali Linux is a well-known OS in the world of ethical hacking. While it is known that the prime focus of Kali Linux is on the summarized use for penetration testing along with security auditing, Kali Linux can also perform several other tasks apart from these two. Kali Linux has been designed in the form of a framework as it comes with various forms of tools which can cover various types of use cases. Some of the tools of Kali Linux can also be used in combination at the time of performing penetration testing.

For instance, it is possible to use Kali Linux on various types of computers such as on the system of the penetration tester, on the servers of the administrators of the system who wants to monitor their own network, on the systems or workstations of the analysts related to system forensics and also on the embedded form of devices generally along with the ARM CPUs which can be easily dropped in the range of the wireless network or which can also be plugged in the system of the targeted user. Many of the devices

related to ARM also perform as great machines for the purpose of attacking which is mainly because of their small factors of formation along with the requirement very low power.

You can also deploy Kali Linux directly in the cloud for the purpose of quickly building a large farm of machines which are able to crack passwords and on the mobile phones along with tablets for allowing an efficient form of portable testing of penetration. But it does not end here; the penetration testers also require servers. The servers are required for using a software of collaboration within a large group of penetration testers, for setting up the web server to be used for campaigns related to phishing, for the purpose of running the tools related to vulnerability scanning and for various other interconnected jobs.

Once you are done with booting of Kali Linux, you will find out that the main menu of Kali Linux has been organized in accordance to various themes across the different forms of activities and tasks which are relevant to the penetration testers and other professionals of information security.

Tasks that can be performed with Kali Linux

Kali Linux helps in performing a wide range of tasks. Let's have a look at some of them.

- **Gathering of information:** Kali Linux can be used for collecting various forms of data related to the targeted networks along with the structure of the same. It also helps in identifying the systems of computers, the operating systems of such computers along with all the services that the computer system runs. Kali Linux can be used for identifying the various potential sensitive parts within the system of information along with the extraction of all forms of listings from the services of a running directory.

- **Analysis of vulnerability:** You can use Kali Linux for the purpose of quick testing of whether a remote or any local system has been affected by any known vulnerabilities or any form of configuration which is not at all secure in nature. The scanners of vulnerability use the databases which contain several signatures for the purpose of identifying the potential threats and vulnerabilities.

- **Analysis of web application:** It helps in the identification of any form of misconfiguration along with weaknesses in the security system of the web applications. It is a very crucial task to identify and then mitigate such issues given that public availability of such applications makes the same the ideal form of targets for all the attackers.

- **Assessment of database:** Database attacks are the most common form of vector for the attackers that include attacks such as SQL injection to attacks in the credentials. Kali Linux provides various tools which can be used for testing the vector of attacks which ranges from data extraction to SQL injection along with analysis of the same.

- **Password attacks:** The systems connected with authentication are always vulnerable to the attacks of the attackers. A wide array of tools can be found in Kali Linux which ranges from online tools of password attack to the offline tools against the systems of hashing or encryption.

- **Wireless form of attacks:** Wireless networks are pervasive in nature. This means that they are always a common vector of attack for the attackers. Kali Linux comes with a wide range of support related to various cards of the network which makes

Kali Linux an obvious choice for the attacks in opposition to the several wireless network types.

- **Reverse engineering:** reverse engineering is a very important form of activity which is being used for various purposes. In providing support for the various forms of offensive activities, reverse engineering is one of the prime methods which is being used for identification of the vulnerabilities and also for tracking the development of exploitation. On the side of defense, it is also being used for analyzing the malware which is employed for the targeted attacks. Within this capacity, the aim is to identify the prime capabilities of a given set of tradecrafts.

- **Tools for exploitation:** Exploitation is the act of taking advantage of any form of existing vulnerability in a system which allows the attacker to gain complete control of a remote form of device or machine. This form of access can also be used by the attackers for further privileges of escalation of attacks which are done either on any form of machine which is accessible to the local network or on the machine which has been compromised. This category of Kali Linux function comes with various tools along with utilities which help in simplifying the overall process writing up your very own form of exploits.

- **Spoofing and sniffing:** Gaining overall access to that packet of data which is travelling across any network is always advantageous for the attackers. Kali Linux can provide you with various tools for the purpose of spoofing which will allow you to imitate any legitimate user along with the sniffing tools which will allow you to analyze and also capture the available pool of data directly from the network wire. When spoofing as

well as sniffing tools are used together, it can turn out to be very powerful.

- **Post exploitation:** Once you have been successful in gaining all-over access to the target system, you might want to maintain the same level of accessibility to the system along with extended control simply by moving laterally over the network. You can find various tools in Kali Linux for assisting you in your goals regarding post exploitation.

- **Forensics:** The live boot environments of Forensic Linux have been very famous in the recent years. Kali Linux comes with a large number of very popular tools of forensics which are based on Linux which will allow you to perform everything, starting from the initial stage of triage to imaging of data along with full analysis of the system and lastly management of case.

- **Tools of reporting:** A test of penetration can only be declared as successful once all the findings of the test have been properly reported. This category of tools from Kali Linux helps in composing the collected data which has been gathered by the use of tools for information gathering, finding out various non-obvious form of relationships and also bringing together everything in several reports.

- **Tools for social engineering:** When the technical aspect of a system is secured properly, there are chances of exploiting the behavior of human beings as a vector of attack. When provided with the perfect influence, human beings can be induced frequently for taking various actions which ultimately leads to the compromising of the security of a system environment. Did the USB drive which was just now plugged in by the secretary contain any form of harmful PDF? Or did the UDB drive just installed a form of Trojan horse backdoor? Was the website of

banking which was used by the accountant just now was a normal expected form of website or a copy of a website for the purpose of phishing attack? Kali Linux comes with various tools that can help you in aiding all these forms of attacks.

- **Services for system:** Kali Linux can provide with tools which will allow you to initiate and also stop various applications which run in the background as the services for the system.

Coordinating tasks of Kali Linux

Kali Linux helps in coordinating several tasks and also helps in balancing the coordination between the software and hardware of a system.

The first and foremost task of Kali Linux is to control the hardware components of the computer system. It helps in detecting along with figuring out the various hardware components when the computer turns on or also when any new device is installed. It helps in making the hardware components available for the various higher level of software with the help of a simplified form of program interface so that the applications can take all-round advantage of the connected devices without the need of addressing any detail like in which extension slot is the option board plugged in. The interface of programming also comes with a layer of abstraction which allows various software to work seamlessly with the hardware.

What makes Kali Linux different from others?

Kali Linux has been specifically designed for gearing up the functioning of the penetration testers and also for the purpose of security auditing. For achieving this, various core changes have also been implemented for Kali Linux which reflects all of these requirements:

- **Root access by design, single-user:** Because of the normal nature of the audits regarding system auditing, Kali Linux has been designed in such a way which can be used in the scenario of single root access. Most of the tools which are employed for the purpose of penetration testing needs escalated form of privileges and as it is typically sound policy for enabling the root privileges when required, during the use cases to which Kali Linux is aimed to, this whole approach might turn out to be a huge burden.

- **The services of network disabled by default:** Kali Linux comes with systematic hooks which disables the services of a network by default. Such hooks allow the users to install several Kali Linux services while also making sure that the distributions also remains completely safe and secure by default no matter which type of packages has been installed. Other additional services like Bluetooth are also kept in the blacklist by default settings.

- **Custom kernel of Linux:** Kali Linux uses up upstream form of the kernel which is patched for the purpose of wireless injection.

- **A set of trusted and minimal repositories:** The absolute key of Kali Linux is to maintain the integrity of a given system, given all the goals and aims of Kali Linux. With the prime aim in mind, the complete collection of sources of upstream software which are used by Kali Linux is kept as minimum as possible. Many of the new users of Kali Linux gets tempted to add the extra repositories to the sources.list. But, by doing so, it leads to the risk of breaking the installation of Kali Linux.

It is not correct to suggest that everyone should be using Kali Linux. Kali Linux is been designed particularly for the security specialists. It comes with a unique nature because of which Kali Linux is not a recommended distribution for those who are not at all familiar with the functioning of Linux or are looking out for some general form of Linux distribution for their desktop, for gaming, designing of website and many more. Even for the experienced users of Linux, Kali Linux might come along with certain challenges which are generally set up due to preserving the security of the systems.

Chapter 10: Different Tools of Kali Linux

As we know that Kali Linux is an open source form of distribution which is completely based on Debian, it helps in providing various tools for the purpose of security auditing along with penetration testing. It has been developed by Offensive Security and is also among some of the most well-known distributions and is being widely used by the ethical hackers. The best thing that comes with Kali Linux is that it does not need to be installed as the OS in your system. Instead of that, you can simply run the iso file which can be loaded in the memory of RAM easily for the purpose of testing the security of a system with the help of around 600 tools.

Kali Linux provides users with various forms of tools like information gathering tools, tools for analysis of vulnerability, web application tools, wireless attack tools, tools for forensics, sniffing along with spoofing tools, hardware hacking tool and many more. Let's have a look at some of the most popular tools from Kali Linux.

Tools from Kali Linux

- **Nmap:** Nmap can be regarded as the most popular network mapping tool. It allows the user to find out the active hosts available within a network and also gathers relevant information in relation to penetration testing. Some of the main features of Nmap are:
  1. It comes with host discovery and helps in identifying the available network hosts.
  2. Nmap comes with the feature of port scanning which allows the users to calculate the total number of open ports on the remote or local form of host.

3. It helps in fetching the OS of a network and also finds out various information about the connected devices.
4. It allows the user to detect the version of the application and also determines the name of the application.
5. It helps in extending the default capabilities of Nmap by the use of NSE or Nmap Scripting Engine.

- **Netcat:** As the name goes by, Netcat functions just like a cat and helps in fetching details about a network. It functions as an application for network exploration which is not only used in the field of security industry but is also famous in the network administration and security administration field. It is generally used for the purpose of checking outbound and inbound network and also for port exploration. It can also be used for conjunction with various languages of programming such as C or Perl or also with bash scripts. The main features of Netcat are:
  1. Port analysis of UCP and TDP
  2. Sniffing of inbound and outbound network
  3. Forward and reverse analysis of DNS
  4. Scanning of remote and local ports
  5. Integration with the standard input of terminals
  6. TCP and UDP tunneling mode

- **Unicornscan:** This is one of the finest tools of infosec which is being used for the purpose of data correction along with gathering. It also offers the users with UDP and TCP scanning along with super beneficial patterns of discovery which helps in finding the remote hosts. It can also help in finding out the software which is running in each of the hosts. The main features of Unicornscan are:
  1. Asynchronous scan of TCP
  2. Asynchronous scan of UDP

    3. Asynchronous banner detection of TCP
    4. Application, OS and system service detection
    5. Capability of using customized sets of data
    6. Supports relational output for SQL

- **Fierce:** Fierce is a tool from Kali Linux which is used for the purpose of port scanning along with network mapping. It can also be used for discovering the hostnames and non-contiguous space of IP across any network. It is somewhat similar in features just like Unicornscan and Nmap but unlike these two, Fierce is specifically being used for the corporate networks. After the target network has been defined by the penetration tester, Fierce runs various tests in opposition to the domain which are selected for retrieving important information which can be used for the further analysis and post exploitation. The features of Fierce include:
      1. Scanning of internal and external IP ranges
      2. Capability of changing the DNS server for the purpose of reverse lookup
      3. Scanning of IP range and complete Class C
      4. Helps in logging capabilities into a file system
      5. Discovery of name servers and attack of zone transfer
      6. Capabilities of brute force by using the custom list of texts

- **OpenVAS:** Also known as Open Vulnerability Assessment System, is a free software which can be used by anyone for the purpose of exploring the remote or local vulnerabilities of a network. This tool of security helps in writing and also integrating the customized plugins of security to the platform of OpenVAS. The main features of OpenVAS are:
      1. It works as a port scanner and network mapper
      2. It helps in discovery of simultaneous host
      3. It supports OpenVAS protocol of transfer

4. It comes integrated with databases of SQL such as SQLite
   5. It performs weekly or daily scans
   6. It helps in exporting the results into HTML, XML or LateX formats of files
   7. It comes with the capability of resuming, pausing and stopping the scans
   8. It is fully supported by both Linux and Windows

- **Nikto:** Nikto is written in Perl and is a tool which is included in Kali Linux, it works as a complementary tool to OpenVAS and to other tools of vulnerability scanner. It allows the penetration testers along with the ethical hackers to carry on with scanning of a full web server for the discovery of vulnerabilities along with flaws in security. This tool gathers all the results of security scanning by finding out the insecure patterns of application and files, server software which has become outdated and the default names along with misconfiguration of software as well as of server. It also supports various proxies for SSL encryption, authentication based on host and many others. The main features of Nikto are:
   1. It helps in scanning multiple ports which are available on a server
   2. It comes with evasion techniques of IDS
   3. It provides the output results in XML, TXT, NBE, HTML and CSV
   4. It comes with the enumeration of Apache and cgiwrap username
   5. It performs scans for the specified directories of CGI
   6. It can identify the software which is installed in the system via the files, favicons and headers
   7. It uses up custom files of configuration
   8. It helps in debugging and providing verbose output

- **WPScan:** WPScan is used for the purpose of auditing the installation security of WordPress. With the help of WPScan, you can easily find out whether or not the setup of your WordPress is susceptible to any form of attack or not or whether if it is giving out too much information in the core, theme files or plugins. This tool also allows the users to find the weak passwords for each and every registered user and can also run a brute force attack for finding out which one can be cracked. The features of WPScan are:
  1. Enumeration of WP username
  2. Security scans of non-intrusive nature
  3. Enumeration of WP plugin vulnerability
  4. Cracking of weak password and brute force attack of WP
  5. Scheduling of WordPress security scans

- **CMSMap:** CMSMap is an open source form of project which is written in Python. It helps in automating the task of vulnerability scanning along with detection in Joomla, WordPress, Moodle and Drupal. This tool can also be used for running a brute force attack and also for launching the various exploits once the vulnerabilities have been discovered. The main features of CMSMap are:
  1. It supports multiple threats scan
  2. It comes with the capability of setting customized header and user agent
  3. It supports encryption of SSL
  4. It saves the output file in the form of text file

- **Fluxion:** This tool functions as an analyzer of Wi-Fi which specializes in attacks of MITM WPA. It allows the users to easily scan the wireless form of networks, search for any form of security flaw in the personal or corporate networks. Unlike the other tools for Wi-Fi cracking, this tool does not perform

any form of brute force attack for cracking attempt as it takes generally a lot of time. Instead of launching a brute force attack, this tool spawns a process of MDK3 which makes sure that all the users who are connected with the targeted network are deauthenticated. After this has been done, the user gets a prompt screen for connecting with a fake pint of access where they are required to enter the Wi-Fi password. Then the tool sends the password of Wi-Fi to you so that you can easily gain access to the same.

Other than all these tools, there are several other tools from Kali Linux such as Aircrack-ng, Kismet Wireless, Wireshark, John the Ripper and many others.

Chapter 11: How can Kali Linux be Used For Hacking?

As we all know by now that Kali Linux has been designed especially for the purpose of penetration testing and security auditing, it can also be used for the purpose of ethical hacking which is required while performing penetration testing and other security checks. Kali Linux comes packed with a huge number of tools which helps in the venture of security infrastructure testing and other forms of testing for securing an organization or company.

Who all uses Kali Linux and why?

Kali Linux can be regarded as the most unique form of OS which can be found today as serves as a platform which can be used up by both the good guys and the bad guys. The administrators of security along with the black hat hackers all use this platform for meeting their needs. One uses this

system for the purpose of preventing and detecting breaches in security infrastructure while the other uses this OS for identifying and thereby exploiting the security breaches. The huge number of tools which comes packed with Kali Linux can be regarded as the Swiss Knife for the toolbox of the security professionals. The professionals who widely use Kali Linux are:

- **Security Administrators:** The administrators of security come with the responsibility of safeguarding the information and data of the concerned institution. The security administrators widely use Kali Linux for the purpose of ensuring that there are no forms of vulnerabilities in the environment of the security infrastructure.
- **Network administrators:** The network administrators come with the responsibility of maintaining a secure and efficient network. Kali Linux is used by the network administrators for the purpose of auditing of network. For instance, Kali Linux can easily detect the access points of rogue.
- **Architects of network:** Such people are responsible for the designing of a secure environment for a network. They use Kali Linux for auditing the internal network designs and makes sure that nothing has been misconfigured or overlooked.
- **Penetration testers:** The penetration testers use Kali Linux for auditing the security environments and also perform reconnaissance for the corporate environments which they are bound to take care of.
- **CISO:** The Chief Information Security Officer takes help of Kali Linux for the purpose of internal auditing of the environment of their infrastructure and finds out if any new form of application or configurations of rogue has been installed in the environment.
- **Forensic engineers:** Kali Linux comes along with a mode of forensics which allows the forensic engineers for performing

discovery of data along with data recovery in various instances.

- **White hat hackers:** The white hat hackers or the ethical hackers are similar to the penetration testers who use Kali Linux for auditing and for finding out vulnerabilities which might be present within a security environment.
- **Black hat hackers:** The black hat hackers use Kali Linux for finding out vulnerabilities in a system and then exploiting the same. Kali Linux comes with various applications of social engineering which can be easily used by the black hat hackers for compromising an individual or an organization.
- **Grey hat hackers:** The grey hat hackers also use Kali Linux just like the black hat as well as the white hat hackers.
- **Computer enthusiasts:** It is a very generic form of term but any person who is interested in getting to know more about computers and networking can use the system of Kali Linux for the purpose of learning more about networking, information technology and common form of vulnerabilities.

Process of hacking

Kali Linux is very popular as a hacking platform. The word "hacking" might not always be negative as it is also being used for various other jobs other than exploitation. By gathering immense knowledge about the process of hacking with Kali Linux, you can learn how to perform for vulnerability check and how to fix them as well in case you want to choose ethical hacking as your career option. The process of hacking with Kali Linux is similar to that of a general hacking process in which a hacker tries to get into the server of an organization or company and thereby gain all forms of access to the data which is stored in the servers. The process of hacking can be divided into five different steps.

- **Reconnaissance:** This is regarded as the very first step while starting with the process of hacking. In this step, the hacker tends to use all the available means for the purpose of

collection of all forms of information about the targeted system. It includes various phases such as target identification, determining the target IP address range, available network, records of DNS and many others. In simple terms, the hacker gathers all contacts of a website or server. This can be achieved by the hacker by using various forms of search engines like maltego, researching about the system of the target, for instance, a server or website or by utilizing various other forms of tools like HTTPTrack for the purpose of downloading a complete website for enumeration at later stages. After the hacker is done with all these steps, he can figure out the employee names, the positions of the employees along with the designated email addresses of the employees.

- **Scanning:** After the collection of all forms of information regarding the target, the hacker starts with the second phase which is scanning. The hackers utilize several forms of tools in this phase such as dialers, port scanner, network mappers, scanners of vulnerability and many others. As Kali Linux comes pre-loaded with a huge bunch of tools, the hackers won't even face any form of difficulty during this phase. The hackers try to find out that information about the target system which can actually help in moving ahead with an attack such as IP addresses, the accounts of the users and computer names. As the hackers get done with basic information collection, they start looking out for the other possible avenues of attack within the target system. The hackers can select various tools from Kali Linux for the purpose of network mapping such as Nmap. The hackers try to find out automated email reply system or simply by basing on the information which has been gathered by them. The hackers move to the next step which includes emailing the staffs of the company regarding various queries,

such as mailing the HR of a company about a detailed enquiry on job vacancy.

- **Gaining overall access:** This phase is regarded as the most important of all when it comes to hacking. In this phase, the attacker attempts to create the design of the network blueprint which has been targeted. It is created with all the relevant information which has been collected by the hacker. After the hackers finish the phase of enumeration and scanning, the step that comes now is gaining access of the targeted network which is based completely on the information collected. It might happen that the hacker wants to use phishing attack. He might try to take it safe and thus use only a very simple attack of phishing for the purpose of gaining access. The hacker can decide to get into the targeted system from the IT department of the organization.

The attacker might also get to know that some recent hiring has been done by the company and it can help in speeding up the procedure. For the phishing attack, the hacker might send out emails of phishing by using the actual address of email of the CTO of the company with the use of a unique form of program and will send out the mails to all the technicians. The email which will be used for the purpose of phishing will be containing a website which will help in gathering all the required user ids and passwords for the purpose of logging in. The hacker can also use other choices like phone app, website mail or some other platform for the purpose of sending out mail of phishing to the users and then asking the individuals for logging in to a new Google portal with the use of their provided credentials.

When the hackers decide to use such a technique, they have a special type of program which runs in the background in their system which is called Social Engineering Toolkit. It is used by attackers for sending out the emails with the address of the server to the users directly after masking the server address with the help of bitly or tinyurl. The attackers can also use other methods for gaining access to the system such as by making a reverse

TCP/IP shell in the PFD format file which can be created by the use of Metasploit. The attackers can also employ overflows of buffer for the attacks which are based on stacking or hijacking of the sessions which ultimately results in gaining overall access to the targeted server.

- **Maintaining the access to the server:** After the hacker has gained access to the target server, he will try to keep the access to the server as it is and keeping it safe for future exploitation and attacks. When a hacker gets access to an overall system, he can use the hijacked system as his own personal base and use the same for launching several other attacks to the other systems. After a hacker gains access to a targeted system and ultimately owns the same, the hijacked system is called a zombie system. The hacker gains access to a whole new array of email addresses and accounts and can start using those for testing other form attacks right from the same domain. For the purpose hiding in the system, the hacker also tries to create a brand new administrator account and tries to get dissolved in the system.

For safety purposes, the hacker also starts to find out and identify those accounts in a system which has not been used by the organization for a long period of time. After the hacker finds out such form of accounts, he changes all the login passwords of the old accounts and elevates all form of privileges right to the administrator of the system like a secondary account for the purpose of having safe access to the network which has been targeted. The hacker can also begin to send out various emails to the other users within an organization which might contain exploited form of files in the PDF format with the reverse shell scheme for extending his all-round access within the system. After all these, the hacker waits for some time to make sure that no form of disturbance has been detected in the system and after getting sure of the same, he starts to create copies of the available pool

of user data like contacts, files, messages, emails and various other forms of data for using them in the later stage.

- **Clearance of track:** Before starting a system attack, the hackers plans out their whole pathway for the attack along with their planning for identity so that if any discrepancy occurs no one can trace them up. The hackers start doing so by altering their MAC address and then run the same system across a VPN so that their identity can be covered up easily. After the hackers have achieved their target, they begin with clearance of their pathways and tracks. This complete phase includes various things such as clearing of the temp files, mails which has been sent, the logs of the servers and various other things. The hacker also tries to make sure that there is no form of alert message from the email provider which can alarm the targeted organization regarding any form of unauthorized or unrecognized login in the system.

A penetration tester follows all these steps for the purpose of testing the vulnerabilities of a system and making sure that those which are available in the system are mended properly.

Chapter 12: Techniques of Port Scanning using Kali Linux

Identification of the open ports on the targeted system is essential for defining the surface of attack of the target. The open ports of the target correspond with the networked services which are running on the system. Errors in programming or flaws in implementation can result in making all these services very much susceptible to the attacks and can also lead to compromise of the overall system. For the purpose of determining the most probable vectors of attack, you are required to enumerate all the ports which are in open condition on all the systems of remote form within the scope of the project. The open number of ports also corresponds with the services which can be easily addressed with the help of either TCP or UDP traffic.

Both UDP and TCP are protocols of transport. TCP or Transmission Control Protocol is the one which is more commonly used than UDP and also provides communication which is connection-oriented. UDP or User Datagram Protocol is a protocol which non-connection oriented in nature which is also sometimes used along with the services in which transmission speed is more important than the integrity of data. The form of penetration testing which used for the purpose of enumerating such services is known

as port scanning. Such technique helps in yielding enough amount of information for the purpose of identifying whether the service is being associated with any port on the server or on the device.

UDP Port Scanning

As TCP is more frequently used than UDP as a protocol 0f transport layer, services which are operated by UDP are most often forgotten. In spite of the normal tendency of overlooking the services of UDP, it is also critical for these services to be enumerated for acquiring an overall understanding of the surface of attack of any form of target. The form of scanning with UDP might often turn out to be tedious, challenging and time-consuming as well. For gaining overall insight into the functioning of these tools it is very essential to understand the two exactly different approaches of UDP scanning which is used.

The first technique which is used is to rely on the ICMP port unreachable responses exclusively. This form of scanning relies on those assumptions which the UDP ports which are not linked with the live service will return ICMP port unreachable response. Lack of this response is taken as the indication of a live form of service. Although this form of approach might turn out to be very effective in various circumstances, there are also chances of the same of returning inaccurate form of results in the cases in which the host is unable to generate port unreachable response or the replies of port unreachable is either filtered by any form of firewall or are rate limited.

It also comes with an alternative in which service specific probes are used for attempting soliciting of a response which will indicate that the service which was expected is running on the port which is targeted. Although this form of approach might turn out to be very effective, it is also very time-consuming at the same time.

TCP Port Scanning

TCP port scanning includes various approaches such as connect scanning, stealth scanning along with zombie scanning. For understanding how all these techniques of scanning work, you need to understand how the connections of TCP are established and also maintained. TCP is a form of protocol which is connection-oriented. Data is transported over TCP only after a successful connection has been created in between the two systems. The process which is associated with the creation of connection of TCP is often called three-way handshake. This term alludes from the three different steps which are involved in the process of connection.

A packet of TCP SYN is sent from that device which wants to establish connection along with the device port which it wants to connect with. If the associate service with the port which the device wants to connect to accepts the connection, the port will be replying to the system which is requesting the connection with a packet of TCP that comes with both ACK and SYN bits activated. The connection is successful when the requesting system responds back to the port with a response of TCP ACK. These three steps in total sums up the three-step process which is required for the establishment of a session of TCP between two systems. All the techniques of TCP port scanning will be performing some sort of variation of this entire process for the purpose of identifying the live services on the remote form of hosts.

Both the process of stealth scanning and connect scanning are quite easy to understand. The process of connect scanning is used up for establishing a complete TCP connection every port which is being scanned. This is done for each of the ports which are scanned for completing the three-way handshake. When a connection is established successfully, the port is determined to be in the open state. However, in the case of stealth scanning, a full connection is not established. Stealth scanning is often referred to as SYN scanning or also half open scanning.

For each and every port which are scanned, one single packet of SYN is sent out to the port of destination and all the ports which replies with a

packet of SYN+ACK are taken as to be running the live form of services. As no final form of ACK is sent out from the system which initiated the connection, the connection is left out as half open. This is known as stealth scanning as the solutions of logging which only records the connections which are established do not record any form of evidence of the performed scan.

The final method which comes with TCP scanning is the zombie scanning. The prime goal of zombie scanning is to map all the open form of ports on a system of remote nature without even producing any form of evidence which you have had an interaction with the system. The principles on which the functioning of zombie scanning depends are complex in nature. You can carry out zombie scanning by following these steps.

- Start by identifying the remote system for the zombie. The system which you are going to identify needs to have these characteristics:
    1. It is in idle form and it doesn't actively communicate with the other systems which are available on the network.
    2. It needs to use an incremental form of IPID sequence.
- Then you will need to send in a packet of SYN+ACK to the zombie and then record the initial value of IPID.
- Send in a packet of SYN along with a source of the spoofed IP address of the system of zombie to the target system of scan.
- Depending on the port status on the target scan, any of the following will happen:
    1. In case the port is in open state, the scan target will be returning a packet of SYN+ACK to the host of zombie which it thinks sent out the original request of SYN. In such a case, the host of zombie will be responding to the unsolicited form of SYN+ACK packet with a packet of RST and then increment the value of IPID by one.

2. In case the port is in the closed state, the scan target will be returning a response of RST to the host of zombie which it thinks sent out the original request of SYN. The packet of RST will be soliciting no form of response from the host of zombie and the value of IPID will therefore not be increased.

Send in another packet of SYN+ACK to the host of zombie and then evaluate the final value of IPID of the RST response which has been returned. In case the value has been increased by one, the port on the target scan is closed and in case the value has been increased by two the port on the target scan is in open state.

For performing a zombie form of scan, an initial request of SYN/ACK is required to be sent to the system of zombie for the purpose of determining the current value of IPID within the returned packet of RST. A spoofed packet of SYN is then sent out to the scan target along with a form of source IP address of the system of zombie. As the zombie actually did not send out the initial request of SYN, it will be interpreting the response of SYN/ACK as being unsolicited and then send a packet of RST back to the system of target and thus increasing the value of IPID by one. At the final stage, another packet of SYN/ACK needs to be sent to the system of zombie which will return a packet of RST and then increase the value of IPID by one.

Chapter 13: Penetration Testing

Each and every infrastructure of IT comes with some weak points which can ultimately lead to some serious attack and can be used for the purpose of stealing and manipulating data. Only one thing can be done in such situations which can help in preventing the hackers from entering the system. You need to perform regular checks of the infrastructure of your security and make sure that there is no form of vulnerabilities present in the structure. Penetration testing helps in finding out the vulnerabilities along with the several weak points in a system. As the owner or administrator of a network, they can always have some advantage over the hackers as they are bound to know the topology of network, the components of infrastructure, the services, the probable points of attack, the executed services and many more.

Penetration testing is done within a real and secure environment so that in case any vulnerability is found, you can mend the same and secure the

system.

Penetration testing in details

As the name goes by, penetration testing is the process of testing a system to find out whether penetration by any third party is possible in the system or not. Penetration testing is often mixed up with ethical hacking as both are somewhat similar in features and functioning. The motive of are also the same but a very thin line differentiates the two. In penetration testing, the tester scans for any form of vulnerability in the system, malicious form of content, risks and flaws in the concerned system. Penetration testing can be performed either in an online network or server or a computer system as well. Penetration testing comes with the goal of strengthening the security system of an organization for the motive of properly defending the security of a system. Unlike hacking, penetration testing legal in nature and is done with all forms of official workings. If used in the proper way it can do wonders. Penetration testing can be considered as a significant part of ethical hacking.

Penetration testing needs to be performed at regular intervals as it comes with the power of improving the capabilities of a system and also improves the strategies related to cyber security. Various types of malicious content are created for the purpose of fishing out the weak points which are available within a program, application or system. For effective testing, the malicious content which is created is spread across the entire network for vulnerability testing. The process of penetration testing might not be able to handle all the concerns related to security; however, it can help in minimizing the probable chances of any form of attack. It helps in making sure that a company is safe from all forms of vulnerabilities and thus protecting the same from cyber attacks. It also helps in checking whether or not the defensive measures are enough for the organization and which of the security measures are required to be changed for the motive of decreasing the vulnerability of the system.

Penetration testing is really helpful in pointing out the strengths along with the weaknesses in the structure of an organization at any one given point of time. You need to note that this whole process if not at all casual in nature. It includes rigorous planning, granting of the required permissions from the concerned management and then initiating the process.

Security scanners

The process of penetration testing starts after an overview of the complete organization has been collected and then the process of searching for the specific weak points starts. For performing all these, you are required to use a security scanner. Depending on the type and nature of the security scanners, the tools can actually help in checking an entire system or network for the weak points which are known. One of the most comprehensive forms of tool for security scanning is OpenVAS. This tool comes with the idea of a huge number of vulnerabilities and can also check for the defenses. After the OpenVAS tool has identified all the open form of tools, you can easily use Nmap for discovering the details. A tool like Wireshark will allow you to find out any form of content which is critical in nature along with any critical form of network activity which can point out certain patterns of attack.

The classic form of Wireshark tool is also useful in identifying the bottlenecks which can indicate the attacks of the hackers and also requires a continuous check. In the world of corporate organizations, the applications which are based on the web often depend on MySQL, Apache and stack of PHP. All these platforms dominate the entire landscape. Such platforms are the favorite targets of the hackers as they usually come with great opportunities of attacks. Kali Linux comes with around two dozen tools which specialize in web application testing. Such scanners can be easily found in the menu of Web Application Analysis. The w3af and Burp Suite are regarded as the best tools of the lot.

Burp Suite helps in the identification and testing of the vulnerabilities and is quite easy to use. Brute force attack can be launched from the module of the intruder which takes help of the request records which are grouped in the proxy intercept tab for the purpose of injecting the required payload in the system of the web. It also helps in detecting configurations of poor security. Configuration of incorrect nature of the security settings can take place at any of the levels of the stack of application. For the purpose of detecting these vulnerabilities, Burp Suite starts with the identification of the target and then executes the Spider command from the menu of context. The outputs which can be found from the scans can help you in finding out the misconfigurations in the system.

Generally, a great amount of caution is needed at the time of product system analyzing with the security scanners which are not designed in a way for getting handled by the kid hands. Although various actions serve for identifying the points of attack, you can expect that the concerned system which is being tested might also get affected. So, you are required to perform all these tests within mirrored form of systems. Generally, the mirrors of the systems are secured by the firewall and IDSs of the system of production, so you can also check the overall effectiveness of the protection mechanisms which are existing already. Several forms of tools can run in various modes which might make it difficult for the IDSs to properly detect the scans. While running in the intelligent modes, they often fail to get detected.

Sounding the weak points

After you have found out where are the gaps in the, the next step that you need to perform is to sound all of them out. An important portion of the penetration tests is the use of the tools which helps in stimulating as many patterns of attack as possible. Metasploit can be regarded as the most

widely used form of tool for penetration testing and is also a great tool for the penetration testers.

Chapter 14: VPN

VPN or virtual private network is the method of connection which is used for adding privacy along with security to any public or private form of network like the internet or hotspots of Wi-Fi. VPNs are most widely used by the corporations for the purpose of protecting various forms of private and sensitive data. However, in the recent years, the craze of using private VPNs is increasing day by day. This is mainly because of the fact that all those interactions which were face to face in the beginning now transformed to the internet form of communication. Privacy increases with the use of VPN as the IP address of the initial system get replaced with the IP address which is provided by the provider of a virtual private network. The subscribers can get an IP address from any city they want from the provider of VPN service. For example, you are living in San Francisco but with service provided by virtual private network, you can look like that you live in Amsterdam, London or any other city.

Security

Security is the prime reason for which the corporations have been using the services of VPNs for years. There are various simple ways on which data can traveling to a network can be intercepted. Firesheep and Wi-Fi spoofing are the two easiest ways in which information can be hacked. For a better understanding of the concept of VPN, a firewall helps in protecting a system along with data on the computer while a VPN helps in protecting all forms of data on the internet or on the web. VPNs cater with the help of various advanced forms of encryption protocols and techniques of tunnel security for the purpose of encapsulating all the transfers of online data.

The most computer savvy users will never connect to the internet without a proper firewall and updated system of antivirus. The evolving number of security threats along with the increase of reliance on the internet, it has made virtual private network a very important part of a well-designed security infrastructure. The checks of integrity ensure that no form of data is lost along with ensuring that the connection which has been established is not hijacked. As all the traffic gets protected, VPNs are always preferred more than the proxies.

Setting up a VPN

The setting up of a VPN is a pretty simple job. It is most often as easy as entering the username and password. The dominant nature of smartphones can easily configure VPNs by using the L2TP/IPsec and PPTP protocols. All forms of major OS can configure VPN PPTP connection. L2TP/IPsec and OpenVPN protocols need a small application which is of open source nature and the certifications to be separately downloaded.

Protocols of VPN

The available pool of protocols along with the features of security tends to grow with the flow of time. The most widely found protocols are:

- PPTP: This form of protocol has been in the world of VPN since the early days of Windows 95. The major advantage of PPTP is that it can be easily set up on any form of major OS. In simple words, PPTP helps in tunneling point to point connections over the protocol of GRE. However, the security concerning PPTP has been recently called out into several questions but it is strong enough although it is not the one which is the most secure of all.
- L2TP/IPsec: L2TP/IPsec is much more secure when compared with PPTP and it also comes with several other features. L2TP/IPsec is the method of implementing two different types of protocols all together in proper order for gaining the overall features of all. For instance, the protocol of L2TP is being widely used for creating tunnel and IPsec protocol helps by providing a secure form of channel. These features of the protocols make them a highly secure form of package.
- OpenVPN: OpenVPN is a virtual private network which is based on SSL and is gaining popularity day by day. The software which is being used for this protocol is open source in nature and is also highly available. SSL is a form of mature protocol concerned with encryption. OpenVPN can easily run on any single TCP or UDP port and thus makes this extremely flexible in nature.

How can VPN help?
The concept behind the working of a VPN is quite simple. It helps in connecting PC, smartphone or any other form of device with another computer or server directly on the platform of the internet. It also allows users to surf the contents which are available on the internet by using the same internet connection of the computer. So, when the computer system

with which you are connecting to for the purpose of internet surfing is from some other country or region, it will be showing that the user who is connecting is also from the similar country as the server. So, VPN can actually help you in connecting with all those sites with which you normally can't. You can use VPN for several tasks such as:

- Bypassing all the restrictions on those websites whose access are restricted only according to geography, mainly for the purpose of streaming audio and video.
- It can help in protecting the users while connecting to any form of unknown Wi-Fi hotspot.
- You can watch online streaming of media directly with the help of VPN such as Netflix and Hulu.
- You can gain a considerable amount of privacy online as VPN helps in hiding the actual location of your system.
- It can help you by protecting your system from scans at the time of using torrent.

The use of VPN today is mostly found for bypassing the restrictions on geography for the motive of watching various forms of restricted contents on the internet simply by taking into use the network of some other torrent or country. It is really helpful at the time of accessing public form of Wi-Fi like the ones which can be found at the coffee shops.

How can you get a VPN for yourself?

Getting a VPN is not that tough and you can get it for yourself depending on your needs. You can start by creating a VPN server for yourself or you can also VPN server. In case you want to create a VPN for your workplace you can do that as well. However, in most of the cases, the use of VPN can be found for the surfing of those contents which are restricted according to the geography of an area such as for torrent which has been banned for many regions and countries. You can download VPN online if you require it only for bypassing the restrictions.

How does VPN work?

When you connect any device such as tablet, smartphone or PC with the VPN, your system of device will start imitating like the local network of the VPN. The traffic of the network will be sent through a secure form of connection directly to the VPN. As the system of the user starts behaving like it is from the same network, the user can easily access all the resources from the local network when the user is seating at some other point of the world. You can also use VPN for imitating as if you are at the same location as of the VPN network. Such a feature gets into play at the time of accessing the websites which are geo-restricted.

As you start surfing the internet after getting connected with your desired VPN, your device will be establishing connection with the website through the connection of the VPN which stays encrypted throughout the connection. The VPN will be carrying forward your request to the website server and will also bring back the response via the same channel.

VPN and its uses

VPN has turned out to be a hot topic in the recent years, especially after the restriction of various websites and contents according to the geography of an area. VPN can be used for various jobs. Let's have a look at some of the most basic uses of VPN.

- You can access your business network at any time while you are on the move. VPN is being used by all the travelers who need to travel around for the purpose of business. Such people need to access the resources of their business network. VPN can be used for business network access along with access of the local network resources at the time of travelling. The local network resources are not needed to be exposed to the internet directly and thus it helps in improving the all-round security of the connection.
- You can also use VPN for accessing your home-based network while you are travelling. For this, you will need to create a VPN for accessing your personal network while travelling. This will

allow you to access a kind of remote desktop access which is possible directly over the internet with the use of VPN. The users can use this feature for sharing the local area files, for playing online games by imitating that your device is also on the similar network as of the VPN.

- You can use VPN for hiding your activities of browsing along with ISP. Suppose you are using a Wi-Fi network which is public in nature. When you browse anything by using such network, the websites which are not of HTTPS nature will be easily available for all those users who are also using the same network in case they know how to carry on such activities. While using public Wi-Fi, it is always safe to hide all your activities of browsing as it also provides a great amount of privacy on the network. VPN can be used for such purposes. Anything that you request over the internet will be passing through the VPN connection and thus providing you great amount of privacy. This technique is also useful for the purpose of bypassing the connection monitoring by the ISP.

- VPN is widely used today for the purpose of bypassing of censorship which can be found widely over the internet. With the use of VPN, you can use the firewall of the local area network and then access the internet with the firewall of the VPN network.

- You can browse all those websites which are geo-blocked with the help of VPN. VPN can help in easily accessing all those websites which are restricted for several regions and countries. You can also use VPN for watching online streaming of media when you are not in your country like Hula, Netflix and several others. VPN is also used for file transfers.

Chapter 15: Firewall

As the rate of cyber crime tends to increase day by day which has turned out to be a threat for most of the businesses all over the world, firewall security is the ultimate thing that can help for securing your organization. The term firewall can actually be compared with a physical form of wall which can help in preventing all forms of unwanted parties across it. Firewall security in the world of computer works like a network device which helps in blocking various forms of network traffic and thus creates a huge barrier between the untrusted and trusted form of networks. It is also relatable to the physical walls in the sense that it tries to block the spread of computer attacks.

Firewall and its types

With the increase in the percentage of cyber attacks, the types of firewall are also evolving with time. There are several types and forms of firewalls that can be found today. Let's have a look at some of them.

Stateful firewall

Stateful firewall is a type of firewall which is somewhat intelligent by nature. It helps in keeping a detailed track of all the connections which are in the active state to make sure that the user can easily customize the firewall management rules in a way which will allow the return packets which are in real a part of the established form of connection. However, this form of wall cannot differentiate in between the bad and good form of network traffic. It comes with prevention form intrusion followed by a complete blockage of the harmful web attacks.

Packet filtering firewall

This form of firewall is somewhat similar to that of the stateful firewall. It comes with various rules for the security of firewall and it comes with the capability of blocking that traffic of internet which is based on the port numbers, IP addresses and IP protocol. The only bad thing about this type of firewall is that it allows all forms of network traffic including the ones which can actually call about an attack. In such cases, the users of such firewall require intrusion prevention with firewall security. By this method, it will be able to filter out the bad and good web traffic. Packet filtering firewall cannot also differentiate between the authentic form of return packet and the one which imitates the actions of a legitimate data packet. So, it is evident that the packet filtering form of firewall will be allowing all forms of return packets within your network.

Application-aware firewall

This form of firewall is capable of understanding the different forms of protocols and also defines the same for the purpose of addressing the particular sections of the protocol by the signatories or rules. It helps by providing a flexible form of firewall protection for the systems of computers. It also permits the rules to be both particular and comprehensive at the same time. It helps in improving the overall functioning of the deep packet form of inspection, however, there are certain types of attacks which might not be noticed by this firewall because the routines defining of the firewall is not strong enough for managing the actual traffic variations.

Deep packet inspection

This is a type of firewall which helps in examining the packets of data in real. It also looks after the types of attacks over the application layer. This type of firewall comes with the capability of performing various functions in relation to the prevention of intrusion. This form of firewall comes along with three forms of different admonitions. At first, the definition of deep

inspection might extend to particular depth for some of the vendors within the data packets and thus will not be examining the entire data packet. This can also conclude in skipping out some of the most dangerous forms of attacks.

Secondly, this type of firewall depends greatly on the form of hardware. So, the hardware involved in a system might not come with the required power of processing the deep inspection of the data packets. You will need to ensure the capacity of bandwidth which the firewall can manage easily while inspecting the packets. Lastly, the technology which is related to the firewall management might not be having the needed percentage of flexibility for managing all the attacks.

Application proxy firewall

This form of firewall might perform as the mediator for various forms of applications like HTTP or the web traffic which intercepts each and every request. It also validates each of them just before allowing them entry. This form of firewall comes with some prime features of intrusion prevention. It is, however, difficult to apply this type of firewall in its complete state. Each of the proxies can handle only one protocol just like the incoming form of web or email. In order to get the ultimate protection of this firewall, it is required to accept all the protocols for the purpose of getting ahead with the protocol violation blocking.

Firewall security and its importance

In this world of today where cyber attacks can take place any time, firewall security is of utmost importance for all the servers and computer systems. The prying eyes are always looking around for the susceptible form of devices which remains connected with the internet. The devices which connect with the internet can be easily attacked by the hackers by implementing any form of harmful code or malware into the device via the

gateway of the internet. Malware attack can result in breaching of data and exploitation as well. Firewall is really important in such situations as:

- It helps in protecting the systems of computers from all forms of unauthorized access.
- It helps in identifying harmful content and also blocks the same.
- It helps in establishing a secure working environment for a network which gets used by several people at one time.
- It helps in protecting all types of confidential and personal data.

Chapter 16: Cryptography

With a sudden increase in the percentage of cyber attacks, it is has turned out to be really important to protect all sorts of sensitive data to the maximum extent as possible. Leakage of data in this world of today might result in some serious losses for many of the businesses or might also result as a threat for someone individual like stealing bank details, credit card details, login ids and passwords and many more. Cryptography is the process which is which to convert simple and plain text into a form which is unintelligible in nature. This technique makes the task of storage as well as transmission of confidential data super easy. Cryptographic texts can only be read by that person who is meant to get the message and read the same. It helps in data protection as well as in authentication of data.

Cryptography is often linked with the security of all sorts of information which also includes the techniques for communication and those which are derived from the mathematical concepts. The technique uses a particular set of rules along with calculations which are also called algorithms. They are used for message transformation into such a form that it turns out to be

super tough to decipher the message. The algorithms are also used for key generation of cryptography along with digital form of signing for the purpose of data privacy, securing website browsing and for the sensitive forms of communication such as credit card transactions, bank details and email.

Cryptography and its techniques

The technique of this process is also linked along with features of cryptology and cryptanalysis. The technique uses various other techniques which include words merging with pictures, usage of mircodots and various other steps which helps in hiding out the information which is to be transported over a network or is meant to be stored in the same. The plain form of text is converted into a coded form of text which is often called ciphertext. It is done with the process of encryption. It can be deciphered with the process of decryption at the receiver end.

Cryptography and its objectives

Cryptography comes with various objectives. Let's have a look at them.

- Cryptography comes with the goal of maintaining data integrity. The piece of information or data which is to be transmitted between the sender and the receiver or which is meant to be stored in the network cannot be altered or changed in any way. In any case, if such things happen, both the parties in communication get notified.
- It also comes with the objective of protecting all forms of sensitive as well as personal data. Its main aim is to secure the data for all the concerned individuals. The data or information which is to be transmitted across a network or to be stored in the same cannot be analyzed by any other third party out of the network.

- The creator and sender of the message will not be permitted to step back from their intention at a much later stage at the time of transportation or creation of data. This act is known as non-repudiation.
- Both the sender and the receiver will be able to confirm the identity of each other before sending out and before receiving the information.

Cryptography and its algorithms

The system of cryptography functions with the help of a set of procedures which are also known as cryptographic algorithms or ciphers. These are used for the purpose of both encryption as well as decryption of message for the motive of protecting and securing the process of communication among various devices, computer systems as well as applications. One cipher suite uses up three forms of algorithms. It uses one algorithm for the process of data encryption. The second algorithm is being used for the purpose of message authentication. And, the third algorithm is used for the process of key exchanging. This entire process remains embedded in the protocols and is also written within the programming language of the software which is used on the operating system together with the systems of computer which are network-based in nature. It includes public generation along with the generation of the private key. The private key is required for both encryption and decryption of information, authentication of message as well as for digital form of signing with the program of key exchange. In simple terms, algorithms can be regarded as the core of cryptography.

Conclusion

As you have completed with the teachings of the entire book, now you can easily create a clear image about the processes which are linked with hacking. You will also be able to gain a lot of knowledge about the functioning of Kali Linux. By now, you must have created a clear perception of all the required tools and components which you need for creating a safe and secure network server for your business and also for personal use. You are the one who is responsible for everything. You alone can secure up an entire system and strengthen up the infrastructure of security.

With Kali Linux and all its tools, you can easily have complete control over the network security interface related to your business as well as personal network. This book is not solely about Kali Linux as you have also learnt a lot about some of the basic networking components with the security of the same. You can use all the tools from Kali Linux for securing your system. The prime benefit that you can get after using Kali Linux is you can also perform a wide range of tests related to system security. This will ultimately help in wiping out all sorts of susceptibilities and security gaps within your infrastructure of information technology.

What you can do for the security of your system and network depends completely on you. You are the one who can either make it or break it. Ensure that you start using all the steps which you have learnt in this book for securing your system.

BOOK 3: COMPUTER NETWORKING FOR BEGINNERS

The Complete Guide to Network Systems, Wireless Technology, IP Subnetting, Including the Basics of Cybersecurity & the Internet of Things for Artificial Intelligence

Download the Audio Book Version of This Book for FREE

If you love listening to audio books on-the-go, I have great news for you. You can download the audio book version of this book for **FREE** just by signing up for a **FREE** 30-day audible trial! See below for more details!

Audible Trial Benefits

As an audible customer, you will receive the below benefits with your 30-day free trial:

- FREE audible book copy of this book
- After the trial, you will get 1 credit each month to use on any audiobook
- Your credits automatically roll over to the next month if you don't use them
- Choose from Audible's 200.000 + titles
- Listen anywhere with the Audible app across multiple devices
- Make easy, no-hassle exchanges of any audiobook you don't love

- Keep your audiobooks forever, even if you cancel your membership
- And much more

Click the links below to get started!

[>> For Audible US](#)

[>> For Audible UK](#)

[>> For Audible FR](#)

[>> For Audible DE](#)

Introduction

Understanding the concept of computer is a potent endeavor to anyone who seeks to make good use of networking resources and services; in the current world where networking in all aspects of human interaction is heavily linked to the use of computers. From social networking through to digital media and business networking, the world has indeed become a global village thanks to the trillion-dollar idea of PC networking. In everything that we do, it is unavoidable to mention the Internet as the biggest fruit of PC networking. With a lot of strides having been made in computer networking, communication systems are becoming more and more efficient day in, day out.

In Computer Networking, readers get to familiarize with the fundamental aspects of computer networking starting with the very basic networking nitty-gritties that primarily aim at setting the tempo towards more advanced concepts of computer networking. Though briefly explained, the information provided for every topic is sufficiently covered to suffice the essential needs of a networking beginner.

The first chapter introduces the reader to the basic concepts of computer that include a simple description of a computer network, network topology and network types. In addition, there is also a brief overview of wireless networks as well as deep discussion into network security.

In subsequent chapters, other important networking topics come to the fore including wireless technology, peer-to-peer networking, router and server basics and IP addressing. Internet basics and functions are also covered with special attention on the worldwide web. Furthermore, the book explores an overview of the CCNA course coverage as well as the essentials of machine learning.

The book introduces the reader to the exciting networking experience: it stirs up readers' curiosity to venture further into the more advanced coverage of computer networking by offering a firm foundation in the larger field of PC networking.

Chapter 1: Computer networking: An Introduction

Networking Essentials

What is a computer network?

Computer network is a term that refers to any group of computers that are linked to one another allowing for communication among them. A network also allows member computers to share applications, data and other network resources (file servers, printers, and many others).

Computer networks may be differentiated according to size, functionality and even location. However, size is the main criterion with which computer networks are classified. Thus, there are a number of unique computer networks that we shall discuss individually under the subtopic *Types of Computer Networks.*

Network Topologies

A network topology refers to the arrangement and the way components of a network are interconnected. Two forms of computer network topologies exist:

Logical Topology

Logical topology defines how linked network devices appear in a network. It is the architectural design of a network's communication mechanism among the different devices.

Physical Topology

Physical topology can be said to be the way all the nodes of a network are geometrically represented. The following are the various types of physical topologies:

Bus Topology

In this topology, all hosts on a network are linked via a single cable. Network devices are either linked directly to the backbone cable or via drop cables.

When a node wants to relay some message, it relays it to the entire network. The message is received by all the network nodes regardless of whether it is

addressed or not.

This topology is primarily adopted for 802.4 and 802.3 (Ethernet) standard networks.

Bus topology configuration is simpler in comparison with other topologies. The backbone cable is a "single lane" through which messages are relayed to all the nodes on the network.

Bus topologies popularly rely on CSMA as the primary access method. CSMA is a media access control that regulates data flow in order to maintain data integrity over the network.

Advantages of Bus Topology

- The cost of installation is low since nodes are interconnected directly using cables without the need for a hub or switch.

- Support for moderate data speeds by use of coaxial and twisted pair cables that allows up to 10 Mbps only.

- Uses familiar technology which makes its installation and troubleshooting a walk in the park since tools and materials are readily available.

- There is a great degree of reliability since failure of a single node does has no effect on the rest of the network nodes.

Disadvantages of Bus Topology

- Cabling is quite extensive. This may make the process quite tedious.

- Troubleshooting for cable failures is mostly a pain to most network administrators.

- Chances of message collision are high in case different nodes send messages simultaneously.

- Addition of new nodes slows down the entire network.

- Expansion of the network causes attenuation-loss of signal strength. This may be corrected with the use of repeaters (to regenerate the signal).

Ring Topology

The only difference between ring topology and bus topology is that in the former the ends are connected; while in the former, ends are open.

When one node gets a message from the sender, that node sends the message to the next node. Hence, communication takes place in one direction-it is unidirectional

Each and every single node on the network is linked to another node without a termination point. Data flows continuously in one loop-endless loop.

Data flow always takes a clockwise direction.

Ring topology often use **token passing** as the main access method.

Token passing: an access method in which tokens are passed from station to another.

Token: a data frame that moves around the network.

Token Passing in Action

- A token moves around the network-from one node to another till the destination.

- The sender puts an address plus data in the token.

- The token passes from one node to the next-checking the token address against the individual addresses of each and every node on the network until it finds a match.

- The token is used as a carrier-for data (and the destination address).

Merits of Ring Topology

- Network management is relatively easy since faulty components can be removed without interfering with the others.

- Most of the hardware and software requirements for this network topology are readily available.

- The installation cost is quite low since the popular twisted pair cables that required in plenty are quite inexpensive.

- The network is largely reliable since it does not rely on a single host machine.

Demerits of Ring Topology

- Troubling may be quite a task in the absence of a specialized test equipment. Detection of a fault in cable is normally a serious challenge.

- Failure in one node leads to failure in the entire network since tokens have to through each node for a complete cycle of communication from sender to destination.

- Addition of new network devices slows down the entire network.

- Communication delay increases with increasing nodes/network components.

Star Topology

In this topology, a central computer, switch or hub connects all the nodes on the network. The central device is the server while the peripherals are clients.

Coaxial cables or Ethernet's RJ-45 are favored for connection of the network nodes to the server. Switches are hubs are preferred as the main connection devices in this topology.

This is by far the most widely used topology in network implementations.

Pros of Star Topology

- There is ease of troubleshooting since problems are handled at individual stations.

- Complex network control features can be implemented with ease at the server side-which also allows the automation of certain tasks.

- There's limited failure since an issue in one cable does not translate into an entire network problem. The fault in a cable may only affect a single node on the network since the nodes are not interconnected via cables.

- Open ports on a switch or hub allow for easy expansion of the network.

- The use of inexpensive coaxial cables makes star topology highly cost effective to implement.

- It has the capacity to handle data speed of up to 100Mbps. Hence, it supports data transmission at very high speeds.

Cons of Star Topology

- If the central connecting device fails of malfunctions, then the entire network goes down.

- Use of cabling at times makes routing an exhausting exercise-cable routing is normally difficult.

Tree Topology

This topology puts the features of bus and star topologies in one basket.

In this topology, all computers are interconnected, but in a hierarchical manner.

The top-most node in this topology is referred to as a **root node** whereas all the others are descendants of the root node.

There exists just a single path between two nodes for the transmission of data-forming a parent-child hierarchy.

Merits of Tree Topology

- It supports broadband data transmission over long distances without issues of attenuation.

- Star topology allows for easy expansion of a network since new devices can be added without to an existing network with little difficulty.

- Ease of management-networks are segmented into star networks that make it relatively easy to manage.

- Errors can be detected and corrected with ease.

- Malfunctioning or breakdown of a single node does not affect the other nodes on the network. Thus, there is limited failure on tree topology networks.

- It supports point-to-point wiring of each and every network segment.

Demerits of Tree Topology

- It is always difficult to handle issues in respect of a fault in a node.

- It's a high cost network topology since broadband transmission can cost an arm and a leg.

- Failure or faults in the main bus cable affects the entire network.

- There is difficulty in reconfiguring the network when new devices are added onto the network.

Mesh Topology
All computers are interconnected via redundant connections in this topology. It offers different (multiple) paths from one node to another.
In mesh topology, there are no connecting devices like switches or hubs. For instance, the Internet.

WANs normally are implemented with mesh topology since communication failures are of serious concern. It is also largely implemented in wireless networks.

The formula for forming mesh topology is shown below:

Number of Cables = (z*(z-1))/2

Where;

z = the number of nodes on the network

There are 2 categories of this topology:

Partially Connected Mesh

In this topology, not all the network devices are linked to the devices with which they have frequent communications. The devices are only connected to some of the devices with which they are normally in constant communication.

Full Mesh Topology

Each network device has a link to every other device in the network in full mesh topology. In simple words, all computers are connected to one another via the redundant connections.

Merits of Mesh Topology

- Mesh topologies are highly reliable since a breakdown in one single connection does not affect the working of the nodes in the network.

- Communication is fast since each computer has connections with all other computers on the network.

- Addition of new devices has no effect on other devices on the network-making reconfiguration quite easy.

Demerits of Mesh Topology

- Mesh topology networks has the capacity to accommodate more devices and transmission media than any other network

topology. This translates to a high cost of setting up mesh networks than all other networks.

- Mesh topology networks are normally too large to manage and maintain effectively.

- A lot of redundancy on the network reduces the network efficiency significantly.

Hybrid Topology

The amalgamation of different network topologies (at least two of them) result in another topology that is conventionally referred to as hybrid topology. It is a connection among different links and computers for data transmission.

A hybrid can only be formed by a combination of dissimilar topologies. For instance, a combination of bus and star topologies. However, a combination of similar topologies does result in a hybrid topology.

Advantages of Hybrid Topology

- An issue in one part of the network does not mess with the entire network.

- Hybrid topology allows network to be scaled further by addition of more devices without messing with the existing network.

- This network topology is quite flexible. An organization can customize the nature of its network to suit its specific network needs and interests.

- The network topology is highly effective since it can be designed in a way that network strength is maximized, and the limitations of the network are minimized.

Disadvantages of Hybrid Topology

- The network topology is quite complex. Thus, it is too difficult to come up with a suitable architectural design of a network.

It is highly costly since hubs used in this sort of computer network are different from the ordinary hubs. The hubs used in this topology are more expensive. Besides, the overall infrastructure is highly costly since a lot of cabling is required plus many more network devices.

Networks Types

The following are the four major classifications of computer networks based on size:

Local Area Network (or simply, LAN)

A LAN refers to any group of computers that are linked to one another a small area like an office or a small building. In a LAN, two or more computers are connected via communication media like coaxial cables, twisted pair cable or fiber-optic cable.

It is easy and less costly to set up a LAN since it can do just fine with inexpensive network hardware such as switches, Ethernet cables and network adapters. The limited traffic allows for faster transmission of data over LANs.

Besides, LANs are easy to manage since they are set up in a small space. Thus, even security enforcement is also enhanced through closer monitoring of activities within the network's geographical location.

Personal Area Network (PAN)

This network is arranged and managed in the space of its user(s)-normally a range not exceeding 10m. It is typically used to connect computer devices for personal use.

Components of a personal area network include a laptop, mobile phone, media player devices as well as play stations. Such components are located within an area of about 30ft of a person's space.

The idea of PANs was born by one Thomas Zimmerman-the first lead research scientist to conceive the idea of personal area networks.

There are 2 classes of PANSs:
- **Wired PANs:** a wired personal area network is created when a person uses a USB cable to connect two different hardware devices. For instance, it is common practice nowadays to

connect a phone to a computer via a USB cable-to share files, access the Internet, and many other things.

- **Wireless PANs:** a wireless PAN is set up by the use of existing wireless technologies such as Bluetooth and Wi-Fi. This is basically a low-range technology network type.

Examples of PANs

There are 3 common types of personal area networks:

1. **Body Area Network:** it moves with an individual. A good example is a mobile network-when one establishes a network connection and then makes a connection with a different device within their range.

2. **Offline Network:** it is also called a home network. It can be set up in a home-linked computer, TV, printers and phones-but is not connected to the internet.

3. **Small Home Office Network:** different devices are connected to the Internet and corporate network via VPN.

Metropolitan Area Network (or simply, MAN)

A MAN is a type of network that extends over a larger geographical area by interconnecting different LANs to form a bigger network of computers. Thus, it covers a wider area than a LAN.

MANs are ideally set up in cities and big towns. Hence, the name metropolitan area network. It is often used by government agencies to connect with citizens some big institutions; communication among banking institutions within a given city; in big institutions of higher learning located in a metropolis; and even used for communication in military bases within a city/town.

The commonly adopted Metropolitan area network protocols include Frame Relay, ISDN, RS-232, ADSL, ATM and OC-3, among others.

Wide Area Network (or simply, WAN)

This is a network that stretches over large geographical regions-cities, states and even countries. It is bigger than LAN or MAN. It is not restricted to a particular geographical location. It spans over large geographical locations by the use of telephone lines, satellite links or fiber optic cables. The Internet is a perfect example among the existing WANs globally.

WANs are widely embraced for education, government and business activities.

WAN Examples

Mobile Broadband: 3G or 4G networks are widely serving people in a big region, state or even country.

Private Network: banks create private networks that link different offices established in different locations via a telephone leased line that's obtained from a telecom company.

Last Mile: telecommunication companies offer internet services to thousands of customers in different cities by simply connecting homes, offices and business premises with fiber.

Advantages of WANs

- WANs cover large geographical locations reaching out to masses of the human population. The impact of the Internet in people's lives globally sums up the advantages of a wide area network.

- Centralized data: WANs support centralization of data/information. This eliminates a need for individuals to buy back-up servers for their emails and files.

- Getting updated files: programmers get updated files within seconds since software work on live servers.

- Quick exchange of message: WANs use technologies and sophisticated tools that enable message exchange to happen

faster than on most other networks. Communication via Skype and Facebook are two good examples of quick message exchange, thanks to the Internet-one of the popular WANs in the world.

- WANs allow sharing of resources and software. It is possible to share hard drives, RAM, and other resources via wide area networks.

- Business without borders: presently, even people separated by the Pacific can still conduct thriving business without moving an inch from their current location because of the Internet. The world is indeed a global village.

- High bandwidth: use of leased lines for companies increases bandwidth. This in turn increases data transfer rates thereby increasing productivity of the company.

Disadvantages of WANs

- Security issues are escalated as the network size increases. Thus, the issue of insecurity is of more concern on a WAN than it is on a LAN or MAN.

- High installation cost: setting a WAN requires the purchase of many, costly equipment as well as software applications to manage and administer the network. Routers, switches and mainframe computers that are needed to serve the network all cost a fortune.

- Network troubleshooting is often a big concern since the network spans large geographical locations.

Neural networks

A neural network refers to a set of algorithms that are used for pattern recognition. The algorithms are loosely molded after the brains of humans.

Neural networks are most crucial in the interpretation of sensory data via machine perception, clustering and labelling raw data.

A neural network recognizes numerical patterns that are found in vectors. All data-time series, images, text and sound-are translated in the vectors.

Functions of Neural Networks

The following are the 3 most vital roles that can be performed by neural networks:

Classification: labeled data sets are the key factors in any classification. Humans have to transfer their own knowledge to the labeled dataset so as to enable a neural network to get the correlation data and the labels in a process known as supervised learning.

A neural network can achieve the following classification tasks:

- Face detection, people identification in images and recognition of facial expressions.

- Object identification in images.

- Voice recognition, speech transcription into text, sentiment recognition in voice, and speaker identification (in a dialogue/conversation).

- Text classification-fraudulent text in insurance claims and spam in emails.

Clustering: this is also referred to as grouping. It involves the detection of similarities. This can be achieved without labeling in a process known as unsupervised learning. The process involves the following:

- Search: images, documents or sounds are compared to obtain related items.

- Detection of anomalies: clustering also seeks to detect unusual behavior among grouped data. This is highly essential in the

detection and prevention of certain undesirable items and activities such as fraud.

Predictive Analysis: this is commonly known as regression.

Deep learning relies on data classification for the establishment of correlations between objects. This can be simply be referred to as static prediction.

Deep learning has the ability to establish correlations between current and future events.

Predictive analysis is most crucial when it comes to the following:

- Health breakdowns.

- Customer churn.

- Hardware breakdowns.

- Employee turnover.

Elements of a Neural Network

Deep learning is another term for 'stacked neural networks'.

Stacked neural networks are networks that are made up of a number of layers.

Each network layer is made up of different nodes-a computational point patterned on the neuron of the brain of a human.

A node puts together all data inputs with the weights (set of coefficients). The weights can either dampen or amplify the input. This, in turn, gives significance to the inputs that concern the task on which a given algorithm is trying to pick up.

In summary, the following are the key elements of a neural network:

- Layers

- Nodes

- Set of coefficients (weights)

The OSI Model

OSI is short form for Open System Interconnection. The model offers a description of the way information and data from a software application is transmitted through a physical media to another software application in a totally unrelated computer.

This reference model is comprises of seven layers. Each layer has a specific role to play.

The OSI Reference model was born in 1984 by the International Organization (ISO). In modern days, this is taken to be the basic architectural model for communication between different network hosts.

In the OSI model, whole tasks are broken down into 7 smaller and manageable chunks. Layers are assigned distinct roles-each layer is assigned a specific task to handle. Also, each layer is sufficiently equipped to handle its tasks independently.

Features of the OSI Model

The OSI model is broadly divided into 2 layers: upper and lower layers.

The upper layer of this model primarily handles issues related to applications. Those issues are executed in the software. The layer that is closest (or the uppermost layer) to the user is the application layer. The end user interacts with software application just as the application software does.

When a layer is said to be an upper layer, it is said so in reference to another. An upper layer is a layer that lies right above the other one.

The lower layer of this model handles issues of data transport. The implementation of the data link as well as physical layers occurs in software and hardware. In this model, the physical layer stands as the lowest layer. It is also the nearest to the physical medium. Primarily, the physical layers provide the necessary information to the physical medium.

Roles of the 7 Layers of the OSI Model

We're going to focus on the functions of the unique layers of the OSI Reference model from the lowest to the uppermost.

Physical Layer

- **Data Transmission:** it defines the mode of transmission between two network devices-whether it is full-duplex, half-duplex or simplex mode.

- **Line Configuration**: it offers a clear definition of the way two or more network devices are physically linked.

- **Signals**: the physical layer determines the nature of signals used to transmit information.

- **Topology**: the physical layer offers a comprehensive definition of the arrangement of network devices.

Data Link Layer

This layer is charged with the task of ensuring error-free data transfer of data frames over the network. It also defines data format on the network.

The data link layer ensures that there is reliable and efficient inter-network device communication. It is responsible for the unique identification of each device that is found on the network.

Data link layer comprises of the following two layers:

1. **Logical link control layer:** it transfers packets to the destination's network layer. Besides, it identifies the specific address of the network layer of the receiver from the packet header. Furthermore, flow control is implemented in this layer.

2. **Media Access Control Layer:** this refers to a link that exists between the physical layer and link control layer. This is what transfers data packets over a network.

The Data Link Layer's Actual Functions

- **Framing**: the data link layer does the translation of the physical layer's raw bit stream into data packets referred to as frames. It adds a header and trailer to the data frame. The header contains both the destination and source addresses.

- **Physical addressing:** the physical addressing layer enjoins a header to the frame. This header has the address of the receiver. The frame is relayed to the receiver whose address is indicated on the header.

- **Data Flow control:** this is the data link layer's primary role. It maintains a constant data rate so that no data is corrupted while on transit.

- **Error control:** this is achieved by addition of a cyclic redundant check (CRC) on the trailer that is put onto the data packet before being sent to the physical layer. In case of any errors, the receiver can request for the retransmissions of the corrupted frame.

- **Access control:** this layer determines which of the available network devices is given top priority over the link at a particular moment.

The Network Layer
This is number 3 on the 7-layer OSI Reference model. It handles device addressing and keeps track of device location on the network. Based on network conditions, the layers determine the most favorable path for data transfer from sender to receiver. Another condition that is considered in determining the best path is service priority, among others.
This layer is charged with the responsibility of routing and forwarding packets. Routers some of the devices on layer 3. The routers are specified in the network layer and are used to offer routing services in a computer internetwork.
Protocols that are used in the routing of network traffic include IPv6 and IP.

Network Layer Functions

- **Addressing**: this layer ensures that the destination and source addresses are added to the header of the frame. Addressing is helpful in the identification of devices on a network.

- **Internetworking**: the network layer offers a logical link between network devices.

- **Packetizing**: the network layer receives frames from upper layers and turns them into packets in a process that is conventionally referred to as packetizing. It is realized by the Internet protocol.

The Transport Layer

This is layer number 4 in the model.

The layer ensures that follows the order in which they are sent. It makes sure that duplication of data does not occur. This layer's core business is to ensure that data is transferred totally.

The physical layer receives data from the upper layers and subdivides them further into smaller chunks that are referred to as segments.

The layer provides communication between destination and source-from end to end- for data reliability. It can also be termed as end-to-end layer.

There are two protocols that are implemented at this layer:

Transmission Control Protocol

This is a standard protocol which allows systems to share messages/information over the internet. The protocol establishes and preserves the link between hosts.

TCP divides data into smaller units referred to as segments. The resulting segments do not travel over the internet using the same route. They reach the destination in no specific. However, TCP reorders the individual segments at the destination to reconstitute the original message.

User Datagram Protocol (or simply, UDP)

This is also a transport layer protocol. As opposed to what happens in TCP, the source does not receive any acknowledgement when the destination receives data. This renders the protocol quite unreliable.

Transport Layer Functions

- **Service-point Addressing:** service-point addressing enables computers to run multiple applications simultaneously. It also allows data transmission to receiver not only from one machine to another machine, but also from one process to the next process. The transport layer adds a port address or service-point address to the packet.

  Whereas the network does the transmission of data from one machine to another, it is the transport layer that ensures data transmission to the appropriate processes.

- **Segmentation and reassembly:** this layer receives a message from its upper layer. It then splits the whole message into several small chunks. The layer assigns sequence numbers to each segment for identification.

  At the receiving end, the transport layer reconstitutes the segments based on the sequence numbers to form the original message.

- **Connection control:** there are 2 services that the transports offers: connectionless service and connection based.

  A connectionless service considers each segment to be a distinct packet. The packets travel through different routes to the destination. On the other hand, the connection-based service makes a connection with the destination machine's transport for before packets are delivered. In the connection-based service, all packets move in a single route.

- **Error control:** Just like in data control, this is achieved on an end-to-end basis-not across a single link. Transport layer at the source ensures that the message gets to its destination error-free.

- **Flow control:** this layer also ensures data control. The data control is done from end to end, but not across a one dedicated link.

The Session Layer

This layer is used for the establishment, maintenance and synchronization of interaction between communicating network devices.

Session Layer Functions

- **Dialog control:** this layer serves as a dialog controller. The layer achieves by initiating dialog between two processes. Alternatively, the layer can be said to authorize communication between one process and another. This can either be half-duplex or full duplex.

- **Synchronization:** the session layer adds checkpoints in a sequence during data transmission. In case of errors along the way, retransmission of data takes place from the specific checkpoint. The entire process is referred to as synchronization and recovery.

The Presentation Layer

This layer primarily deals with the language and formatting of information that is transferred between two network devices. It is the network's "translator."

The presentation layer is a section of the operating system. It is the portion of the operating system that does the conversation of data from a given presentation format to another presentation format.

This layer is also referred to as the syntax layer.

Role of the Presentation Layer

- **Translation**: processes in different systems exchange information as character numbers, character strings, and many more. Different encoding techniques are applied on different computing machines. It is the presentation layer that handles interoperability between then unlike encoding techniques.

The layer does the conversion of data from the sender-based formats into common formats into receiver-dependent formats at the destination computers.

- **Encryption**: the presentation layer performs encryption to ensure the privacy of data.

Encryption is the process that involves the conversion of information transmitted from the sender into another unique form that is then transmitted over the network.

- **Compression**: the presentation compresses data before its transmission. The compression involves the reduction of the number of bits. This is process is essential especially in the transmission of multimedia such as video and audio files.

The Application Layer

This layer offers the interface for users and applications to access resources on the network. It handles network issues like resource allocation, transparency, and many more. This is not an application. It simply plays its application layer role. It provides network services to end-users.

Application Layer Functions

- **Access, transfer, and management of files:** this layer allows users to access files remotely, retrieve them and still manage them remotely.

- **Mail services:** this layer offers an email storage and forwarding storage facility.

- **Directory services:** this layer offers the distributed database bases. This is essential in the provision of important information about different objects.

Computer Network Components

These comprise of network hardware and network software components that constitute a computer network. In this section, we're typically concerned with the major hardware components that are crucial for the installation of a computer network.

Computer network components include computers, cables, network interface card (NIC), switches, modems, hubs and routers.

Computers

Computers may be desktop computers, laptops as well as portable devices (smart phones and tablets) plus their additional accessories such as portable hard drives, CD Players, keyboards and mice. They are the major hardware components of any computer network.

Computers are the primary components without which a network is just but a dream. Computers offer the platform for users to perform their different tasks on the network. In case of a centralized system, computers serve as a link between users and the dedicated network server.

The Network Interface Card

The NIC (as it is commonly called) is a hardware component that links one computer to another on the same network.

The NIC supports network transfer rates from 10Mbps through to 1000Mbps.

All network cards have a unique address assigned by the IEEE. These unique addresses are referred to as the physical/MAC addresses, and are used to recognize each computer on the network.

There are two unique forms of network cards: Wireless and Wired NICs.

Hub

A hub divides a network connection into several devices. A hub connects all computers on a network via cables. Every computer sends a request to the network through the hub.

When the hub gets a request from a particular computer, it broadcasts that request across the network to all network devices.

Each network device checks the request to determine if it belongs there. If not, the request is subsequently discarded.

The downside to this process is consumption of more bandwidth and communication is highly limited. Presently, a hub is as good as obsolete due to the hype with routers and switches.

Switch

A switch links a number of devices on a computer network. This important connection device is technologically more advanced than a hub.

A switch has an updated that determines the destination of transmitted data. The switch transmits a message to the desired destination as per the physical address on each incoming request.

Unlike the hub, it does not transmit data to all devices across the network. Thus, there is increased data transmission speeds since individual computers communicate directly with the switch.

Router

A router gives an internet connection to a local area network. It receives, analyzes and forwards incoming packets to another computer network.

It operates in Layer three of the OSI Reference model-simply referred to as the network layer.

Packet forwarding is based on the information contained in the routing table. A router is smart enough to choose or decide the most appropriate path for the transmission of data from all available paths.

Modem

A modem is an acronym that stand for Modulator/Demodulator. It changes digital data into analog signals over a telephone line.

The modem makes it possible for a computer to establish a connection to the Internet via an existing telephone line. It is installed on the PCI slot of the motherboard-not on the motherboard itself.

Connectors and Cables

A cable is a physical transmission medium which is used to transmit a signal.

The cables used for transmission include the coaxial cables, twisted pair cables and fiber optic cables.

Testing the Network Wiring

The first step to test the network wiring is to make a chart to help keep track of the testing progress. A floor plan showing where the wires are or a room-by-room listing will work fine. Check each one off when it tests okay.

Once the chart is done, get someone to help you, posting that person at the distant end of the cable from you. In a large building, use two-way handheld radios or cell phones to communicate; this will move the testing project along faster.

To test your cables, use a cable tester or, if you don't have a cable tester, a volt-ohm meter. If you are going to test a lot of end-point jacks, then borrow or invest in a cable tester. When the connections are good, with all four pairs of wires in the cable conducting properly, lights will illuminate on both ends of the tester. Having this tester makes checking a large number of cables an easy job.

If you opt to use a volt-ohm meter, you will need to build a patch cable that can be used to sort out the respective pairs on one end of the cable run so you can check for continuity on the other end. A shorting cord for the distant end of the cable connects together the appropriate pairs on RJ-11 or RJ-12 or on the Ethernet connector.

If you are going to use this more primitive method instead of a cable tester, you will need three shorted patch cables; one four-position plug will work for RJ-11 and 12. If you have RJ-25s, you will need to test them with a 6P6C plug; to test Ethernet cables, you will need an 8P8C plug with the appropriate pairs shorted out. You will then to need to fabricate another plug and patch cord with the wire stripped back 1⁄2 inch for connecting to

the leads on the ohm meter. Each pair should trigger the alarm on the continuity check or just read a few ohms of resistance, which will vary with distance. The resistance reading should not exceed 19.7 ohms over 100 meters of cable length.

Basic Network Troubleshooting

Network troubleshooting refers to all the measures and techniques assembled to identify, diagnose and resolve network issues. The process is systematic and primarily seeks to restore normalcy to the functionality of a computer network.

Network administrators are charged with the responsibility of identifying network problems and repairing it with the aim of ensuring a smooth run of operations in the network. They also do whatever it takes to ensure that the network is operating at optimal levels.

The following are among the many computer network troubleshooting processes:

- Configuration and reconfiguration of switches, routers or any other network component.

- Identifying any network issues and figuring out a way to fix it.

- Installation and repair of network cables as well as Wi-Fi devices.

- Getting rid of malware from the network.

- Getting firmware devices up to date.

- Installation and uninstallation of software as is necessary.

Network troubleshooting can be done manually, or as an automated task-especially when it has to do with network software applications. Network diagnostic software is a valuable tool when it comes to the identification of network issues that may not be easy to detect with the human eye.

Network troubleshooting includes both hardware troubleshooting and software troubleshooting.

Hardware Troubleshooting

This is a form troubleshooting that takes care of issues with hardware components. It may include:

- Removal of faulty or damaged RAM, hard disk or NIC.

- Dusting of computer and other network devices-dust accumulation sometimes leads to malfunctioning of devices.

- Tightening of cables that connect different network components.

- Updating or installation of important hardware drivers

Hardware troubleshooting begins with the discovery of a given hardware issue, the cause and, finally, taking the necessary remedial action.

Software Troubleshooting

Software entails a set of measures for scanning, recognizing, diagnosing and offering solutions to issues with software in the network. It includes issues with network operating systems, diagnostic software as well as software applications installed on individual network computers.

Chapter 2: Network Management

Effective network management must address all issues pertaining to the following:

- Hardware
- Administration and end-user support
- Software
- Data management

Hardware Management and Maintenance

Hardware maintenance can be performed as per the following routines and considerations:

Cleaning

Every two weeks clean all network equipment. Doing so will help keep your equipment cool and make other maintenance tasks easier to perform. When cleaning, dust the equipment, shelves, and nearby areas. A small vacuum should be used to vacuum keyboards and the computer vent and fan openings. Additionally, you should use the vacuum to gently suck dust out of removable media drives. Unused wall jacks and empty equipment jacks in dust-prone environments can be vacuumed on occasion as well.
For printers and plotters, follow the manual instructions for cleaning print heads on ink jets and vacuuming paper dust from laser printers. Monitors can be wiped down with eye-glass cleaning solution and glasses-cleaning cloths.

Performing Inspections

Keeping a close eye on the condition of all hardware is essential. For this reason, you should inspect all hardware at least once per month. This inspection should include the following:

- Make sure cooling vents are not blocked or excessively dusty.

- Listen to and feel the vents to make sure cooling fans are operating.

- Sniff the area. When power supplies and other parts are near failing, they may emit an odd odor from excessive heating. A burnt smell means trouble is imminent or has already occurred.

- Check all power cables, peripheral cables, and network cables for tightness in their sockets.

- Check all power cables, peripheral cables, and network cables for fraying or other damage.

- Check the server area for proper operation of heating, venting, and cooling systems to be sure they are operable- even if those systems are not needed at the time of the inspections.

Upgrading Firmware

"Firmware" refers to any program that is resident in a chip. For example, a computer's BIOS is firmware. Sometimes, manufacturers release updates for firmware to fix flaws or to enable the equipment to work with some newly released hardware device or operating-system upgrade. You should check the manufacturer's Web site or help desk for all network equipment at least quarterly to determine whether any firmware upgrades are available for your equipment.

If so, be sure to adhere to the maker's instructions to the letter for loading new firmware and firmware updates. Firmware loads often require low-level booting from a DOS or maintenance disk, although some will be compatible with the computer's operating system.

Upgrading Hardware

Two factors drive hardware upgrades:

- Performance issues due to changes in applications or the addition of new applications may necessitate a hardware upgrade or the addition of new features that are linked to the hardware's capability or capacity. For example, adding memory and installing an additional hard drive for more file space are typical upgrades performed to support those changes.

- You may opt to upgrade hardware on a purely optional basis-for example, adding a bigger monitor, higher-quality sound card, a TV card, or a similar device.

Repairing Hardware

As the person responsible for the network, you must assess your willingness and ability to perform hardware repairs-before a hardware component stops working. To that end, you should go through your entire hardware inventory and determine the following:

- Is the equipment still under warranty? If so, take advantage of that warranty in the event the equipment stops working.

- Would it be more cost-effective to simply replace a piece of hardware if it breaks? Given the high cost of technical labor, repairing a low-cost item, such as a printer that can be replaced for $50, may not be justified. It might even be best to replace rather than repair PCs purchased for less than $600 if you've used them for more than 10 months. Don't get me wrong: I am not advocating short equipment life cycles or unnecessarily adding to scrap piles.

- For big-ticket items, you may want to transfer the repair risk to someone else by arranging for service and support contracts-assuming your budget can support this.

Administration

Large networks often have one or more staff members dedicated exclusively to performing network administrative tasks. For smaller networks, the manager must wear various hats and perform multiple roles to support the network. Over time, he or she must rise to the level of journeyman-or at least experienced apprentice-to be successful.

Primary or routine network-administrative tasks fall into one of the following categories:

- Administering and supporting end users

- Adding workstations and peripheral devices

- Maintaining system-wide documentation

Maintaining System-Wide Documentation

Maintaining system-wide documentation might seem like a task you could skip, but you shouldn't. Without complete documentation, a lot of person hours can be wasted when something goes wrong, or when you are trying to add hardware to a server or applications to network hosts or workstations. Regrettably, for some technicians and network managers, checking the documentation prior to making system changes is not priority one as it should be. Good documentation practices are not a bane because they take time; they are a benefit to the network manager with little time to waste. Network documentation should include all operation and maintenance booklets as well as manuals for all the hardware.

Administering and Supporting End Users

As the network administrator, you will likely be responsible for administering and supporting end-users. Examples of tasks you'll need to perform may include the following:

- Vetting new users for security purposes

- Adding, deleting, and changing end-user accounts

- Creating and administering group, role-based, and individual access controls

- Providing technical support

- Adding workstations and peripheral devices

Adding Workstations and Peripheral Devices

There will likely be times when some software-based administrative chores must be completed in order to add new workstations and peripheral devices to the network. Examples are hardcoding an IP address into a new workstation or printer or attaching a new printer to a print server's queue. In addition, users may need to be assigned rights to access new equipment such as printers, along with access passwords for new workstations on the network. For more information, consult the documentation provided with

the new equipment and your own documentation of necessary steps from previous changes.

Virtualization in Cloud Computing

What is cloud computing?

Cloud computing refers to delivery of IT resources via the internet as per the demand. It is normally implemented on a pay-as-you-go pricing basis. The concept of cloud computing seeks to offer a solution to users' needs of IT infrastructure at a low cost.

Why cloud computing?

For small as well as big IT companies that still rely on traditional methods to operate primarily require a server to carry out their different tasks. The setting up of a server room requires skilled personnel, different servers, modems, switches, and lots of other networking resources-plus a lot more of other non-IT requirements that contribute to the completeness of a working office.

Traditional methods require a lot of human input, expensive equipment and a lot of other logistical necessities. These things require large sums of money. In order to set up a fully functional server, the organization or individual must be willing to break the bank. However, that is no longer thanks to the concept of cloud computing. Cloud computing helps individuals to cut down on infrastructure costs by eliminating the need for the purchase of expensive equipment and spending a lot of funds on hired personnel for the administration and management of IT resources.

Features of Cloud Computing

- Cloud computing operates in a distributed computing environment. This makes resource sharing to happen quickly.

- Cloud computing minimizes chances of infrastructural failure due to the existences of many servers. This makes it a more reliable infrastructure for IT operations.

- Cloud computing allows for large-scale, on-demand provision of IT resources without the need for engineers and many other

professional that would otherwise come in handy.

- Cloud computing enables multiple users to share resources and work more efficiently by sharing the same infrastructure.

- Cloud computing eliminates physical location or distance concerns since users can access systems and resources regardless of their geographic location.

- Maintaining cloud computing applications is easier since they do not need to be installed on each user's computer.

- Cloud computing reduces the operating cost of an organization since it eliminates the organization's need to set up its own infrastructure-this turns out to be quite an expensive undertaking for most organizations. Besides, it allows an organization to only pay for a service or resource when needed.

- Cloud computing allows for pay-per-use mode for different services. It is a convenient way to use especially when a user needs to use a resource only once.

Benefits of Cloud Computing

The following are the major benefits of cloud computing:

- A person can conveniently use a low-cost computer to perform complex computer tasks that would otherwise require powerful machines-machines that fetch high prices in the market. Applications run on a cloud, not on the user's machine.

- Low-cost IT infrastructure is can sufficiently meet an organization's needs. There is no need of investing in high-cost IT infrastructure to handle big servers.

- Low-cost maintenance due to reduced infrastructural requirements-in terms of both hardware and software.

- Instant software updates for web-based applications-no need to worry about obsolete applications and high upgrade costs.

- The execution capacity of cloud servers is very high. This increases computing power and leads to efficient task execution due to high speeds of data processing and message delivery

- A cloud provides users with very large storage capacities-at very low costs.

Disadvantages of Cloud Computing

Some of the most conspicuous limitations of cloud computing are listed below:

- You cannot access resources on a cloud without internet connectivity. It is mandatory that the user is connected to the internet in order to access resources (or use the IT infrastructure) offered on a cloud.

- Low internet connectivity may be greatly frustrating when trying to perform important tasks on a cloud. Most web-based applications run effectively on a lot of bandwidth. Low bandwidth may at times cripple the execution of some tasks.

- There is no guarantee of security and confidentiality of data stored on a cloud. Unauthorized users can gain access to your data/information stored on a cloud.

What is virtualization?

Virtualization is a process by which a virtual version of some actual thing is created. In computing, this may involve virtualization of an operating system, network resources, server, storage device or even a desktop.

Technically, we can refer to virtualization as a technique which permits sharing of one instance of a physical resource or application among multiple users or groups.

The technique involves the assignment of a logical name to a physical storage of a given resource or application and offering a pointer to the specific resource or application as is required.

The Concept behind Virtualization

The process of creating a virtual machine over an existing hardware and operating system is referred to as Hardware virtualization.
Hardware virtualization creates an environment which is separated logically from the actual machine.
The machine on which the creation of a virtual machine occurs is referred to as the host machine. On the other hand, the created virtual machine is technically referred to as a guest machine.

Types of Virtualization
The following are the different types of virtualization

- Server virtualization
- Storage virtualization
- Operating system virtualization
- Hardware virtualization

Server Virtualization
When virtual machine manager (VMM)-virtual machine software-is directly installed on the server then the process is referred to as server virtualization.
Why Server Virtualization?
Server virtualization is essential because it is possible to subdivide a physical server into multiple servers on a demand basis, and also for load balancing.

Storage Virtualization
This is the process that involves the grouping of multiple physical storage devices in a network so that they appear like a single storage device. Software applications are also used for the implementation of storage virtualization.
Why storage virtualization?

This is crucial for recovery and back-up reasons.

Operating System Virtualization
In this case, the virtual machine software (VMM) is installed directly on the operating system of the host machine. Unlike in hardware virtualization, VMM is not installed on the hardware.
Why operating system virtualization?
Operating system virtualization comes in handy when there is a need to test applications on a different operating system platform.

Hardware Virtualization
In hardware virtualization, the virtual machine software is installed directly on the hardware system.
Hypervisor is charged with the responsibility of controlling and monitoring the memory, processor and hardware resources.
We can install different operating system on the system and use it to run a lot of other applications-after virtualization of the hardware system.
Why hardware virtualization?
Hardware virtualization is largely important for server platforms since the control of virtual machines is not as difficult as the control of a physical server.

How Virtualization Works in Cloud Computing
Virtualization is a greatly potent concept in cloud computing. Normally, in cloud computing, users need to share resources available in the clouds. For instance, applications and files are some of the sharable resources that may be stored in the clouds. With virtualization, users are provided with a platform that making sharing of such resources a practical experience.
The primary goal of virtualization is to offer applications with their standard versions to users in the clouds. When new application versions are released, users look up to software developer for the new release. This is possible but may turn out to be quite a hectic affair if all users have to download the new version from a central server. To resolve that issue,

virtualized servers and software can be maintained by third parties at fee, but cloud users can efficiently access the new software releases.

In summary, virtualization primarily means running several operating systems on one machine that share all hardware resources.

This technique is hugely helpful because it makes it possible to pool network resources and share them to different users conveniently, and at less cost.

Chapter 3: Computer Network Communication Technologies

How computers communicate in a network

Communication among different computers in a network takes place when data is transmitted from one machine to another.

The transmitting computer is referred the sender (source). The machine to which data is transmitted is referred to as the receiver (destination).

For communication to take place, data is transmitted in the form of data packets. For data to get to the desired destination, data packets are assigned source and destination address. The source address identifies the sender. On the other hand, the destination address identifies the receiver (destination).

There are different ways in which data may be transmitted from one machine to another. Notably, there is more than one way in which data transmission happens.

The way in which data can be transmitted from one computer to another is referred to as data transmission mode. Transmission mode is also referred to as communication mode.

The different data transmission modes include the following:

- Simplex mode: communication goes on in one direction. That is, the communication is unidirectional. A device can only receive data but can't send, and vice-versa.

- Half-duplex mode: communication is bidirectional. That is, a device can send and receive data. However, it cannot send and receive at the same time.

- Full-duplex mode: communication is bidirectional. Unlike in half-duplex mode, communicating devices can transmit and receive at the same time.

Addressing

A message needs to originate from a known sender and be clearly addressed to a known recipient.

For instance, traditional post office requires that the sender identifies himself or herself by name and location. The sender must also indicate the

recipient's name and exact location (address).

Similarly, computers in a network are identified by their MAC addresses and IP addresses. MAC addresses are embedded on the computers' NICs while IP addresses are manually assigned during the configuration process; or the IP addresses are dynamically assigned by enabling the Dynamic Host Configuration Protocol (DHCP) on each machine. Each and every network host has a unique MAC address and unique IP address that identifies the given machine.

Understanding Ethernet

Ethernet network architecture is the most widespread of all network architecture all over the globe. We're going to examine the depths of this architecture and most likely find out why this architecture is as popular as it is.

Most network peripheral components have a built-in NIC. As a result, they can be easily plugged into an Ethernet wall outlet. It must be noted that the standard predetermined Ethernet length of wire of 100m from a hub or switch remains so even when it comes to NIC-equipped print servers and printers; just as it is the case with workstations.

Printers that do not have a built-in NIC can still be used on a network by getting a connection with a network print server through a parallel, serial or USB port, or onboard NIC.

Ethernet Network Access Strategy

Suffice to say, Ethernet is a passive network architecture that embraces the wait-and-listen approach. It is also referred to as contention-based architecture since all computers on the network have to contend with the time of transmission on a given network medium.

Access to Ethernet networks is via CSMA/CD. This simply means that the network hosts have to listen to the network until the transmission medium is clear so that they can also transmit. Basically, they have to "sense" and determine that the line is indeed clear to initiate their own data transmission processes. A network host only sends out its data once it "feels" that the transmission is clear. In case there are multiple transmissions, a collision or collisions take place on the transmission medium. The machines sense the collisions and immediately halt their transmission processes.

One of the machines starts the retransmission as the others wait for the line to clear before they can retransmit their data. This process happens until all the network have completed their transmissions.

In a similar fashion, hosts wait and listen on the line for data meant for them. When a particular host senses that incoming is mean for them, they

open door for its reception and actually does receive the data onto its NIC interface. Ethernet is characterized by frequent collisions. As a result, some devices have collision to prompt you when a collision happens. In fact, collisions are the main limitations of the Ethernet architecture. On the hand, Ethernet is the most affordable of all other network architectures.

Note:

- Collisions slow down the network.

- Excess collisions may bring down a network completely.

Fast Ethernet

The traditional Ethernet has a speed of 10Mbps. Fast Ethernet offers a speed that is higher than the original 10Mbps. It has a 100Mbps transfer rate. The throughput is higher than the traditional Ethernet standard since the time it takes to transmit data over a network medium has been minimized by whopping factor of 10. Thus, Fast Ethernet works at a rate that is 10 times the traditional speed of 10Mbps.

Traditionally, hubs and other connecting devices were designed to accommodate the 10 Mbps transfer rate. For such devices, Fast Ethernet is not supported. Fortunately, many connecting devices are being with NICs that can comfortably handle both 10Mbps and 100Mbps transfer rates. That means that the devices can accommodate both the original 10Mbps Ethernet as well the Fast Ethernet.

Gigabit Ethernet

This is another version of Ethernet that is even faster than Fast Ethernet. It uses the same data formats and IEEE Ethernet specifications just like the other Ethernets-10Mbps and Fast Ethernet.

With Gigabit Ethernet, users are able to enjoy 1000Mbps transfer on a network. Unlike Fast Ethernet that operates on both twisted-pair cables and fiber-optic cables, Gigabit Ethernet was initially restricted to fiber—optic cabling. This required that a LAN be set up with specialized servers and

high-speed switches. Gigabit Ethernet was considered to be a backbone for large LANs that required high transmission speeds.

Currently, anyone can practically enjoy the amazing high speeds of Gigabit Ethernet since it is being bundled out in network cards that can be conveniently installed in network servers and network clients.

Ethernet IEEE Cable Specifications

The following is a list showing some of the Ethernet specifications:

- 802.3 for Ethernet LAN

- 802.5 for Token-Ring LAN

- 802.7 for Broadband TAG

- 802.8 for Fiber-Optic TAG

- 802.9 for Data Networks and Integrated Voice

- 802.10 for Network Security

- 802.11 for Wireless Networks

  Note: TAG stands for Technical Advisory Group

The following points must be taken into account:

- Ethernet is well-defined by the IEEE specifications of 802.3.

- It works at the Data Link layer of the OSI mode.

- A number of the various IEEE types of Ethernet are available depending on the nature of cabling preferred on the given computer network.

These types of Ethernet-Gigabit Ethernet and Fast Ether- are designated by 3-part names, like *10BASE-T*. The first section of the name describes the transmission speed. For instance, 10 specifies 10Mbps Ethernet.

The second part of the name, which is "base" for all the different forms of Ethernet, indicates that the Ethernet signal is *baseband*. This means that the

data drifts in a stream as one signal. This type of data transmission cannot allow the transmission of multiple channels of data or information as can the alternative-the *broadband.*

The last part of the Ethernet type name specifies the type of cable in use. For instance, in 10BASE-T, the *T* indicates a twisted-pair cable, and it is presumed to be unshielded twisted-pair cable.

10BASE-T: This type of Ethernet works with a twisted-pair cable (unshielded twisted cable). The maximum cable length (without signal amplification) is 100m. 10BASE-T is operable on a star topology.

10BASE-2: This type of Ethernet works with a fairly flexible coaxial cable (RG-58A/U I or a *thinnet*), with a maximum cable length of 185m (this is rounded off to 200. Thus, the *2* in 10BASE-2). With the use of T-connectors to link the cabling to the network hosts' network cards, 10BASE-2 uses a bus topology. Though 10BASE-2 has always been the most pocket-friendly option for the Ethernet implementation, 10BASE-T setups are presently the most widespread.

10BASE-5: This is type of Ethernet that uses a large-gauge coaxial cable (also referred to as *thicknet*), and the hosts on the network are linked to a main trunk line. The cables from the network hosts join the main trunk cable using vampire tabs, which pierce the primary trunk cable.

100BASE-TX: This is the type of Fast Ethernet that relies on the same Category 5 UTP cabling that is available on 10BASE-T Ethernet. This enactment can also employ 100-Ohm shielded twisted pair cable. The maximum cable length in the absence of a repeater is 100 meters.

100BASE-T4: This is the sort of Fast Ethernet that runs over Category 5 cabling, as can the 100BASE-TX. However, it can as well run over lower-grade twisted-pair cabling like Categories 3 and 4. In this type of Ethernet, the maximum cable run is the standard 100m length.

100BASE-FX: This is the sort of Fast Ethernet that spans over fiber-optic cable with a maximum length of 412m.

1000Base-T: This is the kind of Gigabit Ethernet that delivers 1000Mbps over Category 5 twisted pair cables.

10Gigabit Ethernet. This is the kind of Ethernet that delivers 10 billion bits per second over fiber optic cables.

Peer-to-Peer Communication

The peer-to-peer network setup offers a simple and cost-friendly networking solution in which the basic networking functions are needed. Peer-to-peer is ideal where file sharing and other basic resources such as printers. In this networking arrangement, there is no need for a dedicated server. The elimination of a network from the budget makes this kind of networking a highly pocket-friendly venture. It must be understood that network servers are quite expensive. They significantly contribute to the overall cost of a network. Therefore, its omission from the shopping list is a great way of cutting down on the installation cost of a network, as well as overall management of the network. It is the most ideal option for the implementation of small networks.

Features of Peer-to-Peer Networking

The following are the main features of Peer-to-Peer Networking:

- All network computers have equal privileges. They behave like peers.

- Each network host acts like both client and server. Essentially, a network can send out requests to another network hosts. Similarly, each host can receive requests from other any other host on the network.

- There is centralized administration of resources. This is due to the omission of a dedicated server from the setup. This also forms a concrete basis for connoting a peer-to-peer network as a workgroup. The peers collaborate freely without any centralized control or regulation.

- You only need to install an operating system on each peer and then physically connecting the peers via the NICs (NICs are not even necessary when working with Macintosh computers).

Merits of Peer-to-Peer Communication

The following are the main advantages of a peer-to-peer network implementation:

- Peer-to-peer network implementations are easy install since they only involve the installation of operating systems on peer computers and physically connecting them.

- They are less expensive since a costly server computer is not needed. Furthermore, a Network Operating System (NOS) is also not needed.

- All required is readily available, most of which comes packaged in your operating system.

- The network implementation is reliable since failure or malfunction of a single peer does not lead to failure of another, or failure of the entire network.

Demerits of Peer-to-Peer Communication
The following are the main limitations of a peer-to-peer network:

- If multiple users access a printer that is connected to your computer, you're your computer's processing resources are used as the printer serves your peers.

- There are challenges when it comes to data backups since there is no centralized location for sharable files.

- Resource management is difficult since resources have to be managed one by one.

- It is quite difficult to manage resources since resources are scattered all over the network.

- The security of the network might be easily compromised. Also, users might have to keep track of multiple network access credentials (usernames and passwords).

Chapter 4: The Internet

Internet basics

This section covers some of the basic technology concepts that makes the Internet work and discusses various options for connecting to the information superhighway so that everyone on your network can surf the Internet, communicate via e-mail, share digital pictures with others, conduct research using countless online resources, make purchases online, download movies and music, video conference, and more.

Internet Technical Terms

Just as you don't necessarily need to know the inner works of a combustion engine to drive a car, it's not imperative that you understand every aspect of how the Internet works in order to take advantage of all that it offers. That said, it never hurts to examine, however briefly, the various terms and concepts that relate to the Internet.

TCP/IP

TCP/IP—short for Transmission Control Protocol/Internet Protocol—is a group of rules called protocols that define how devices, be they similar or diverse (i.e., computers, routers, and modems), connect and communicate with each other. (In this context, a "protocol" describes technical details about how any two communication devices will interact and work together to move digital data from one device to another.)

TCP/IP works by determining the best available transmission path for data to travel. Rather than sending all the data in one large chunk, however, the protocol breaks the data into small packets.

These packets can travel over any number of different paths to reach their destination; when they arrive, they are reassembled in order.

To ensure that packets arrive at the correct destination, each one contains both the destination address and the source address. This information is stored in each packet's "envelope," or "header."

The TCP part of the protocol controls the breakdown of data on the sending end and its reassembly on the receiving end, while IP handles the routing of

the data packets.

Think of it this way: Sending data via TCP/IP is not unlike sending letters via the U.S. Postal

Service. Each letter you send by post contains the sender's address (i.e., the source address) and the recipient's address (i.e., the destination address). The difference is that with snail mail, you send the whole letter in one package or envelope (packet). If you were to send that same letter over the Internet, it would be sent in hundreds if not thousands of packets (envelopes) to get to its destination, after which it would be electronically reassembled.

Internet protocols in use under the TCP/IP banner include UDP, PPP, SLIP, VoIP, and FTP.

Sub-net Mask

A sub-net mask is a number applied within a host configuration file that allows for the division of an IP class C network into separately routable networks. For home networks on an ISP's larger network, the sub-net mask will most often be 255.255.255.0, because home networks are not usually split into physically separate segments with internal routers. In office buildings and business environments, sub-nets are used to detach traffic onto physically isolated networks to retain the data traffic on the low and to enhance performance for access to peripherals and local servers. Data traffic destined for another sub-net or to the WAN will have to pass through the router.

DNS

Just as it is easier to remember someone's name than it is to remember her phone number, so, too, is it easier to remember the location of a Web site by its domain name rather than its IP address. For example, suppose you frequently visit the Web site of Ford Motor Company. Chances are, you will probably remember the site's domain name-i.e., Ford.com-and not its IP address. Your computer's Web browser, however, operates in the exact opposite way. It needs to know Ford.com's IP address in order to connect with the site.

That's the point domain name system comes in. When you enter the domain name of a site you want to visit (Ford.com), your Web browser initiates a session with a DNS server either locally or on the Internet to locate the IP address associated with that domain name. DNS servers perform a hierarchical lookup for the IP addresses using domain name associations for registered domain names to locate the IP address of the site you want to visit. If the DNS server your computer is linked to cannot determine the IP address linked with the domain name you entered, the DNS server will then look up the number on successively higher-level DNS servers until it finds the entry (or errors out).

Once the IP address is found, your computer can locate and communicate with the computer housing the Ford.com Web site. The first DNS server stores the association in memory for a time in case you or someone else it serves needs to visit that site again. The DNS server stores only frequently used associations because it can look up the ones it does not know on the higher-level DNS servers.

Assessing Internet Service Plans

Two things are necessary to establish home access to the Internet: at least one Internet-capable computer on your network and the purchase of an Internet service plan from an Internet service provider (ISP). What plans are available will vary somewhat by geography (with suburban and rural areas having fewer options than urban ones), the communication media you want to use, and the options put forth by your ISP.

Some critical plan features include the following:

- Price

- Internet speed

- Equipment provided by ISP

- Customer service support

- Nature of IP address provided-static or dynamic

- Transmission media used

- Email addresses

- Webpage hosting

- Complimentary Wi-Fi access

Making the Connection

To connect your computer to the Internet, you must choose from the service-plan options available in your area. Once you have evaluated the connection plans and media options in your area and have selected an ISP, review some of the following considerations below for guidelines on how to set up Internet access.

Connecting with Dial-Up

Dial-up is, for the most part, obsolete from a speed perspective, but in some rural areas, it is the only available low-cost Internet-connection option. When connecting your computer using dial-up over a plain old telephone service (POTS) line, there are three common scenarios:

- Hooking up a computer or laptop with a built-in modem
- Using an external dial-up modem connected via a USB port
- Using a modem that will connect to a 9-pin serial port.

Connecting with Cable

A popular Internet-connection choice in many areas is cable. In fact, your home or small office may already have a cable connection for television service, making the addition of a cable modem to the mix fairly simple. Cable Internet service is high speed-much better than that offered by dial-up. In addition, many cable-based packages bundle increased television channels for viewing and Internet phone service.

Connecting with Wireless (Wi-Fi)

Connecting wirelessly to the Internet is fairly simple, but your network must include a gateway or router designed for wireless connections. In addition, any computers on your network must have Wi-Fi capabilities built in or, in the case of a laptop or notebook computer, a slot for a wireless Wi-Fi card.

If your computer or workstations are not configured for Wi-Fi, fear not. There are hosts of manufacturers making devices to support wireless connections—essentially, these are portable wireless NICs that can be plugged into either an Ethernet port or a USB port.

Connecting with DSL

Using DSL to connect to the Internet over standard phone lines has an advantage of accessing the internet are higher speeds than the dial-up option

(assuming you live in an area where DSL service is available). Moreover, whereas a dial-up connection relies upon the audio/analog band on a phone line, data on a DSL Internet connection passes over the wire pair at a frequency that is higher-meaning that users can still use their phone lines while at the same time using the Internet (and, by extension, keep your Internet connection live 24/7).

Network Address Translation

Network address translation (NAT) is an important feature on Internet connection devices and gateways that allows a computer to have an IP addresses that is not visible on the Internet, yet still receive and send data packets over the Internet. These addresses are hidden and are assigned from a different set of IP addresses-called private IP addresses-from the addresses that are seen or exposed on the Internet. These private addresses are assigned to computers inside the firewall, enabling them to use TCP/IP protocols for communicating to internal devices and to hosts on the Internet without being seen-thereby making it harder to hack into the internal computer. Using NAT is the first tier in firewalling or protecting your network computers from unwanted intruders anywhere on the Internet.

Private IP addresses also extend the connectivity of the Internet to more computers than there are available IP addresses because the same private, internal network IP address can be used at hundreds, thousands, or even millions of locations.

It works like this: When you open a browser to reach, for example, Yahoo.com, the data packet reaches your Internet gateway/firewall, which in turn starts a session to keep track of your MAC address and IP address. It then replaces your private IP address from the data packet with its own visible IP address in the data packet and sends the request to Yahoo.com. When the information is returned from Yahoo for your session, the process is reversed; the Internet gateway/firewall strips out its own IP address, re-inserts your computer's private IP address and MAC address into the packet header, and passes the packet down the network wire to your computer.

When this happens, your internal IP address is said to have been "network address translated"-although a better term might be "network address substituted." By default, most home network gateways use NAT and assign private IP addresses to all the computers on the home network.

Private Networks

Private networks are IP networks with host computers that hide behind a device that provides NAT. The computers on these networks are assigned IP addresses outside of the pool of numbers used on the Internet. Essentially, any number in the private address range can be assigned locally to a computer or host.

Private network IP addresses begin with any of the following numbers:

- 10

- 172.16–172.31

- 192.168

A complete example might be 192.168.11.4 or 10.101.101.1.

Worldwide Web: Window to the World

Like a living organism, the Web is constantly changing as networks are added or changed. The growth of the Internet both in geographic reach and audience presents every connected entity with opportunities to communicate like never before in history. If your use of the Web is limited to simply downloading information and receiving e-mail, you are hardly scratching the surface of what can be accomplished over the Web. Ways to use the Web to inform, educate, and exchange ideas, goods, and services with a worldwide audience are limited only by one's imagination and creativity. This chapter merely skims the surface of what you can do on the Web.

Leveraging Your Connection to the Web

Connecting your network-or a sub-network of your network-to the Internet stretches the reach of your home or office network to the far corners of the earth. For under $120 per month in most markets around the country, you can obtain a connection to the Internet that runs at decent speeds and includes up to five static IP addresses.

These addresses can significantly enhance your ability to garner the most benefit from your connection to the Internet. That's because in order to

make Web servers, Web cams, and other resources available on the Web, you need at least one static IP address that is visible on the Internet. Additionally, a static IP address can be used to enable VPN clients to connect to your network resources. Without a static IP address, much of your communication to the outside world is limited. With a static IP address, however, your network can become a Web site, client-services provider, radio station, TV station, or blog-just to name a few.

The Web really is a window on the world. Not only can you see out, obtaining incredible amounts of data from the Web, so too can others anywhere in the world see in, enabling you to share information of your choosing with a worldwide audience. Adding your own resources to the Web-the ultimate unfettered two-way, free-speech forum-can both provide value to you and your organization and increase the utility of the Web for others.

Popular Uses of the Web

The following are the main uses of the web:

Finding or Publishing Information

Most people use the Internet to obtain information-which is why some people call it the largest library in the world. The best way to obtain information online is to enter keywords or phrases into a search engine like Yahoo, Google and Ask.

When you type a keyword or phrase into the search field on any one of these sites, it returns any number of links to Web pages that relate to the word or phrase you entered. Ask yourself or your organization's management: What information about you, your family, or your company should be posted to a Web server?

There is more to getting your information found or your voice heard on the Internet than simply getting a domain name such as thisismywebsite.com.

To ensure that the information on your site can be found when someone performs a related search, you enter key search words into your document headings and possibly pay to register your site with various search engines. Learning key search words and adapting your document headings and labels accordingly is a science in itself. And even if you master it, your business Web site might be listed at the top of the search results one day and slip to 100 or 1,000 the next. Like the Wild West, there are few rules on the Internet, and anything goes when it comes to getting noticed.

Communication

This takes place in the following ways:

E-mail

The most popular Internet communication tool is e-mail-that is, messages are sent electronically from sender to host on the Internet, potentially forwarded to other hosts, and ultimately downloaded at the recipient's convenience.

One way to obtain an e-mail account is from your Internet service provider (ISP); most plans include the use of at least one e-mail address. Alternatively, you might run your own home or office e-mail server under a domain name you own. You access messages received via these accounts through special software called an e-mail client.

Another option is to use any one of several free Web browsers–accessible e-mail services, such as the following:

- Yahoo! Mail (http://mail.yahoo.com)

- Gmail (http://www.gmail.com)

Instant Messaging (IM)

Another way to communicate over the Internet is via instant messaging (IM). IM provides instant communication; there is no middleman to store or forward the message. Both end-users must be online to IM; when they do, the text they type is transmitted instantly from one to the other in back-and-forth fashion the second the Send button (or similar) is clicked. You can IM

using an IM client on your desktop or, in some cases, a Web browser. Popular instant-messaging applications include the following:

- Yahoo! Messenger

- Window Live Messenger

Video Conferencing

Video conferencing gives users the rare chance of conducting virtual meetings, thereby saving on a lot of travel expenses. To do a video conference over the Internet, at least one participant ought to have a static IP address visible to the Internet. Additionally, each participant should have service with an upload speed of at least 400Kbps to maintain quality communications, particularly if you're using the video component. To video conference, you must have access to a Web cam of some sort.

Blogging

Blogs, short for Weblogs, are sites on which people can share information with other interested or likeminded individuals. Think of a blog as a digital journal that can be read by people around the world.

Entertainment and Media

The Internet boasts a plethora of entertainment options, including the following:

- Interactive gaming

- Music

- Video

- News

- Internet radio

- Internet television

Engaging in Commerce

Commerce represents one of the most common uses of the Internet. Business-related activities include (but are not limited to) the following:

- Banking

- Advertising

- Retail sales and marketing

- Auctions

Downloading Software

Many major software publishers-including Microsoft, Corel, and Sun-offer users the ability to download what would otherwise be boxed commercial off-the-shelf software (COTS). All you need is a good Internet connection and a PayPal account, credit card, or in some cases a checkbook to pay the fee. There is also a wide variety of trial software, freeware, and shareware, as well as open-source software, available for download online.

Surveillance

Setting up surveillance cameras to be viewed over the Web is nearly a plug-and-play operation, provided you have the necessary IP addresses to support the camera or Web servers. This technology allows, for example, monitoring of your home or office while away or, say, checking on your summer house while you are at home.

Business owners can set up cameras at their place of work to monitor events at the office or keep tabs while away.

Chapter 5: Router and Server Basics

Routers and servers are very important network devices. This section seeks to outline some of the fundamental concepts of routing as well as the client/server architecture. We're also going to examine a VLAN and how to configure it.

Router: what is it, and what does it do?
A router is just another networking device that primarily connects different networks. A router plays the role of forwarding data packets based on what information is contained in the header of a data packet.

This is a device that operates in the network layer of the OSI model. In the TCP/IP model, a router operates in the internet layer.

Routing refers to the process of determining the best path along which data transmission takes place-from source to destination. Routing is done by a router, which has been defined above.

Routing algorithms are responsible for actualizing the routing process. The routing algorithms refer to a piece of software that works behind the scenes to ensure that the most appropriate path is selected for the transmission of data from sender to receiver.

The routing algorithms are also responsible for the initialization of the routing table. They are also responsible for the maintenance of the routing table.

Routing metrics are used by routing protocols in the determination of the best path for data transmission. Routing metrics include hop count, delay, bandwidth and current load among others.

Routing Metrics and Costs
Metrics and costs play a key role in determining the best path. Metrics refer to network variables that are considered in the determination of the best path. Routing metrics include the following:

- Delay: this refers to the time that a router takes in the queuing, processing and transmitting of data to a given interface. The path with the lowest delay value is unquestionably taken to be the best path.

- Hop Count: this refers to a metric that offers a specification of passes through a connecting device like a router. The path with the lowest hop count is preferred to any other available path if routing protocols consider the hop as a primary variable.

- Bandwidth: this refers to the link capacity. It is given in bits per second. The transfer rates of all links are compared. The link with the highest transfer rate is embraced as the best path.

- Reliability: the reliability value is determined dynamically. Some links are more vulnerable to malfunctioning than others. Besides, some links are more easily repaired than others-after a breakdown. Whatever the case, a more reliable link is preferred to a less reliable link. The system administrator is charged with responsibility of assigning reliability values which are numeric in nature.

- Load: this is the degree of how busy a network link is at any given moment. It may be in the form of packets that are processed per unit time; processor utilization or memory use. The load increases with increasing traffic. In routing, the link with a lighter load is considered to be the best path for data transmission.

Routing Types

Routing appears in the following classifications:
Static Routing
This is also referred to as non-adaptive routing. The administrator has to add routes in the routing table manually. Packets are sent from source to destination along a path that's defined by the administrator. Routing does not depend on network topology or network state. It is the job of the administrator to decide the routes along which data are transmitted from source to destination.

Merits of static routing

- There is no overhead on router CPU usage.

- There is more security since the administrator has control over a particular network only.

- There is no bandwidth usage between different routers.

Limitations of Static Routing

- It is quite exhausting to come up with a routing table for a big network.

- The administrator must be highly knowledgeable in networking and particularly in the network topology he or she's dealing with.

Default Routing

In this technique, router configuration is done in a way that a router sends all data packets to a single hop. It does not matter the network on which the hop is found. Packets are simply relayed to the machine on which it configured by default.

This technique is most ideal when a given network has to handle a single exit point. However, a router would choose another path that is specified in a routing table and ignore the one that's set by default.

Dynamic Routing

This is also referred to as adaptive routing. In this approach, a router determines the routing path as per the prevailing condition in the network. Dynamic protocols the heavy lifting when it comes to discovering of new routes. These protocols are RIP and OSPF. Automatic adjustments are meant when particular routes fail to function as expected.

Features of Dynamic Protocols

The following are features of dynamic protocols:

- Routers must have the same protocols to exchange routes.

- A router broadcasts information to all connected routers in whenever it discovers an issue or issues in the topology or network status.

Merits of dynamic Routing

- They're quite easy to configure.

- The best option when it comes to determining the best paths due to changes in network status and topology.

Limits of Dynamic Routing

- It's a lot more costly when it comes to bandwidth and CPU usage.

- It's not as secure as default and static routing.

Important Notes!

- A router filters out network traffic not merely by packet address, but by a specific protocol.

- A router does not divide a network physically. It does so logically.

- IP routers divides networks into a number of subnets to ensure that specific network traffic meant for a particular IP address can be allowed to pass between specified network segments. However, this intelligent data forwarding leads to decreased speeds.

- Network efficiency is higher with the use of routers in complex networks.

Network Servers

A network server refers to a software application that runs a remote network machine to offer services to other machines on a given network.

Client computers in a network make requests to the server whenever they need certain services. The offers an open window for client request, but does not, at given moment, initiate a service.

Servers are infinite programs that, when started, runs infinitely unless an issue pops up. A server always stands in wait for requests from client machines on a given network. The server responds appropriately to all incoming client requests.

Some networks do not have a dedicated server to control communication on the network. Given such arrangements, network devices communicate directly with one another. There are, however, a number of merits as well as demerits of servers in network operation.

Merits of Servers

The following are pros of using a server to handle client request on networks:

- Centralized administration ensures more security in in respect of resource sharing.

- Use of a server provides a centralized back-up system since important data is stored in a server computer.

- There is increased network speed in respect of resource sharing.

- There is a higher scalability level since it's possible to expand the network in terms of clients and server separately.

Limitations of Using Servers

- Traffic congestion is always a big issue when many clients have to send requests simultaneously.

- There is no robustness in a network since a breakdown or malfunction of a server derails all request handling features of a network.

- Specific hardware may be required at the server-side since a client/server network is greatly decisive.

- There may exist a resource in server computer, but not in a client computer.

Servers are available in different forms. The following is a list of different servers that play highly significant roles in computer networks:

Access Servers

Remote (LAN) access offers network connectivity to remote users who may be otherwise constrained by geographic limitations of a LAN. An access server makes use of a telephone line to connect an office or user with an office network.

Network Time Servers

Network time servers are servers that handle all network timing information from different sources including radio broadcasts and satellites. These servers then avail the gathered information (from different sources) the given network.

Time servers rely of NTP and UDP/Time protocols to communicate with other nodes. In so doing, there is proper synchronization of coordinated activities in the network.

Device Servers

A device server refers to a specialized and network-based network device that meant to perform one or more server functions. A device has three main features that include client access and minimal operating architecture. There is no per seat operating system license in the minimal operating architecture. Also, client access is independent of any proprietary protocol or operating system.

Besides the above two features, a device server is a "closed-box" server. This means that it requires minimal maintenance, is easy to install, and is remotely manageable via a browser.

Examples of device include network time servers, terminal servers and print servers. These device servers are designed to handle perform specific tasks. Each device server is characterized by a unique set of configuration features in software and hardware. The unique features help these servers to work optimally.

Multiport Device Servers

These servers allow sharing of devices between terminals and hosts locally and all over a given network. One terminal can be connected to multiple hosts and can conveniently switch among the different hosts. Multiport device servers can as well be used on device that have serial ports only.

A multiport device server can convert between known protocols such as TCP/IP and LAT. This is possible primarily due to a multiport device server's natural ability of translation.

Print Servers

Print servers make it possible for different network users to share network printers. A print server may support either a serial or parallel interface. As a result, print server accepts requests from all network users using appropriate protocols. A print server also manages printing jobs on every network printer.

Understanding VLAN

VLAN is an acronym for Virtual Local Area Network (normally referred to Virtual LAN). It refers to a switched network that is segmented logically using a project team, application or function. The logical segmentation is done without consideration of users' physical locations.

VLANs are more or less same as physical LANs. The only difference is that VLANs allow end stations to be grouped regardless of whether they are on the same physical segment or not.

A VLAN can accommodate any form of switch module port. Multicast, broadcast and unicast data packets can be forwarded and flooded to end stations only in a given VLAN.

Each VLAN is taken as a logical network. Packets destined for stations outside of a VLAN must be forwarded through a router to reach its destination. Notably, a VLAN can be associated with an IP sub-nets.

Supported VLANs
Conventionally, we identify VLANs with a number ranging from 1 to 4094. The following must be noted:

- 1002-1005 VLAN IDs are reserved for FDDI and Token Ring VLANs

- VLAN IDs > 1005 are not found in the VLAN database since they are extended range.

- Switch module supports extended-range and normal range VLANs (1005).

- Number of configured features, SVIs and routed ports affects functioning of the switch module hardware.

VLAN Configuration Guidelines
It is important to understand the following facts:

- 1005 VLANs are supported on the switch module.

- Numbers between 1 and 1001 are used to identify normal-range VLANs.

- 1002 -1005 are reserved for FDDI and Token Ring VLANs.

- Switch module has no FDDI and Token Ring support.

- 1-1005 VLAN IDs are normally stored in the VLAN database as well as the file containing the switch module configuration information.

- 1006-4094 (extended-range) VLAN IDs are limited to private LAN, RSPAN VLAN, MTU, and UNI-ENI VLANs. These VLAN IDs are not saved in the VLAN database.

The following steps will help you create or modify a VLAN:

1. Use the [**configure terminal**] command to enter the global configuration mode.

2. Enter the [**vlan** *<vlan-id>*] to enter VLAN configuration mode.

   -Use an existing VLAN ID to modify an existing VLAN

   -Choose a new ID to create a new VLAN

3. Use the command [**name** *<vlan-name>*] to give your VLAN a name.

   Though this is optional for normal-range VLANs.

4. Use the [**mtu** *<mtu-size>*] to set the MTU size.

   This also optional.

5. Use the command [**end**] to return to privileged EXEC mode.

6. Use the [**show vlan {name** *vlan-name* | **id** *vlan-id***}**]

7. Use the [**copy running-config startup config**] command to verify entries.

8. To delete a VLAN, use the command [**no vlan *vlan-id***]

Note that VLAN 1 and VLANs 1002-1005 cannot be deleted.

Chapter 6: IP addressing and IP sub-netting

IP Address

What is an IP address?

An Internet protocol (IP) address is a four-octet, eight-bit digital address (32 bits total) that, when written out, looks like this: 10.156.158.12. Evidently, an IP is a unique set of numbers that are separated by dots. The set of numbers is used to identify a computer (or network device) using Internet Protocol (IP) for network communication. In an IP address, the value of any of the octets-the numbers between the periods-can be from 0 to 255.

An IP address is not entirely different from a phone number. If you know someone's phone number-say, your Uncle Mike-you can call her by dialing her number on your telephone's keypad. Then, your phone company's computers and switching equipment go to work to connect your phone with the phone belonging to Uncle Mike over an audio communication channel.

Once connected, you can speak with buddy Bradley, even if he is many miles away. When you do, the audio signal carrying your voice will typically travel over a pair of copper wires from your house to a switch at your local phone company.

From there, the signal might be converted to a light wave in order to travel over a fiber optic cable to another switch. From this second switch, the audio signal might be converted to a radio-wave signal in order to travel from one microwave tower to another. Eventually, as the signal nears its destination-Uncle Mike's house-it will be converted back to an audio analog signal, traveling over a pair of copper wires from Uncle Mike's phone company to her house. (This scenario assumes the use of land lines. If cell phones are involved, then this process will vary in the details, but not in the concept.)

What is the Function of an IP Address?
In a similar fashion to how phones use numbers to connect on a local, regional, national, or international scale, an IP address facilitates connections between computer hosts as well as routing equipment. Put another way, if two computers on the Internet have each other's IP address, they can communicate. But unlike phones, which use switching equipment

to connect, computers connect to each other over the Internet through the use of routing equipment, which shares the communication paths with hundreds or thousands of other computers.

When data is sent from a computer to a router, the router's job is to find a short, open communication path to another router that is both close to and connected to the destination computer.

The router accomplishes this either by using default routes or by dynamically learning and recording tables, called "routing tables," that keep track of which IP addresses are present on any one of the router's many open, up-and running communication ports. Because all the routers connected together on the Internet resemble a spider's web, data can travel over many different routes or paths if necessary, to get to its intended destination. If one of the routers or some other connecting link goes offline, the other routers trying to move the data search for an alternative route to the destination.

In order to facilitate this dynamic communication method, routers are also assigned IP addresses so they can find each other.

IP Sub-netting

Routed IP environments require that your pool of IP addresses be sub-netted. This allows each sub-net to see itself as a separate segment of the larger internetwork. The router then ties together the various sub-nets into one network. The router knows how to route traffic to the correct segment because it builds a routing table. The routing table is basically the networks roadmap.

IP sub-netting is fairly complex, and so to make this discussion informative but still digestible at an introductory level, we will limit our exploration of sub-netting to one class of IP addresses; we will look at an example of sub-netting a Class B range of IP addresses. The mathematical tricks that we use to sub-net the Class B network can also be used to sub-net a Class A or Class C network (although sub-netting Class C networks greatly limits the number of usable IP addressed that you end up with).

Sub-netting is a two-part process.

First you must determine the sub-net mask for the network (it will be different than the default sub-net masks; for example, the default for Class B is 255.255.0.0). After figuring out the new sub-net mask for the network, you must then compute the range of IP addresses that will be in each sub-net.

IPv4 vs. IPv6

Currently, IP version 4 (IPv4) addresses are the Internet IP addresses of choice. As mentioned, these addresses are composed of four sets of eight bits. In the future, we will likely adopt the IP version 6 (IPv6) address scheme. IPv6 differs in form and substance from IPv4 in two ways:

- IPv6 addresses have eight 16-bit numbers (128 bits total), usually expressed in four-digit hexadecimal form. The range of a single 16-bit number is greater than that of an eight-bit number, spanning from zero to 65,535.

- The 16-bit numbers in an IPv6 address are separated by colons rather than periods.

Why make the switch? Because under IPv4, there are not enough numbers available to assign one to every computer or device on the Internet that needs one. IPv6 solves this problem, offering 2 raised to the 128th power addresses; in contrast, IPv6 offers only 2 raised to the 32nd power-although masking and private-address strategies have been used to extend the number of available IPv4 addresses on the Internet.

Chapter 7: Introduction to Cisco System and CCNA Certification

What is CCNA?

CCNA stands for Cisco Certified Network Associate. It is a widely held certification among computer network enthusiasts. This certification is valid for any computer network engineer that starts with the low-level network engineers, network support engineers and network administrators through to network specialists.

The certification program was founded in 1998. An estimated 1 million certificates have been awarded to qualified persons ever since.

CCNA program spans numerous crucial networking concepts. It is an essential foundational course for interested candidates in the preparation for current and emerging networking technologies.

CCNA Scope

The following are the major topics that feature greatly in CCNA certification program:

- Networking essentials-which includes definition of network; description of network components; and the basic network setup and configuration requirements.

- IP addressing

- The OSI Reference Model

- IP Routing

- Routers-routing protocols (OSPF, EIGRP and RIP)

- WLAN and VLAN

- Network device security

- Network security and management

- Network troubleshooting

Why CCNA?

- The CCNA certificate is a trusted validation of a networking professional's ability to manage and administer a routed and medium-level network. It is the validation of an individual's understanding of a network, and their ability to operate and configure it. It also certifies a person's ability to troubleshoot the network.

- The course offers lessons on how to meet network users' requirement by proper determination of a network topology.

- Candidates learn how to make point-point network.

- Protocol routing for connecting networks is also realized through CCNA certification.

- Network address construction is adequately explained in the CCNA certification course.

- The course offers a comprehensive explanation on the establishment of a connection with a remote network.

- CCNA certification is offered using easy-to-understand study material.

- The course is a prerequisite for other important CISCO certification programs such as CCNA Wireless, CCNA Security and CCNA Voice, among others.

- CCNA certification program equips learners with the necessary knowledge and skills for the installation, configuration and management of a small LAN and WAN service networks.

Different Forms of CCNA Certifications

There are two main approaches to fulfilling a CCNA certification course:

1. Combined CCNA Exam

2. ICND1 Exam and ICND2

Note that every CCNA certificate is only valid for a period of 3 years. Every holder of a CCNA certificate is expected to take a fresh examination after every 3 years.

Chapter 8: Fundamentals of Network Security

Network security is one of the most important aspects in overall computer. Now, more than ever, having an adequate security protocol which can combine both functionality and protection is essential for all types of users. Amid the large amount of threats out there, users must have a system which they can rely on when going about their usual day to day business. In that regard, the need to integrate these aspects motivates users to find the ideal solution to their individual needs.

In this chapter, we are going to be taking a look at the fundamentals of network security, guidelines and best practices, as well as, the most common threats found lurking today. In addition, we are going to be discussing the ways in which the average user can take better care to avoid becoming the victim of unscrupulous folks out there.

The only thing to fear, is fear itself

It's natural for the average user to fear hackers. This may lead the average user to spend a pretty penny on security measures which may or may not offer the combination of security and functionality that such users need. In fact, the most common side effect of top-notch security measures is a slow system. For example, a solid online security system will essentially scan every single bit of information that goes through their system. However, this can lead to an overall slowing of internet connection speeds.

As a result, users may feel that they have been short-changed by the performance of their security system. Other systems may sacrifice certain safety features in exchange for faster performance and internet speeds. Yet, these tradeoffs may leave critical information and transactions vulnerable to unwanted folks.

That is why the first thing to keep in mind is that breaking the bank out of fear of being vulnerable may lead you to overspend on a system that, while keeping your system safe, may end up costing you more in the end. That is why finding the right balance is essential; striking a balance between features and performance is the ultimate goal of all users. Therefore, we

will be taking an objective look at what threats are out there, which ones you will most likely be vulnerable to, and what you can do to protect yourself.

Please keep one thing in mind: of all the horror stories out there, some may not necessarily apply to so long as you don't engage in the types of practices that attract hackers and fraudsters such as handle a large volume of financial transactions on a daily basis. Consequently, taking the proper steps to secure the transactions that you do make, will go a long way toward keeping your information, and your money, safe.

What to be on the Lookout for

Threats are out there, and they are real. However, the most important thing to keep in mind is that hackers and fraudsters love feeding on low-hanging fruit.

What does that mean?

It means that cheaters are looking for vulnerable people who don't know any better. Hence, they prey on these individuals. This is why the elderly are a common target for phone scammers. As such, online cheaters are looking for folks who have left windows unlocked or have neglected to ensure their information is safe.

Also, please keep in mind that a lot of the fraudsters and hackers out there will essentially go on a phishing expedition. That's right, we mean "phishing" and not "fishing" since hackers tend to cast a wide net in search of unsuspecting victims.

As a matter of fact, phishing is one of the most common means of victimizing individuals. A hacker will go the extra mile to produce an official-looking email in which they feign to be part of some organization, such as a bank, in which you are registered with. Now, they have no way of knowing if you are actually a customer of that bank. So, the email needs to look as natural as possible so that, in the event that you are actually a customer of that bank, you will believe that it is an official communication. The scammers then trick the user into providing their username and password under the pretense that the bank is undergoing security updates

and so on. When the unsuspecting victim falls for the scam, it is quite plausible that the crooks will then proceed to empty the victim's bank account.

This scam was so predominant, that virtually every company out there took the necessary steps to ensure that their customers would not be subject to these attacks. In the end, phishing has been essentially crushed. Nevertheless, there are one, or two, hackers out there who still give it the old college try.

That being said, the following threats outlined in this chapter are examples of what you need to be on the lookout for.

Network Intruders

Intruders are lurking about. There is no doubt about that. It should be noted that most intrusion events happen from within. That means that intruders are generally people who have access to the network and seek to gain unauthorized access to other parts of the network.

Thus, it is important to have a clear picture of who has access to what and how sensitive information ought to be protect. For instance, if a company handles personal information belonging to their customers, great care needs to be taken in order to ensure that such information does not fall into the wrong hands. This is especially critical if such information could potentially become profitable.

A good practice in this regard is to update the roster of individuals who have access to your network, or your personal computer for that matter. That way, if a breach should happen to occur, it will be easier to pinpoint who has access to what and where the breach may have originated.

Outside intruders may attempt some kind of remote access to your network, but they would at least need to have some kind of insight into usernames and passwords. Although, it's worth pointing out that hackers need only a healthy list of usernames, because as we will see in a moment, they can use other means to break through a password.

Social Engineering

Phishing is an example of social engineering. As such, social engineering seeks to use clever tactics in order to extract as much information as possible from potential victims. That is why all users need to take care with sensitive information in all means of online interaction. For example, hackers may be circling the water on social media, looking for unsuspecting prey to pounce on. If you are getting visions of sharks swimming about, then you are right on the money. The fact of the matter is that hackers will often pose as customer service agents pretending to help customers regain access to their accounts or open a new one. Even if the customer does not

reveal all of their information, the hacker may get just enough to breach an account.

The saying, "if it's too good to be true, it probably is" still rings true when it comes to social engineering. So, always keep an eye out for any suspicious activity out there. If you are even in doubt, just stay away from suspicious users and websites.

Password hacking

Virtually every type of network requires two things to access it: a username and a password. Even if hackers happen to get their hands on your username, they still have to crack your password. This can be rather complex. Yet, you can make it easy for hackers if you don't pay close attention to the types of passwords you are using.

Currently, most sites and network protocols require passwords to have a combination of letters, numbers and special characters (@, #, & or *). In short, the longer your password is, the better. In fact, the best passwords make use of random characters and numbers. This makes it virtually impossible for password cracking software to get its hands on your password.

For instance, "happy" isn't exactly the safest password. It may take a password cracking software minutes to break it. However, a password like "H@ppy" is far more complex and may take password cracking software days to or even weeks before it is able to get anywhere. By then, the hacker will have given up and moved on to the next target.

If you happen to use a password generator software, please bear in mind that you will get a password of around 16 characters using random letters numbers and special characters. Perhaps the best piece of advice here is to make sure that your passwords don't have any type of logical connection to you. They need to be as random as possible. That way, you can ensure that your account will be as safe as it can be.

Packet sniffing

This is a complex endeavor, but if successful, a hacker can have access to potentially unlimited amounts of data. This requires the hacker to install software that can read the information traffic going through your network. If your information is not encrypted, then you are just a sitting duck. That is why most network routers come with their own built-in encryption. Furthermore, free webmail services also use their own encryption. That way, if a hacker happens to intercept your package, the only thing they will be able to get is meaningless gibberish. It may end up taking them week or months before they are able to even get close to breaking through the encryption. While it is not impossible, it will certainly deter hackers from wasting their time on your data.

Exploiting vulnerabilities
Nothing is perfect; and software is no exception. That is why the shark metaphor is so apt when it comes to hackers. They are consistently circling the water looking for backdoors, loopholes and other coding errors that may allow then access to a network. In fact, an entire industry has been built around hacking large corporations and then extorting them out of large sums of money. If the software you are using happens to have one such mistake, you might be vulnerable to an intrusion. These issues are generally addressed by software manufacturers and solved as promptly as possible. However, software manufacturers don't generally become aware of the problem until someone's network has been breached. Nevertheless, always be on the lookout for software updates and security patches. That way, you can improve the overall level of protection in your network.

Malware
This has got to be the most common means of granting unauthorized access to wrongdoers. Malware consists of a program that latches onto your computer files and opens a door for hackers to walk through. When installed, it may be very difficult to detect its presence. Most of time, you

will pick up on it until it's too late. The most common form of malware is a program commonly referred to as a virus.

Now, malware is completely ineffective unless one thing happens: the user needs to allow the malware package to enter their computer. This can be done through a download, inserting a pen drive or installing software. Often, malware poses as another type of software thereby tricking the user to open the file, installing the program or downloading the virus. In such cases, you can still kill the program from ruining your computer though you would have to kill it immediately. A good antivirus program will help keep you safe especially if you usually download information online.

The best thing you can do to protect yourself in this case is to avoid opening any email attachments, downloading software or installing anything that isn't from a trusted source or that might look fishy.

Denial of Service (ransomware)

Ransomware works in the same manner as malware does but with one particular twist: malware generally looks to snoop on your activity and steal information such as account numbers, passwords and usernames whereas ransomware will lock up your computer and deny you access to your network. What the hackers are after in this case is a payment in order to liberate your system and/or files. However, once you pay, there is no guarantee they won't try it again. If you should ever fall prey to a ransomware attack, you need to drastically overhaul your security settings. Otherwise, it will only be a matter of time before hackers try the same shenanigans again.

What can be done about these threats?

Fortunately, there is plenty which can be done to ensure the security of your network. We have already pointed out some of the best practices which you can implement as a part of your usual activity. Moreover, there are other measure which you can take so that you can avoid becoming a target of unwanted attacks.

Network Security Areas or Zones
In a perfect world, there would be a single solution to all of the threats we have outlined. However, engineers are yet to produce a one-size-fits-all solution. The main reason for this is that hackers like keeping up with the times, that is, as soon as a security measured is implemented, hackers are looking for a workaround.
That being said, let's take a look at the various network areas, or zones, which comprise a typical network apparatus.

Logical Security Zones
The most common type of network is a small, home-based network. Typically, these networks consist of internet access coming from an ISP into a router installed in a customer's home. Then, various devices are connected to this network in order to access the internet.
Now, in general terms, most home networks are relative safe given the type of security measures that come preinstalled with the equipment that the average user purchases. Unless users attempt to access the dark web or frequently download software from untrusted sources, then risks should be minimal.
There are two types of traffic that you will find in this type of setup, intranet traffic and internet traffic. Intranet traffic is all of the traffic that happens within the network itself. This traffic cannot be accessed by anyone that does not have access to the network. Unless the network is somehow compromised, the data contained within is rather bulletproof.

Internet traffic is any data that comes in or goes out of the network. This is where things get tricky. If care is not taken to encrypt data or restrict access, then the network may be compromised. One simple example of compromise could be removing your router's password thereby leaving it as an open network. That means that anyone can access your network and even play with your router. This could lead to packet sniffing and interception of your data.

Therefore, great care needs to be taken to restrict the access to wireless data points. If you are not careful to restrict network access through the use of a username and password, you would be opening up the door to unwanted intruders.

Data security areas or zones

The next level of security is protecting the data itself. So, even if the network is comprised and subsequently breached by an unwanted intruder, the data package will be useless to them unless they are able to break the encryption. High-level encryption is virtually impossible to break. This can only be done through the use of the algorithm which was used to encrypt the data in the first place. So, unless the hacker actually has access to the algorithm, there is not much they can do to break it.

It should be noted that if the data is encrypted with a low-level algorithm, then there is a chance that a very clever hacker could break the code and subsequently crack into your data. However, sites and networks that use 128-bit encryption are essentially bulletproof as cracking through encryption that complex may take forever. Unless a hacker is somehow unusually determined to break through, they will give up when they see that breaking through may take them years to achieve.

Physical Access areas or zones

This is the next line of network security. It is also arguably the most important since a breach in the physical security of a network can lead to

catastrophic results. This is the reason why you see armed guards at the door of server rooms, or at the entrance of building housing physical equipment.

While physical security doesn't always need to go to such extremes, it is important to protect access to physical equipment. Often, the biggest breaches of security don't occur as a result of hack, but rather, they are the work of an individual who gains access to physical computers and is able to download information into a pen drive or a CD. Some of the standard practices which many companies implement include disabling USB ports or disk drives. Also, restrictions on data transfer can be set in place such as the requirement of administrator passwords for file copying and accessing shared folders.

For the average user, adding password protection to a computer is often enough to keep snoops out. Also, logging out of a session on a shared computer is a great way to avoid unwanted access to file folders. One common mistake is allowing a web browser to save the password to your accounts on a shared computer. If someone figures out your username, all they need to do is enter it and the browser takes care of the rest.

Understanding access to data

One common practice is to assign levels of access to data. This means that certain data sets may have "open data" meaning that anyone within the organization, or with network access, may be able to see it. In other cases, the definition of open access may refer to certain information being of public domain. This kind of information can be downloaded from a website or accessed by request. This kind of information may lack encryption in order to facilitate access.

"Restricted data" refers to data which is not freely available to all users. Now, the definition of "restricted" is rather broad in the sense that it may only apply to users with network access, or it might be restricted to users with password access. This is commonly found in shared folders and cloud-based storage applications. These data sets may also have some kind of

encryption attached to it thus limiting the abilities of unwanted users to read the information.

Lastly, "confidential data" is the type which contains sensitive information of an organization. In this type of data, there are extreme security measures attached to it including high-level encryption and password access. This type of data may also be stored in "secret" drives and folders which only a limited number of users have access to. By understanding the tiers of sensitivity attached to data, you can take the appropriate measures that you need in order to protect your files and sensitive information. One such example "confidential data" could personal information belonging to customers while "open data" may be information about the products and services that a company offers to its customers.

Network security best practices

Here are the best practices which you can put into practice when looking to protect your data.

- Take control of physical access points such as USB and disk drives

- Don't neglect password restrictions to any sensitive information, folders and drives

- Ensure that access to wireless networks is protected by password access while wired connections are accessible to only those users who have permission to use them

- Avoid the use of external pen drives, disks and any other media which has not been previously scanned or cleared

- The usage of high-level encryption will ensure that your data is bulletproof

- Avoid the exchange of information over open networks or hotspots; these could be prone to packet sniffing

- Shut down network access if there is a suspected network breach

- Update passwords regularly (once a month is a good rule of thumb)

- Discourage password sharing among co-workers

- It is better to spring a little extra for top-level encryption in order to ensure that confidential data is safe even if intercepted

- On the whole, network security is a matter of ensuring that you have the proper procedures in place. Also, it is important to have a clear idea of what to do in case of a data or security breach. Most firewall and antivirus programs will offer functionalities that will alert you of coming from an unknown

source. In general terms, these alerts happen in real time. Consequently, you will have the opportunity to promptly shut down any potential attacks on your network

Chapter 9: Wireless Technology and Security

Fully functional wireless networks and devices seemed like a dream a decade ago. In fact, many experts in the field did not believe that that entire computer networks could be run on wireless connections. Yet, the dependence on wired connections has been drastically reduced over the last few years. Nowadays, a great deal of networks are run on wireless connections. For instance, cellphone communications are run entirely on wireless communications. Everything from calls to internet access; there is no need for any wired connections.

In a nutshell, a wireless network is connection among devices which does not require the use of cables to connect the devices, either to the network or amongst themselves. In that sense, wireless networks offer a high degree of flexibility and simplicity as the physical limitations of having to install large amounts of wires and cabling is not needed. In addition, the network can span over longer distances and does not demand that users be physically present in a given location in order to gain access to the network. At present, wireless networks have simplified both access and mobility. Users can be on the move constantly and not miss a beat. This has enabled a multitude of users to experience new ways of communicating especially

when moving away for long distances. Indeed, wireless technology has revolutionized the way we interact with the world around us.

However, the use of wireless connections has also opened up a new set of threats and considerations which need to be taken into account. In general, wireless access requires both providers and users to exercise increased security practices. This is a stark contrast to a traditional wired connection.

With wired connections, unauthorized users needed to have access to both a physical terminal and a port to connect to. This means that breaking into the network itself was a rather complex task. Of course, an authorized user could perform a remote hack by gaining access through the use of a username and password. But most times, there had to be someone physically present in order to gain access to the network itself.

This aspect of wired networks is lost in wireless communications. Anyone within range of the Wi-Fi signal can potentially gain access to the network. This is why free Wi-Fi hotspots can become so unsafe for the average user. The fact of the matter is that anyone can gain access to the network especially if it lacks encryption or the use of authentication.

As a result, security experts advocate the use of authentication either by means of a username and password, or a captive portal in the case of hotspots. Furthermore, encryption is highly recommended in order to avoid potential loss of information due to a package interception. With that in mind, most home networks are rather safe, that is, a typical residential network that provides wireless internet access to a group of devices is rather safe from unwanted users. Yet, there is always the risk of someone being on the prowl looking to find a vulnerable network.

In the case of business networks, wireless access needs to offer a greater degree of security. In many cases, enterprises deal with sensitive and confidential information. What this means is that networks administrators need to ensure that unwanted users stay as far away as possible. Still, employees need to gain access to the network so that they can perform their day to day functions. With this in mind, awareness of the need to protect access to such connections truly becomes of the utmost importance.

What to consider when setting up a wireless connection

First and foremost, security is the biggest concern for wireless networks. In a world in which information is power, having access to information is a valuable commodity. As such, the question of security boils down to access to information and not access to the network itself. This means that the biggest concern in the mind of network administrators is not that unwanted users log on to the network; their biggest concern is that by logging on to the network, unwanted users will be privy to potentially sensitive information.

In that regard, restricting access to unwanted users in the number one priority for network administrators. Simple measures such as making sure that all users have a username and password is often enough to stifle the average hacker.

The next item that network administrators look to when setting up their wireless networks is simplicity. Simplicity implies the ease with which users can log on to the network and then use it to complete their objectives. With that in mind, administrators need to consider the bandwidth that will be required in order to accommodate the number of users on that particular network.

Often, it is easy to underestimate the requisite bandwidth. In fact, it is quite common for an administrator to estimate a given bandwidth and then realize that it is not enough to handle a full load. Other times, an administrator might make an appropriate assessment of the bandwidth that is required for the current load of users but does not take into account future users. Therefore, it is essential that an administrator consider both present and future bandwidth requirements.

Another important aspect to consider is the actual, physical infrastructure that is needed to produce the wireless network or provide Wi-Fi access to users. This physical infrastructure comes in the way of routers and signal extenders. The most challenging part of the physical infrastructure is directing the signal to where users will be. This can be a potential headache when homes and buildings are designed in a free-flowing style. Wi-Fi signal can often get stuck going around corners or up and down stairs. Even

concrete buildings can absorb the signal rather reflect to the devices that will connect.

One other key consideration when setting up a wireless network is understanding the devices that will connect to it. For example, if the network will mainly consist of smartphones and tablets, then a larger number of devices can be accommodated on to the network. In contrast, if a large number of PCs (which tend to be hungrier for bandwidth) connect to the network, then arrangements need to be made in order to accommodate the larger number of devices.

On the whole, setting up a wireless network is far easier than a standard wired connection. The configuration of software is less complex while there is a definite savings in terms of money; there are far fewer components to purchase and cabling that needs to be installed.

So, if you are looking to set up a network that won't break the bank and will allow multiple users to long on easily, then a wireless network can certainly fulfill those needs.

Drawbacks of a wireless network

For all its merits, a wireless network also poses a series of drawbacks that need to be taken into account. For starters, Wi-Fi access is far less reliable than wired access. The main reason for this is that signal strength on wireless networks fluctuates a lot more on wireless connections than it does on wired connections. In fact, there are many factors which can influence the quality of a wireless connection's signal strength. Factors such as the weather or interference from other devices (phones, TVs and radios) may play into reducing the quality of a wireless signal. As a result, close monitoring of signal quality is essential in order to ensure proper functioning of the network.

Another drawback is the connection's speed. Generally speaking, a wired connection will always run a lot faster than a wireless one. The reason for this is that a Wi-Fi signal can disperse throughout the environment where it is located. In the case of a wired connection, it has nowhere else to go, but

the cable where it is being transmitted. This is what makes wired connections a lot more reliable and faster than wireless ones.

Also, wireless connections are dependent on the wireless adapter used by the devices that are connecting to the network. So, even if the network is reliable and free from obstructions, devices that do not have a solid and functional wireless adapter may actually experience decreased performance. Therefore, it is important to ensure that the devices connecting to the network have the best available wireless hardware.

As you can see, wireless networks, even with their drawbacks, offer a serviceable solution to virtually all residential users and most enterprises. Ultimately, the network administrator, or home user, needs to determine if the characteristics of a wireless network meets their needs, or if a wired network might be more suitable. Yet, wireless networks offer a good solution across the board.

Types of wireless networks and connections

Thus far, we have focused on wireless networks under the assumption that the network is designed to connect to the internet. This assumption is valid since the vast majority of networks will require some kind of internet access at some point. Most of the time, the reason for the network is so that it can connect to the internet.

Yet, there are other types of wireless connections which devices can use to communicate among each other.

To start off, Bluetooth is a common tool used to connect a small number of devices together. Usually, it consists two devices which are paired with each other. A common example of this wireless headphones connected to a smartphone. Also, a printer can connect to a computer or a phone may synch with a computer and so on.

Bluetooth is far slower than other types of wireless communication. Yet, it is useful for linking devices which aren't too far apart. Perhaps the biggest drawback is security as Bluetooth connections don't have much in the way of encryption to speak of. So, a skilled hacker may be able to connect to a device which happens to have Bluetooth enabled.

Beyond a run-of-the-mill Bluetooth connection, wireless networks can be rather simple and straightforward, or they can get rather complex.

A Wireless Local Area Network (WLAN) can be set up to enable communication among a small number of computers with or without access to the internet. In most cases, WLANs will have some form of access to the internet. Although, very confidential connections may be confined to a small workgroup using a WLAN without necessarily having access to the internet. In fact, there may be a separate WLAN that does have internet access. That way, the risk of a security breach can be limited. This type of connection can be limited to a home, individual office or perhaps an entire building.

The next level is a Wireless Wide Area Network (WWAN). In this type of network, there is a much larger area to be covered. If the physical area exceeds the usual range of a wireless network, internet access may be required in order to join multiple connections. This type of connection can be used to link larger geographical areas such as several city blocks.

A larger type of network is known as a Wireless Metropolitan Area Network (WMAN). This is the kind of network which can be used to link an entire city. Unless there are several repeater points that can distribute the signal over large physical areas, the most effective way to connect devices over a large area is to link them over the internet. In that case, any number of users can log on to the network as the need may be.

WMANs are used to connect entire cities. These are the types of networks which are used by large organizations. For instance, companies with several locations spread out over a city may rely on this type of configuration to help them manage communication among all of their users. These networks are also commonly used for commercial purposes such as advertising and broadcasting information.

One important factor that all of these connections have in common is that there is a central hub from which the signal is generated. This hub then sends out the signal to the various devices that will connect to it. Also, repeaters and signal extenders may be used in order to ensure that the signal

covers the entire physical space that the administrator is looking to grant access.

Based on this logic, an ad-hoc connection is one that is set up without any kind of planning. For example, a laptop computer uses Bluetooth to connect a wireless printer and a smartphone. Then, the network will be disbanded once the tasks are complete. Another good example is when a smartphone is used as a wireless hotspot. In this case, the phone enables various devices to connect to it and thereby access the internet. Once the need for internet access is gone, then the network is disbanded.

As you can see, an ad-hoc network serves a very specific purpose and then when it is no longer required, it disappears. Experienced users are good at building ad-hoc networks from scratch, either to help them when they are in a tough spot, or as means of streamlining work.

Lastly, a hybrid connection is the kind which combines both wired and wireless access. This kind of network is commonly seen in enterprise settings. This type of network addresses all of the needs of a company in that wired connections are generally used for desktop PCs while wireless connections are used for laptops and mobile devices.

Furthermore, a hybrid network offers a combination of mobility and security. Secure connections can be achieved on the wired connection while the wireless network can provide mobility to users. Since wireless connections tend to be less reliable than wired ones, the most sensitive tasks and information is carried out over the wired connection leaving less complex and sensitive tasks to the wireless network.

Other uses of wireless technology

In the past, more rudimentary wireless technology such as shortwave radios provided the opportunity to connect over certain distances especially when traditional broadcasting was unavailable. To this day, shortwave radios are essentially used by emergency broadcast systems and aficionados. However, this type of technology is still seen in law enforcement and emergency crews. While shortwave technology does not have a great deal

of range, it is often enough to cover larger areas such as a medium-sized city.

Also, traditional radio communications are a great alternative since they provide users with the ability to connect without having the need for an expensive setup. With that in mind, user don't have to rely on the internet or even electrical energy to communicate.

As a matter of fact, military grade radios use solar powered cells to recharge batteries. This enables military units to remain operational for extended periods of time especially when deployed in battle. These types of radios are also used in marine communications as well as airborne surveillance.

In addition, companies that have a great deal of moving parts use radios to communicate among its various parts. Trucking companies, factories and warehouses all use some type of radio communication to stay in the loop. Hospitals and emergency services utilize radio communications as a means of ensuring capabilities.

Radio technology has also been applied to what is now known as Radio Frequency Identification (RFID). This type of technology is used as a wireless network mainly deployed to identify and track units such as people, animals and vehicles. It is commonly used by trucking and cab companies to keep tabs on their vehicles. Additionally, researchers use RFID to track animals in the wild.

There are two parts to an RFID system, an emitter and a receiver. The emitter, or transmitter, sends out a radio frequency that is picked up by the receiver. The receiver can then translate that signal into a digital signal which can be utilized to track movements on a map or pinpoint the location of a unit. This system has been deployed in search and rescue operations as well as scientific research.

Some critics of RFID claim that it is used to illegally surveil individuals. However, there is not conclusive evidence of its use outside of tracking people and objects for legitimate purposes. For example, people who are under house arrest that must wear an ankle bracelet have an RFID system

attached to them. This allows law enforcement to determine their position in case they choose to flee.

One other interesting application of wireless technology can be seen in satellite communications. Traditional cellular telephony makes use of the Global System for Mobile (GSM) network. This network uses a SIM card with a unique ID on it to identify a mobile phone's number. This enables the mobile phone to place calls and have access to the internet.

However, in the case of a satellite communication, mobile receivers and transmitters don't necessarily use the traditional GSM network. In fact, they may bypass this system altogether and connect to any of the other communications satellites in the Earth's orbit. These so-called satellite phones have coverage in virtually all corners of the Earth.

One great example of this kind of communications is military satcom. Harris radios, for instance, use UHF and VHF frequencies to communicate among aircraft and seaborne vessels. Ultra-high frequency (UHF) and Very High Frequency (VHF), can be used to triangulate communication between an airborne vessel and a ground crew. Given the advancements in this type of technology, air crews can send live video in HD to ground crews who can then assess a situation as seen from the aircraft.

This type of technology has been widely employed for aerial reconnaissance, for example in border protection, as well as, search and rescue and disaster relief. There is also a weather application for the use of the technology through the use of Unmanned Aerial Vehicles (UAV). These types of vehicles can track weather patterns to a point where a manned crew would never be able to reach. All of the communications are relayed from the aircraft up to the satellite and then broadcast back down to the ground crew. Likewise, the ground crew can then communicate with the aircraft in mid-flight.

These examples all show how wireless communications have a plethora of applications. There ensure that communication does not breakdown and can be maintained over vast areas of physical space. While the more robust capabilities of such wireless communication is reserved for the domain of

military and law enforcement, civilian applications have led to the mapping of remote areas of the planet and the discovery of previously unknown features of the Earth.

One of the most common issues that arises when discussing UHF and VHF communications is encryption. It should be noted that these types of networks use military-grade encryption meaning that it uses the most robust algorithms known to date. These algorithms are essentially impossible to crack since they would require a vast amount of computing power that the average hacker would never be able to procure.

If you are keen on using these types of communication networks, there are civilian versions of satellite phones which can be purchased freely. They come with a subscription that grants access to the network of satellites that orbit the Earth. They are essentially the same satellites that the military uses. However, you would not have access to the same channels and frequencies that the military does.

The Global Positioning System

The Global Positioning System (GPS) is essentially a wireless network that is used as a navigation aid. It is used in all types of transportation. In fact, it is so common nowadays, that most smartphones come equipped with a GPS application that enables users to pinpoint any location on the planet.

One great example of such application is Google Maps. This application can help drivers navigate practically any city in the world with frightening accuracy. Technically, this application does not need internet access since it uses its satellite link and not its internet signal. Yet, the application does not function appropriately since it uses an internet connection to download maps for its current location. What this means is that your phone determines your location by using the GPS network, but needs internet access to download the map for that location.

Google Street View is an example of how accurate satellite mapping can be. The satellite orbiting the Earth can literally take pictures of an average

street while it is hundreds of miles above. This is certainly an incredible and very useful feature.

The GPS system uses the same system as satcom or GSM technology. It essentially works in a triangulation system, that is, the satellite, the receiver and the tower. This triangulation is what enables the pinpointing of precise locations.

In the case of marine applications in which there are no towers in the middle of the ocean, GPS can maintain triangulation by using other satellites in orbit. This is why military applications of the GPS system enable units to navigate any corner of the world. For the average user though, GPS is a lifesaver when driving in a new city or through an unknown path.

Bringing it all together

Wireless networks are here to stay. They make linking devices, and by extension users, a lot easier than through traditional wired connections. For all of the ease of use and mobility, questions about the overall reliability of wireless networks still abound. The fact of the matter is that this is technology which is yet to be perfected. What this means is that there are bugs still to be worked out.

As far as the average user is concerned, wireless has come a long way. Most cell carriers have a decent record of reliability. So, unless there is a natural disaster that disrupts coverage, wireless networks are usually reliable.

The other aspect to wireless networks is security. High-level encryption has increased information security tremendously over the last few years. In fact, standard encryption as offered by free web-based email servers is good enough to keep the average intruder away from your information. There are more robust subscription services out there which offer closet to military-grade encryption. These services are commonly used by companies and individuals who handle a great deal of sensitive information. On the whole, wireless technology is easy to use and very flexible. It will certainly meet the needs of the average user while enabling more advanced users to get their job done. By following the security guidelines which we have outlined

earlier in this book, you can rest assured that your information will be safe from the attacks of unscrupulous folks. After all, hackers love to feed off low hanging fruit. They generally run away from a serious challenge. Thus, you can make it hard for them by beefing up your security measures.

Chapter 10: Introduction to Machine Learning: A Computer Networking Perspective

What is Machine Learning?

The above analogy is merely intended to draw our attention to the primary topic issue that concerns this study session: machine learning.

Can computers really learn? Better still, can they learn from their experiences with the different tasks that they handle on a day-to-day basis? Machine learning is a discipline that can sufficiently provide answers to the above questions.

In a layman's language, machine learning can be described as a process of acquiring skills that have been accumulated over time through observation. We can deduct from this simple description that machine learning starts by observation. The observation goes for a certain period.

In the process of observing, skills are acquired while others are sharpened even further-learning from experience. This is one example of a normal human's learning throughout their time here on planet earth.

A more concrete definition of machine does exist. Let's say that machine is a process that involves the acquisition of specialized skill(s) computed/accumulated from data over a period of time. Still, this sounds like we're talking about humans.

But we need to talk about machine learning in terms of machine 'behavior', and more specifically, as regards computers.

In Computing Terms

Machine learning is classified as a subset of Artificial Intelligence. As a subset of Artificial Intelligence, machine learning entails the development of algorithms that enable computers to learn from accumulated data and functions performed in the past. This concept was conceived in 1959 by Arthur Samuel.

Sample historical data (training data) offer a basis for machine learning algorithms to deliver models that aid in decision-making as well as predictions. This is done without any explicit programming.

Creative predictive models are products of the union between statistics and computer science.

How Machine Learning Works

Machine learning system achieves its mandate by following the following steps:

- Thorough examination of (or learning from) historical data that has accumulated over a given period.

- Building of prediction models.

- Prediction of output (on reception of new data).

The amount of data largely determines how accurate the predicted output turns out to be. A huge sample of historical is necessary for creating better prediction models that guarantee high accuracy of predicted output.

Features of Machine Learning

The following are the main characteristics of Machine Learning:

- The use of data for pattern detection in datasets.

- Automatic improvement after learning from historical data.

- Machine learning is technologically data driven.

- It has appreciable similarities with data mining-dealing with huge data amounts.

Why Machine Learning?

There is a day-by-day increase in the need for machine learning. One key factor for embracing machine learning with unimaginable seriousness as it is now, obviously, is the ability of machines to handle tasks that are highly sophisticated; better than humans. We can read, understand and interpret data, but we are only limited to a few megabytes of data. Machines can handle terabytes of data, or even more, with a lot of accuracy.

With machine learning, the world can now boast of self-driven cars. Friend suggestion on social media platforms is also real, thanks to machine

learning. Furthermore, face recognition, among other big advancements, are outcomes after big strides in machine learning.

In summary, the need for machine learning is anchored on the following key observations:

- Increment and rapid production of data, in respect of the widespread adoption of information systems.

- Handling highly sophisticated tasks that are practically impossible to a human being.

- Critical decision-making needs with regards to large and complex data or information. For instance, in finance and economics projections for individual businesses, companies and even governments.

- Extracting patterns and information that are not easily visible to the human eye

Classification of machine learning

Machine learning is classified as follows:

Reinforcement learning: this is a feedback-based mode of learning. A learning agent (machine) is penalized for every wrong pattern prediction and undesirable outcome. On the other hand, the same learning agent (machine) is rewarded for a job well done.

Unsupervised learning: a machine trains on outcome prediction without any supervision. Given data sets are not labeled, but the machine has to learn and eventually predict an outcome. Unsupervised machine learning is categorized into clustering and association.

Supervised learning: labeled data is provided to a machine. The machine is trains on pattern prediction therefrom and eventually gives a prediction on that basis.

There are two categories of supervised learning: classification and regression.

Machine Learning Applications

Machine learning is a crucial to each and every sector be it economic, social or administrative sector. Thus, the application of machine learning can be summarized under the following three key areas:

Machine Learning in Analytics

Statistics is every aspect of life. Sometimes we use it without knowing that we're actually using it. For instance, a person who wonders about a terrible experience would expectedly learn from it and take a totally different course of action given similar circumstances. Knowingly or otherwise, the person assesses the different events leading to a particularly horrible incident against what would have led to a different turn of events. The process may seem straightforward, but the truth is there is a lot of analytical work happening in the busy mind of the victim.

In normal life, politics invest so much in statistics to gauge the popularity of political candidates and help in making crucial political decisions. Similarly, all sectors use statistical results based on historical data to predict various events in the future.

Machine learning takes up this whole business to a whole new level by eliminating so much of human effort by merely allowing algorithms to analyze data and offer prediction of patterns and likely outcome based on accumulated data.

Machine Learning in Management

Management rely heavily on statistical data in decision-making. In business, the management is charged with the responsibility of making decision regarding employ recruitment and laying off; employ remunerations; budgetary projections; and the structure of leadership within a given organization or institution. In formal settings, such decisions are not arrived at hastily, or without thought. In most cases, it takes months or even years to analyze relevant masses of data before deciding what is good for the organization or institution. With machine learning, the analysis of such data is more efficient since machines have the ability to handle huge data amounts that would otherwise take years to be done by a human. In fact, some analytical work is way beyond the ability of the smartest man on earth.

Machine Learning in Security

Security has been enhanced greatly with the adoption of machine learning systems. CCTV cameras, metal detectors, security alarms and other security devices are perfect examples of machine learning in use. For instance, face recognition services are particularly important in ATM machines since they massively add to the security of ATM centers.

In summary, machine learning is an important subset of artificial intelligence; especially during the present times when we have to deal with a lot of data and information as a result of the information systems euphoria. Healthcare, financial sector, education sector and, and governments need statistical data to make crucial decisions for sustainable operations. With machine learning, such needs can be met more efficiently.

Conclusion

A computer network can comprise of two computers. As long as the computers are connected, can communicate and share resources, the network is complete. However, there is no limit as far as the number of networked computers. The internet is a perfect example of how big a network can be.

Resource sharing and communication are the two biggest functions of a network. But that does not mean there is not much beyond communication and sharing of resources on computer. In fact, many people make connections to the internet merely for entertainment. Others get linked to networks for collaborative work, research and many other tasks.

Networks, notably, do not comprise of computers alone. There are a lot of more other network components that make computer much more interesting. Printers, cables, routers and many other hardware devices add to computer networks' collection of important hardware requirements. Besides, software applications and network protocols all gang up to make computer networks what they really are.

As important as computer networks are, running and managing them is never a walk in the park. Network security is one concern that gives concerned network users and administrators sleepless nights. Networks attacks are real, and they pose serious threats to the security and safety of user information as well as network resources. Being in position to deal with network threats effectively is key maintaining the integrity of the network.

In summary, computer networks are invaluable assets that must be guarded keenly. They make communication more effective besides offering a

platform for many other important functions. In this IT-oriented era, computer networking is the key to effective communication, research and collaboration.

BOOK 4: Python Programming

The ultimate crash course for learning Python quickly, with practical examples and coding language tricks for beginners.

Computer programming for data science and machine learning.

Introduction

The following chapters will discuss many of the different topics that you need to know when it is time to learn how to work with the Python programming language. There are a lot of companies and individuals who want to be able to learn how to code, and the Python language is going to be one of the best options for you to choose to go with. This guidebook will make sure that you are able to get some of the basics of this process down, so you can write your own codes down in no time, no matter what you would like to use these for.

To start, this guidebook is going to take a look at some of the basics that come with Python. We are going to spend some time looking at the benefits of working with Python and why it is such a good language to work with, and then we will spend some time exploring what this language takes to get installed on the various different operating systems that are out there. From there, we can look at a few of the basic parts that come with this code, including the objects, comments, keywords, functions, and more.

Now that some of those basics are out of the way, it is time for us to take a look at some of the coding that we are able to do when it is time to work with this language. The best way to learn anything about coding and how to make this happen, we first need to look at some of the codes and get some practice with those. Here we are going to learn how to work with the different data types that are there, how to work with the functions, the importance of classes in the Python language, and how to write out some of our own conditional statements. All of these are going to provide us with some really great codes that we are able to write with the help of the Python language.

After we have had some time to learn those different parts of the coding world, it is time to look at some of the tips and tricks that we are able to do when it comes to learning how to code. No matter how much we prepare, there are times when things are not going to work the way that you want. These tips are going to show you some of the things that you can do when your code is not behaving the way that it should.

The end of this guidebook is going to spend some time looking at the basics of data analysis and why this is an important part of the process. Many companies are jumping on board and looking at how to work with data science and data analysis easier than ever before. You will quickly be able to keep up with the benefits of working with this, the different steps that you need to take in order to work with a data analysis for your own projects, and some of the best algorithms that we are going to be able to work with when it comes to the Python language and data analysis. There is so much that we are able to work in order to get started with the Python language, and this guidebook will spend some time talking about the most important parts of this process so that you are able to make it work for all of your business need. When you are ready to learn more about Python programming and how you can work with this on data science or a data analysis project, make sure to check out this guidebook to get started!
Every effort was made to ensure it is full of as much useful information as possible; please enjoy it!

Chapter 1: What is the Python Language, and Why Should I Use It?

There are a lot of different coding languages out there that we are able to work with. We may find that some of these are going to work the best when they help us create webpages and work online. Others are going to be able to handle some of the more technical aspects of our work, whether we are working with math or science, and can tell us a lot about the world, the trends around us, and more. And still, others are able to be there to help a beginner get into the world of coding.

If you are just getting started with the idea of coding, then the best option that you can choose to learn how to code is with the Python language. This language is simple to learn while having all of the features and power that you need to get coding done. This language is going to come with the standard library and plenty of extensions to get the work completed. And as

we go through this guidebook, you will quickly see why this language is actually pretty easy for you to learn, even as a beginner.

There are a lot of programmers out there who love to work with the Python language and wouldn't choose anything else to work with for their coding needs. And this could easily become you with a little bit of knowledge and practice along the way. when you are ready to learn a bit more about the Python language and how it works, make sure to check out the rest of this guidebook to see what we are talking about!

How to Use Python

While we are here, we need to spend some time taking a look at some of the major uses that come with the Python language. Many companies are more than willing to learn how to work with the Python language for some of their big computer tasks. You will find that Python is already making a bit different in some of the methods that companies use to work well, and towards making sure that the bottom line is not going to bottom out in the process. Here are some of the different ways that we are able to work with the Python language in order to see just all of the capabilities that can come with this language.

The first way that we are able to work with Python is to help us to develop some applications. This can include things like a software development application, network programming, games, web applications, GUI based applications, and more. This is due to the fact that Python is able to make an interface that is interactive and is easy to develop applications with.

Python is also going to be able to help out with multiple programming paradigms. Python is going to provide us with some of the support that is needed to handle these. As it is able to support OOP, as well as support for structured programming, Python is going to come with a bunch of features that allow it to support the concepts and more that comes with a functional

programming language. We are able to use this language to help out with a system that is dynamic, and even with automatic management of the memory.

We will also see that Python is going to have a robust and fairly large library that is going to be useful for developing any applications that you would like. It is also going to make it easier for a developer to rely more on Python compared to the other languages as well. The standard Python library is going to help us use a different range of modules that are available so that we can complete any of the tasks that you would like.

Next on the list is that the Python language is going to be compatible with a lot of major systems and platforms. This is going to make it easier for us to complete a lot more projects because we will be able to finish them no matter what platform or system we are working with. Python is going to be so compatible because it is used mainly for developing some new applications.

With the help of the interpreter, the code is going to be run on the right platform, and it has the tools that are needed to support most of the major operating systems that are out there. this makes it easier to pull up the programs and more that come with Python and get them to work no matter what you plan to do with them or where they need to work.

If you plan to do some work with databases while you are doing your coding, then Python is going to be a good one to use to help with this. You will find that the Python language is going to make it easier for us to access the databases that we would like. Python is going to help us to customize the interfaces that come with most of the common databases, making it easier for us to get some of the work done that we would like in the process.

These, of course, are just a few of the ways that you are able to work with the Python language and see some results in the process. When you are

ready to handle some coding, no matter what kind of coding you are looking to work with, you will find that the Python language is going to be one of the best choices to go with. Python is going to have all of the tools and the features that you are looking for, and it is easy to use as well. This all comes together and ensures that you are going to be able to finish any project that you would like with the help of this coding language.

The Benefits of Python

When it is time to start working with a coding language, you will find that there is no one out there that is better than the Python coding language. There are so many benefits that you will be able to receive when it is time to work with Python that it will not take long before you are able to see why it is the one that most programmers are going to choose for their needs. Let's take a look at some of the benefits that you will be able to get when you work with this kind of coding language.

The first benefit is that this language is designed to be easy to read and write. Unlike some of the options that have been around for a long time and are more traditional coding languages, Python wanted to be easy for even beginners to learn how to use. Whether you have been a coder in the past and have some experience with making all of this work, or this is the first time that you have ever done anything with coding, you will find that the Python language is the best choice to go with.

Even though there is a nice learning curve when it is time to work with the Python language, you will find that this is not going to take away any of the functionality or the capabilities that you will see with this language. Python is a language that has power, and it often has more power than those who are first exploring it will think. In addition to doing some basic coding that you would like and that we will explore more in this guidebook, Python comes with the extensions and more that will help you handle data analysis, data science, machine learning, and more.

The large community that comes with Python is another added bonus that you are likely to enjoy as well. This is a community of all kinds of programmers found all throughout the world. You will be able to work with beginners all the way up to advance, with those who are just beginning with

the work of coding for the very first time, all the way to professionals who have been doing it for a very long time.

This is a great thing for someone who is just getting started with Python for the first time. This allows you to find a bunch of other programmers, many with some experience in coding, who can answer your questions, show you how to try something new with the coding that you want to do, and more. Any time that you get stuck or want to learn something new with Python, this community will be able to help you out.

Python is also known as an OOP or object-oriented programming language. This is good news for someone just starting because it has a nice logical way that you are able to work with organizing your code. You will find that this is going to make it easier for you to store all of your objects, or parts of the code, into the right class, and then everything will be right where you left it when it is all done. This will make more sense when we explore it a little bit later, but it is basically going to make things easier when it is time to write the code.

All of the libraries that come with this kind of language are going to be nice to work with as well. The traditional or standard library that comes with Python is able to take on a lot of different tasks and can help with some of the basic codings that you want to do. But if you would like to do something that is a little more complex, such as what we see with data analysis and machine learning, then you may need to add something on to Python to make this work.

The good news is that Python comes with the capabilities and extensions that you need to make all of this happen. There are a ton of data science and machine learning libraries out there that match up perfectly with the Python coding language. This makes it easier for you to work on creating the kinds of codes that you need with Python, no matter how simple or how complex they may be.

As we can see, there are a lot of different parts that come with the Python language, and understanding what these parts are and how they work together is going to be so important when it is time to work on some of your codes. With this in mind, let's take some of those basics and explore some of the neat things that you are able to do with the Python language.

Chapter 2: How Can I Install Python on My Computer?

Now that we have had a chance to talk about Python a little bit, and some of the reasons why you would want to choose to work with this coding language rather than one of the others that are out there, it is time to get to work. Before we are able to jump in and start doing some of the codings that need to be done, we first need to take some time to learn the proper steps needed in order to install Python on your system

The good news, as we mentioned a little bit in the past chapter, is that Python is able to work with any kind of operating system or platform that you would like to focus on. This is good news because it ensures that we are able to go through and work with this language without changing up our

system, no matter what system is the one that we are using, or the one that we will want to work with for this project.

The first thing that we need to consider here is which version of Python we would like to work with. There are a few different options here, which is going to ensure that we are going to be able to pick the version that is right for you. Python is an open-sourced language, which means that it is free to install and use, and can be modified to fit your needs. There is still a development team that works on fixing the bugs and making the updates that are necessary. This is good news for you, but it does mean that you have to figure out which version is the right one for you.

There are Python 2 and Python 3 right now, with a few updates that fit into each one. They are similar, but they will often depend on the kinds of features that you would like to use. If your computer is already providing you with a version of Python 2, then this is a good one to work with, and you don't have to update it if you do not want to in this process. It is going to be pretty similar to Python 3 and some of the codes that we will do in this guidebook.

However, most programmers are going to work with Python 3. This is the newest version of the Python language that is out there, and it is going to have some of the best features and more that you need for coding. And it is likely that at some point, the developers who work on Python 2 will stop providing support for that version, so going with the newer option is going to be the best option to go with.

You will be able to find both of these downloads online to work with overall, so you have to choose which one is the best for you. You can also find versions of these that will work on all of the main operating systems so it should not be too hard to add these onto whichever system you are working with.

Another thing to keep in mind here is that with the coding we are going to look at below, and with the steps that we will look at in order to install Python on all three of the major operating systems, is that we are going to get our files from www.python.org. Python is going to come with a compiler, a text editor, and an IDE, along with some other files as well. These are all important to ensure that the program is going to run properly, and if one is missing, then you will not be able to get any of the codes that you want to work on.

There may be some other third-party sights that you are able to give a look at and then download those instead. This is an option, and depending on what you would like to do with the Python language when you are done, you may want to do this. But not all of them are going to provide you with all of the extras that are needed. You will have to go through and download each of the parts to make sure that you are not missing something.

On the other hand, when you download from the link that is provided at www.python.org, you will be able to know for a certainty that all of the files and parts that you need are there. you just need to go to the link there, click on the operating system that you are working with, and then move on to let it download on your system without any problems along the way at all.

Now that we have had some time to take a look at some of the things about installing Python on your computer, no matter which operating system you will use, it is time to take a look at the specific steps that are needed to make this happen.

Installing Python on Mac OS X

The first option that we are going to take a look at when it is time to install Python on our computers is the Mac OS X. If you are working off a computer that will rely on the Mac operating system, then you should notice

early on that there will be some kind of version of Python 2 already on the computer. The exact version is something that can change based on how long you have had your computer, but we can easily check on which version by opening up our terminal app and work with the prompt below to get it done:

Python – V

This is going to show you the version you get so a number will come up. You can also choose to install Python 3 on this system if you would like, and it isn't required to uninstall the 2.X version on the computer. To check for the 3.X installation, you just need to open up the terminal app and then type in the following prompt:

Python3 – V

The default on OS X is that Python 3 is not going to be installed at all. If you want to use Python 3, you can install it using some of the installers that are on Python.org. This is a good place to go because it will install everything that you need to write and execute your codes with Python. It will have the Python shell, the IDLE development tools, and the interpreter. Unlike what happens with Python 2.X, these tools are installed as a standard application in the Applications folder.

You have to make sure that you are able to run both the Python IDLE and the Python shell on any computer that you are using. And being able to run these can be dependent on which version you have chosen for your system as well. You will need to go through and check on your system which version of Python is there before proceeding and then moves on from there to determine which IDLE and shell you need to work with.

Installing Python on a Windows System

The second option that we are going to take a look at when it is time to work with the Python language on an operating system is on a Windows computer. There are many people who rely on the Windows operating system, so it is important to know how to go through this process and how to get Python set up. This process is going to take a bit longer than the other two, mainly because Windows already has its own coding language pre-installed, so we have to take some of the extra steps that are needed to get this done.

The good news is that even though Python is going to not be the automatic option on your computer doesn't mean that it won't work great. You will find that this coding language will be fine on Windows, and there will be a lot that you are able to do with it on these systems. It is just there to give you some notice that you have to take on a few extra steps.

Now that we have that out of the way, some of the steps that you will need to follow in order to make sure that the Python language is set up and ready to go will include the following:

1. To set this up, you need to visit the official Python download page and grab the Windows installer. You can choose to do the latest version of Python 3, or go with another option. By default, the installer is going to provide you with the 32-bit version of Python, but you can choose to switch this to the 64-bit version if you wish. The 32-bit is often best to make sure that there aren't any compatibility issues with the older packages, but you can experiment if you wish.
2. Now right-click on the installer and select "Run as Administrator." There are going to be two options to choose from. You will want to pick out "Customize Installation."

3. On the following screen, make sure all of the boxes under "Optional Features" are clicked and then click to move on.
4. When you are still under the part for Advanced Options, you need to spend a few moments figuring out where you would like the Python files to be installed. You can then click on the Install button. This will take a few minutes to finish, so have some patience and wait. Then, when the installation is done, you will be able to close out of it.
5. After this part is done, it is time for us to go through and create the PATH variable that we need for the system. We want to have it set up so that it includes the directories and all of the packages and other components that are needed later on. The steps that we need to use to make all of this fit together will include:

   a. First, we need to get the Control Panel opened up. We are going to do this by clicking on the taskbar and then typing in the Control Panel. Then we are going to click on the icon.
   b. Inside of this same Control Panel, do a search for the table of Environment. Then we can click on the button for Edit the System Environment Variables. When we are on that part, we are able to click on the part that says Environment Variables.
   c. Inside of the Environment Variables, we are going to find a tab for User Variables. You get two choices here of either editing the PATH variable that you find there, or you can create a brand new on.
   d. If there isn't a variable for the PATH already on the system you are using; then it is time for us to go through and create our own. You can do this by clicking on the button for New, make a name for the

variable of that PATH, and then add in the directories that you would like.

Then it is time to click or close out of all the dialogs that have come up for the Control Panel and go to the next step.

6. When you reach this part, you should be able to open up the Command Prompt. This is easy to do if you go to the start menu. Then click on Windows System and then on Command Prompt. Type in the word python. This will help us to get the interpreter of Python up and running and ready to go for us.

When we have the steps above done, which will really only take a few minutes overall and won't be as bad as it may seem when you first get started, you will be able to use the Python program on your Windows system. You can then choose to open up the interpreter of Python and anything else that you need, and writing out all of the codes that you need in no time.

Installing Python on a Linux Operating System

And finally, we need to take a moment to see some of the steps that are needed to make sure that we can properly install the Python language on our Linux operating system. Linux is not quite as big of a household name as we will find with the other two operating systems that we spent some time within this guidebook. But there are so many uses for it, and it is able to provide us with a bunch of neat features and more. There are many programmers who already use this language, and many more who are likely to do it in the future, so it is a good thing to know a little bit more about when it is time to install our Python coding language eon it.

The first step that we need to take is to see whether or not there is already a version of Python 3, or really even any version of Python found on our

operating system yet. The best way that we can get all of this done is to open up the Linux command prompt, and then write out the following code to see what information we are able to get:

- $ python3 - - version

If you are on Ubuntu 16.10 or newer, then it is a simple process to install Python 3.6. you just need to use the following commands:

- $ sudo apt-get update
- $ sudo apt-get install Python3.6

If you are relying on an older version of Ubuntu or another version, then you may want to work with the deadsnakes PPA, or another tool, to help you download the Python 3.6 version. The code that you need to do this includes:

- $ sudo apt-get install software-properties-common
- $ sudo add-apt repository ppa:deadsnakes/ppa
- # sudo apt-get update
- $ sudo apt-get install python3.6

The good news here is that if you are working with other distributions of Linux, it is likely that you already have Python 3 installed on the system.
If not, you can use the distribution's package manager. And if the package of Python 3 is not recent enough, or not the right one for you, you can go through and use these same steps to help install a more recent version of Python on your computer as needed.

And there we have it! We have now gone through and downloaded Python onto the three major operating systems that are out there. This is going to make things so much easier when you are ready to handle some of the codings that you want. And if you used the codes that we described, and

went through the other steps that we talked about, you will find that working with the Python language is going to be easy. You can now just open up the right files and all of the parts that you need, and it will work well for you!

Chapter 3: The Basics of the Python Code

At this point, we are going to know a bit about the Python code and why it is so important to helping you get a lot of projects done. We also know the steps that are needed in order to download and install this code, and all of the necessary files, on your computer. Now it is time for us to move on to some of the basic parts that come with the Python code and how we are able to use these to help us go through all of the more advanced codes that we will need to do later on.

The Keywords

The first thing that we need to take a look at is the Python keywords. These are basically going to be the words that we need to reserve. These are going to be the commands that you will need to give the compiler when you would like it to do. We need to make sure that we are using the keywords in the proper manner and ensure that we are not using them in the wrong place, or the compiler will really get confused and raise an error.

There are a number of keywords that are already found in the Python code, and it is important to use them in the right place when we are working on some of the codes that we want to do with this program. Remember that the

compiler is using these as commands to tell it what to do. So, treat them as something special along the way as well.

Looking at the Comments

The second thing that we need to take a look at when it is time to start writing some codes in the Python language is working on the comments. These comments are going to be important when it is time to leave a little note or a message inside of your code. This is helpful when you would like to leave this message, and you want to make sure that it is not going to interfere or mess with any of the rest of the code that you are writing.

To make this happen in Python, we just need to add the # symbol before any of the messages that we would like to use. You are able to add in as many of these as you would like, but you need to make sure that you keep it as limited as possible. This will help us to keep some of the coding as nice and neat as possible and will make it easier for no problems to show up because of the coding that you are trying to complete.

These comments are a nice way to work with the code and leave the messages and more that you are looking for. This will make it easier for you to write out something that you need to remember or that you would like another coder to know about before getting started, and all you need to do is add in the right symbol in front of the work that you are writing.

The Importance of Variables

The Python variables are also going to be important when it is time to handle some of our codes. And in specific, we need to spend some time learning how to assign a value over to the variable that we are working with. The variable is just going to be a blank piece of the memory if we don't make sure that some kind of value is assigned over to it. This means

that we will not be able to get something to show up when we try to pull that part of the code. If the variable assigns some value to it, then it is going to be able to react in this manner when you run the code.

As we work with some of these variables in our code, you may find that there are going to be three main options that you have available to use. Each of these is going to be useful, and it will depend on the type of code that you plan to use and the value type that you want to assign to that variable. The main variables that you are able to work with will include:

- Float: this would include numbers like 3.14 and so on.
- String: this is going to be like a statement where you could write out something like "Thank you for visiting my page!" or another similar phrase.
- Whole number: this would be any of the other numbers that you would use that do not have a decimal point.

As we do some work with these variables, we have to remember that we do not have to spend our time creating some declaration to save up this particular spot in memory. This is needed in some of the other coding options that you use, but it is going to happen automatically when working in Python right when you assign a value over to the variable with the help of the equal sign. If you would like to make sure that this is going to happen, you are able to look to see that this equal sign is there, and then you know that it is going to work the way that you want.

The good news here is that assigning a value over to the variable is going to be pretty easy. Some of the ways that we are able to do this inside of our code will include:

x = 12   *#this is an example of an integer assignment*
pi = 3.14   *#this is an example of a floating-point assignment*

customer name = John Doe   #this is an example of a string assignment

Another option that you can do here, and that we have mentioned a little bit in this chapter already has a variable assigned to two or more values. There are some instances when you are writing code, and you will need to take two values and place them with the same variable.

To do this, you simply need to use the same procedure that we talked about above. Just make sure that you add in an equal sign to each part to help the compiler know that these are all associated with the same variable. So, you would want to write out something like a = b = c = 1 to show the compiler that each of these variables equals one. Or you could have 1 = b = 2 to show that there are two values to a variable.

The most important thing to remember here is that the variable needs to be assigned to a value in order to work in the code and that these variables are simply spots that are reserved in the memory for those values. Then, when the compiler needs to call up those values, it simply will call up the variable to bring everything out.

Bringing in Python Strings

For some of the codes that we are going to write out in Python, there is going to be a type of data that is known as a string. These strings are going to just be a series of characters that are found in your code and can help you to get different things done in the process as well.

We also will be able to see that there are a variety of operators that we are able to work with when we want to make sure that the strings are going to work with our Python code. An operator is going to just be a symbol that is able to perform a specific operation in the code that you are writing. In this

case, the operation is just going to be done on the string and nothing else. Some of the operators that are going to be nice for helping us to get this done and to ensure that the code will behave will include:

1. The concatenation operator: This is going to be the operator that will come in use when you would like to concatenate some of your strings.
2. Repetition operator: This is going to be the best operator to use when it is time to create many copies of the same string. You will be able to make some decisions on how many different times you would like to repeat that string and how many copies you would like.
3. The slice operator: This is the best operator for a programmer to use when they want to be able to go through a specific string and then pull out a character that they would like. Any time that we are working with this option, we are going to have to remember that 0 is going to be the first character of the string, and then you can move on from there.
4. Range slice operator: This is a good operator that we are able to work with when we wish to a range of characters from the index, rather than just picking out one character. This would work if you would like to show off one word, one part of the sentence, or one sentence.
5. In operator: This operator is going to search for a specified character in a target string. If the character is present somewhere in the string, then you will get an answer of True returned to you. If that character is not inside the string, then you will get an answer of False returned to you.
6. No in operator: This is the operator that will work in the opposite manner as the in operator. It is going to search for a specified character in your string. But if the operator is not able to find that character in the string, then you will get the True

answer returned. If that character is found in the string, then it is going to return False.

The Python Functions

Another part of our basics that we are going toned to look at is the Python functions. These functions are just going to be a set of expressions that we will know as statements as well. They are going to be some of the first objects that fit int classes, which means that you will not have to concentrate on a ton of restrictions as well. There will be a lot of freedom on what you are able to use. You will instead be able to use these functions how you would like, and they are going to work in a manner that is similar to things like strings, numbers, and more. And these are going to also come with some attributes that we are able to bring out as long as we are working with the prefix of dir for our function.

One thing that you are going to find is that with these functions, you are going to find a ton of diversity that comes with them. And there are also a lot of attributes that are there for you to use to create and bring up these functions. Some of the different choices that you are going to see with these functions will include:

- __doc__: This is going to return the docstring of the function that you are requesting.
- Func_default: This one is going to return a tuple of the values of your default argument.
- Func_globals: This one will return a reference that points to the dictionary holding the global variables for that function.
- Func_dict: This one is responsible for returning the namespace that will support the attributes for all your arbitrary functions.
- Func_closure: This will return to you a tuple of all the cells that hold the bindings for the free variables inside of the function.

There are different things that you can do with your functions, such as passing it as an argument over to one of your other functions if you need it. Any function that is able to take on a new one as the argument will be considered the higher-order function in the code. These are good to learn because they are important to your code as you move.

The Operators

We can also spend a little bit of time taking a look at the operators that show up in the Python language. These operators are going to help us to make our codes a little bit stronger and will ensure that we are able to see some great results in the codes as well. We will also find that there are many types of operators that we are able to work with, based on the operator that you want to work with.

For example, there are arithmetic operators. These are the ones that allow you to add, subtract, multiply, and divide parts of your code. There are comparison operators that will allow us to have a way to compare one or more parts of the code with each other. And we can even do assignment operators that are going to help us to work on putting a value to the variable. You have to choose which option is the best for the coding that you do.

There are a lot of different parts that you are able to work with when it comes to some of the basics of the Python code. Knowing how to make some of these parts work for your needs and the codes that you want to write is going to be important. Even some of the more complex codes that you will work on later, you will find that working with some of these basics is going to help us to get through that part of the code.

Chapter 4: The Different Data Types in Python

Data is going to be a really important part of the Python language. No matter what kind of coding you are trying to work on at the time, you will find that Python is going to be able to take control over it, and ensure that you are able to see some amazing results in the long run. So, with that in mind, we need to spend some of our time learning a bit more about the different data types that are found in Python, and how we are able to use these for our needs.

For the most part, the types of data will be able to help define the format, will set up the upper and lower bounds of the data to help the program use it appropriately and more. But we will find that the data types that we use in

Python are going to be a bit more than that. In Python, we don't even need to spend time declaring the variable by explicitly mentioning the data type. This is known as dynamic typing and can really save us a lot of time and hassle in the process.

Python is able to determine the type of a literal just by looking at the syntax that is found at runtime. For example, the quotes are going to be there to mark the declaration that is found in a string value, the square brackets will show us a list, and the curly brackets are going to be for the dictionary. Another example of this is when the decimals that are not numbers are going to be assigned to a type of integer, but then the ones that have a decimal in them will be seen as a floating number instead.

Everything, including the modules, the functions, and variables will be a Python object. Another interesting fact here that we can look at is that the variables are not going to have types like they do in other languages. Instead, we will find that they are just going to be labeled. It is the value that we are going to associate over with a variable that will have a type. This is why the same variable can talk about different Python data types.

With this in mind, there are quite a few different types of data that we are able to use in the Python language. We are going to take a look at some of the most common ones below:

First, we are going to have a Boolean. This is going to be a data type that is found in almost all of the programming languages that we can work with. The Boolean that we work within Python is going to come with two values, the True and the False. These values are going to be constants, and we are able to use them to either compare or assign a Boolean value.

Another type of data that we are able to work with is the numbers. These are actually going to be one of the most prominent of the data types in Python. Unlike what we are going to see with some of the other languages

out there, the ones that will rely on just floats and integers, Python is going to introduce a few different parts, such as the complex, to help us handle these numbers. The things that we need to remember when it comes to working with numbers in this language will include:

1. The numbers that we see in Python are going to be classified with some of the keywords that you may have seen already, including complex, float, and int.
2. Python is going to have a function known as type() that is already built-in, and that will determine the data type that comes with a variable or with a value we are using.
3. Another function instance() is going to be built in and will help us to test for the type that we have with an object.
4. We are also able to work with imaginary and complex numbers when it comes to working in Python. We just need to add in the "j" or "J" to help with this.
5. To help us to form a complex number, we are able to use the type as a constructor if we would like.
6. The integers in this language are not going to come with a size limitation at all, as long as we already have some of the required memory in place.
7. Afloat type number is able to have some precision. This can't go on indefinitely, though. But we can see it go up to 15 decimal places if we would like.

Another data type is the string that we talked about a bit before. This is going to be a sequence that has one or more characters that we are able to enclose within one single quote or in a double quote. We are able to put any kind of number, letter, or symbol into the string that we want to work with. You will find that the Python language is also able to support a multi-line string. You just need to make sure that there is a triple quotation mark right at the beginning of this string and then one at the end.

Next on the list is the byte. This is going to be considered a data type that is immutable. It is able to store a sequence of bytes, which is each going to be 8-bits, which will range from between 0 to 255. Similar to what we see with an array, it is going to be able to fetch the value of one byte when we use the index. But we will not be able to modify the value that we are working with.

Now that we compared the byte to a string a little bit, we need to make sure that we know how they are different. Some of the main differences that come with these include:

1. The byte objects are going to hold onto a sequence of bytes. But the string is going to be responsible for storing a sequence of characters.
2. The bytes are going to object that can be read by a machine. But the strings are going to be in a form that humans are able to read.
3. Since the byte is something that the machine is able to read, it can be stored right onto a disk without more work. But the strings have to go through and have some encoding done before we can change them to be on a disk.

One of the situations where we are going to find that these bytes really do matter is when we are carrying out the I/O operations with buffering enabled. We may have, for example, a program that is always receiving all of the data that comes over the network. This program is going to parse the data after waiting for the terminators of the headers of the message to appear in the message. It is going to keep appending the incoming bytes to a buffer.

Now we are able to move on to the lists. The Python list is going to be a construct that is similar to an array, and it is able to store some of the arbitrarily typed objects in an ordered sequence that you would like the

most. It is going to be a very flexible option when it comes to working with data types, and it doesn't have to come in at a fixed size. The index is a list that is going to begin on 0 when we are in Python.

Another definition that we are going to see when we are working with lists in this language is that these are going to be a heterogeneous collection of items of varied data types that we are able to work with. For example, when we have a list object, we are able to store the files onto a folder, or the employee data in a company and know where it is later on. These lists and the objects that are found in them are going to be mutable. Python is going to allow us to make some modifications to a list, and any of the elements that are in it through assignments and we can work with some of the list methods that come with Python as well.

We are also going to spend some time looking at the Tuples in this language. A tuple is a bit different, but there may be some similarities that we need to focus on. A tuple is basically going to be a heterogeneous collection of objects in Python that are separated out with commas. This means that the objects of different types of data are able to be together in one tuple. The tuple and the list are going to be similar because there are a lot of traits that they share together, including the following:

1. Both of the objects are in an ordered sequence.
2. They are able to store values of different types
3. Nesting is something that we are able to do in these.
4. They are going to allow us to work with indexing and repetition as much as we would like.

However, the tuple and the list are going to be different. The biggest reason that they are so different is that the tuple is going to be seen as immutable. Python does not allow us the option to make modifications to the tuple once it is created. This means that we are not able to remove or add any kind of element at a later time. If we do need to go through here and make some

changes, then we will need to go through and create a brand new one with an updated sequence of elements to get this done.

At some point, you may find that you need to work with something that is known as a set in Python. This is the one option that is going to support some of the different operations of math that you may need to do like symmetric difference, intersection, and union. If you plan to do some mathematical work with the use of Python, then this is the data type that you want.

To keep it simple, a set is going to be an unordered collection of immutable and unique objects. Its definition is going to have some enclosing braces with it, which allows us to have the items separated out with a comma inside of there. since the set is going to get some of the implementations from the Set in mathematics, it is also going to be able to have more than one occurrence within the same element.

You will find that there are many codes that you can write that will need a set. The set type is going to have a big advantage over some of the lists that you may want to write as well. It is able to implement a highly optimized method that is then responsible for checking whether the container is hosting a specific element or not the mechanism here is based on a data structure that we can call a hash table.

And the final data type that we are going to spend some time on in this chapter is the dictionary. This is going to be an unordered collection in Python that has some key-value pairs we can call it a built-in mapping type in Python where the keys are going to provide us a map to the values that we are working with. These pairs are there in this language because they will help us to have a more intuitive method to store the data that we have.

The dictionary is so important because it is going to come in and help us solve the problem that comes up with efficiently storing a large set of data.

We are going to spend some time later in this guidebook talking about data analysis and what it is able to do for your business if you let it. But the fact is that most companies are already gathering a huge amount of data to use for some of their own needs. And the dictionary is the perfect tool to make all of this happen. Python has already set it up so that the dictionary object is going to be optimized as much as possible to retrieve the data that you would like in the process.

As we can see here, there are a lot of different types of data that we are able to use when it comes to working with the Python language. Each of these data types is going to be so important to what you would like to accomplish with this kind of process, and you will find that you will have your choice of them based on what is going on in your code at that time.

Take some time to look over the different types of data and what they are able to do for you, and then see how valuable they will be to your needs.

Chapter 5: The Python Functions

Another topic that we need to spend some time looking at when it comes to the Python language is that of the functions. Blocks of code or some logic in the code that is able to perform a specific operation that you would like are going to be known as a function in Python. We have already spent a bit of time in the previous chapters taking a look at what these functions are all about and how we are able to get them to work for our needs. But now we need to take a bit closer look at what these functions can do and how they are able to help us get the best results in no time.

The Different Types of Functions

There are going to be two main types of functions that we are able to find in the Python language. Knowing what these are and how we are able to make them work is going to be so important when it is time to work with this language. The two main types of functions that show up here will include:

1. The library functions: These are going to be some of the built-in functions that show up in the python language. We do not need to worry as much about some of the logic that is in the library function because these are the ones that Python is able to handle on its own. You will find that some of these library functions that will be there often in your codes may include examples like round() and print().

2. User-defined functions: Instead of relying on some of the built-in functions, this language is going to allow a user to create some of their own functions as well. So, if you would like to do something like performing a mathematical calculation in these codes, then you would be able to place them in a separate

function with the right name, and then you are able to call up that function many times.

The Advantages of Python Functions

With that understanding in place, we need to take a look at some of the advantages that we are able to see when we are working with our functions in Python. There are many of these, which is why you are going to see these functions show up in a lot of the work that you do and in many of the codes that are needed for this language.

The first advantage that we are going to see with these functions is that it is able to take some of the larger programs that you are writing out, and then divide them into smaller groups. This may seem like a bad idea, but it is done so that we are able to gain a better understanding of the code in a faster manner, and it will really make the process of debugging the program easier and better in no time.

Another benefit is that some of these Python functions are going to prevent us from writing out the same kind of logic more than one time. This is good news because it helps us to wrap up the logic in just one function, and then we can call it up the same many times over.

Then there are two more benefits that are going to come up when we are able to work with some of the Python functions. First, we will find that these functions are going to allow more than one person to work on the exact same program at the same time. this is because we can assign different functions to each person on the team and let them get to work on that. It is also able to help us out because we can call up the same function with different inputs over and over again, as many as we would like, and can save a lot of time and hassle in the process as well.

The Syntax of the Function

We also need to spend some time taking a look at the way that this is going to appear in some of the codings that we do. To start, we need to look at the syntax that comes with Python functions, and then we can move on to an explanation of some of the different parts that are in it, so we know more about how this is supposed to work.

def Function_Name (Parameters): Local Variable Declaration Programming Logic Executable Statement 1 …… return

1. Def: This is going to be the keyword that will introduce the function definition that we are working with. Keep in mind here that the def keyword needs to be followed right away with the Function_Name.
2. Function_Name: This is going to be any name that you would like to give other than the reserved keywords that we talked about before.
3. Parameters: Every function of Python is going to accept any amount of parameters that you would like. You can have none, or you can have many. It is going to be dependent on what the requirements of the user are all about.
4. The local variable declaration: Sometimes we may need to go through and use a temporary variable. These are going to be the kind that is only needed for that function then we can declare them inside of the function that we are working with. It is not something that is mandatory here, and it is going to depend on the kinds of requirements that you want to have for your work. Remember that these variables are going to be available just to one particular function, and you will not be able to access them when you are outside of that function.
5. Logic: This is going to be anything mathematical or any type of code or calculation that you would like to be able to implement on that kind of function.

6. Executable statement: This is going to be any of the print statements that tell the code to print some data from the function that you would like.
7. Return: This is another keyword that we are able to focus on because it is going to be needed if we would like to return something form our function. For example, we can use this to help us return the sum of our two integers and more.

There are a lot of things that we are able to do when it is time to bring out the functions and get them to work the way that we would like. When we are able to bring out those functions and use them properly in the code, they are going to ensure that we get the best results out of this possible and that our code stays strong and secure as well.

Chapter 6: How to Write Your Own Conditional Statements

The next thing that we are going to take a look at is a conditional statement. These conditional statements are going to make it easier for the system to respond to any kind of answer that the user is able to give, without the programmer having to know what this is ahead of time. it is pretty much impossible for the programmer to go through and know what answers the user is going to provide in all situations, so working with these conditional statements will help the program run in a proper manner to the information that the user is going to provide.

There are a lot of different programs that are going to rely on these conditional statements. These can often be seen as something that is a bit more simplistic to work with, but this is always something that we are able to add to our code and see some great results. The three conditional statements that you are likely to see along here is the if statement, the if else statement, and the elif statement. Let's take a look at each one and who it is able to work.

Starting with the If Statement

The first of these conditional statements that we need to take a look at is the if statement. You will quickly find that this is one of the most basic of the three conditional statements because it is only going to let the program that you have written proceed forward when the user provides the right answer based on your condition. With this one, if the user ends up putting in an answer that is seen as false based on what you have put into your code, then the code is going to stop, and nothing is going to happen. If the answer

though, is seen as true on your conditions, then your code is going to display some kind of message on the screen, or there is going to be some other task that you assigned.

As we can see already, there are going to be some times when if the statement is going to cause problems in the codes that you are writing, but it is still going to be an important thing for a beginner to learn how to use these statements. A good example of the syntax that you are able to write out to handle some of these conditional statements will include:

age = int(input("Enter your age:"))
if (age <=18):
     print("You are not eligible for voting, try next election!")
print("Program ends")

There are a few things that will show up with this code. If you have a user go to the program and state that their age is under 18, then the program will work and display the message that is listed there. The user can read this message and then end the program right there.

But, things can go wrong if the user puts in that their age is above 18. This is true for the user, but because it doesn't meet the conditions that you wrote out, the program will see it as false. As the code is written right now, nothing is going to happen because it isn't set up to handle this. The user will just see a blank screen any time they enter an age that is over 18.

Moving to the If Else Statements

The second type of conditional statement that we need to take a look at is the if else statement. We can already see that while the if statement is a good one to practice with, but they're probably are not too many times in the program when you would be able to use this kind of statement. It is

likely that the codes you are trying to write will benefit if the user is able to work with whatever answer the user puts in, no matter what their age is. And it is not likely that you want to leave your code so that nothing shows up all of the time.

A better option to use is the if else statement. These statements will provide an output to the user, no matter what they provide as their input. With the example above, the user may get the previous message if they state their age is 16, but then the code would also have a response if the user says their age is 32. It will respond to any answer that it is given, helping the code to keep moving through the program and ensuring that nothing just stops.

With the voting example that we had above, you can implement the following code to make an if else statement:

age = int(input("Enter your age:"))
if (age <=18):
    print("You are not eligible for voting, try next election!")
else
    print("Congratulations! You are eligible to vote. Check out your local polling station to find out more information!)
print("Program ends")

With this option, you are adding in the else statement, which will cover every age that doesn't fall under 18. This way, if the user does list that as their age, something will still pull up on the screen for them. This can provide you with more freedom when working on your code, and you can even add in a few more layers to this. If you want to divide it up so that you get four or five age groups and each one gets a different response, you simply need to add in more if statements to make it happen. The else statement is at the end to catch everything else.

There are a lot of options that we are able to use with this one. For example, you can work with the code syntax that we had before and then ask your

user what their favorite color is. you can then work with the if else statements to help you cover up some of the basic colors that you would like. If the user is able to put in one of these colors, then the statement that you put in it is going to show up on the screen. The else statement is going to be the part that we can add to hold onto any of the other colors that you did not plan on, and gives us a good statement to go with it as well.

As we can see, there are a ton of different benefits that are going to come with using the if else statement. These are going to come in to help us really get some results, no matter what input we are getting from the user. There are a lot of programs where you will be able to use the if else statements and get the best results.

Finishing Off with the Elif Statements

And the third type of conditional statement that we are going to work with is the elif statement. These elif statements are going to add on another level, but they are going to be easy to work with as well. You are able to add in as many of the elif statements to the code as you would like to help with the code, and this can help us to cover up any of the different conditions or statements that we want to cover in the code. Think of this elif statement in a manner similar to what we see with some old games, the ones where a gamer is able to pick from options on a menu to decide what actions they would take. The elif statement is going to work in a similar manner.

From this, we are going to be able to work with as few or as many options as the elif statement. You are able to go through and just two or three of these, or you are able to expand out and have thirty of these if you want. A good rule here, though, is that less is more in many cases. You don't want to have too many of these in place, or you are going to end up with a mess, and that is a lot of code to write out in the process as well.

Now that we have had a moment to discuss what these elif statements are about, and how they differ a bit from the if statement and the if else statement, let's take a look at some of the syntax that we are able to work on when it comes to programming with these statements:

Print("Let's enjoy a Pizza! Ok, let's go inside Pizzahut!")
print("Waiter, Please select Pizza of your choice from the menu")
pizzachoice = int(input("Please enter your choice of Pizza:"))
if pizzachoice == 1:
    print('I want to enjoy a pizza napoletana')
elif pizzachoice == 2:
    print('I want to enjoy a pizza rustica')
elif pizzachoice == 3:
    print('I want to enjoy a pizza capricciosa')
else:
    print("Sorry, I do not want any of the listed pizza's, please bring a Coca Cola for me.")

With this option, the user is able to choose the type of pizza they want to enjoy, but you can use the same syntax for anything you need in your code. If the user pushes the number 2 in the code, they are going to get a pizza rustica.
If they don't like any of the options, then they are telling the program that they just want to have something to drink, in this case, a Coca Cola.

This is a simple way to use the elif statements to give the user a set of choices. In the other options, the user could add in any choice that they wanted, but in the elif statement, they can either pick that they want one of the choices that you provide, or they will have to go with the default option at the end. This can work well for many games that you may want to pick, for some tests online, and other programs where you want to limit the choices that the user gets to pick from in the code.

And this is how the conditional statements are going to work for your needs. You are able to set up any kind of condition that you would like to work with to ensure that the program is going to work the way that you want.

You can set it up with one true condition like the if statement, you can set it up so that you can handle all of the answers that the user can give and still get some kind of response out of the process, or you are able to work with the elif statements and do something that is similar to how a menu works in a game. Each of these can be important, and it is easy to add these conditions into our code to add in some power.

Chapter 7: The Python Classes and How to Write Your Own

One thing that is really neat about working with the Python coding language is handling some of the classes that come with this language. These classes are going to be the containers in the language that is able to hold onto all of the objects, or the different parts, that come with some of the code that you are writing. Being able to write these classes and get the right kinds of codes into them is an important part of keeping your code as organized and easy to use in this language as possible.

That is why we are going to spend some time in this chapter taking a look at how we are able to work on creating a new class. This is one of the best ways to organize the code and will make it so that nothing will get lost or move around in the wrong manner when it is time to execute the code that you want to write. To get started with the process of making one of these classes though, we need to make sure that we use the right keywords.

With the classes, you have some freedom in naming them anything that you would like. We just need to make sure that we are not using a keyword as the name, and that the class name is going to show up right after the keyword that we are using. And it is usually a good idea to pick out a name for the class that you will be able to remember and pull up later.

After you have had some time to go through and name your class, we also need to spend some time naming the subclass. This is the part that is found inside of the parenthesis. And then add in a semicolon at the end of this line so that you match up with some of the good coding protocols in Python that are available to make things easier.

With all of this running through your head, it is likely that this process, and the idea of creating your own class, is going to seem really complicated. That is why we need to stop for a few minutes and look at an example of how we would want to write all of this out in Python. The code that we are going to use to help us to write out a class in the Python language will include:

```
class Vehicle(object):
#constructor
def_init_(self, steering, wheels, clutch, breaks, gears):
self._steering = steering
self._wheels = wheels
self._clutch = clutch
self._breaks =breaks
self._gears  = gears
#destructor
def_del_(self):
    print("This is destructor….")

#member functions or methods
def Display_Vehicle(self):
   print('Steering:' , self._steering)
  print('Wheels:', self._wheels)
   print('Clutch:', self._clutch)
   print('Breaks:', self._breaks)
   print('Gears:', self._gears)
#instantiate a vehicle option
myGenericVehicle = Vehicle('Power Steering', 4, 'Super Clutch', 'Disk Breaks', 5)

myGenericVehicle.Display_Vehicle()
```

If you would like, you can try out this code. Just open up your text editor and type the code inside. As you work on writing this out, you will notice that a few of the topics we have already discussed in this guidebook show up in this code. Once you have a chance to write out and then execute this code, let's divide it up and see what happened above.

Now the first part here is the class definition. This is the part where we have to instantiate the object, and then we end up with the definition of the class. This is going to make it easier to know the right syntax of the code. During this time, pay some attention to the code because it is the part that will tell the compiler what you would like it to do, and it is going to highlight the commands that you think are the most important there. if you would like to take the time to bring on a new kind of class definition to the code, then you will be able to work with the functions of object_method or object_attribute to help.

Then we are able to work on some of the special attributes. These are found in many of the codes that you want to write out, including the one above. These attributes are good because they will provide the programmer with some peace of mind, ensuring that this attribute is going to be used in the proper manner and that you will not mess it up. There are a few options with these that are found in the code that we wrote before, but there are a few other special attributes that you can consider in the codes that you would like including:

__bases__: this is considered a tuple that contains any of the superclasses
__module__: this is where you are going to find the name of the module, and it will also hold your classes.
__name__: this will hold on to the class name.
__doc__: this is where you are going to find the reference string inside the document for your class.
__dict__: this is going to be the variable for the dict. Inside the class name.

And finally, we need to take a few minutes to look at the steps that are needed in order to access the members of one of your classes here. You want to make sure that while you are writing the code, both the text editor and the compiler are set up to recognize any of the classes that you are writing. This will ensure that your code is going to be executed in the proper manner.

But for the compiler to execute the code in the right manner, then the code is going to need to be written properly. There are going to be a few different methods that the programmer is able to choose to make this work, and all of them can help you get the work done. However, we are going to focus on working with the accessor method because it is a common option to work with, and it is often going to be the most efficient to get the work done.

To help us to understand the process that is needed in order to access the different members of your created class, the first thing that we need to do is take a look at the code below:

```
class Cat(object)
    itsAge = None
    itsWeight = None
    itsName = None
    #set accessor function use to assign values to the fields or member
vars
    def setItsAge(self, itsAge):
    self.itsAge = itsAge

    def setItsWeight(self, itsWeight):
    self.itsWeight = itsWeight

    def setItsName(self, itsName):
    self.itsName =itsName
```

```
    #get accessor function use to return the values from a field
    def getItsAge(self):
    return self.itsAge
    def getItsWeight(self):
    return self.itsWeight

    def getItsName(self):
    return self.itsName

objFrisky = Cat()
objFrisky.setItsAge(5)
objFrisky.setItsWeight(10)
objFrisky.setItsName("Frisky")
print("Cats Name is:", objFrisky.getItsname())
print("Its age is:", objFrisky.getItsAge())
print("Its weight is:", objFrisky.getItsName())
```

Before we move on, type this into your compiler. If you have your compiler run this, you are going to get some results that show up on the screen right away. This will include that the cat's name is Frisky (or you can change the name to something else if you want), that the age is 5, and that the weight is 10. This is the information that was put into the code, so the compiler is going to pull them up to give you the results that you want. You can take some time to add different options into the code and see how it changes over time.

Keep in mind during this process that there are often a lot of classes that you are able to work with inside of your Python code, and they are meant to make life easier in coding, not more difficult. These classes can come into use when you would like a better way to take care of your information and make sure it is in the right order so that it makes a lot more sense. You are able to create the class that you would like and add in the objects that match the most.

But do keep in mind that when someone else comes and takes a look at some of the code that you are writing, they should be able to look into the class and understand why some of the objects are found there. The objects don't need to be identical with this, but they do need to make sense as to why they are going to be found in the same class.

There are a lot of times when a class is going to be useful inside your code. And it is definitely worth your time to take a look at them and learn how to use them. Most of the codes that you are going to work on and write in Python are going to rely on these classes, so take some time to look through this code and see how you are able to write out some of the classes that you want to do.

Chapter 8: Handling Files in Python

Working with files is going to be a topic that comes up in the Python language. When you are working in Python, there are going to be times when we would like to take some of the data that we are working on and then store it, making it accessible for you to use later when the data is needed. You are able to choose to save this data in a manner that you are able to find it later on and then see it show up at the right time in any code that you are working with.

This data is something that you are able to save, just like a file on a disk. Or you can make the decision to reuse the data as often as you would like, doing something like an inheritance, as long as you make sure that the correct steps are done along the way. While inheritances are an important part of a lot of the code that we want to write, we need to focus more now on how to create and work with some of the files that are found on the Python code.

Now, there are going to be a few things that we are able to work with when it is time to handle the files in Python. To make sure that this is as easy to understand as possible, we can imagine the files as similar to what we are going to see in a Word document that you have written in, and then at some point, you will want to save them. These files are going to work in a similar manner, but rather than saving the pages, you will need to go through and save some of the code that you are writing.

In this chapter, we are going to take a look at a few of the different things that you are able to handle when it is time to work with some of these files in Python. We are going to take a look at how we are able to write out a new code onto a file that we were able to create before. We are going to

look at how we can seek out or move a file to a brand-new location. We are going to look at how we are able to create a file. And then we can finish up with some information on closing and saving a file that we worked on.

Creating Our Own New Files

One of the first things that we need to focus on when it comes to working with these files is how to create a new one when you would like so that you can add in some code to that file. If you want to make your own file and create it with some code on it, then it is important that we open it up inside of the IDLE. You will then need to choose the model that you would like to use when you write some new code. As you work with these, there are going to be three main options for the codes, and you can pick the right one based on what your code is doing. The three modes that are present will include appending (a), write(w), and mode(x).

Any time that you want to make some changes to the current file that is opened, you can use the (w) or write option because this is often the easiest one to work with. Any time that you are trying to open up a file and then write a new string in that file, you will work with binary files and will need the write() method. This is going to work well because it ensures that you will get the right characters returned at the end.

The write() function is really easy to use and allows you to make as many changes to the file as you would like. You can add some new information to the file, or you can make some changes to one that you opened up. If you are interested in doing the writing in the code, you can use this example to make things easier:

#file handling operations
#writing to a new file hello.txt
f = open('hello.txt', 'w', encoding = 'utf-8')

f.write("Hello Python Developers!")
f.write("Welcome to Python World")
f.flush()
f.close()

Before we take the time to go through and learn more, we need to type the code above onto the compiler and let it execute. This code is helpful here because it will ensure that you can create a new file and get all of the information that you want into the right directory. Keep in mind that the default directory is known as the current directory at all. It is always possible to go through and switch out the directory that you want to use to store the information, but this is something that you should do before writing any code in it because it is much easier to accomplish.

Whatever directory you are in when you get started is the one where you will need to find yourself back in when it is time to search for that file. If you want to find that file in a new directory, then it is time to move it over there before you save. With the option that we wrote out above, when you go to the right directory, and you open it up, you are going to get the message to show up that is there.

At this point, we have taken some time to write out this program, and we may be happy with it. But then there are going to be times when we would like to do some overwriting of the program. This will change things up so that we will end up with a new statement, rather than the one above, that shows up on the screen when we open that file. This is possible to change things up; even on a previously created file, we just need to make a few changes to the syntax that we are working with. A good example of what we need to write out in order to overwrite our previous code will include:

#file handling operations
#writing to a new file hello.txt
f = open('hello.txt', 'w', encoding = 'utf-8')

f.write("Hello Python Developers!")
f.write("Welcome to Python World")
mylist = ["Apple," "Orange," "Banana"]
#writelines() is used to write multiple lines into the file
f.write(mylist)
f.flush()
f.close()

The example above is a good one to use when you want to make a few changes to a file that you worked on before because you just need to add in one new line. This example wouldn't need to use that third line because it just has some simple words, but you can add in anything that you want to the program, just use the syntax above and change it up for what you need.

Can I Create a Binary File?

We are going to take a little bit of a detour here and explore for a minute what we are able to do with binary files in Python. This is going to be a bit different, but it is simple and can make a difference in some of the codes that we are writing. You can complete this process in Python because you just take the data that you are working with and then make sure that it turns into a sound or an image file, rather than writing it out as a text file like before.

You can change any of the text files that you write out in this language over to a binary file if that is easiest. In fact, it doesn't matter what kind of file you are working with, whether it is a picture, sound, or text file. The most important thing that we have to remember when we are changing things into binary files is that we need to supply our data inside of the object to make it easier to expose a bit later on. We can then use the syntax below to help us get our files turned into binary files:

 # write binary data to a file
 # writing the file hello.dat write binary mode
F = open('hello.dat', 'wb')
 # writing as byte strings
f.write(b" I am writing data in binary file!/n")
f.write(b" Let's write another list/n")
f.close()

How to Open a File

We have taken a look at how to create a file and how to turn any file that we have over to a binary file. These are both really important topics when we are exploring what can happen with the files in Python. But now we need to take some time to look at the process of opening up our files.

At some point, you will need to close down a file and move on to something else. Then at a later time, you will want to open up one of the files again and see what is in it or get it to execute it. Once you have taken the time to open up that file, it is going to be easier to work with again. The code that we are able to work with here will include:

 # read binary data to a file
#writing the file hello.dat write append binary mode

with open("hello.dat", 'rb') as f:
*   data = f.read()*
*   text = data.decode('utf-8'(*
print(text)

the output that you would get form putting this into the system would be like the following:

Hello, world!

This is a demo using with
This file contains three lines
Hello world
This is a demo using with

This file contains three lines.

As we can see here, this is going to be a really simple way that we are able to open up the file that we are working on, and then we can go through and overwrite it as we talked about before, or just write something brand new as a continuation of what we were talking about before.

How to Seek One of the Files

As we go through this chapter, you shall find that these files are a great way to work with and keep the Python script as clean and organized as possible. We have taken some time to really explore a few of the things that we are able to work with when it is time to see results with these files. All of these things can be useful to ensure that the files are going to be opened, written in, and saved in the manner that we would like along the way.

We have taken the time to write out a new file, how to overwrite some of the data that we already wrote in, and even how to open up the files. There may also be some times when you want to move a file to a new place. For example, if you are working on your file, and you find that things are matching up the way that you want, such as misspelling the name of your identifier the wrong way, or you place it in the wrong directory, you may want to use the seek option to help you find that file and then make the changes.

You will be able to go through and change up the position where the file is so that it goes to the right spot, or even so that it becomes a lot easier for it

to find. You just need to tell your code where you would like to find the code and then make the changes that are needed.

Working with files in the Python language will help you out a lot when you are trying to get things in order inside your code when you want to make changes to what you wrote, and so much more. Try out a few of the codes that are inside this guidebook and see just how easy it can be to create a new file, make changes, and get the files to work the way that you would like.

Chapter 9: Tips and Tricks to Get the Most Out of Python
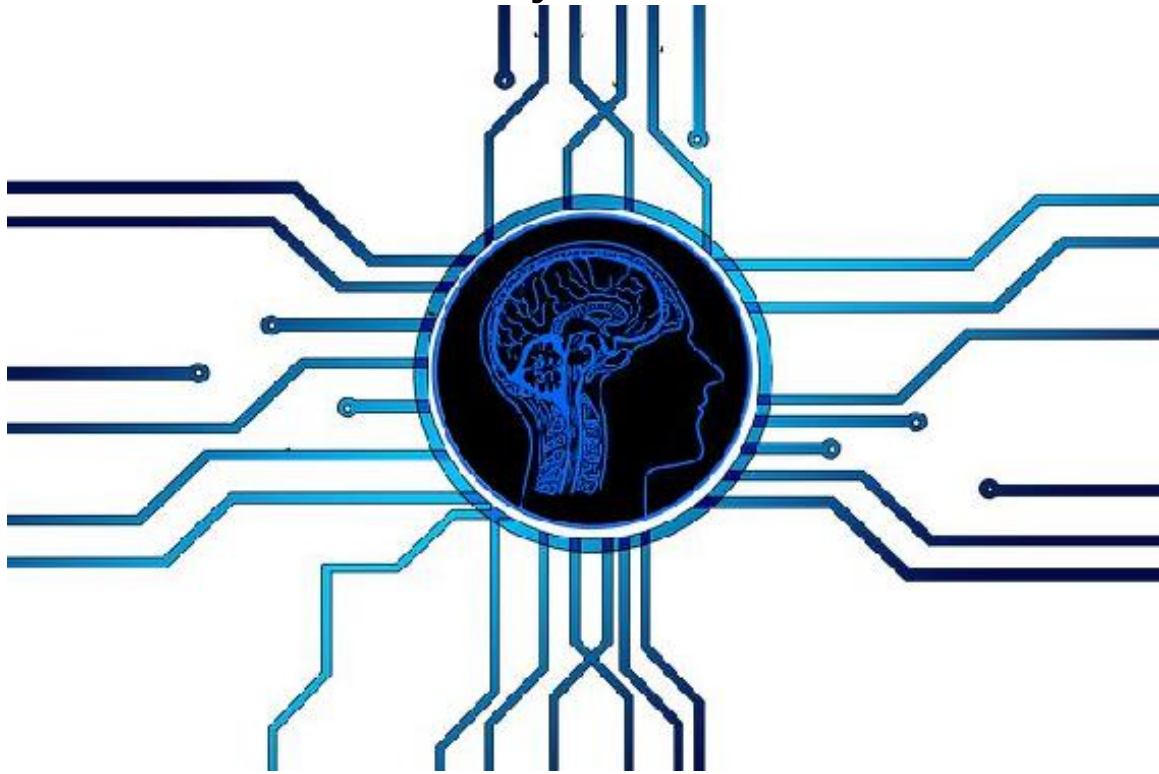
Now that we have had some time to talk about the Python code and all of the different things that we are able to do with that kind of code, it is now time for us to work on how to get the most out of this code. In this chapter, we are going to really focus on some of the things that you can do when your code is not behaving the way that you want, or to help make coding, in general, a bit easier as well. Some of the tips that you as a beginner can follow when working with Python in this manner include:

Comment Out the Code

Every type of coding language that you will work on is going to have the comment, which is basically a way that you can leave a note in your code without the compiler going through and trying to execute these notes as part of the program. This is advantageous to you because you can choose to comment out of a code that you don't want to run right now, but which you don't want to lose track of. You just need to put the # in front of that line that you want to comment out of.

If your script is long, you can comment on some parts of the code that are not related to the changes that you want to work on. This can sometimes make it run faster, and it can make it easier to search through the rest of the code to help find that mistake. You need to be careful when doing this though because it won't help you to comment out of parts that set the variables that the program needs to use later on. If you do this, then you are going to have issues getting the code to run.

When you are done testing out the code, and you have gotten it all organized and ready to go, you must go back through and remove these comment characters. This helps you to turn the whole program back on so you can see if it works the right way.

Print Things Out

When there are times when you are not able to figure out what is wrong with the code that you are working on, then you can take some time to print out a lot of things to see what is going to work. If you are unsure about which values are ever attached to one of your variables, then print it out. if you are uncertain about anything that is going on in the code that you are writing, then it is time to print it out. Then, when you are actually going through and running the program, you are able to look at your screen and see how the values change or what has been a bit different thanks to that.

You may find that this is going to be helpful to print out some strings before you go through and print out the variable. This is a good way to make sure

that your print statements are going to help you to learn how to make the program work for your needs.

Work with the Code You Know Will Behave

When you are in doubt about things, you may want to take some time to practice codes that are in existence already, ones that you know are going to work, and you don't have to worry about whether they are going to do what you would like or not. This is going to help you to figure out some of the syntax and the structure that has to come with these codes, and then you can work from there to get the best results. This is a good thing for a beginner to just gain some familiarity so that they are able to use that later on as they work on their own codes, or on things that are more complex.

If you do decide to work on your own project, then you can Google around to find a script that has a lot of what you need. It doesn't have to be perfect, but sometimes having that background in place and some of the syntaxes can make a big difference. Make sure that you take the time to run the code before you try to implement any changes. This ensures that the code works before you make any changes. Then you can add in some small changes to that code, and stop to test it often to see if any of the changes that you implement cause the bugs.

Read All of the Error Messages

This one is sometimes hard as a beginner because you may assume that you will have no idea what the error messages mean in the first place, so why should you take the time to learn them and actually read them? But, actually taking your time to read through those and learn how they work, and what they are telling you, can make it easier to fix up some of your code when things are not working well.

You will quickly find that a lot of your error messages are accurate and will actually be pretty descriptive for you. The language runtime tried to execute the program, but it ran into a problem. This would mean that you skipped a step, you had a typo, or something else is missing from your code.

There may be times when you do not understand the message that comes up with an error, but it does try to tell you what went wrong with the code. At a minimum, there will be information on a line number the error is on, and you can head to that part of the code and look for where the bug might be located.

Run the Code Often

Another thing that you need to consider doing when it is time to work with the Python code is run it on a regular basis. There is nothing worse, especially when you are a beginner, then sitting down and writing code for hours. You may feel accomplished, but when you go and try to run that code, you are going to be in for a surprise. There will be a ton of errors that you need to go through and fix, and then you will feel really frustrated with yourself in the process.

Instead of having hours of code that you need to sort through and try to sort through a bunch of errors, it is best if you try to run the code as often as possible. If you are able to go through and run it every few minutes or after you finish each little part that you are working with, then this is often going to be the best. It will ensure that you are catching some of the mistakes and

issues early on in the work, and will make it so much easier for you to really figure out what is going on in the code, and fix the errors.

It is a lot easier to find the errors and the mistakes in a small amount of code than it is to try and do this in a large block of code that you worked on for hours. It may seem like a waste of time and that it is slowing you down to run the code so often. But in the long run, it is going to end up saving you a ton of time and energy and can ensure that you are able to find the solutions that you are looking for. And it makes it so much easier to be able to fix any of the errors that show up because you know exactly where they are going to fall along the way as well.

Take a Break When Needed

Sometimes, working on coding can be a breeze. Everything is going to fit in just as you need it to and you can spend hours working on this part and getting it all to match up the way that you would like. And other times, things are not going to run as smoothly for you and will not work out as well. When the latter starts to happen, it is time to take a break.

Often, we work on the code for far too long. We want to get it done or get something figured out with it, so we will spend hours and hours working on it. And then we get frustrated. Something or other on the code is not working out the way that we want, and we can't seem to get it fixed and working well again. And no matter how hard we try, and how much work we put into it, we end up just making the whole situation worse in the process.

While it is sometimes hard to just let go of the coding that we are doing and take a break, sometimes this is just what we need. When we are tired and have been staring at something for a long time, it is just going to be worse if we continue to stare at it and try out too many things. When we take a break

and learn how to walk away from what we are doing on the computer, it is going to become so much easier.

You may have to force yourself to take this break a bit. When we are frustrated is usually the time when we wan to dig our heels into it a bit more, and we don't want to leave all of the work that we are doing at the time. But even an hour break or more can be nice to allow us to get away and rethink things. And when we come back, we can find the mistake or the error or the problem that we couldn't before, and we can get it taken care of.

Ask for Help

As a beginner, there may be some times when you will need help with the coding that you are doing. It is hard for some people to ask for help because they want to be able to do it all on their own. But learning how to code takes some time and patience, and then there are going to be those times when we just need some help because we aren't sure what steps to take next.

When you are asking for help, there are a few steps that you should consider taking in order to make it easier. These help to make sure that the other person knows what you have tried and doesn't waste time going through all of that again as well. Some of the things that you can consider knowing and explaining to the person who is going to help you out with your code will include:

1. Explain what you want the program to do for you and where the error is occurring.
2. Show the other person the code that is sending you an error so they can see it for themselves.

3. Show the other person the stack trace, all of it, including the message you got stating the error.
4. Explain everything that you have already tried on the error. This helps the other person have a good idea of what you have tried and what they should try to get the code to work.

You may find that the process of thinking through these items in your mind and then getting them written down on paper can help make a brand-new solution obvious. If that doesn't work, try to find someone who may be able to look through the code and tell you what is going wrong and what you can do to fix the error.

There are a lot of parts that need to come together when you are working as a programmer for the first time. being able to get all of these to work with one another, and ensuring that you are able to learn how to fix some of the problems that come up in your own code can be so important when you are a beginner trying to figure all of this out. Make sure to check out some of the different steps and tips that you can do when working in Python to make sure that you are coding correctly and that your program or application is going to work for your needs.

Chapter 10: A Quick Introduction to Data Analysis

Many of the topics that we have spent time talking about in this guidebook are going to lead us to a good look at data analysis and what this can entail for us. There are many companies who would like to learn how to work with the Python coding language because they know it is going to help them figure out how to better serve their customers, figure out the best options when it comes to which products to release to their customers, and so much more.

There are a lot of benefits that will come with working on data analysis, but before we do that, we need to first learn a little bit more about data analysis and what this process is all about. Data analysis is simply going to be a process that we can use in order to take our raw data and order and organize

it in a meaningful manner. This is going to make sure that we are able to take out some of the important information that is found in that set of data, and then use it for our needs.

The process of organizing and then thinking about our data is going to be so important in helping us figure out what the data is going to contain and what we are able to learn from it. There are going to be methods that we are able to use in order to approach one of these kinds of analyses for their own needs, but the most important thing that we need to remember here is to keep our biases and any manipulation out of the mix.

It is really easy for someone to get in there and add in some of their own biases and more in the mix. This is sometimes done on purpose if the person wants to make sure some of their ideas are going to be the ones picked. But this is also something that can happen on a more unconscious level as well. No matter where it comes from, it is important to pay attention to the presentation of the data analysis and to really be critical about the process that is used along with it.

Another thing to remember here is that raw data can easily take on a lot of different forms. And we are going to see the data come at us in places like observations, social media, survey responses, and measurements. In the raw form that we usually collect this data from, the information is going to prove to be really useful. But there are also problems with it being overwhelming as well.

Over the course of the process that you take for this data analysis, we need to take all of that raw data, no matter how much of it there is, and then order it in the manner that will make it the most useful for us. For example, we may go through and tally some of the survey results that you are going to see after sending these out to your customers. this will give you a better idea of how many people actually went through and answered the survey,

and can help you see how people responded to the different questions that were there.

In the course that you take, in general, to organizing the data, there are going to be some trends that will emerge. And sometimes, we are going to see that these trends are highlighted in the writeup that you see of the data. This is done so that the readers are going to take note. In a causal survey over something, it may be interesting to note that men like one thing better than women, or that women are more favorable to something than men. This is important for marketers to know sometimes, and can make it easier to learn how to market and reach each one.

In addition, modeling the data with the help of things like mathematics and some other tools is going to help us here. These can make sure that the right points are exaggerated, the ones that are of the most interest to us, which makes them a bit easier for the researcher to see and pay some attention to as they go through the process.

Another thing to add to any data analysis that we are working on is that the charts, textual writeups, and graphs are all going to be important forms of data analysis. These methods are going to be important because they are meant to distill and refine the data in a manner that readers are able to glean the most interesting parts of the information without having to sort through the data all on their own.

Summarizing the data and what all has happened during the analysis is often going to be pretty important when you would like to support any of the arguments that you made with that data, as is presenting the data in a manner that is clear and understandable. The raw data is something that you may choose to include in the appendix in some cases, because this allows those who need it, and why maybe using the information to make important decisions, a way to look through the specifics of the data, of the analysis, and more, if they so choose along the way.

With this in mind, it is time for us to take a look at some of the steps that are going to come to us when it is time to work on our data analysis. These steps are going to be pretty similar no matter what kind of project we are performing or working on in data analysis, and we will be able to see the Python language help us get it all done. Some of the different steps that you are able to take to complete a data analysis successfully will include:

Define the Question

In the business or organizational data analysis that you are working on, we need to make sure that we are jumping in on the right kinds of questions. It is never a good idea to start out with the idea that you will just sort through the data and hope that some good insights will pop out at you. There is just too much data to work with any more. And doing this is going to waste a lot of time and energy. You will end up in a rabbit hole if you do this and not be able to get anything done.

The questions that you rely on to get this work done need to be measurable, clear, and concise. You should also make sure that the questions are going to be there to either qualify or take away some of the potential solutions to the opportunity or problem that we are working with.

For example, a good way to start with this is to make sure that you know your clearly defined problem ahead of time. Maybe you are currently working for the government as a contractor, and you find that a lot of your costs are rising. This makes it a lot harder for you to submit contracts that are competitive, and you might lose your work. One of the questions, out of many that you are able to use in order to solve this business problem, could include whether or not you can reduce your staff without compromising on quality.

Pick Out the Important Measuring Options

Once we know what kind of business problem we would like to focus on, it is time for us to step back and work on creating the clear measurements and the priorities that we want to go with them. This is a step that we are able to break down into two parts, which makes it easier for us to work through. These steps are going to include deciding what you need to measure and then deciding which steps we need to take in order to measure it.

So, with the first step, we have to decide what exactly we would like to measure in the first place. Going back to that government contractor that we were looking at, we would want to focus on what kind of data is needed to help answer these key questions. For this one, it would be a good idea to know how many staff members you have, how much they cost you, and the percentage of their time that they are spending right now on functions that are pretty necessary to the business.

In answering these questions, you are going to then come up with some sub-questions as well. For example, do you see that your staff is not being utilized to their full potential, and are there some improvements to the process that you could do to help with this and reduce costs? There are always a lot of options and questions that you are able to work with on this one, and the more you focus on these questions, and the further down you go with them, the better you will be able to make a decision as well.

And the final thing that we need to look at when it is time to decide what to measure out of your process, you need to make sure that all of the reasonable objections that may come up if you do go with one action over another. For example, if you do decide to cut down on a few staff, how will you be able to respond if there is a big surge in demand for your services.

You need to then move on to the second step that is going to come with this process. Thinking about how you are going to measure out this data is going to be just as important as all of the other steps, especially before you go through the phase for data collection. The reason for this is that the measuring process that you are working with is either going to be able to back up the analysis that you do later on, or it is going to discredit the analysis that you are doing. There are a few different key questions that you are able to work with here, and they will include:

1. What is the time frame that you want to follow? Think about how long you want to search for the information, how much historical data you want to use, and how long you expect these changes to last. If you are just dealing with a temporary decline, such as a seasonal one, maybe finding another way to reduce costs is better than laying people off.
2. What is the unit of measure that you are going to work with? This could b a comparison of Euro and USD for example.
3. What factors do you think should be included in this process. You could check out the cost of your employees with just their annual salary. However, this is not going to provide you with the full picture because you will be able to learn a lot more if you work with the annual salary, and the cost of benefits that you are providing to your employees can add up as well.

It is important to have a good plan in place before you try to do anything with your data. Even collecting it is something that can be done after you have your measurement criteria in place, and you are sure that you know what steps to take next.

Collect the Data

At this point, we have had some time to go through and clearly define the question that we would like to solve for our business. And then, we also took some time to learn which measurement priorities we need to set as well to ensure we get the right kind of data and more. Now it is time to collect the data. As we go through and organize and collect the data, there are going to be a few points that we are able to keep in mind that will make it easier for us to really see some of the results that we want.

First, before we go through and collect the new data, we need to look at the information that we already have. if you already have some of your own sources and databases that have been collecting data, then you need to look at those first. This can help you to really take a look at this data and save yourself a lot of time and money in the process. Depending on how much data is collected, you may have all of the data that you need without having to look in other locations.

Then we need to move on and determine how we are going to handle the data that we bring in. if we look at some of our own resources and find that we need some more information, we then need to think of a good naming and file storing system ahead of time. This is going to make it easier for those who are on the team to work with this to collaborate with one another. This process is going to save your team some time and can prevent the members from going through and collecting the same information more than one time.

If you end up needing to go through the process of gathering up the data and using interviews and observations to finish this up, then you need to make sure that there is a good interview template that everyone is able to use. This is going to make sure that there is a level of consistency in what you will see, and can save some time.

And finally, we are going to make sure that we keep all of the data that we are collecting through this process as organized as possible. The method

you use is going to depend on the data that you are using and how you plan to use it for your needs. You could do something like having a log with the collection dates, and then add in any of the sources notes that you need to keep this all in line. this practice is important to spend some time on because it is going to help validate some of the conclusions that you are making later on.

Collecting the data that you need to finish this process will help us to really get the data analysis done. You need to make sure that when you collect this data, you are picking it out from some reputable and good sources along the way. the higher the quality that you can put with the data that you collect, the better it is going to be later on. When you perform the analysis, this data is going to give you some of the results and predictions that you are going to put your decisions on, so being a bit selective with this can help.

Look Through and Analyze the Data

And this brings us to our fourth step, which is where we are going to take all of that data that we have been collecting so far, and we will analyze it. After we have been able to collect the data that is able to help us answer the question that we came up with in the first step, it is time for us to go through a deeper analysis of the data. There are a few options and steps that we are able to follow to make this happen, and we are going to spend some time discussing those now.

Begin this with manipulating the data such as plotting out that data and seeing what kind of correlation is there. or you can even create your own pivot table to work within Excel. This pivot table can work for you because it will allow you to do some sorting and filtering of your data by some different variables. It is also going to make it easier to figure out the maximum, mean, standard deviation, and a minimum of the data that you have.

As you are working on manipulating the data, you will sometimes find that you already have the exact data that you need right from the start. But in most cases, you will still need to either gather more data to help or make some revisions to the original question that you were asking. Either way, this initial analysis of the correlations, outliers, variations, and trends will make sure that you are able to focus the analysis that you are doing on answering the question that you originally posted. It is also a good way for us to handle some of the objections that others may have worked raised against you.

During this specific step, the tools and software that come with data analysis are going to be really helpful. There are a few different advanced processes that you are able to choose from that will help us handle these advanced statistical data analyses that we want to complete. In most cases, though, you are able to keep things simple because you can work with Microsoft Excel and get a lot of the same decision-making tools that you need. There are a lot of features and functions that come with Excel, and you will find that you are able to use this to help you get the best results as well.

The analysis that we do during this step is going to take some time. Picking out the tools that we want to use, and the right algorithm is going to make it a lot easier. We will talk about some of the algorithms that are available to help us complete this analysis and often it is going to depend on how much data you want to work with, and what answers you are hoping to get out of it as well. There are also a few ways that we need to go through and separate our data.

These two methods are going to be the training set and the testing set. Before we are able to get through and actually work on the analysis, we need to train and test the algorithm that we would like to use on the proper methods and what it needs to do overall. The algorithm is not going to come already prepared and ready to go. Instead, we have to show it how to

behave and do some work with it, and then we can put through the data we want and get the predictions.

To start, we need to split our data into two sets, the training, and the testing set. The training set needs to be a lot larger than the testing set. This is because we are basically using this as a way to teach the program how we want it to behave, and ensuring that it has plenty of time to see the relationships that are there and figure things out. We need to also make sure that the data we are working with is high quality so that it is going to work for learning faster and helping us later.

After we have had a chance to go through and send the training data into the algorithm, it is then time for us to work with the testing set. We send this one through to see how well the algorithm was able to retain the information that is gathered in the last set. The accuracy the first time is not going to be perfect, but since the algorithm is going to learn more times than it does the process, this is not a big deal. It will get better.

The higher the accuracy that you are able to get, the better. But if you end up getting an accuracy level that is below 50 percent, then this is a sign that something went wrong. Your algorithm should be able to make random guesses on the testing and get 50 percent or more right. So, if you are getting less than this, it is a sign that the algorithm is not set up right, or maybe that the data that you are using in that is not very good, and you need to restart with something else.

You will need to go through the training and testing phases a few times if you are able to. This will make it easier for us to figure out the best way to train the algorithm and ensure that it is set up properly. The more information and data that you are able to send through to the algorithm, the easier things get, and the higher the accuracy will become overall. When you do this a few rounds, you should notice that when you finish with each of the testing grounds, the accuracy rate is going up.

We do not need to have 100 percent accuracy with our algorithm, but the higher that we are able to get this, the better. When we get to an accuracy level that we are comfortable with, we can then move on to actually using this algorithm for some of the information that we need. It is then ready for us to put the data we want into it and learn how to make our decisions and have some good predictions in the process.

Interpret the Results

The final thing that we need to take a look at here is interpreting the results that we get out of the analysis. After we have spent some time analyzing the data, and maybe doing some more research if it is needed, it is time for us to go through and interpret the results that we get from this. As you spend some time interpreting the analysis, keep in mind that you will need to go through and see whether the original hypotheses that you formed was the right one or not. Sometimes it is going to be right, and other times it is going to be wrong. And when the hypotheses are wrong, you need to go through and make some adjustments to figure out how to get things back on track.

As we are going through and interpreting some of the results that we are able to get out of the data, there are a few questions that we need to ask for the data. We need to first start out by asking whether the data is able to answer our original questions. Think back to that question that we had in the first section of this, and then determine whether the analysis that we just went through was able to answer the original question and how it was able to help out with this.

The next thing that we need to look at is whether the data is going to be strong enough and sound enough to help us defend against any of those objections that we had before. Think back to all of the objections that you had heard when we got started with the project, and think of what you are

able to do in order to allay any fears and worries that are going to come up with this.

We also need to make sure that this is the right conclusion that we came up with and that it is going to be the right one for us to work with. For example, we need to take a look and see if there are any limitations on the conclusions that we have, and any other angles that we may not have looked at yet concerning this. The more in-depth that we are able to get with this, the better it is going to be for us when we finally do come up with the conclusion that we want to use.

If you find that the interpretation that you are able to get of this data after finishing the analysis is able to hold up under the questions that we just did and more, and they meet with the considerations that we just had, then you likely have a conclusion that is productive and that you are able to use. The only step that is left for this one is to use the results of the analysis in order to figure out the best course of action to work with.

This is a process that takes some time to accomplish and is not always as easy and straightforward as we may like. It is supposed to be this way. The algorithms are meant to come into this process and speed it up, though, much more than we would be able to see if we tried to go through this process in a manual manner.

When we take the time to follow the five steps that we listed out above to help with data analysis and all of the parts that come with this process, it is easy to see how we are able to make better decisions for your business or any company you are working with. The reason that your decisions are going to be stronger and smarter than ever before is that they are going to be backed by data that has been collected robustly and analyzed so you know you are able to trust them.

With some practice, and with some careful consideration of what is found in the data that you are looking at, you will find that your analysis of the data is going to only get faster and more accurate. And this is going to mean a

lot of good things for your business because it means that you are able to make better and more informed decisions that will help your business run in the most effective manner possible.

Chapter 11: Some of the Best Python Algorithms for Data Analysis

We have spent some time in this guidebook looking at some of the basics of the Python language, as well as some of the basics of data analysis and what we are able to do with this for some of our own needs. Now it is time for us to take both of these ideas and put them together.

In order to complete a good data analysis in the first place, we need to be able to run the data that we have collected through a number of algorithms, usually, a few, though in some cases, one will be fine. These algorithms are going to help us to learn what patterns and insights are found in that data, and can give us some good predictions in the end.

There are a number of different Python data analysis algorithms that we are able to work with along the way, and it is going to often depend on the what we would like to do with that data to start with, and how we are able to benefit from this as well. With this in mind, let's dive right into some of the basics that come with these algorithms, and learn a bit more about why Python algorithms are the best for data analysis.

Neural Networks

The first kind of Python algorithm that we are able to use for data analysis is going to be known as the neural network. This is a fairly sophisticated type of algorithm to work with in the first place, but it is going to provide you with a lot of the power and more that you need to get some of the work done.

The neural network is going to work in a manner similar to how the brain will function. It is going to form connections to the things that it gets right,

and it can remember this for future use as well. This helps it to learn as time goes on, and makes it easier and faster and more efficient at some of the work that it needs to complete as well. Once it has been able to go through and learn something new, you will find that it will remember it in the future.

In addition, the neural network is really good at helping with things like identifying what is in an image. It does this by going through the different layers of the pictures, one at a time, and then making a good prediction of what is going to be found in that image. The more layers that it is able to go through, the more accurate it will be at guessing what is in the picture as well. This makes it an effective tool to use for many of the projects you are working with.

However, we have to keep in mind that the computational costs are going to be a bit higher with this one. Because we are working with an algorithm that is able to methodically go through a bunch of layers and make accurate predictions, we can see why the costs will be higher with this one. You have to determine ahead of time if this is the right kind of process for you to work with, or if one of the other algorithms out there would work better.

Clustering

There are a number of different clustering techniques that you are able to work with when it comes to handling a data analysis. These are all going to take the data, though and put them together in clusters so you can see more about the distribution of where your points of data are going to fall. There are different ways that you are able to work with clustering, and different ways that you can cluster up your data, but it is a good way to put the information into groups and see some good insights right away.

The number of clusters that you will have in your work is going to depend on what you are focusing on. For example, if you are working to divide

your group into males and females, then you will only need to have two clusters. But if you would like to divide things up by the age of your customers, or by where they live in the country or the world, then you will probably need to have more than two clusters in order to have the best idea of what is going on.

With these clusters, you will be able to plot them on something like a scatterplot, and then figure out where everything is going to fall. This is a great way to figure out the best options to take, whether there is a new demographic that you need to work with, and more. You can just put these clusters on your scatterplot, and you can learn a lot of information in a short amount of time.

The idea of these clusters is pretty simple. When you have some data points that fall into the same cluster, you can safely make the assumption that the points are going to have some similarities with one another. But if the data points are not in the same cluster, then this is a sign that they are going to be different from one another. If we are able to keep this in mind, it is a lot easier for us to work with understanding this kind of algorithm.

Support Vector Machines

These are going to be the algorithms that you are able to use in order to figure out the right course of action to take. When you see that the line and the hyperplane come together, then this is a good idea that the information that is there is going to lead you to the results that you are looking for. This can make picking out the right decision easier, as long as you know the right steps that are going to help you to get the best results as well.

The biggest trick that comes with this one sometimes is figuring out which lines and hyperplanes we should pick and work with. Sometimes, depending on the type of data that you are working with and what is all in

it, you will end up with more than one hyperplane. Sometimes this will not be a big deal because you will be able to see which one is the best for your needs. Other times it may seem like there is more than one answer that is going to work the best for your needs as well.

There are a number of options that we are able to work in order to pick out the right SVM for what we want to do. You will learn how to do this more with experience, but you will be able to learn about your data and then figure this out as well. Sometimes experimenting with a few of these to see which one is the best is another good option that we are able to choose as well.

Naïve Bayes

The Naïve Bayes algorithm is a good choice to go with when you want to do a bit of exploration with the data in the beginning. Maybe you want to see what the best way is to split up the data that you have, or you are not yet certain about what kind of algorithm is going to be the best one for you to focus your attention on yet. In some cases, you may need to show some of the data and some of the information that you have ahead of time, right after collecting it, to those who want to see what is going on, but may not understand all of the more technical aspects that come with it.

This is where the Naïve Bayes algorithm is going to come into play and can really help us. With this option, we are able to take a good exploration of the data that we have, and then determine the best steps to take after. Sometimes this helps us to choose which of the other algorithms are the best ones for us to go with. And other times, it may be a good way to create a beginner's model so that we can show this off before being able to finish all of the work for the final project.

The Naïve Bayes algorithm is usually not going to be the first choice that we make when it is time to handle some of our data, and we will usually go through and make a few other adjustments to the process as well and finish off with another kind of algorithm. But it is definitely a good algorithm to go with because it adds in a lot of the different parts that you need to get a good idea about what the data contains, and what else we are able to do with it along the way.

Decision Trees

Another option that we are able to go with here is the decision trees. These trees are going to help us to make some good decisions and compare some of the likely outcomes of the decisions that we do make. if you are uncertain about which steps to take, especially if you have a few choices to pick from, then the decision tree is going to be one of the best algorithms for you to choose from.

To start, you would be able to ask your question and implement it in the information or the data that you have. Of course, we first need to clean and organize it, but we will assume that this part is already done. When you are ready, the data can tell us which steps are the best, and the likely outcome of each choice that you try to make along the way. This can help you to make some better decisions because you can look at the data in a more clear and organized manner, and you know that it actually makes sound sense, rather than just using your intuition along the way.

In addition, you are able to take a bunch of these decision trees and combine them together into a random forest. The random forest is simply going to be when we combine together a bunch of those decision trees in order to compare more than one option.

Maybe you want to see which is best for you. If you ever want to compare more than one idea or thought for you to take, then working with the random forest, or even a simple decision tree, can get this done. You just

add in the information that you want to work with, and then compare the decisions and their potential outcomes before going.

As we can see, there are a lot of different types of algorithms that we are able to use when it comes to handling our data and making sure that it makes as much sense as possible. A bit point of working with data analysis is to help us to sort through all of that data. Otherwise, that data is just going to sit around and be worthless to us and just sitting there taking up space.

Each of the algorithms that we went through above is meant to help us to figure out different ways that we are able to move around our data, and different ways that we will come to a better understanding of what is inside of our data, and how we can use that.

Whether you want to be able to make smarter decisions for your business, or you are looking for ways to find new customers or beat out the competition, the algorithms above, with the help of the Python code, will be able to help you get there.

Conclusion

Thank you for making it through to the end of *Python Programming*, let's hope it was informative and able to provide you with all of the tools you need to achieve your goals whatever they may be.

The next step is to get started with the basics of programming in Python. You will find that even as a beginner, or someone who has never done any kind of coding in the past, working with the Python language is easy enough to learn. And we spent some time in this guidebook learning more about how this is going to work, and how we are able to code in no time.

There are a lot of reasons why people and businesses are going to want to learn how to work with the Python language, and with some of the codes and more that we looked at in this guidebook will show us just how easy working with this kind of language can be overall. It is one of the best coding languages for basic coding all the way up to data analysis and machine learning, so there is always going to be a lot of things that you are able to do with it.

This guidebook took some time to explore more about what Python is able to do for us. We looked at some of the basics, like how to install this language on your own computer, and then moved on to some of the different codes that you are able to write out with the help of this language. And then, we moved on to how we can use this in a more practical manner with the help of a good data analysis to push it along and actually help your business learn from the data and everything it has been able to do overtime.

When it is time for us to explore more about Python programming and what we are able to do with some simple coding along the way, make sure to read through this guidebook and learn how to get started.