**Chapter 1**

**ABOUT THE COMPANY: CELESTIAL V SOLUTIONS**

****

**1.1 Company Profile**

Celestial V Solutions is an IT Company, based in Bangalore, the silicon city of India. The company delivers IT solutions to a broad range of domestic and international level.

Celestial V Solutions is a provider of end-to-end IT services and solutions designed to help clients and improve efficiency. The specializing in outsourcing and offshore, the systems integration and application development, software and consulting, and quality assurance should be maintained.

The human resources training division turns out complete software professionals through specially designed instructors. The services model is simple. The company provide out of house expertise to the clients, who need to reach beyond their in-house capabilities or accelerate the progress of project under deadline, and the work to gain the clients confidence that the team or leased personnel will provide a rock-solid asset for planning and complete projects in time and on budgets.

To become the leaders in IT Solutions and sustain growth with quality, Innovation and Creativity is the vision of the company.

The mission is to provide highest and reliable quality of software solutions and services globally. Through innovation and creativity for professional satisfaction the company needs to fulfill the organizational goals and societal needs.

The primary goal of Celestial V Solutions would be towards acquiring the technology and know-how for product development in the fields of wireless application, Internet/Intranet technologies, Client/Server, E-Commerce Solutions, Enterprise Resource Planning and the embedded technologies. The major player in this field would be ‘ERP’. Apart from the above it will also act as a customized solution provider and consulting firm.

Providing a fast and responsive service of the highest standards is what company thrive at working with start-ups as well as Fortune 500 companies. Development and management of both short term and long term projects is ability to understand the real world of business and stretch to the maximum. Significant operations and IT experience in a range of industries. Get your projects done On-time and In-budget. The remaining committed to each other’s welfare and success. Company does unto others as it expect others to do unto us.

The staffs at the company are an excellent blend of technical expertise and creative skills. The company has a team of professional web developers and designers who work to create attractive and functional web to make a website popular and make it rank well on the search engines. The KPO team has experienced industry professionals with high analytical and processing skills.

The team shares a passion to think ahead, innovate and work hard to achieve the set goals. They know that clients need one hundred per cent of the attention and work to ensure the same. This is what drives us to product high quality work that not only company but also you can be proud of.

**1.2 Products**

The products developed and delivered to the customers by the Celestial V Solutions are as follows:

Build Soft

Esybill

Esy2send

Esy2send mob app

RARS (Result and Attendance Retrieval system)

The working description of the above products is given below:

**Build Soft**

Web Based application which helps the construction companies to calculate the estimation cost for construction work and to maintain their Payment Transaction details. This system includes MIS Module with reports and graphs which help the management to understand the financial details of the company quickly.

**Esybill**

This application helps most of our retail customers. Using this they can easily generate bills, keeping track of their stocks, inventory, sales and purchase details. It also gives them an option to view the reports on daily, weekly and monthly basis.

**Esy2send**

Esysend is an application which helps us in sending SMS, bulk SMS, e-mail, bulk e-mail and both SMS and e-mail simultaneously. It is mainly developed targeting industries like marketing, schools, colleges, real estates, finance and other marketing industries.

**Esy2send mob app**

Esysend also has its mobile version, using this mobile application you can send SMS or bulk SMS with your mobile on the go when you are away from your system. All you need to use this mobile application is GPRS connection in your mobile and the mobile application of Esy2send which you can download from esy2send website. The biggest advantage in using this application is that you don’t have to carry your laptop or system wherever you go and whenever you need to use esy2send application. By making a single click on your mobile application you can communicate in mass.

**RARS (Result and Attendance Retrieval system)**

It is another mobile from us, which servers at the most for schools and colleges. It helps the parents to know their children marks and results of the tests and examination. Once you download the mobile application on to your mobile, you can start using it and it’s very easy to know your children curriculum on the go over the mobile. As soon as you place the request for details of your children, you will receive the message of the same within 5 seconds.

**1.3 Services**

Celestial V Solutions strive in providing the following services to the customers:

Software services

Web Services

Search Engine Optimizations

J2ME

Embedded system

The explanation of the above services is given in detail below:

**Software services**

Finding the right technologies for your growing business needs should not be a brain drain; after all, you have got more important things to think about – like running your business. The reality is that the scope of technology is growing rapidly, making it nearly impossible for you to be an expert for so many new hardware and software products. That is why the company are here. The companyas pioneers of the software industry provide apt solutions for all your needs. The expert motto is to provide excellent quality of products for all the customer needs in timely manner.

**Web Services**

A website is like having global presence in the virtual market place, Internet. But just having a website deployed cannot get a business what it should from a website, thus it is important to get it done from experts and make it serve the purpose of having it. There are various aspects that need to be worked upon to make a website successful and create professional image of the business owing it. At Celestial V Solutions very well know how to accomplish this.

**Search Engine Optimizations**

Company provide with an excellent solution to the costumers’ challenge in optimizing their property in search engines. With hands on experience with optimizing search results across various search engines, provide extensive reports and advice on how to improve costumers’ property occupancy depending on various factors in different search engines.

**J2ME**

J2ME is a technology which helps us in developing mobile applications, that are not an excuse for mobile applications, and there is a team of skilled developers who have hand in experience, creating many mobile applications.

**Embedded system**

Embedded Systems Design and Development Serviceshelp company to achieve their product development and sustenance initiatives. The technologies committed to providing the highest quality of service to its customers, delivering advanced systems, solutions and services that benefit the businesses, the industry and the society.

**Chapter 2**

**ABOUT THE DEPARTMENT**

**2.1 Research and Development**

In the company, the research is being conducted in the field of networking, for efficient file transfer across the network, with greater security and speed. Research and development is one of the means by which the business of the company can experience future growth by developing new products or processes to improve and their operations. The activities that are classified as R&D differ from company to company, but there are two primary models, with an R&D department either being staffed by engineers and tasked directly with developing new products. In this company, the R&D department is staffed by the engineers.

R&D is an important means for achieving a future growth for the company, with the invention of the new technologies for the efficient growth of the company. The kind type of R&D being conducted in the company is basic research as well as applied research. When the research aims to understand a subject matter more completely and build on the body of knowledge relating to it, then it falls in the basic research category. This research does not have much practical or commercial application. The findings of such research may often be of potential interest to the company.

Applied research has more specific and directed objectives. This type of research aims to determine methods to address a specific customer/industry need or requirement. These investigations are all focused on specific commercial objectives regarding products or processes.

**2.2 Processes Adopted for Software Developments**

Software Development Life cycle (SDLC) was introduced to address the problems faced during the software development process. It is a disciplined and systematic approach that divides the software development process into various phases, such as requirements analysis, design, and coding. The SDLC aims to produce high quality software that meets or exceeds customer expectations, reaches completion within times and cost estimates. The phase-wise development process helps us track schedule, cost, and quality of the software projects.

There are six phases in SDLC:

**Feasibility analysis**

Includes analysis of project requirements in terms of input data and desired output, processing required to transform input into output, cost-benefit analysis, and schedule of the project. The feasibility analysis also includes the technical feasibility of a project in terms of available software tools, hardware, and skilled software professionals. At the end of this phase, a feasibility report for the entire project is created.

**Requirement analysis and specification**

It includes gathering, analysing validating, and specifying requirements. At the end of this phase, the Software Requirements Specification (SRS) document is prepared. SRS is a format document that acts as a written agreement between the development team and the customer. SRS acts as a input to the design phase and includes functional, performance, software, hardware, and network requirements of the project.

**Design**

Includes translation of the requirements specified in the SRS into a logical structure that can be implemented in a programming language. The output of the design phase is a design document that acts as an input for all the subsequent SDLC phases.

**Coding**

Includes the implementation of the requirements specified in the design document into executable programming language code. The output of the coding phase is the source code for the software that acts as input to the testing and maintenance phase.

**Testing**

Testing includes detection of errors in the software. The testing process starts with a test plan that recognises test-related activities, such as test case generation, testing criteria, and resource allocation for testing. The code is tested and mapped against the design document created in the design phase. The output of the testing phase is a test report containing errors that occurred while testing the application

**Maintenance**

Include the implementation of changes that the software might undergo over a period of time, or implementation of new requirements after the software is deployed at the customer location. The maintenance phase also includes handling the residual errors that may exist in the software even after the testing phase.

The company is making use of the waterfall model approach for developing the projects.

**2.2.1 Waterfall approach**

The waterfallapproach describes the software development process in a linear sequential flow. This means that any phase in the development process begins only if the previous plan is complete. The waterfall approach does not define the process to go back to the previous phase to handle changes in requirement. Therefore, different projects may follow different approaches to handle such situations.

The waterfall approach is the earliest approach that was used for software development. Initially, most projects followed the waterfall approach because they did not focus on changing requirements.

Every software developed is different and requires a suitable SDLC approach to be followed based on the internal and external factors. Some situations where the use of Waterfall model is most appropriate are:

* Requirements are very well documented, clear and fixed.
* Product definition is stable.
* Technology is understood and is not dynamic.
* There are no ambiguous requirements.
* Ample resources with required expertise are available to support the product.
* The project is short.

The advantage of waterfall development is that it allows for departmentalization and control. A schedule can be set with deadlines for each stage of development and a product can proceed through the development process model phases one by one. Development moves from concept, through design, implementation, testing, installation, troubleshooting, and ends up at operation and maintenance. Each phase of development proceeds in strict order.

The disadvantage of waterfall development is that it does not allow for much reflection or revision. Once an application is in the testing stage, it is very difficult to go back and change something that was not well-documented or thought upon in the concept stage.

**2.2.2 Version Control System**

Version control system also known as revision system is used in the software development process for recording the changes to project done over time. A version control system is a repository of files, often the files for the source code of computer programs, with monitored access. A component of software configuration management, version control, also called as source control. It is the management of changes to documents, computer programs, large web sites, and other collections of information.

Changes are usually identified by a number or letter code, termed the “revision number”, “revision level”, or simple “revision”. For example, an initial set of files is “revision 1”. When the first change is made, the resulting set is “revision 2”, and so on. Each revision is associated with a timestamp and the person making the change. Revisions can be compared, restored, and with some types of files, merged.

Version control systems most commonly run as stand-alone applications, but revision control is also embedded in various types of software. Software developers sometimes use revision control software to maintain documentation and configuration files as well as source code.

As teams design, develop and deploy software, it is common for multiple versions of the same software to be deployed in different sites and for the software’s developers to be working simultaneously on updates, bugs or features of the software are often only present in certain versions (because of the fixing of some problems and the introduction of others as the program develops). Therefore, for the purposes of locating and fixing bugs, it is vitally important to be able to retrieve and run different versions of the software to determine in which version (s) the problem occurs. It may also be necessary to develop two versions of the software concurrently.

**2.2.3 Bit Bucket**

Bit Bucket is the git solution for professional teams. It is a distributed version control system that makes it easy for us to collaborate with the team. Bit Bucket is only collaborative Git solution that massively scales. The second wave of adoption will be moving the full software development workflow to the cloud for the same reason applications moved: accelerating business velocity. Bit Bucket server gives secure, fast, enterprise-grade controls, like fine-grained permissions and powerful management features.

**2.3 Roles and Responsibilities of Individuals in the Department**

**Software Engineer**

Software Engineer are also known as application programmer, software architect or system programmer/engineer.

The work of the software engineer typically includes designing and programming system-level software: operating systems, database systems, embedded systems and so on. They understand how both hardware and software function. The work involves talking to clients and colleagues to assess and define what solution or system is needed, which means there are a lot of interaction as well as full-on technical work.

**System Analyst**

System Analyst is also known as product specialist, systems engineer, solutions specialist, technical designer.

Systems Analyst investigate and analyse business problems and then design information systems that provide a feasible solution, typically in response to requests from their business or a customer. They gather requirements and identify the costs and the time needed to implement the project. The job needs a mix of business and technical knowledge, and good understanding of people. It’s a role for analyst programmers to move into and typically requires a few years’ experience from graduation.

**Business Analyst**

Business Analyst is also known as business architect, enterprise-wide information specialist.

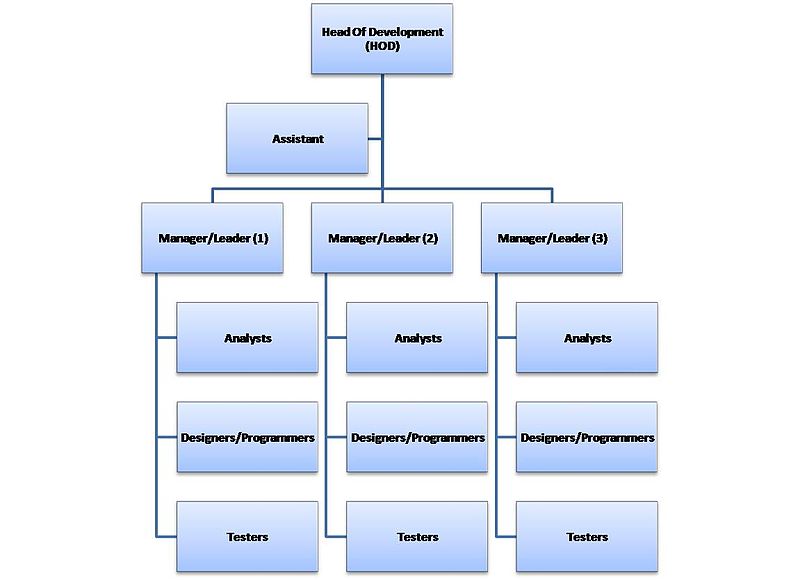
Business Analysts are true midfielders, equally happy talking with technology people, business managers and end users. They identify opportunities for improvement to processes and business operations using the information technology. The role is project based and begins with analysing customers’ needs, gathering and documenting requirements and creating a project plan to design the resulting technology solution. Business analysts need technology understanding, but don’t necessarily need a technical degree.

**Network Engineer**

Network Engineer is also known as hardware engineer, network engineer.

Network engineering is one of the more technically demanding IT jobs. Broadly speaking the role involves setting up, administering, maintaining and upgrading communication systems, local area networks and wide area networks for an organisation. Network engineers are also responsible for security, data storage and disaster recovery strategies. It is a highly technical role and you’ll gather a hoard of specialist technical certifications as you progress.

**Web Developer**

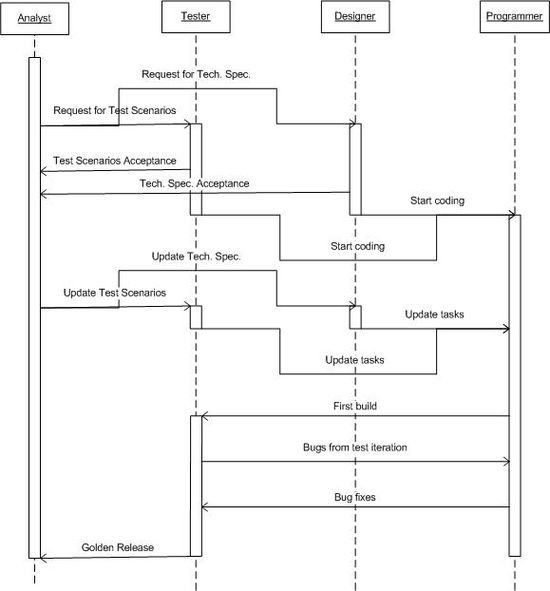


**Figure 2.1 Organization Structure**

The organisation structure is shown in Figure 2.1 where the manager of a software company is usually called the Head of Development who is also called as the Project Manager and reports to the stakeholders. He leads the sub-teams via the Team Leaders who in turn lead the group of analysts, programmers and testers. All the teams are fully independent and they work separately on the different projects. Like for the example in the company, one team is working on the networking projects, whereas the other team is working on the android projects.

The structure that is being followed in the is quite simple and all the employees’ reports to one person, what make the situation quite clear however it is not a good solution in terms of knowledge exchange and optimal usage of human resources.

In this model there are dedicated managers/leaders for each main specialization, "renting" their people for particular projects led by product/project managers, who formally or informally buy the people and pay for their time. This leads to each private employee having two bosses – the product/project manager and the specialized "resource" manager. On one hand it optimizes the usage of human resources; on the other hand it may give rise to conflicts about which one manager has priority in the structure.



**Figure 2.2 Interaction Diagram**

The above Figure 2.2 shows UML diagram for the interaction between the groups in the software company, including the sequence of activities being carried out between the analyst, testers, designers and programmers.

**2.5 On-going Project**

The on-going project in the company is TechBus. The TechBus is school bus tracing and student tracking solutions. Almost every campus is looking to safeguard their campus environment protecting campus from cheating or to provide a safer campus environment. TechBus can help to make the campus environment safer using latest technologies. Convert your buses into smart-buses by using the state-of-the-art GPS tracking technology for the safety of the students. The schools enforcing such systems are preferred by parents for offering such distinguishing services.

School Bus fitted with TechBus GPS tracking device and RFID reader. The location detail is received from the satellite on TechBus GPS tracker module and the location Detail along with Student attendance from RFID is transmitted via GPRS or 3G from the TechBus GPS & RFID devices. Data from the TechBus device is received on our server through cellular network, which is stored for analysis. The alerts are sent via SMS and email to parents and transport managers. The viewing of live Bus location details, student attendance and other reports from anywhere using this TechBus cloud based application.

**Chapter 2**

**TRAINING CONTENT**

**Anaconda**

Anaconda Distribution contains conda and Anaconda Navigator, as well as Python and hundreds of scientific [packages](https://docs.anaconda.com/anaconda/packages/). When you installed Anaconda, you installed all these too. You can try both conda and Navigator to see which is right for you to manage your packages and environments. You can even switch between them, and the work you do with one can be viewed in the other.

Try this simple programming exercise, with Navigator and the command line, to help you decide which approach is right for you.

Your first Python program: Hello, Anaconda!

Use Anaconda Navigator to launch an application. Then, create and run a simple Python program with Spyder and Jupyter Notebook.

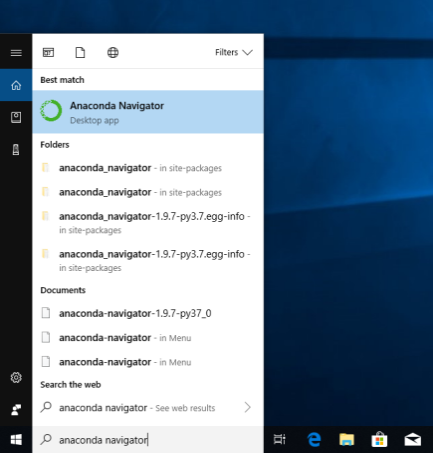
Open Navigator

Choose the instructions for your operating system.

* [Windows](https://docs.anaconda.com/anaconda/user-guide/getting-started/" \l "open-nav-win)
* [macOS](https://docs.anaconda.com/anaconda/user-guide/getting-started/" \l "open-nav-mac)
* [Linux](https://docs.anaconda.com/anaconda/user-guide/getting-started/" \l "open-nav-lin)

***Windows***

From the Start menu, click the Anaconda Navigator desktop app.



***macOS***

Open Launchpad, then click the Anaconda Navigator icon.



***Linux***

Open a terminal window and type anaconda-navigator.

Run Python in Spyder IDE (integrated development environment)

***Tip***

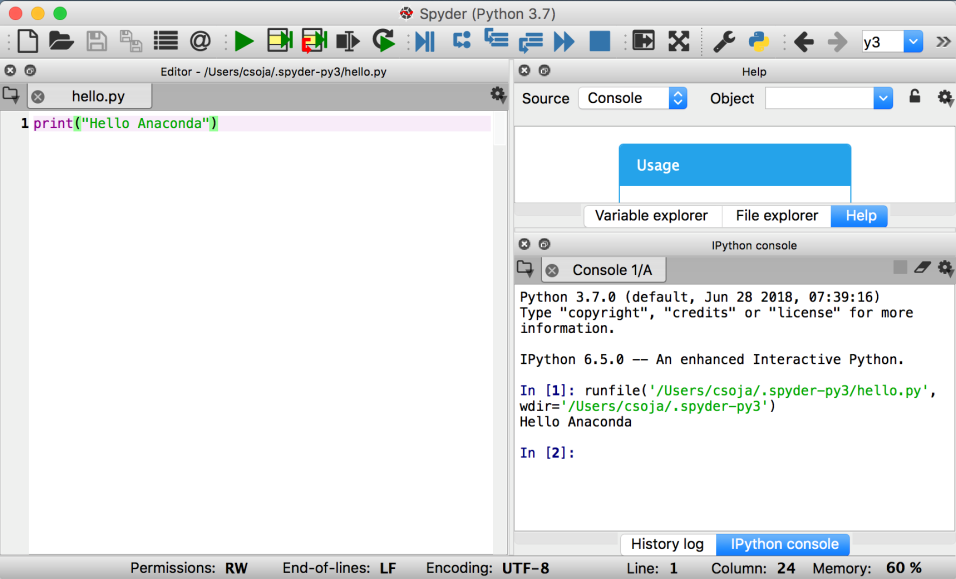
Navigator’s Home screen displays several applications for you to choose from. For more information, see links at the bottom of this page.

1. On Navigator’s Home tab, in the Applications pane on the right, scroll to the Spyder tile and click the Install button to install Spyder.

**Note**

If you already have Spyder installed, you can jump right to the Launch step.

1. Launch Spyder by clicking Spyder’s Launch button.
2. In the new file on the left, delete any placeholder text, then type or copy/paste print("Hello Anaconda").
3. In the top menu, click File - Save As and name your new program hello.py.
4. Run your new program by clicking the triangle Run button.
5. You can see your program’s output in the bottom right Console pane.



Close Spyder

From Spyder’s top menu bar, select Spyder - Quit Spyder (In macOS, select Python - Quit Spyder).

***Run Python in a Jupyter Notebook***

1. On Navigator’s Home tab, in the Applications pane on the right, scroll to the Jupyter Notebook tile and click the Install button to install Jupyter Notebook.

**Note**

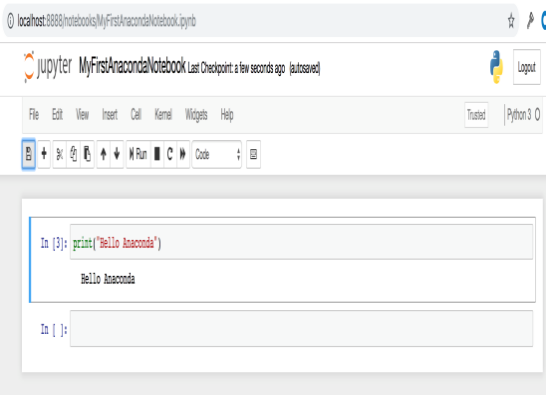
If you already have Jupyter Notebook installed, you can jump right to the Launch step.

1. Launch Jupyter Notebook by clicking Jupyter Notebook’s Launch button.

This will launch a new browser window (or a new tab) showing the [Notebook Dashboard](https://jupyter-notebook-beginner-guide.readthedocs.io/en/latest/what_is_jupyter.html" \l "notebook-dashboard).



1. On the top of the right hand side, there is a dropdown menu labeled “New”. Create a new Notebook with the Python version you installed.
2. Rename your Notebook. Either click on the current name and edit it or find rename under File in the top menu bar. You can name it to whatever you’d like, but for this example we’ll use MyFirstAnacondaNotebook.
3. In the first line of the Notebook, type or copy/paste print("Hello Anaconda").
4. Save your Notebook by either clicking the save and checkpoint icon or select File - Save and Checkpoint in the top menu.
5. Run your new program by clicking the Run button or selecting Cell - Run All from the top menu.



Close Jupyter Notebook

1. From Jupyter Notebooks top menu bar, select File - Close and Halt.
2. Click the Quit button at the upper right of the Notebook Dashboard and close the window or tab.

Close Navigator

From Navigator’s top menu bar, select Anaconda Navigator - Quit Anaconda-Navigator.

Write a Python program using Anaconda Prompt or terminal

Open Anaconda Prompt

Choose the instructions for your operating system.

* [Windows](https://docs.anaconda.com/anaconda/user-guide/getting-started/" \l "open-prompt-win)
* [macOS](https://docs.anaconda.com/anaconda/user-guide/getting-started/" \l "open-prompt-mac)
* [Linux](https://docs.anaconda.com/anaconda/user-guide/getting-started/" \l "open-prompt-lin)

Windows

From the Start menu, search for and open “Anaconda Prompt”:

**Python Introduction:**

Python is an easy to learn, powerful programming language. It has efficient high-level data structures and a simple but effective approach to object-oriented programming. Python’s elegant syntax and dynamic typing, together with its interpreted nature, make it an ideal language for scripting and rapid application development in many areas on most platforms.

The Python interpreter and the extensive standard library are freely available in source or binary form for all major platforms from the Python Web site, <https://www.python.org/>, and may be freely distributed. The same site also contains distributions of and pointers to many free third party Python modules, programs and tools, and additional documentation.

The Python interpreter is easily extended with new functions and data types implemented in C or C++ (or other languages callable from C). Python is also suitable as an extension language for customizable applications.

Python is a high-level, interpreted, interactive and object-oriented scripting language. Python is designed to be highly readable. It uses English keywords frequently where as other languages use punctuation, and it has fewer syntactical constructions than other languages.

* **Python is Interpreted −** Python is processed at runtime by the interpreter. You do not need to compile your program before executing it. This is similar to PERL and PHP.
* **Python is Interactive −** you can actually sit at a Python prompt and interact with the interpreter directly to write your programs.
* **Python is Object-Oriented −** Python supports Object-Oriented style or technique of programming that encapsulates code within objects.
* **Python is a Beginner's Language −** Python is a great language for the beginner-level programmers and supports the development of a wide range of applications from simple text processing to WWW browsers to games.

**Features:**

Python's features include – All these.

* **Easy-to-learn −** Python has few keywords, simple structure, and a clearly defined syntax. This allows the student to pick up the language quickly.
* **Easy-to-read −** Python code is more clearly defined and visible to the eyes.
* **Easy-to-maintain −** Python's source code is fairly easy-to-maintain.
* **A broad standard library −** Python's bulk of the library is very portable and cross-platform compatible on UNIX, Windows, and Macintosh.
* **Interactive Mode −** Python has support for an interactive mode which allows interactive testing and debugging of snippets of code.
* **Portable −** Python can run on a wide variety of hardware platforms and has the same interface on all platforms.
* **Extendable −** you can add low-level modules to the Python interpreter. These modules enable programmers to add to or customize their tools to be more efficient.
* **Databases −** Python provides interfaces to all major commercial databases.
* **GUI Programming−** Python supports GUI applications that can be created and ported to many system calls, libraries and windows systems, such as Windows MFC, Macintosh, and the X Window system of Unix.
* **Scalable −** Python provides a better structure and support for large programs than shell scripting.

**Why python emerging as a leader:**

There’s battle out there happening in the minds of aspiring data scientists to choose the best data science tool. Though there are quite a number of data science tools that provide the much-needed option, the close combat narrows down between two popular languages – Python and R.

Between the two, Python is emerging as the popular language used more in data science applications.

Take the case of the tech giant Google that has created the deep learning framework called tensor flow – Python is the primary language used for creating this framework. Its footprint has continued to increase in the environment promoted by Netflix. Production engineers at Face book and Khan Academy have for long been using it as a prominent language in their environment.

Python has other advantages that speed up it’s upward swing to the top of data science tools. It integrates well with the most cloud as well as platform-as-a-service providers. In supporting multiprocessing for parallel computing, it brings the distinct advantage of ensuring large-scale performance in data science and machine learning. Python can also be extended with modules written in C/C++.

**Where Python becomes the perfect-fit:**

There are tailor-made situations where it is the best data science tool for the job. It is perfect when data analysis tasks involve integration with web apps or when there is a need to incorporate statistical code into the production database. The full-fledged programming nature of Python makes it a perfect fit for implementing algorithms.  
Its packages rooted for specific data science jobs. Packages like Numpy, SciPy, and pandas produce good results for data analysis jobs. While there is a need for graphics, Python’s matplotlib emerges as a good package, and for machine learning tasks, scikit-learn becomes the ideal alternate.

**Why is Python preferred over other data science tools?**

It is ‘Pythonic’ when the code is written in a fluent and natural style. Apart from that, it is also known for other features that have captured the imaginations of data science community.

**Easy to learn:**

The most alluring factor of Python is that anyone aspiring to learn this language can learn it easily and quickly. When compared to other data science languages like R, Python promotes a shorter learning curve and scores over others by promoting an easy-to-understand syntax.

**Scalability:**

When compared to other languages like R, Python has established a lead by emerging as a scalable language, and it is faster than other languages like Matlab and Stata. Python’s scalability lies in the flexibility that it gives to solve problems, as in the case of YouTube that migrated to Python. Python has come good for different usages in different industries and for rapid development of applications of all kinds.

**Choice of data science libraries:**

The significant factor giving the push for Python is the variety of data science/data analytics libraries made available for the aspirants. Pandas, StatsModels, NumPy, SciPy, and Scikit-Learn, are some of the libraries well known in the data science community. Python does not stop with that as libraries have been growing over time. What you thought was a constraint a year ago would be addressed well by Python with a robust solution addressing problems of specific nature.

**Python community:**

One of the reasons for the phenomenal rise of Python is attributed to its ecosystem. As Python extends its reach to the data science community, more and more volunteers are creating data science libraries. This, in turn, has led the way for creating the most modern tools and processing in Python.

The widespread and involved community promotes easy access for aspirants who want to find solutions to their coding problems. Whatever queries you need, it is a click or a Google search away. Enthusiasts can also find access to professionals on Code mentor and Stack Overflow to find the right answers for their queries.

**Graphics and visualization:**

Python comes with varied visualization options. Matplotlib provides the solid foundation around which other libraries like Sea born, pandas plotting, and ggplot have been built. The visualization packages help you get a good sense of data, create charts, graphical plot and create web-ready interactive plots.

## Why Choose Python?

If you’re going to write programs, there are literally dozens of commonly used languages to choose from. Why choose Python? Here are some of the features that make Python an appealing choice.

### Python is Popular

Python has been growing in popularity over the last few years. The 2018 [Stack Overflow Developer Survey](https://insights.stackoverflow.com/survey/2018) ranked Python as the 7th most popular and the number one most wanted technology of the year. [World-class software development countries around the globe use Python every single day.](https://realpython.com/world-class-companies-using-python/)

According to [research by Dice](https://insights.dice.com/2016/02/01/whats-hot-and-not-in-tech-skills/) Python is also one of the hottest skills to have and the most popular programming language in the world based on the.  [Popularity of Programming Language Index](https://pypl.github.io/PYPL.html)

Due to the popularity and widespread use of Python as a programming language, Python developers are sought after and paid well. If you’d like to dig deeper into [Python salary statistics and job opportunities, you can do so here](https://dbader.org/blog/why-learn-python).

Many languages are compiled, meaning the source code you create needs to be translated into machine code, the language of your computer’s processor, before it can be run. Programs written in an interpreted language are passed straight to an interpreter that runs them directly.

This makes for a quicker development cycle because you just type in your code and run it, without the intermediate compilation step.

One potential downside to interpreted languages is execution speed. Programs that are compiled into the native language of the computer processor tend to run more quickly than interpreted programs. For some applications that are particularly computationally intensive, like graphics processing or intense number crunching, this can be limiting.

In practice, however, for most programs, the difference in execution speed is measured in milliseconds, or seconds at most, and not appreciably noticeable to a human user. The expediency of coding in an interpreted language is typically worth it for most applications.

### Python is Free

The Python interpreter is developed under an OSI-approved open-source license, making it free to install, use, and distribute, even for commercial purposes. A version of the interpreter is available for virtually any platform there is, including all flavors of Unix, Windows, macOS, smartphones and tablets, and probably anything else you ever heard of. A version even exists for the half dozen people remaining who use OS/2.

### Python is Portable

Because Python code is interpreted and not compiled into native machine instructions, code written for one platform will work on any other platform that has the Python interpreter installed. (This is true of any interpreted language, not just Python.)

### Python is Simple

As programming languages go, Python is relatively uncluttered, and the developers have deliberately kept it that way. A rough estimate of the complexity of a language can be gleaned from the number of keywords or reserved words in the language. These are words that are reserved for special meaning by the compiler or interpreter because they designate specific built-in functionality of the language. Python 3 has 33 keywords, and Python 2 has 31. By contrast, C++ has 62, Java has 53, and Visual Basic has more than 120, though these latter examples probably vary somewhat by implementation or dialect. Python code has a simple and clean structure that is easy to learn and easy to read. In fact, as you will see, the language definition enforces code structure that is easy to read.

### But It’s Not That Simple

For all its syntactical simplicity, Python supports most constructs that would be expected in a very high-level language, including complex dynamic data types, structured and functional programming, and [object-oriented programming](https://realpython.com/python3-object-oriented-programming/).

Additionally, a very extensive library of classes and functions is available that provides capability well beyond what is built into the language, such as database manipulation or GUI programming.

Python accomplishes what many programming languages don’t: the language itself is simply designed, but it is very versatile in terms of what you can accomplish with it.

**Is Python ‘the’ tool for machine learning?**

When it comes to data science, machine learning is one of the significant elements used to maximize value from data. With Python as the data science tool, exploring the basics of machine learning becomes easy and effective. In a nutshell, machine learning is more about statistics, mathematical optimization, and probability. It has become the most preferred machine learning tool in the way it allows aspirants to ‘do math’ easily.

Name any math function, and you have a Python package meeting the requirement. There is Numpy for numerical linear algebra, CVXOPT for convex optimization, Scipy for general scientific computing, SymPy for symbolic algebra, PYMC3, and Statsmodel for statistical modeling.

With the grip on the basics of machine learning algorithm including logistic regression and linear regression, it makes it easy to implement machine learning systems for predictions by way of its scikit-learn library. It’s easy to customize for neutral networks and deep learning with libraries including Keras, Theano, and TensorFlow.

Data science landscape is changing rapidly, and tools used for extracting value from data science have also grown in numbers. The two most popular languages that fight for the top spot are R and Python. Both are revered by enthusiasts, and both come with their strengths and weaknesses. But with the tech giants like Google showing the way to use Python and with the learning curve made short and easy, it inches ahead to become the most popular language in the data science world.

**Machine learning**

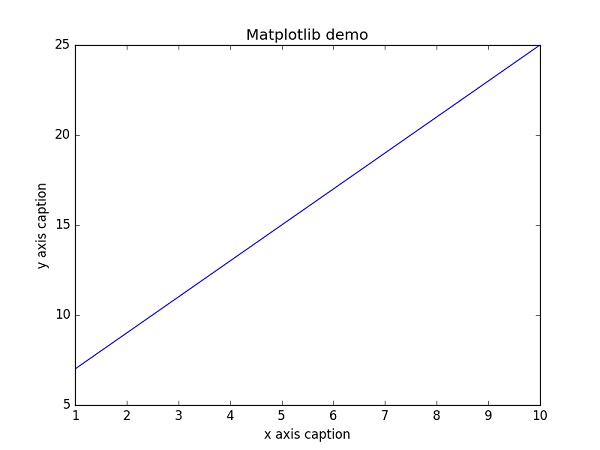
Machine Learning is a method of statistical learning where each instance in a dataset is described by a set of features or attributes. In contrast, the term “Deep Learning” is a method of statistical learning that extracts features or attributes from raw data. Deep Learning does this by utilizing neural networks with many hidden layers, big data, and powerful computational resources. The terms seem somewhat interchangeable, however, with Deep Learning method, The algorithm constructs representations of the data automatically. In contrast, data representations are hard-coded as a set of features in machine learning algorithms, requiring further processes such as feature selection and extraction, (such as PCA).

Both of these terms are in dramatic contrast with another class of classical artificial intelligence algorithms known as Rule-Based Systems where each decision is manually programmed in such a way that it resembles a statistical model.

In Machine Learning and Deep Learning, there are many different models that fall into two different categories, supervised and unsupervised. In unsupervised learning, algorithms such as k-Means, hierarchical clustering, and Gaussian mixture models attempt to learn meaningful structures in the data. Supervised learning involves an output label associated with each instance in the dataset. This output can be discrete/categorical or real-valued. Regression models estimate real-valued outputs, whereas classification models estimate discrete-valued outputs. Simple binary classification models have just two output labels, 1 (positive) and 0 (negative). Some popular supervised learning algorithms that are considered Machine Learning: are linear regression, logistic regression, decision trees, support vector machines, and neural networks, as well as non-parametric models such as k-Nearest

**Matplotlib**

Matplotlib is an amazing visualization library in Python for 2D plots of arrays. Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack. It was introduced by John Hunter in the year 2002.One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals. Matplotlib consists of several plots like line, bar, scatter, histogram etc.



**Jupyter Notebook**

The Jupyter Notebook is an **interactive computing environment** that enables users to author notebook documents that include: - Live code - Interactive widgets - Plots - Narrative text - Equations - Images – Video.

These documents provide a complete and self-contained record of a computation that can be converted to various formats and shared with others using email, [Dropbox](https://www.dropbox.com/), version control systems (like git/[GitHub](https://github.com/)) or [nbviewer.jupyter.org](http://nbviewer.jupyter.org/).

**Components**

The Jupyter Notebook combines three components:

* **The notebook web application**: An interactive web application for writing and running code interactively and authoring notebook documents.
* **Kernels**: Separate processes started by the notebook web application that runs users’ code in a given language and returns output back to the notebook web application. The kernel also handles things like computations for interactive widgets, tab completion and introspection.
* **Notebook documents**: Self-contained documents that contain a representation of all content visible in the notebook web application, including inputs and outputs of the computations, narrative text, equations, images, and rich media representations of objects. Each notebook document has its own kernel.

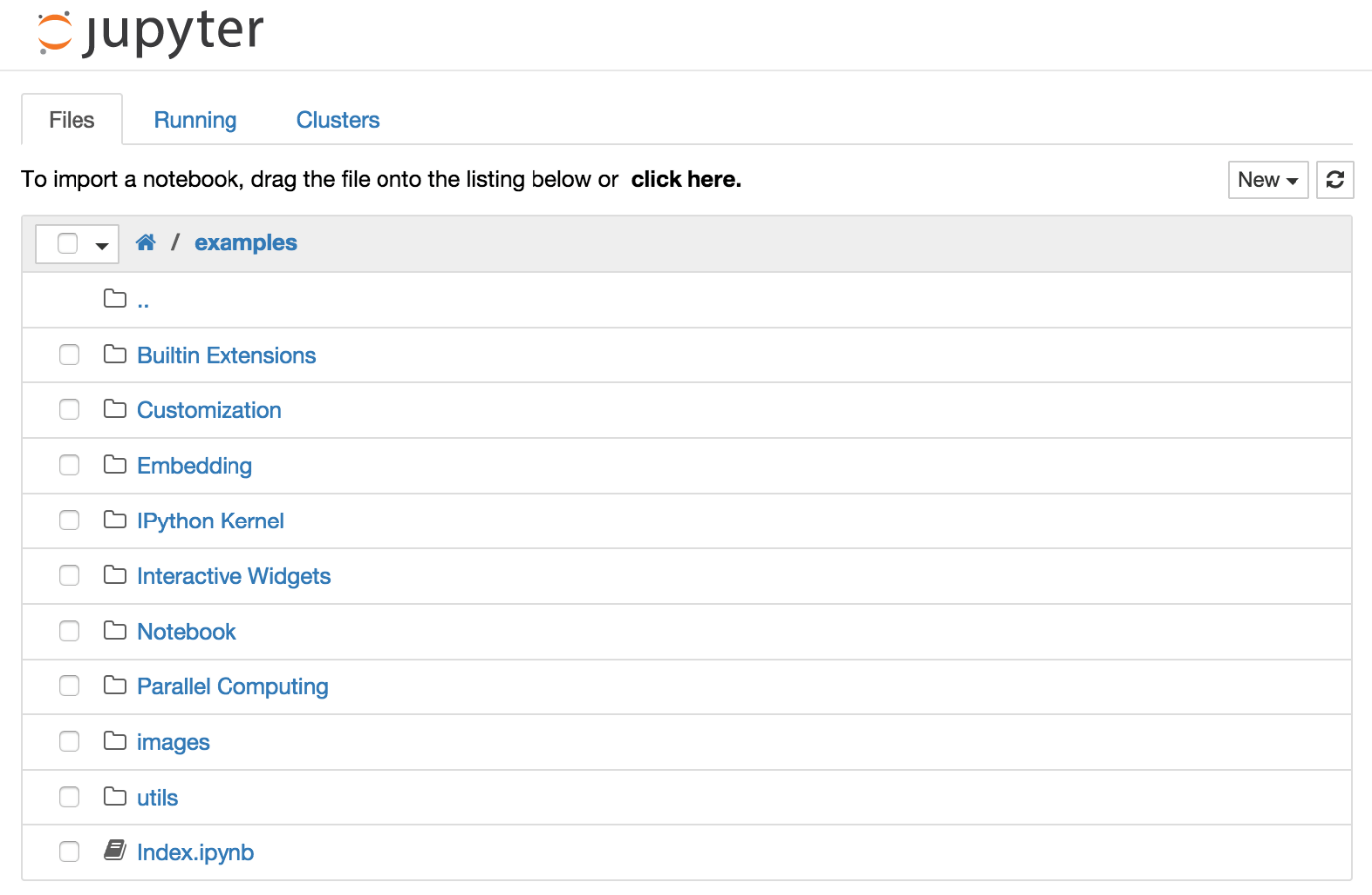
**Notebook web application**

The notebook web application enables users to:

* **Edit code in the browser**, with automatic syntax highlighting, indentation, and tab completion/introspection.
* **Run code from the browser**, with the results of computations attached to the code which generated them.
* See the results of computations with **rich media representations**, such as HTML, LaTeX, PNG, SVG, PDF, etc.
* Create and use **interactive JavaScript widgets**, which bind interactive user interface controls and visualizations to reactive kernel side computations.
* Author **narrative text** using the [Markdown](https://daringfireball.net/projects/markdown/) markup language.
* Include mathematical equations using **LaTeX syntax in Markdown**, which are rendered in-browser by [MathJax](https://www.mathjax.org/).

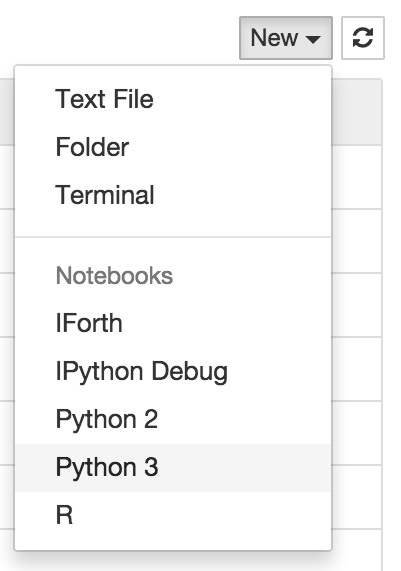
**The Notebook dashboard**

When you first start the notebook server, your browser will open to the notebook dashboard. The dashboard serves as a home page for the notebook. Its main purpose is to display the notebooks and files in the current directory. For example, here is a screenshot of the dashboard page for the example’s directory in the Jupyter repository:



The top of the notebook list displays clickable breadcrumbs of the current directory. By clicking on these breadcrumbs or on sub-directories in the notebook list, you can navigate your file system.

To create a new notebook, click on the “New” button at the top of the list and select a kernel from the dropdown (as seen below). Which kernels are listed depend on what’s installed on the server. Some of the kernels in the screenshot below may not exist as an option to you.



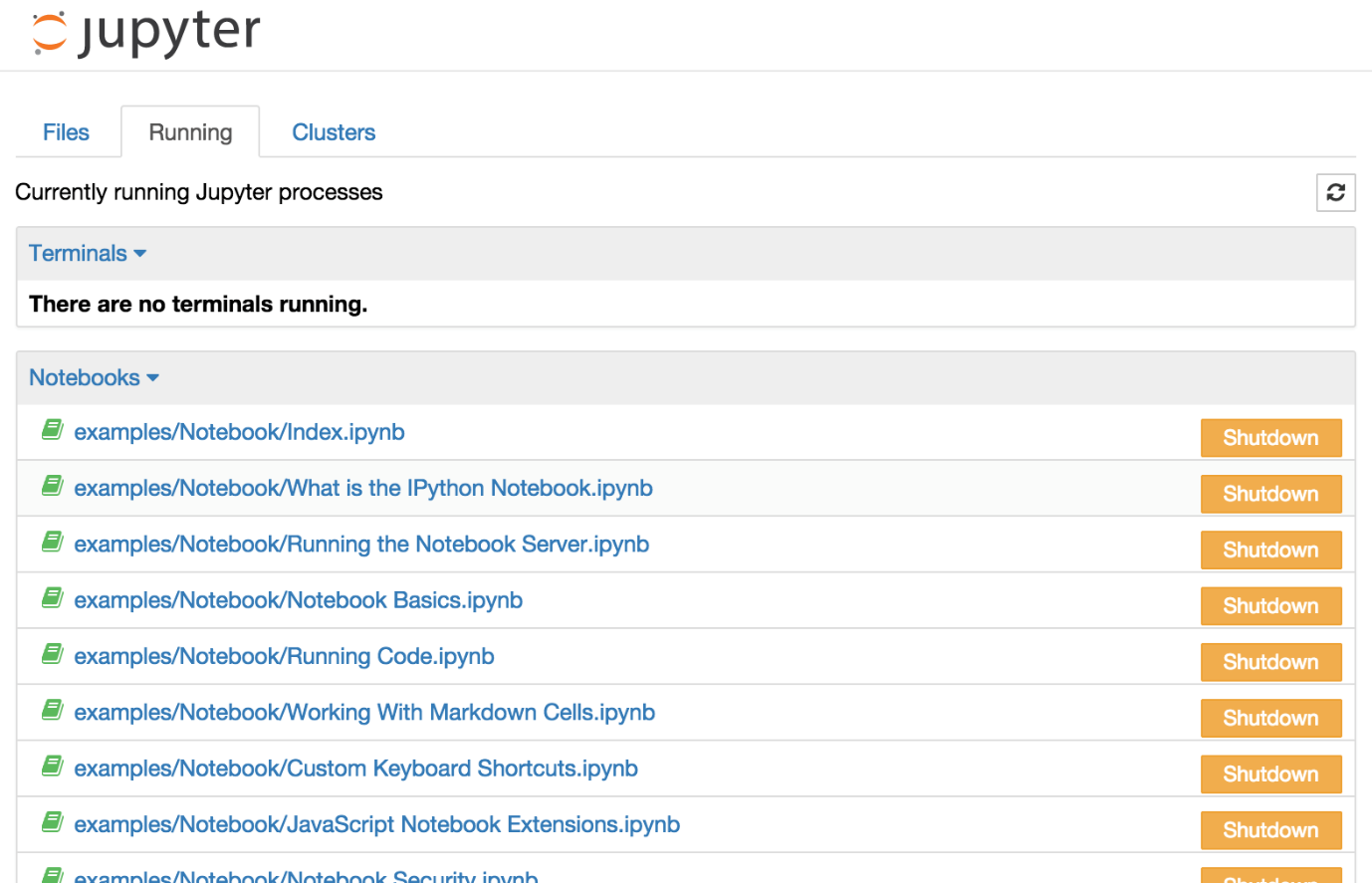
Notebooks and files can be uploaded to the current directory by dragging a notebook file onto the notebook list or by the “click here” text above the list.

The notebook list shows green “Running” text and a green notebook icon next to running notebooks (as seen below). Notebooks remain running until you explicitly shut them down; closing the notebook’s page is not sufficient.



To shutdown, delete, duplicate, or rename a notebook check the checkbox next to it and an array of controls will appear at the top of the notebook list (as seen below). You can also use the same operations on directories and files when applicable.

To see all of your running notebooks along with their directories, click on the “Running” tab:



This view provides a convenient way to track notebooks that you start as you navigate the file system in a long running notebook server.

**Chapter 3**

**TASK PERFORM**

**3.1 Technologies**

**3.1.1 Hardware**

* Processor : i3/i5
* Hard disk : 500GB
* RAM : 4GB

**3.1.2 Software**

* Operating system : Windows 8/10
* Programming Language : Python-3
* Framework : Anaconda
* IDE : Jupyter Notebook
* ML-Libraries : Scikit-Learn

**3.2 Task Performed**

|  |  |  |
| --- | --- | --- |
| **SL NO** | **WEEK** | **TASKCOMPLETED** |
| 1 | **Week1** | * Introduction of Python * Statements & Syntax * Data Structure |
| 2 | **Week2** | * File Operation * Functions * Modules & Packages * Classes |
| 3 | **Week3** | * Numpy * Pandas * Data Mining Concept |
| 4 | **Week4** | * Working with project * Machine learning concept |

**3.4 Working Code**

{

"cells": [

{

"cell\_type": "markdown",

"metadata": {},

"source": [

"## Import Libraries"

]

},

{

"cell\_type": "code",

"execution\_count": 1,

"metadata": {},

"outputs": [],

"source": [

"import warnings\n",

"warnings.filterwarnings(\"always\")\n",

"warnings.filterwarnings(\"ignore\")"

]

},

{

"cell\_type": "code",

"execution\_count": 2,

"metadata": {},

"outputs": [],

"source": [

"import os\n",

"import numpy as np\n",

"import pandas as pd\n",

"import matplotlib.pyplot as plt\n",

"%matplotlib inline\n",

"import seaborn as sns\n",

"pd.set\_option(\"display.max\_columns\",None)"

]

},

{

"cell\_type": "markdown",

"metadata": {},

"source": [

"## Data Loading"

]

},

{

"cell\_type": "code",

"execution\_count": 3,

"metadata": {},

"outputs": [],

"source": [

"df=pd.read\_csv(\"input/women.csv\")"

]

},

{

"cell\_type": "code",

"execution\_count": 4,

"metadata": {},

"outputs": [

{

"data": {

"text/plain": [

"(12134, 22)"

]

},

"execution\_count": 4,

"metadata": {},

"output\_type": "execute\_result"

}

],

"source": [

"df.shape"

]

},

{

"cell\_type": "code",

"execution\_count": 5,

"metadata": {},

"outputs": [

{

"data": {

"text/html": [

"<div>\n",

"<style scoped>\n",

" .dataframe tbody tr th:only-of-type {\n",

" vertical-align: middle;\n",

" }\n",

"\n",

" .dataframe tbody tr th {\n",

" vertical-align: top;\n",

" }\n",

"\n",

" .dataframe thead th {\n",

" text-align: right;\n",

" }\n",

"</style>\n",

"<table border=\"1\" class=\"dataframe\">\n",

" <thead>\n",

" <tr style=\"text-align: right;\">\n",

" <th></th>\n",

" <th>category</th>\n",

" <th>subcategory</th>\n",

" <th>name</th>\n",

" <th>current\_price</th>\n",

" <th>raw\_price</th>\n",

" <th>currency</th>\n",

" <th>discount</th>\n",

" <th>likes\_count</th>\n",

" <th>is\_new</th>\n",

" <th>brand</th>\n",

" <th>brand\_url</th>\n",

" <th>codCountry</th>\n",

" <th>variation\_0\_color</th>\n",

" <th>variation\_1\_color</th>\n",

" <th>variation\_0\_thumbnail</th>\n",

" <th>variation\_0\_image</th>\n",

" <th>variation\_1\_thumbnail</th>\n",

" <th>variation\_1\_image</th>\n",

" <th>image\_url</th>\n",

" <th>url</th>\n",

" <th>id</th>\n",

" <th>model</th>\n",

" </tr>\n",

" </thead>\n",

" <tbody>\n",

" <tr>\n",

" <th>0</th>\n",

" <td>women</td>\n",

" <td>T-shirts</td>\n",

" <td>T-shirt boutonné à manches courtes et imprimé ...</td>\n",

" <td>23.99</td>\n",

" <td>46.99</td>\n",

" <td>USD</td>\n",

" <td>49.0</td>\n",

" <td>313.0</td>\n",

" <td>False</td>\n",

" <td>NaN</td>\n",

" <td>NaN</td>\n",

" <td>ID,MY,PH,SG,TH,VN</td>\n",

" <td>White</td>\n",

" <td>Navy</td>\n",

" <td>https://imgaz1.chiccdn.com/thumb/list\_grid/oau...</td>\n",

" <td>https://imgaz1.chiccdn.com/thumb/view/oaupload...</td>\n",

" <td>https://imgaz1.chiccdn.com/thumb/list\_grid/oau...</td>\n",

" <td>https://imgaz1.chiccdn.com/thumb/view/oaupload...</td>\n",

" <td>https://imgaz1.chiccdn.com/thumb/view/oaupload...</td>\n",

" <td>https://fr.newchic.com/t-shirts-3666/p-1690810...</td>\n",

" <td>1690810.0</td>\n",

" <td>SKUF40137</td>\n",

" </tr>\n",

" <tr>\n",

" <th>1</th>\n",

" <td>women</td>\n",

" <td>Soutiens-gorge</td>\n",

" <td>Plus Soutiens-gorge avec fermeture à l'avant</td>\n",

**Chapter 4**

# Steps for Project Development

# Libraries Importing:

Normally, a library is a collection of books or is a room or place where many books are stored to be used later. Similarly, in the programming world, a library is a collection of precompiled codes that can be used later on in a program for some specific well-defined operations. Other than pre-compiled codes, a library may contain documentation, configuration data, message templates, classes, and values, etc.

A Python library is a collection of related modules. It contains bundles of code that can be used repeatedly in different programs. It makes Python Programming simpler and convenient for the programmer. As we don’t need to write the same code again and again for different programs. Python libraries play a very vital role in fields of Machine Learning, Data Science, Data Visualization, etc.

**Working of Python Library**

As is stated above, a Python library is simply a collection of codes or modules of codes that we can use in a program for specific operations. We use libraries so that we don’t need to write the code again in our program that is already available. But how it works. Actually, in the MS Windows environment, the library files have a DLL extension (Dynamic Load Libraries). When we link a library with our program and run that program, the linker automatically searches for that library. It extracts the functionalities of that library and interprets the program accordingly. That’s how we use the methods of a library in our program. We will see further, how we bring in the libraries in our Python programs.

### Python standard library

The Python Standard Library contains the exact syntax, semantics, and tokens of Python. It contains built-in modules that provide access to basic system functionality like I/O and some other core modules. Most of the Python Libraries are written in the C programming language. The Python standard library consists of more than 200 core modules. All these work together to make Python a high-level programming language. Python Standard Library plays a very important role. Without it, the programmers can’t have access to the functionalities of Python. But other than this, there are several other libraries in Python that make a programmer’s life easier. Let’s have a look at some of the commonly used libraries:

1. **TensorFlow:**This library was developed by Google in collaboration with the Brain Team. It is an open-source library used for high-level computations. It is also used in machine learning and deep learning algorithms. It contains a large number of tensor operations. Researchers also use this Python library to solve complex computations in Mathematics and Physics.
2. **Matplotlib:**This library is responsible for plotting numerical data. And that’s why it is used in data analysis. It is also an open-source library and plots high-defined figures like pie charts, histograms, scatterplots, graphs, etc.
3. **Pandas:**Pandas are an important library for data scientists. It is an open-source machine learning library that provides flexible high-level data structures and a variety of analysis tools. It eases data analysis, data manipulation, and cleaning of data. Pandas support operations like Sorting, Re-indexing, Iteration, Concatenation, Conversion of data, Visualizations, Aggregations, etc.
4. **Numpy:**The name “Numpy” stands for “Numerical Python”. It is the commonly used library. It is a popular machine learning library that supports large matrices and multi-dimensional data. It consists of in-built mathematical functions for easy computations. Even libraries like TensorFlow use Numpy internally to perform several operations on tensors. Array Interface is one of the key features of this library.
5. **SciPy:**The name “SciPy” stands for “Scientific Python”. It is an open-source library used for high-level scientific computations. This library is built over an extension of Numpy. It works with Numpy to handle complex computations. While Numpy allows sorting and indexing of array data, the numerical data code is stored in SciPy. It is also widely used by application developers and engineers.
6. **Scikit-learn:**It is a famous Python library to work with complex data. Scikit-learn is an open-source library that supports machine learning. It supports variously supervised and unsupervised algorithms like linear regression, classification, clustering, etc. This library works in association with Numpy and SciPy.

There are many more libraries in Python. We can use a suitable library for our purposes. Hence, Python libraries play a very crucial role and are very helpful to the developers.

### Use of Libraries in Python Program

As we write large-size programs in Python, we want to maintain the code’s modularity. For the easy maintenance of the code, we split the code into different parts and we can use that code later ever we need it. In Python, *modules* play that part. Instead of using the same code in different programs and making the code complex, we define mostly used functions in modules and we can just simply import them in a program wherever there is a requirement. We don’t need to write that code but still, we can use its functionality by importing its module. Multiple interrelated modules are stored in a library. And whenever we need to use a module, we import it from its library. In Python, it’s a very simple job to do due to its easy syntax. We just need to use **import**.

# Data Loading:

Suppose if you want to start a ML project then what is the first and most important thing you would require? It is the data that we need to load for starting any of the ML project. With respect to data, the most common format of data for ML projects is CSV (comma-separated values).

Basically, CSV is a simple file format which is used to store tabular data (number and text) such as a spreadsheet in plain text. In Python, we can load CSV data into with different ways but before loading CSV data we must have to take care about some considerations.

### Consideration While Loading CSV data

CSV data format is the most common format for ML data, but we need to take care about following major considerations while loading the same into our ML projects −

### File Header

In CSV data files, the header contains the information for each field. We must use the same delimiter for the header file and for data file because it is the header file that specifies how should data fields be interpreted.

The following are the two cases related to CSV file header which must be considered −

* **Case-I: When Data file is having a file header** − It will automatically assign the names to each column of data if data file is having a file header.
* **Case-II: When Data file is not having a file header** − We need to assign the names to each column of data manually if data file is not having a file header.

In both the cases, we must need to specify explicitly weather our CSV file contains header or not.

### Comments

Comments in any data file are having their significance. In CSV data file, comments are indicated by a hash (#) at the start of the line. We need to consider comments while loading CSV data into ML projects because if we are having comments in the file then we may need to indicate, depends upon the method we choose for loading, whether to expect those comments or not.

### Delimiter

In CSV data files, comma (,) character is the standard delimiter. The role of delimiter is to separate the values in the fields. It is important to consider the role of delimiter while uploading the CSV file into ML projects because we can also use a different delimiter such as a tab or white space. But in the case of using a different delimiter than standard one, we must have to specify it explicitly.

### Quotes

In CSV data files, double quotation (“ ”) mark is the default quote character. It is important to consider the role of quotes while uploading the CSV file into ML projects because we can also use other quote character than double quotation mark. But in case of using a different quote character than standard one, we must have to specify it explicitly.

### Methods to Load CSV Data File

While working with ML projects, the most crucial task is to load the data properly into it. The most common data format for ML projects is CSV and it comes in various flavors and varying difficulties to parse. In this section, we are going to discuss about three common approaches in Python to load CSV data file −

### Load CSV with Python Standard Library

The first and most used approach to load CSV data file is the use of Python standard library which provides us a variety of built-in modules namely **csv module** and the reader()function. The following is an example of loading CSV data file with the help of it.

# Data Preprocessing:

Data Preprocessing includes the steps we need to follow to transform or encode data so that it may be easily parsed by the machine.

The main agenda for a model to be accurate and precise in predictions is that the algorithm should be able to easily interpret the data's features.

### Why is Data Preprocessing important?

The majority of the real-world datasets are highly susceptible to missing, inconsistent, and noisy data due to their heterogeneous origin.

Applying data mining algorithms on this noisy data would not give quality results as they would fail to identify patterns effectively. Data Processing is, therefore, important to improve the overall data quality.

* Duplicate or missing values may give an incorrect view of the overall statistics of data.
* Outliers and inconsistent data points often tend to disturb the model’s overall learning, leading to false predictions.

Quality decisions must be based on quality data. Data Preprocessing is important to get this quality data, without which it would just be a Garbage In, Garbage Out scenario.

### Features in machine learning

Individual independent variables that operate as an input in our machine learning model are referred to as features. They can be thought of as representations or attributes that describe the data and help the models to predict the classes/labels.

For example, features in a structured dataset like in a CSV format refer to each column representing a measurable piece of data that can be used for analysis: Name, Age, Sex, Fare, and so on.

### 4 Steps in Data Preprocessing

Now, let's discuss more in-depth four main stages of data preprocessing.

### Data Cleaning

Data Cleaning is particularly done as part of data preprocessing to clean the data by filling missing values, smoothing the noisy data, resolving the inconsistency, and removing outliers.

### Missing values

Here are a few ways to solve this issue:

* Ignore those tuples

This method should be considered when the dataset is huge and numerous missing values are present within a tuple.

* Fill in the missing values

There are many methods to achieve this, such as filling in the values manually, predicting the missing values using regression method, or numerical methods like attribute mean.

# Model prediction:

“Prediction” refers to the output of an [algorithm](https://www.datarobot.com/wiki/algorithm/" \t "_blank) after it has been [trained](https://www.datarobot.com/wiki/training-validation-holdout/" \t "_blank) on a historical dataset and applied to new data when forecasting the likelihood of a particular outcome, such as whether or not a customer will churn in 30 days. The algorithm will generate probable values for an unknown variable for each record in the new data, allowing the model builder to identify what that value will most likely be.

The word “prediction” can be misleading. In some cases, it really does mean that you are predicting a future outcome, such as when you’re using machine learning to determine [the next best action](https://www.datarobot.com/use-cases/next-best-action/) in a marketing campaign. Other times, though, the “prediction” has to do with, for example, whether or not a transaction that already occurred was fraudulent. In that case, the transaction already happened, but you’re making an educated guess about whether or not it was legitimate, allowing you to take the appropriate action.

### Why are Predictions Important?

Machine learning [model](https://www.datarobot.com/wiki/model/" \t "_blank) predictions allow businesses to make highly accurate guesses as to the likely outcomes of a question based on historical data, which can be about all kinds of things – customer churn likelihood, possible fraudulent activity, and more. These provide the business with [insights](https://www.datarobot.com/wiki/insights/" \t "_blank) that result in tangible business value. For example, if a model predicts a customer is likely to churn, the business can target them with specific communications and outreach that will prevent the loss of that customer.

**Chapter 4**

**LEARNING OUTCOME**

•**Skill Development:** What new skills will I learn or hope to learn during the internship? Some skills you may have the chance to develop include: operating office or computer equipment; handling a variety of situations simultaneously; organizing or analyzing data, records, or budgets; or improving teamwork, writing, and speaking abilities. Assignments and work environment will determine the types of skills developed.

•**Personal Development:** How will I benefit personally from the internship experience? During your internship, make a special effort to observe the personal style of supervisors and colleagues. Be able to identify clear examples of leadership styles that either promote good working relationships or hinder a productive work environment. Note how to deal with pressure, tension, and praise in work relationships. It is important to understand how to communicate knowledge to strangers, supervisors, and peers. Be aware of when to speak and when to listen. Try to understand the employer’s organizational chart (formal or informal). What is a typical career path? What sort of education and experience is normally required in each line of work? What can you learning about laws, taxes, and fringe benefits that are important to employees in this field? What grievances are there? What are the personal satisfactions and rewards? Try to obtain knowledge of job opportunities available in this field. Become familiar with occupational literature and professional organizations.

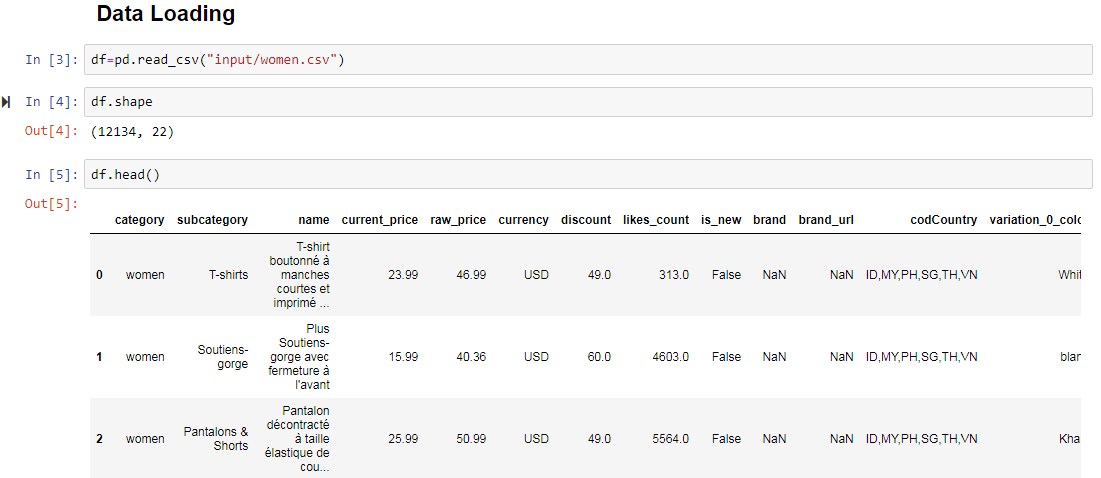
**Chapter 5**

**Result**

Fig 1: Importing library and reading data



Fig 2: Data



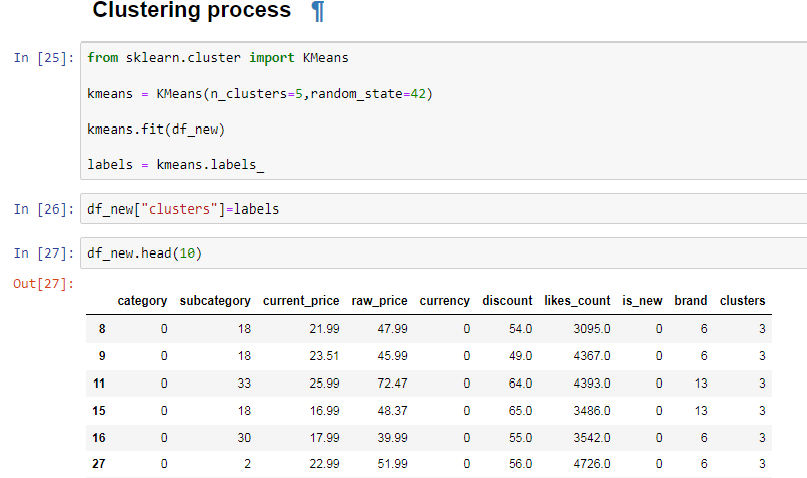
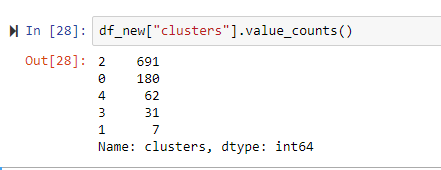


Fig 3: Algorithm



Fig 4: Final Result



**REFERENCES**

1.https://www.geeksforgeeks.org/machine-learning/

2.https://www.python.org/

3.https://www.w3schools.com/python/

4.https://www.coursera.org/learn/machine-learning

5.https://www.techopedia.com/definition/1181/data-mining

6. https://scikit-learn.org/stable/