**INTERNSHIP REPORT**

**A Report of Internship**

**at**

**YBI Foundation**

**Submitted by**

**SUMAN JAISWAL**

**Regd.No.: 21791A32H1**

*in partial fulfillment for the award of the degree*

*of*

**BACHELOR OF TECHNOLOGY**

*in*

**COMPUTER SCIENCE ENGINEERING**

****

****

**SRI VENKATESWARA COLLEGE OF ENGINEERING AND TECHNOLOGY(AUTONOMUS)**

(Approved by AICTE, Affiliated to JNTUA)

(Accredited by N.B.A., New Delhi & NAAC. Bangalore)

**R.V.S. Nagar, Chittoor-517127**

**ANDHRA PRADESH**

DECEMBER 2023

**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SRI VENKATESWARA COLLEGE OF ENGINEERING &TECHNOLOGY(AUTONOMOUS**)

****

**CERTIFICATE**

This is to certify Certified that this **Report of Internship at**

**“.YBI Foundation..”**

Being submitted herewith to the

**SRI VENKNATESWARA COLLEGE OF ENGINEERING & TECHNOLOGY(AUTONOMOUS)**

is the bonafide work of

**.. SUMAN JAISWAL (*21781A32H1*)..**

who carried out the project work under our guidance and supervision**.**

|  |  |  |
| --- | --- | --- |
| **College internship Coordinator** | **Department Internship Coordinator** | **Head of The Department** |

**INTERNSHIP CERTIFICATION**



**ACKNOWLEDGEMENT**

I would like to express my special thanks of gratitude to my Internship Training guide Mrs. Malika Srivastava, Head of the Department DR. M. LAVANYA as well as our principal Dr. M. Mohan Babu who gave me the golden opportunity to do this wonderful Internship at “YBI Foundation”, which provided me an opportunity explore the new horizons.

I sincerely express my gratitude towards the “YBI Foundation”, for providing this opportunity.

I would like to thank internship coordinator Department of CSD for their support and advices.

Secondly, I would also like to thank my parents and friends who helped me a lot in finalizing this report within the limited time frame.

|  |  |
| --- | --- |
| Date: 08/12/2023 | **SUMAN JAISWAL** |
| Place Chittoor | 21781A32H1 |
|  | III CSD- ‘B’ |

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
|  | **TITLE** | **Page No.** |
|  | Cover Page | 1 |
|  | Certificate | 2 |
|  | Acknowledgement | 4 |
|  | Table of Content | 5 |
|  |
|  | List of Figures |  |
|  |  |  |
| **1.0** | **INTRODUCTION** | **1** |
|  | 1.1 Sub section | 1 |
|  | 1.2 Sub section | 1 |
| **2.0** | **COMPANY PROFILE** | **2** |
|  | 2.1 Sub section | 2 |
|  | 2.2 Sub section | 2 |
| **3.0** | **DEPARTMENT PROFILE** | **3** |
|  | 3.1 Sub section | 3 |
|  | 3.2 Sub section | 3 |
| **4.0** | **PROJECT WORK** | **4** |
|  | 4.1 Sub section | 4 |
|  | 4.2 Sub section | 4 |
| **5.0** | **CASE STUDY** | **5** |
|  | 5.1 Sub section | 5 |
|  | 5.2 Sub section | 5 |
| **6.0** | **CONCLUSION** | **6** |
|  | **REFERENCES** | **7** |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Figure No.** | **Title** | **Page No**. |
| Figure 1.1 | Line chart | 21 |
| Figure 1.1 | Pie chart | 22 |
| Figure 1.1 | Output of project | 36 |

**TABLES OF CONTENT**

INDEX

**S.no CONTENTS Page.no**

1. Introduction to python 10

1.1 features of python

2.**Data Science**

1. **Data Collection:**
   * Describe the sources of data used in your analysis.
   * Discuss the data collection process, including any challenges or biases. Data cleaning
   * Data transformation
   * Feature scaling
   * Handling missing data
2. **Data Exploration and Preprocessing:**
   * Explore the dataset and provide descriptive statistics.
   * Discuss any missing data, outliers, or data quality issues.
   * Describe the preprocessing steps undertaken to clean and prepare the data

3.**Exploratory Data Analysis (EDA)**:

* This section should include visualizations and statistical summaries of the data to gain insights and identify patterns Histograms
* Box plots
* Scatter plots
* Pair plots
* Heatmaps

4. **Data Visualization:**

* Matplotlib
* Seaborn
* Plotly
* Tableau
* g. gplot2 (R)

5. **Text Analysis**:

* Natural Language Processing (NLP)
* Text mining
* Sentiment analysis
* Topic modelling

3. Python for machine learning

3.1 Understanding operator

3.2 Variables and Data Types

3.3 Conditional Statements, For and While Loops

3.4 Functions

3.5 Data Structure

4. INTRODUCTION

4.1 Taste of machine learning

4.2 FUTURE OF MACHINE LEARNING

5. TECHNOLOGY LEARNT

5.1 Introduction to AI & ML

5.2 Definition of Machine Learning

5.3 Machine Learning Algorithm

6. TECHINQUES OF MACHINE LEARNING

6.1 SUPERSIVED LEARNING

6.2 TYPES OF SUPERVISED LEARNING

6.3 FEATURES ENGINEERING

6.4 INTRODUCTION TO DEEP LEARNING

7**.Database and SQL:**

* Relational databases (e.g., SQL Server, MySQL)
* NoSQL databases (e.g., MongoDB, Cassandra)

8. REFERENCES

9.CONCLUSION

**WEEKLY OVERVIEW OF INTERNSHIP ACTIVITES**

|  |  |  |  |
| --- | --- | --- | --- |
| S.NO | Date | Day | Name of the topic |
| 1 | 05/07/23 | Tuesday | Introduction of About the topics (Optional) |
| 2 | 06/07/23 | Wednesday | Background need & importance, introduction & installation, writing first program, I/o in python |
| 3 | 09/07/23 | Sunday | Keywords & Variables Data types - Numbers & Strings Operators in python Hands-on implementation |
| 4 | 10/07/23 | Monday | Indentation & scopes if, else & elif blocks introduction to loops for & While loops Break, continue statements |
| 5 | 13/07/23 | Wednesday | Data Structures |
| 6 | 17/07/23  To  18/07/23 | Sunday& Monday | Functions & Exception Handling |
| 7 | 19/07/23  To  19/07/23 | Thursday | Handling Pattern based problems |
| 8 | 20/07/23  To  20/07/23 | Sunday & Monday | map, filter, reduce, lambda zip, enumerate, sorted |
| 9 | 21/07/23 | Tuesday | File handling - I/o Numpy arrays Case studies |
| 10 | 22/07/23 | Wednesday | About the Data Science and Python on it. |
| 11 | 24/07/23  To  26/07/23 | Friday & Monday | Types of Data science and its process |
| 12 | 28/07/23 | Wednesday | Data Collection like collection, processing, cleaning, transforming etc.  Data Exploration and preprocessing |
| 13 | 30/07/23  To  30/07/23 | Friday | Exploratory Data Analysis (EDA): bar chart, line chart, box, pair chart etc |
| 14 | 01/08/23  To  03/08/23 | Wednesday & Friday | Data Visualization: matplotlib, tableau, seaborn etc  Text Analysis: NLP, text mining etc |
| 15 | 05/08/23  To  06/08/23 | Sunday & Monday | About Machine learning and its types and related with python |
| 16 | 09/08/23 | Wednesday | Types of Machine Learning and its algorithm and how it is used in data science. |
| 17 | 09/08/23 | Wednesday | Logistic Regression Evaluation metrics |
| 18 | 11/08/23  To  14/08/23 | Friday & Monday | KNN & SVM Decision tree &Ensemble, learning |
| 19 | 17/08/23 | Wednesday | Problem Statements Unsupervised Learning – I Wine Quality Prediction Diabetes prediction House price prediction |
| 20 | 19/08/23  To  19/08/23 | Friday | Titanic dataset Need of unsupervised learning K-means clustering Training k-means |
| 21 | 22/08/23  To  23/08/23 | Wednesday & Friday | Mean shift clustering K-means v/s Mean Shift clustering Industrial use cases of unsupervised learning Hyperparameter |
| 22 | 25/08/23  To  26/08/23 | Sunday & Monday | Fundamentals & Tools |
| 23 | 30/08/23 | Friday | Normalization & Transformation Cross Validations |
| 24 | 01/09/23 | Sunday | Tableau Introduction Marks Cards |
| 25 | 02/09/23 | Monday | Basics of charts |
| 26 | 03/09/23 | Tuesday | Different charts in Tableau |
| 27 | 04/09/23 | Wednesday | Database and SQL**:**  Relational databases (e.g., SQL Server, MySQL)  NoSQL databases (e.g., MongoDB, Cassandra |

1. **INTRODUCTION**

**Data Science and Machine Learning**

**Data science** is a multidisciplinary field that involves the use of scientific methods, processes, algorithms, and systems to extract insights and knowledge from structured and unstructured data. It combines expertise from various domains such as statistics, computer science, mathematics, and domain-specific knowledge to analyze and interpret complex data sets. It includes the data collection, data processing, Exploratory data analysis, Data models etc. It used statistical and different algorithm to predict the values.

Machine Learning (ML)  is a sub-category of [artificial intelligence](https://www.hpe.com/in/en/what-is/artificial-intelligence.html), that refers to the process by which computers develop pattern recognition, or the ability to continuously learn from and make predictions based on data, then make adjustments without being specifically programmed to do so.

Machine learning is incredibly complex and how it works varies depending on the task and the algorithm used to accomplish it. However, at its core, a machine learning model is a computer looking at data and identifying patterns, and then using those insights to better complete its assigned task. Any task that relies upon a set of data points or rules can be automated using machine learning, even those more complex tasks such as responding to customer service calls and reviewing resumes.

This are used to perform the different prediction and calculate the values as well as it is used in different AI tools like home appliance, Finnace, Marketing, Healthcare etc.

**Training Objective**: To enhance the knowledge of learner and perform well prediction to get the result.

**Student’s Work Assignment**: They have assigned Minor and Major project during training.

1. **COMPANY PROFILE**

YBI Foundation is a Delhi-based not-for-profit EdTech company that aims to enable the youth to grow in the world of emerging technologies. They offer a mix of online and offline approaches to bring new skills, education, technologies for students, academicians and practitioners. They believe in the learning anywhere and anytime approach to reach out to learners. The platform provides free online instructor-led classes for students to excel in data science, business analytics, machine learning, cloud computing and big data. They aim to focus on innovation, creativity, technology approach and keep themselves in sync with the present industry requirements. They endeavor to support learners to achieve the highest possible goals in their academics and professions.



Module-1: Introduction to python

Python is a dynamic, interpreted (bytecode-compiled) language. There are no type declarations of variables, parameters, functions, or methods in source code. This makes the code short and flexible, and you lose the compile-time type checking of the source code.

1.1 Features of Python

1) Relatively easy to learn, grasp and code.

2) Open Source

3) More built-in modules/packages

4) Portable

5) High level language

6) Interpreted language

7) Object Oriented.

8) Platform independent

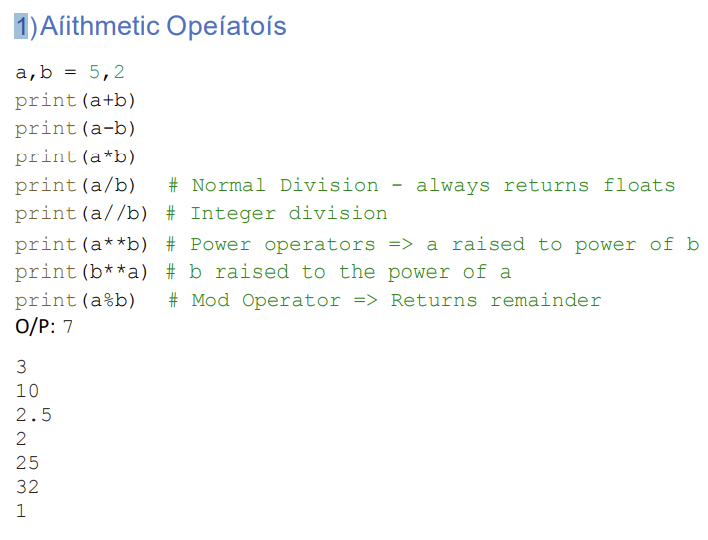
9) Supports GUI

10) Extensible.

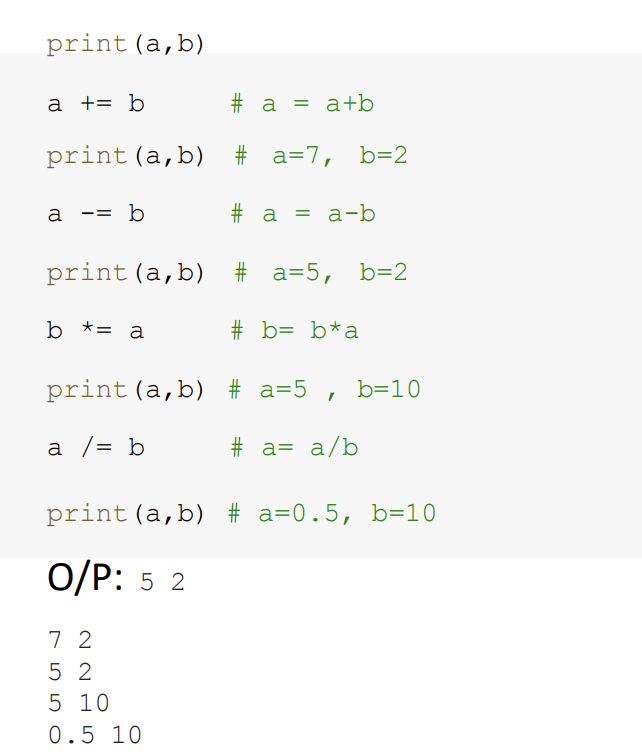
11) Dynamically typed

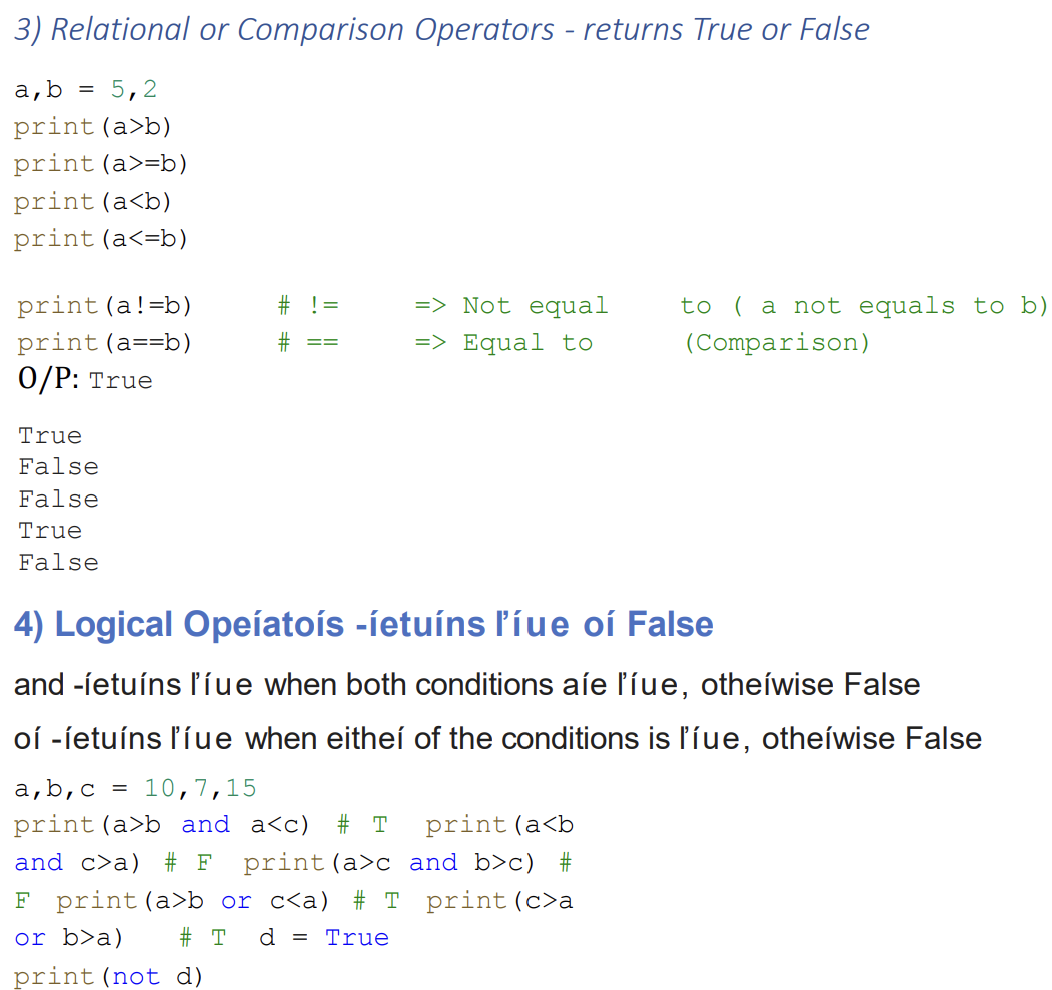
**Module-2: Python for Data Science**

**2.1 Understanding operator**

****

2)Assignment Operator





O/P: True

False

False

True

True

False



6)Identity Operator:(“is” and “is not”)

X=3

Y=3

Print(x is not y)

O/P:FLASE

7)Bitwise Operator:

Bitwise operators are used to performing bitwise calculations on integers

**2.2 Variables and Data Types**

Variable

1)A variable is a container which is used to store some value. The value can be of any data type 2) Multiple variables can be initialized in the same line by separating them by comma

3) Variables can be reinitialized, reassigned, redeclared

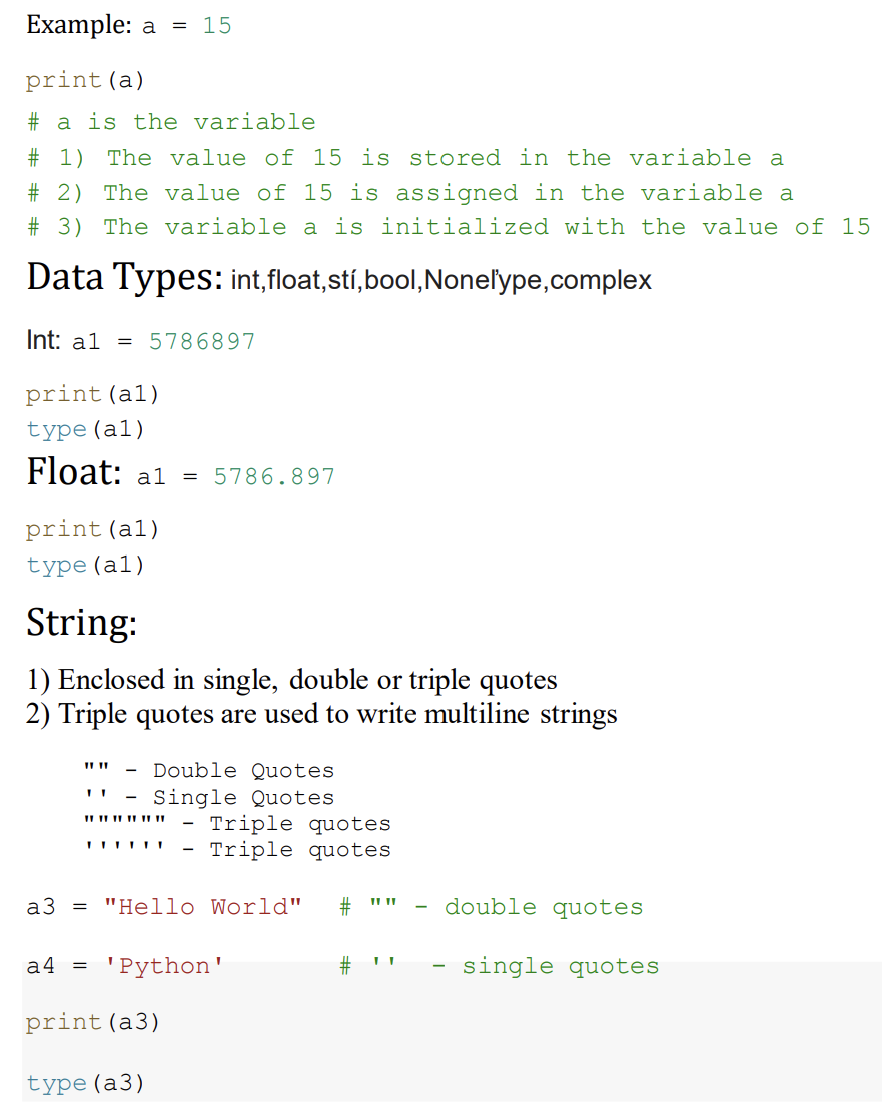
Rules of Naming a variable

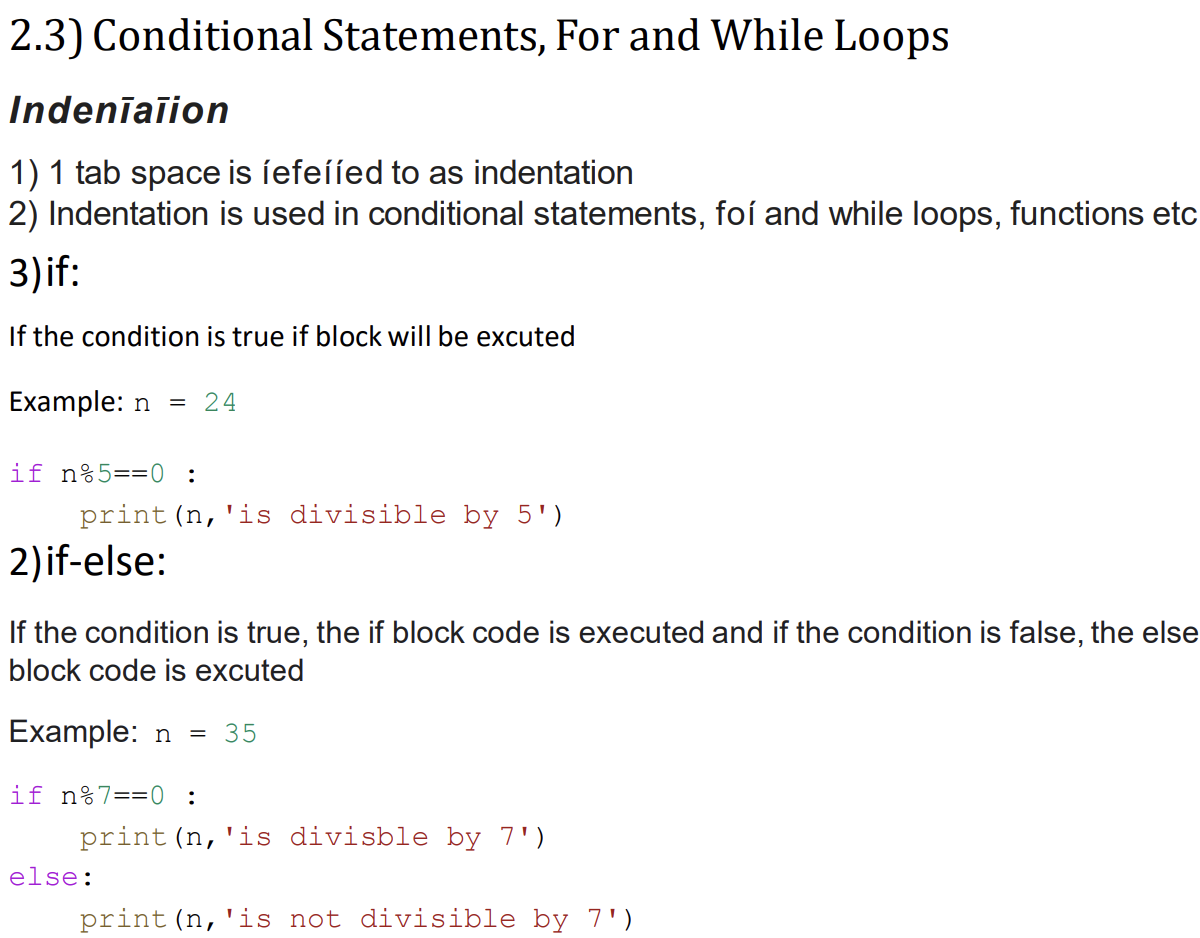
1) Start start with A-Z, a-z or underscore(\_)

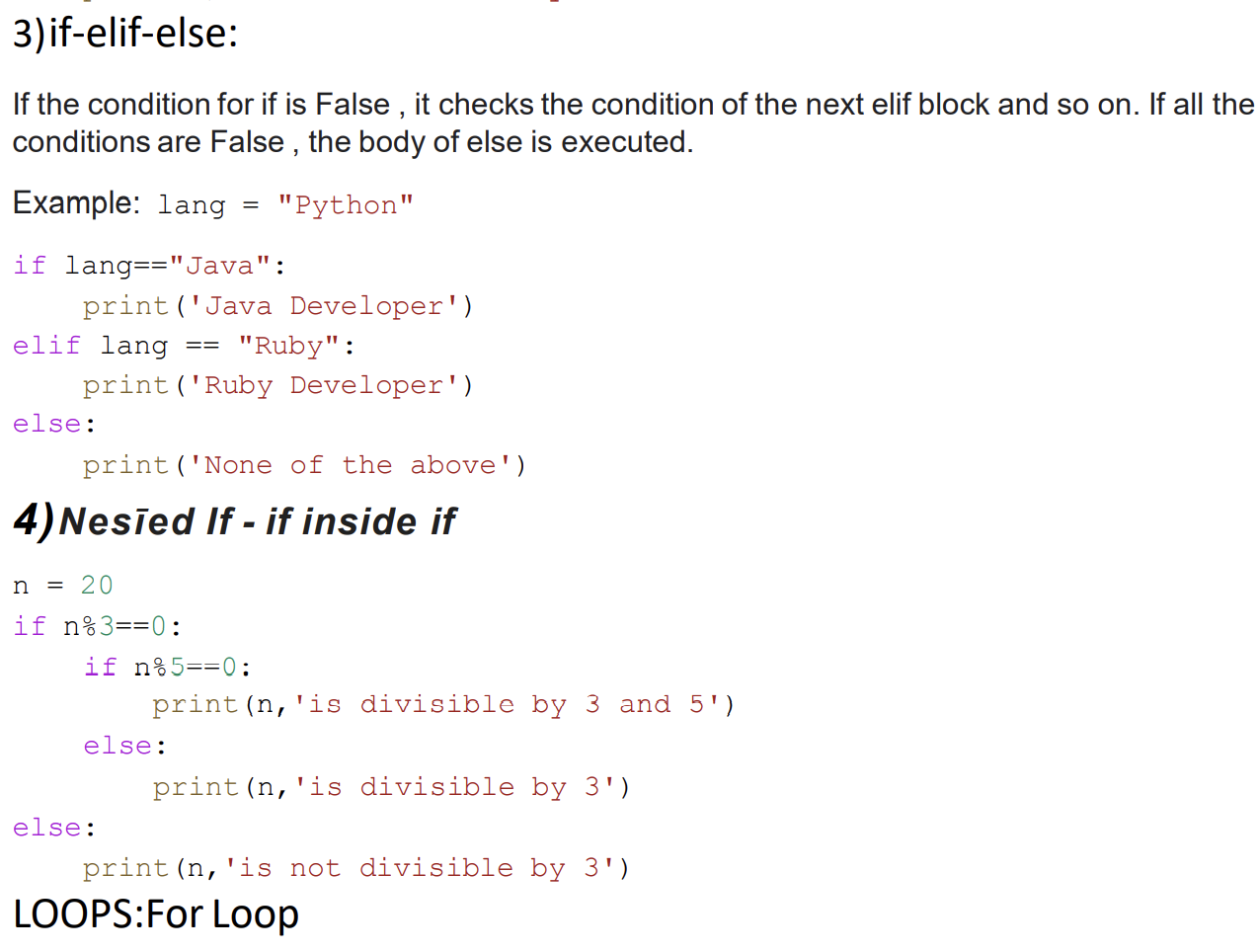
2) Should not start with number

3) Should not contain special characters

4) Should not contain python keywords







For Loops

In range (start, end, step) method 1) In ascending order, iteration proceeds from start till end-1, step size is positive

2) In descending order, iteration proceeds from start till end+1, step size is negative

3) Default step value = +1

4) range () works only on int

5) If only one value is passed in range function, it is considered as end value

6) Default start value is 0, if not mentioned

Example: for i in range (4): # start=0, end=4, step=1 print(i)

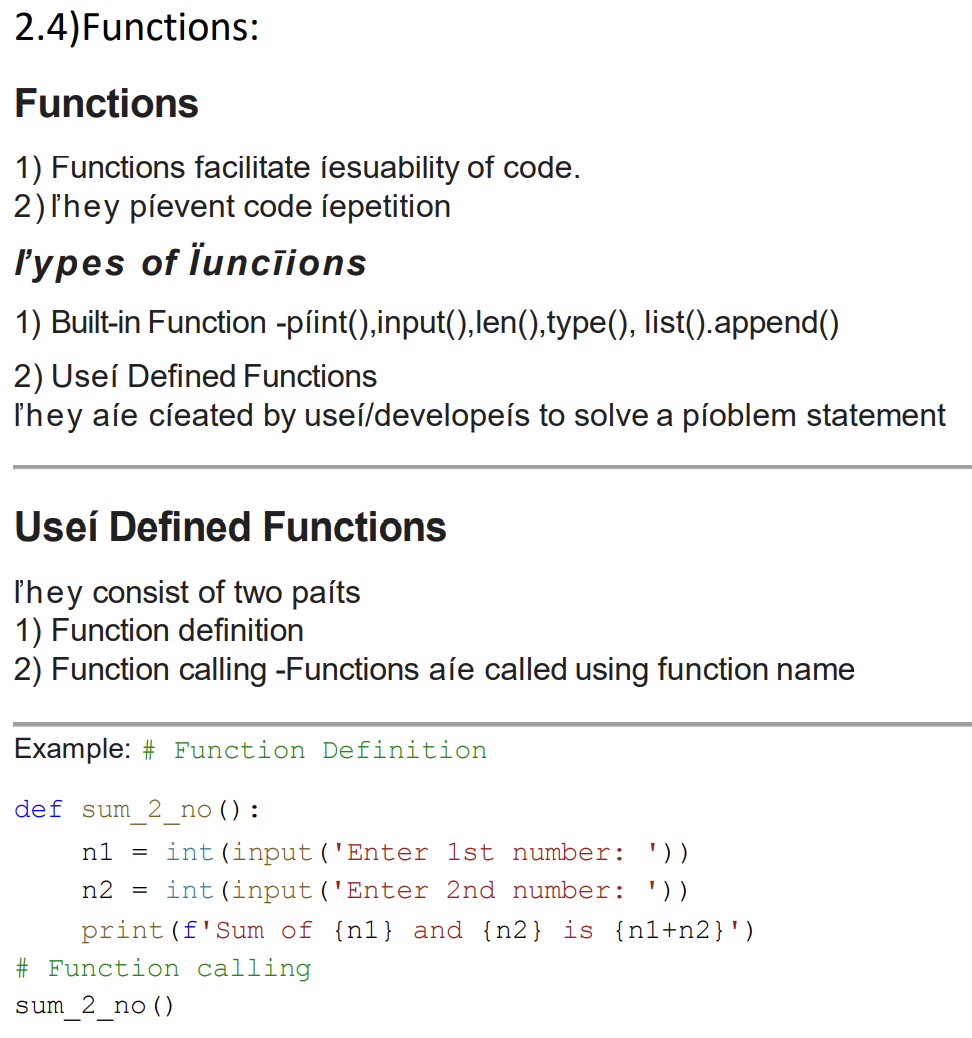
**While Loops**

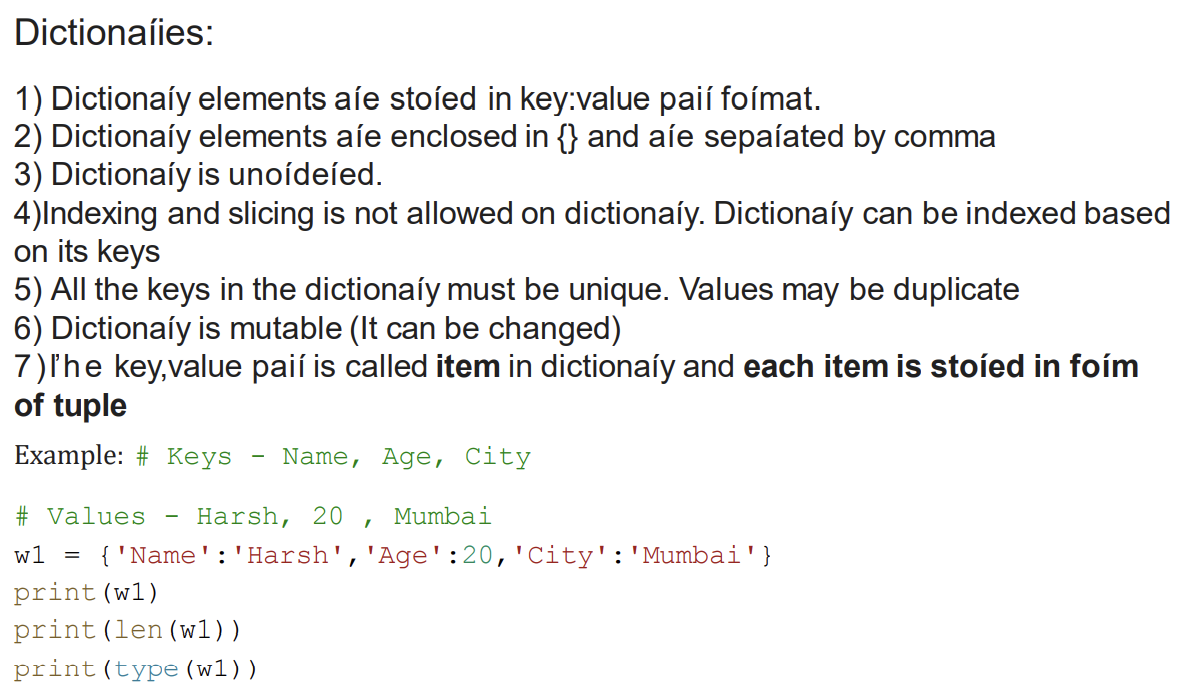
i=0 # start

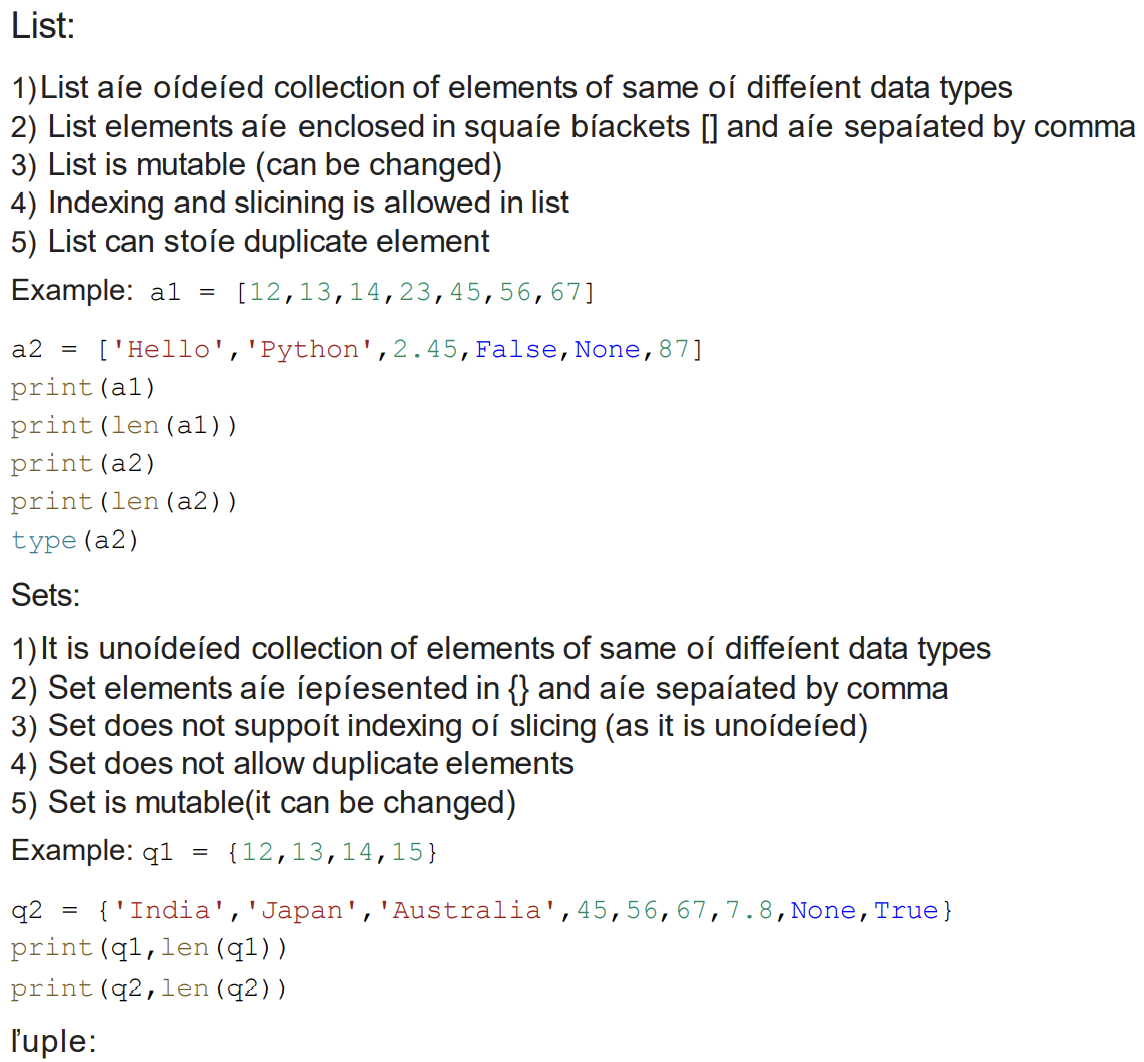
while i<5: #end

print (i,end=” “)

i+=1 #step







1)Tuple elements are represented in () and the elements are separated by comma.

2) Tuple is also an ordered collection of elements of same or different data types

3) Tuple is Immutable (can’t be changed)

4) Indexing and slicing is allowed

5) Duplicate elements are allowed

**Example:** w1 = (12,13,14,15,56,13)

**Data Science**

Data science is a multidisciplinary field that involves the use of scientific methods, processes, algorithms, and systems to extract insights and knowledge from structured and unstructured data. It combines expertise from various domains such as statistics, computer science, mathematics, and domain-specific knowledge to analyze and interpret complex data sets

**Data Collection:**

Data collection is a crucial step in the data science process, as it involves gathering relevant information from various sources to analyse and derive insights. Here are key aspects of data collection in data science:

1. **Define Objectives and Scope:**.
2. **Identify Data Sources:**

3.**Data Extraction:**

**4.Data Integration:**

**Exploratory Data Analysis (EDA**) is a critical phase in the data science process that involves visually and statistically exploring a dataset to gain a deeper understanding of its patterns, characteristics, and relationships. EDA helps in formulating hypotheses, identifying patterns, and guiding subsequent steps in the data analysis process. Here are key aspects of Exploratory Data Analysis:

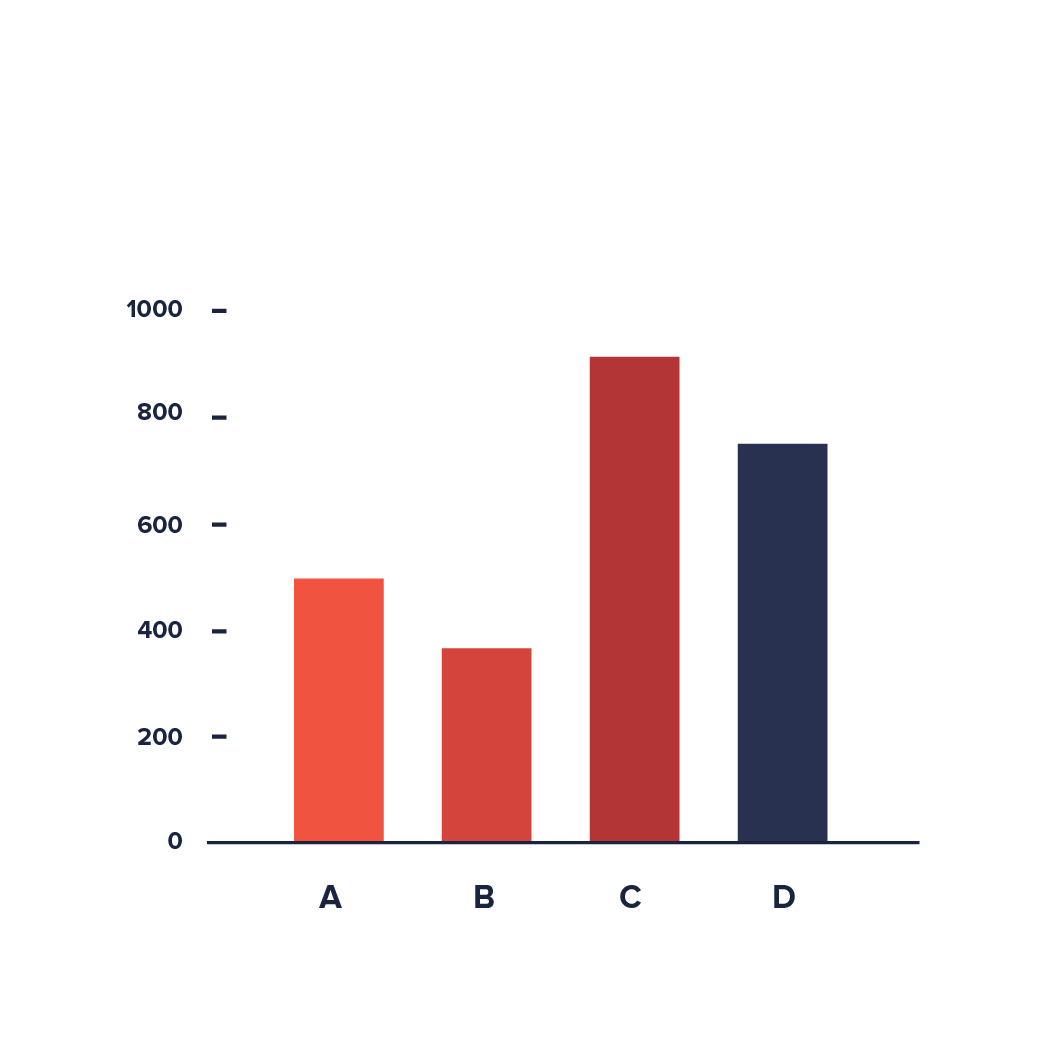
1. **Descriptive Statistics:**
2. **Data Visualization:**
3. **Univariate Analysis:**
4. **Bivariate Analysis:**
5. **Multivariate Analysis:**
6. **Handling Missing Data:**

**Data Visualization:**

Data visualization is the representation of data in graphical or visual formats, making it easier to understand patterns, trends, and insights. Effective data visualization is a crucial aspect of data analysis and communication.

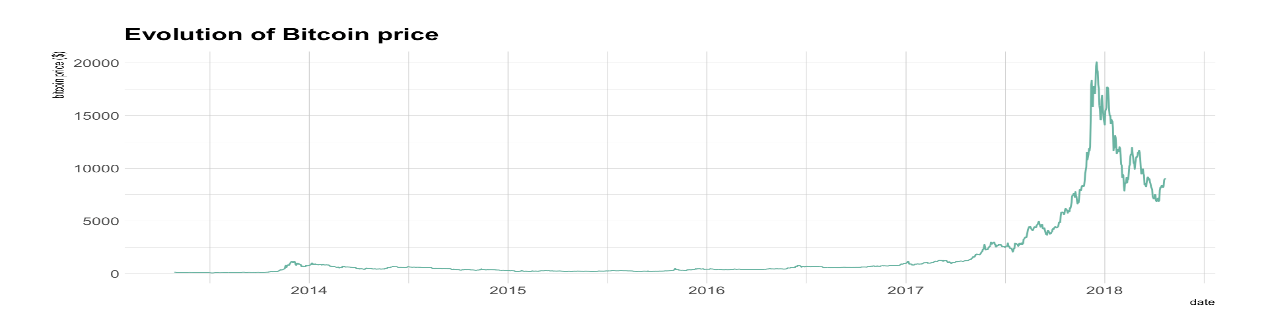
1.Bar Chart:

Bar charts are used to compare two or more categories of data. Each category is represented by a bar, and the length of the bar is proportional to the value of the category. Bar charts are a good choice for showing data that is categorical or discrete.

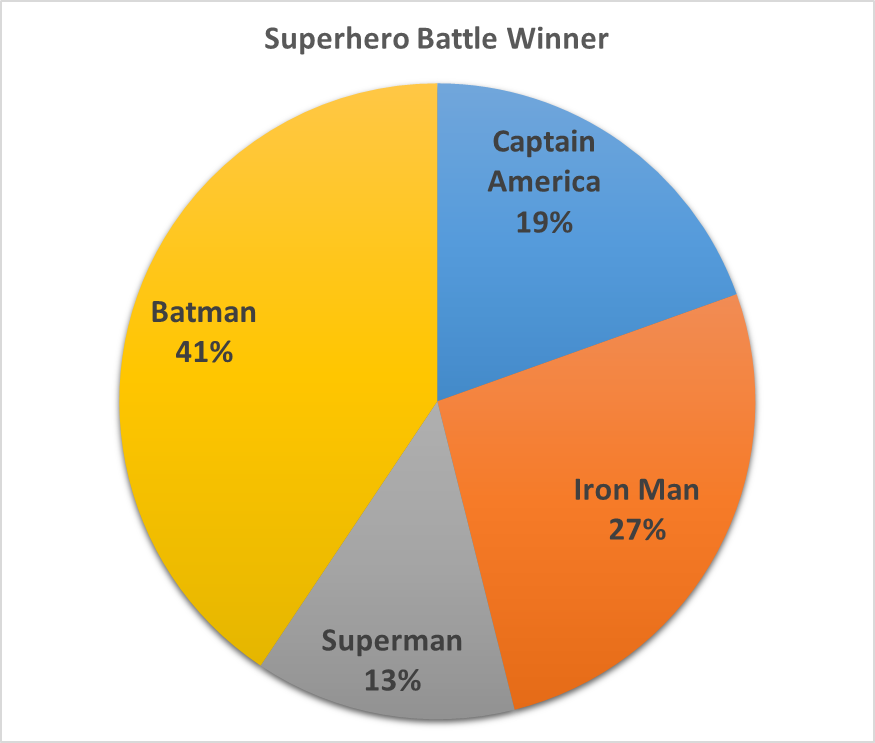
 Fig: bar chart

1. Line Chart:

Line graphs, or line charts, are a simple but effective staple for representing time-series data. They are visually similar to scatterplots but represent data points separated by time intervals with segments joined by a line



1. Pie Chart:

pie charts represent a single variable, broken down into percentages or proportions.

**Machine Learning**

Machine learning (ML) is a subfield of artificial intelligence (AI) that focuses on the development of algorithms and statistical models that enable computers to perform tasks without explicit programming. The primary goal of machine learning is to allow computers to learn from data and improve their performance over time. Here are key concepts and aspects of machine learning:

1. **Types of Machine Learning:**
   * **Supervised Learning:** The algorithm is trained on a labeled dataset, where each input is associated with the correct output. The model generalizes patterns from the training data to make predictions on new, unseen data.
   * **Unsupervised Learning:** The algorithm is given data without explicit labels. The model explores patterns and structures in the data, such as clustering similar data points or reducing dimensionality.
   * **Reinforcement Learning:** The algorithm learns by interacting with an environment. It receives feedback in the form of rewards or penalties, allowing it to learn optimal strategies to achieve a goal.

Supervised learning algorithms - **Regression and Classification.**

**Regression:**

1)It is a predictive modelling technique which investigates the relationship between dependent and independent variables(one or more)

2)Dependent variable is continuous in nature eg -Sales, Weight, Profit, Revenue, Price, Distance, Magnitude,Height, Weight etc

y = dependent variable/output

x = independent variable/input(s)

**Linear Regression:**

1)It is a regression model that estimates the relationship between one independent variable and one dependent variable using a straight line.

2) It has an equation of the form y = ax + b or y = mx + c

Where

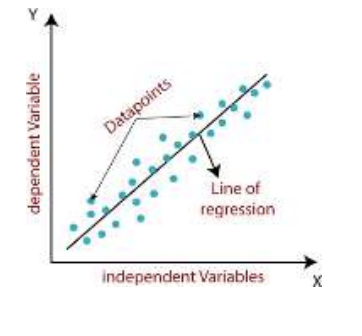
x = independent variable/ input feature/input attribute/input column

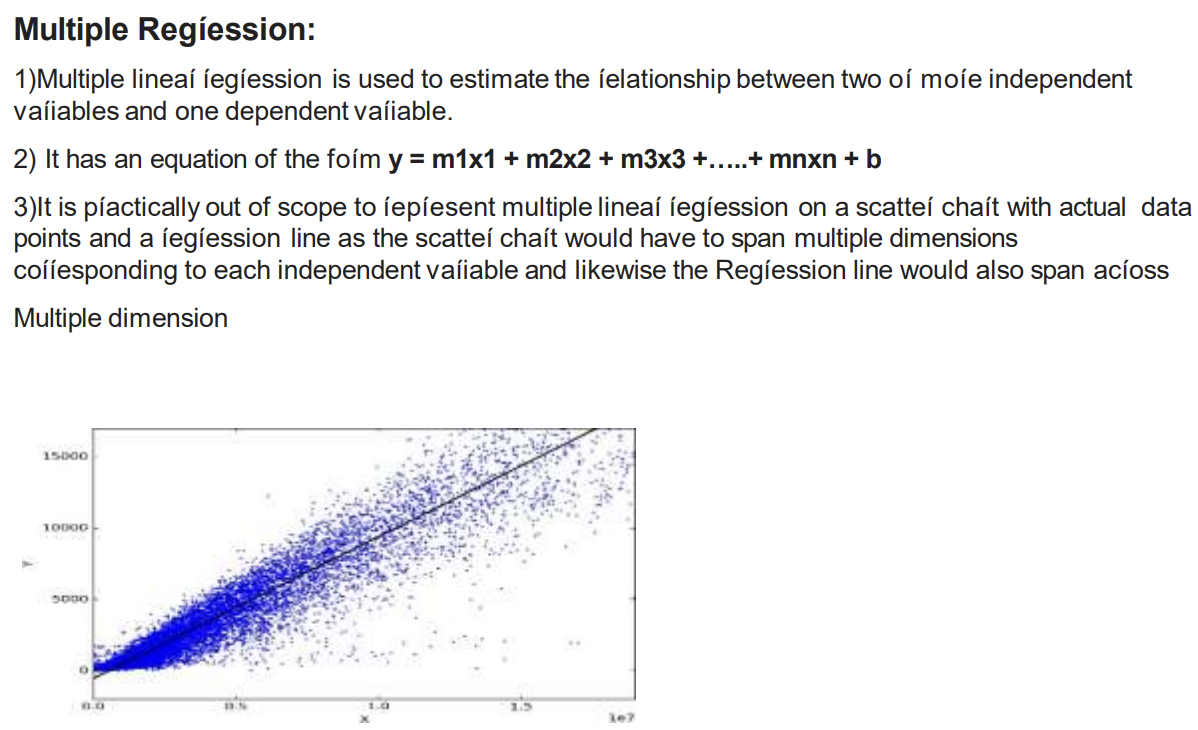
y = dependent variable / output feature/target attribute/ output column

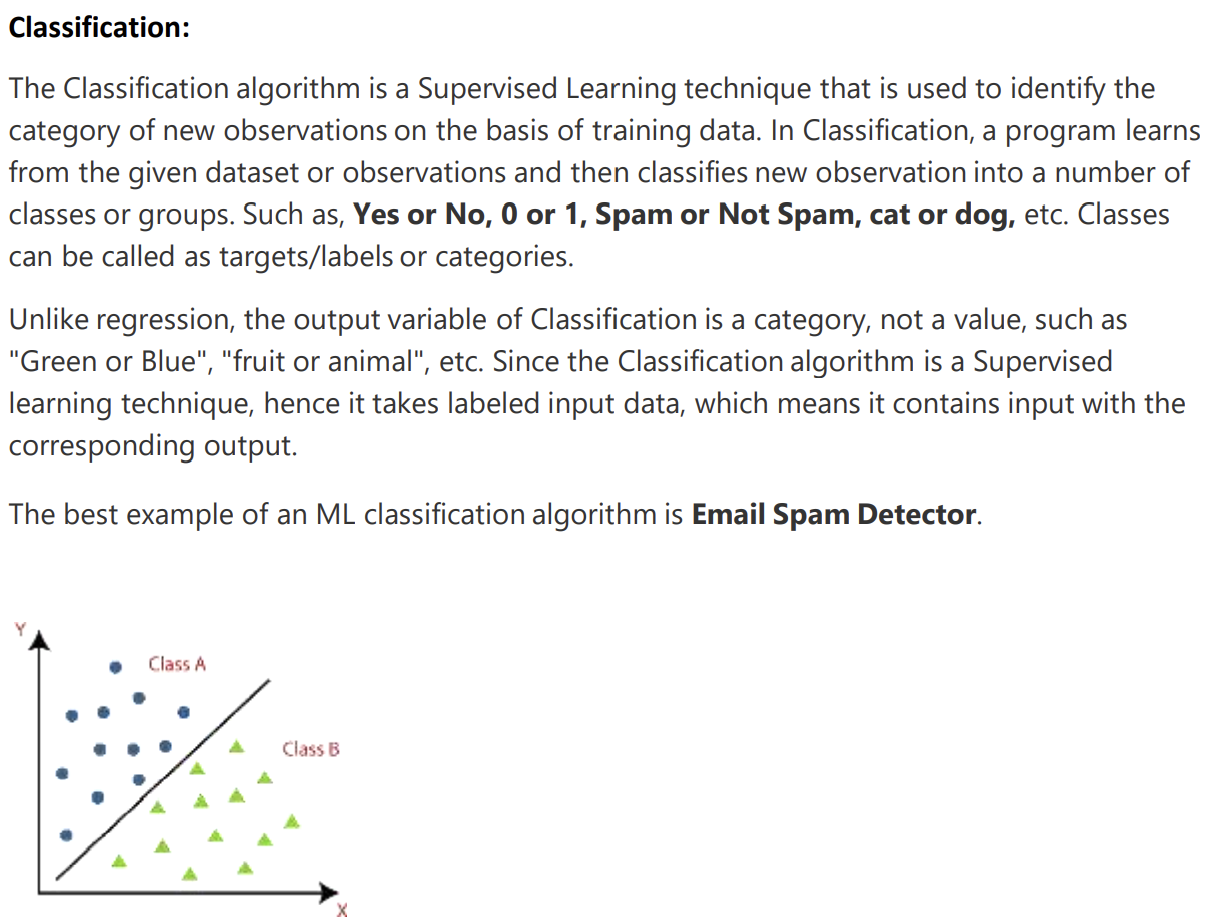
a/m = slope or coefficient or weight or how much we expect y to change as x changes

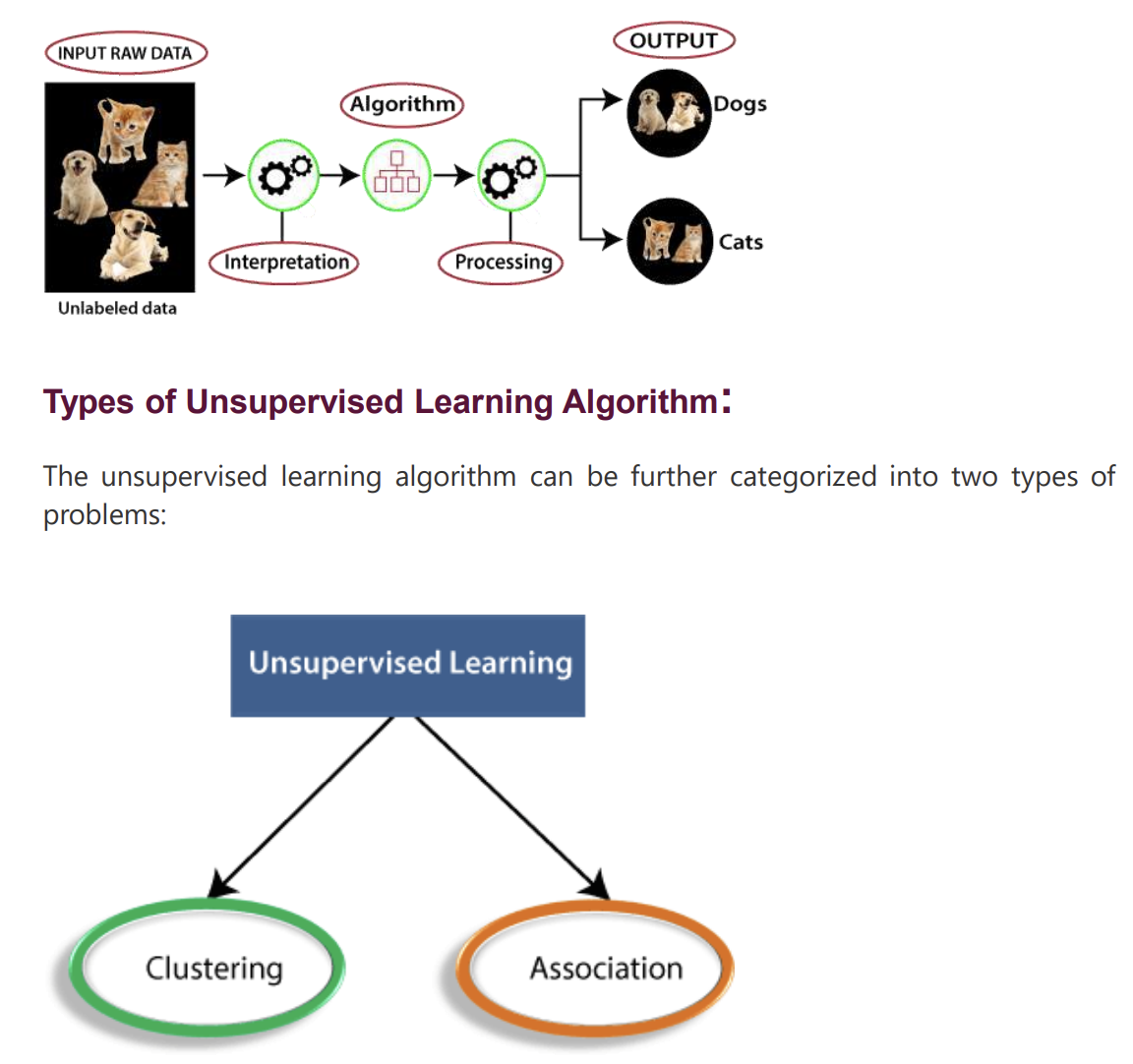
b/c = intercept / constant / bias

In this graph, x = Time spent Studying, y = Marks obtained. The orange dots are the corresponding data points. The blue line is the best fit line for Linear regression(y = mx +c)







****

**Association:**

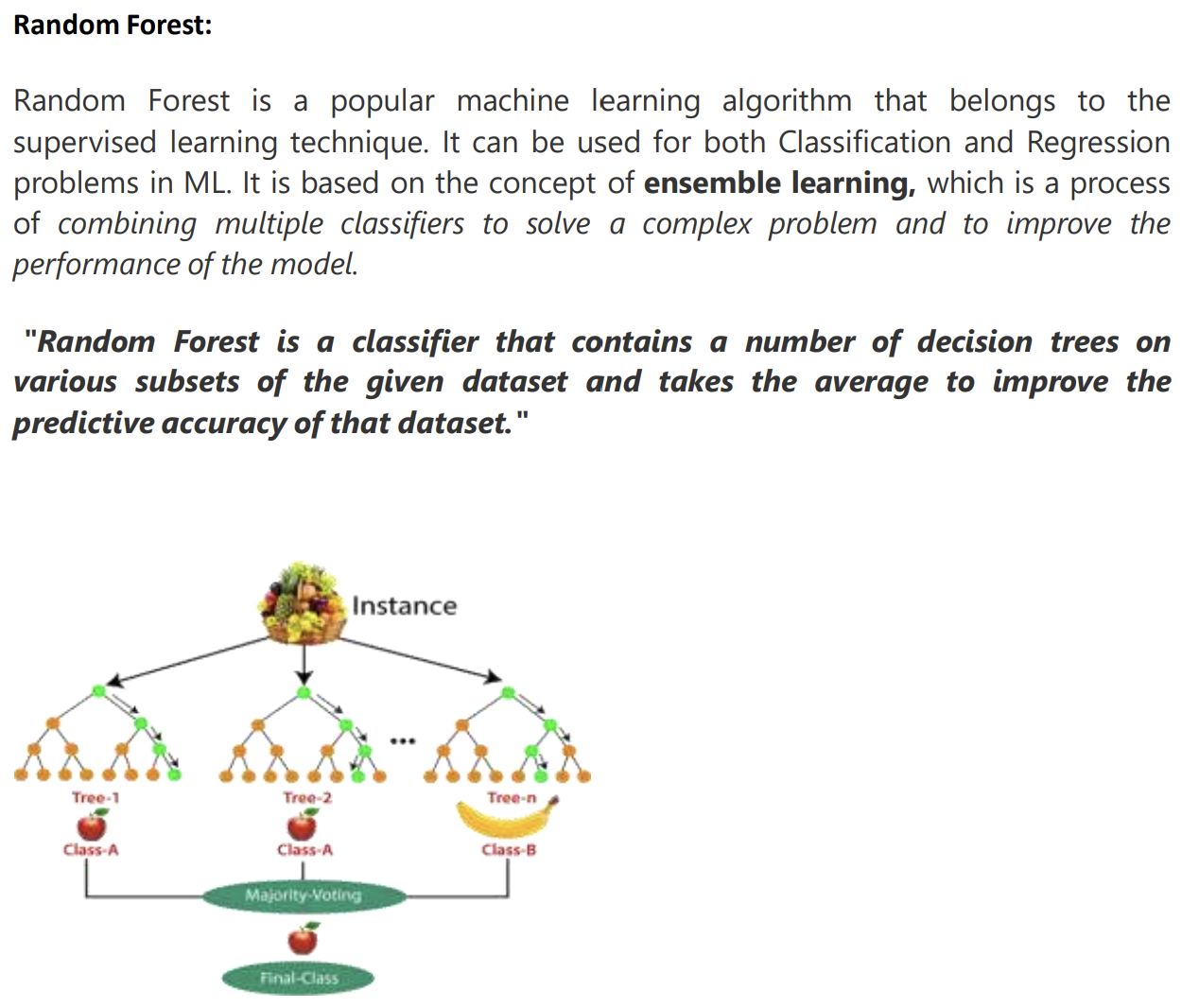
**Definition:** Association in data mining refers to discovering relationships or patterns among variables in large datasets. It focuses on finding associations or connections between items, events, or attributes.

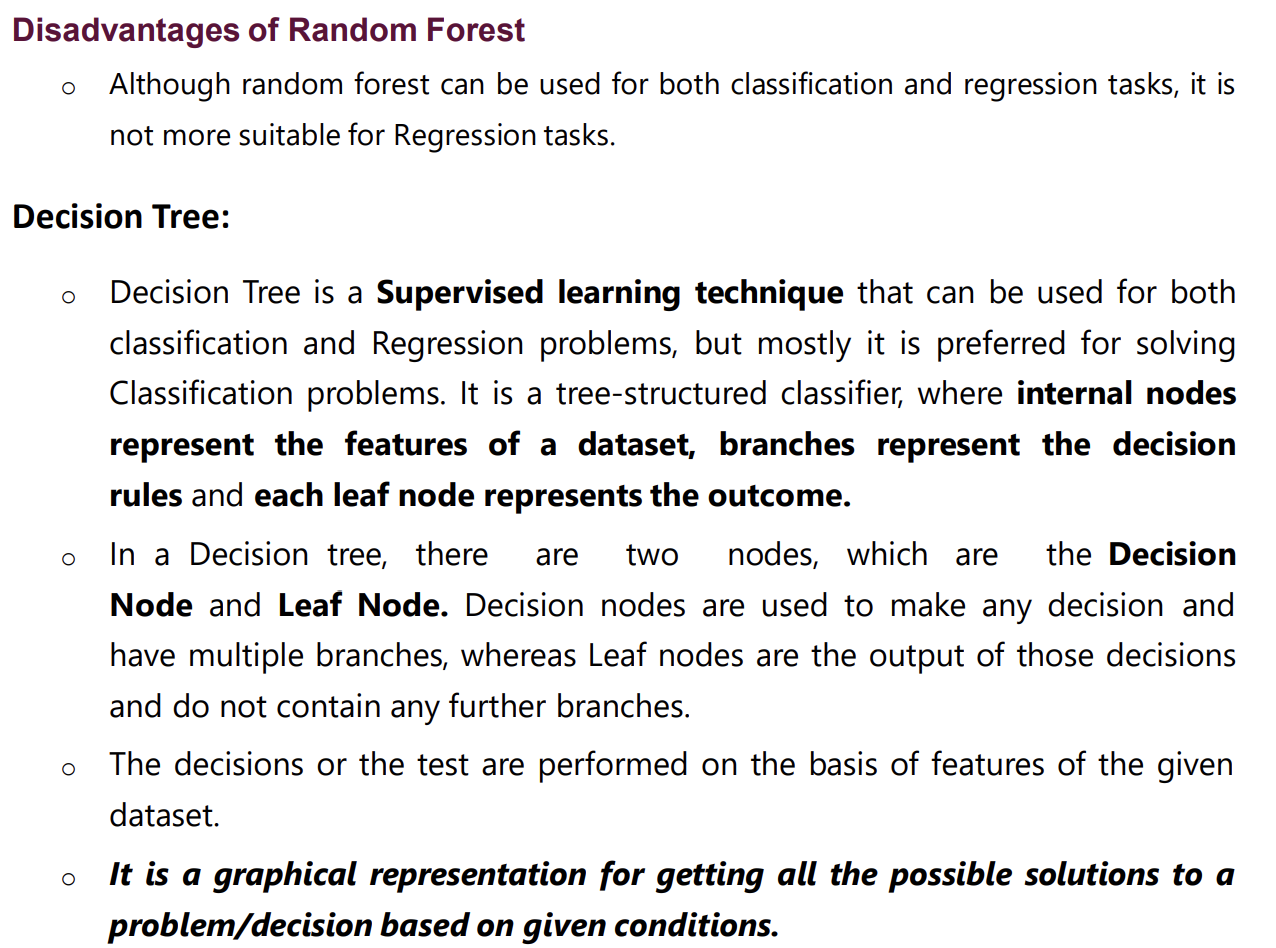
**Key Points:**

1. **Association Rule Mining:**
   * **Support:** Measures the frequency of a set of items in the dataset.
   * **Confidence:** Measures the likelihood that an association rule holds true.
   * **Lift:** Measures the ratio of the observed support to the expected support.
2. **Types of Association Rule Algorithms:**
   * **A-priori Algorithm:** Generates frequent itemset and association rules.
   * **FP-Growth (Frequent Pattern Growth):** Utilizes a tree structure to mine frequent itemset efficiently.

**Clustering:**

Clustering is a method of grouping the objects into clusters such that objects with most similarities remains into a group and has less or no similarities with the objects of another group. Cluster analysis finds the commonalities between the data objects and categorizes them as per the presence and absence of those commonalities.





It may have an overfitting issue, which can be resolved using the Random Forest algorithm.

For more class labels, the computational complexity of the decision tree may increase.

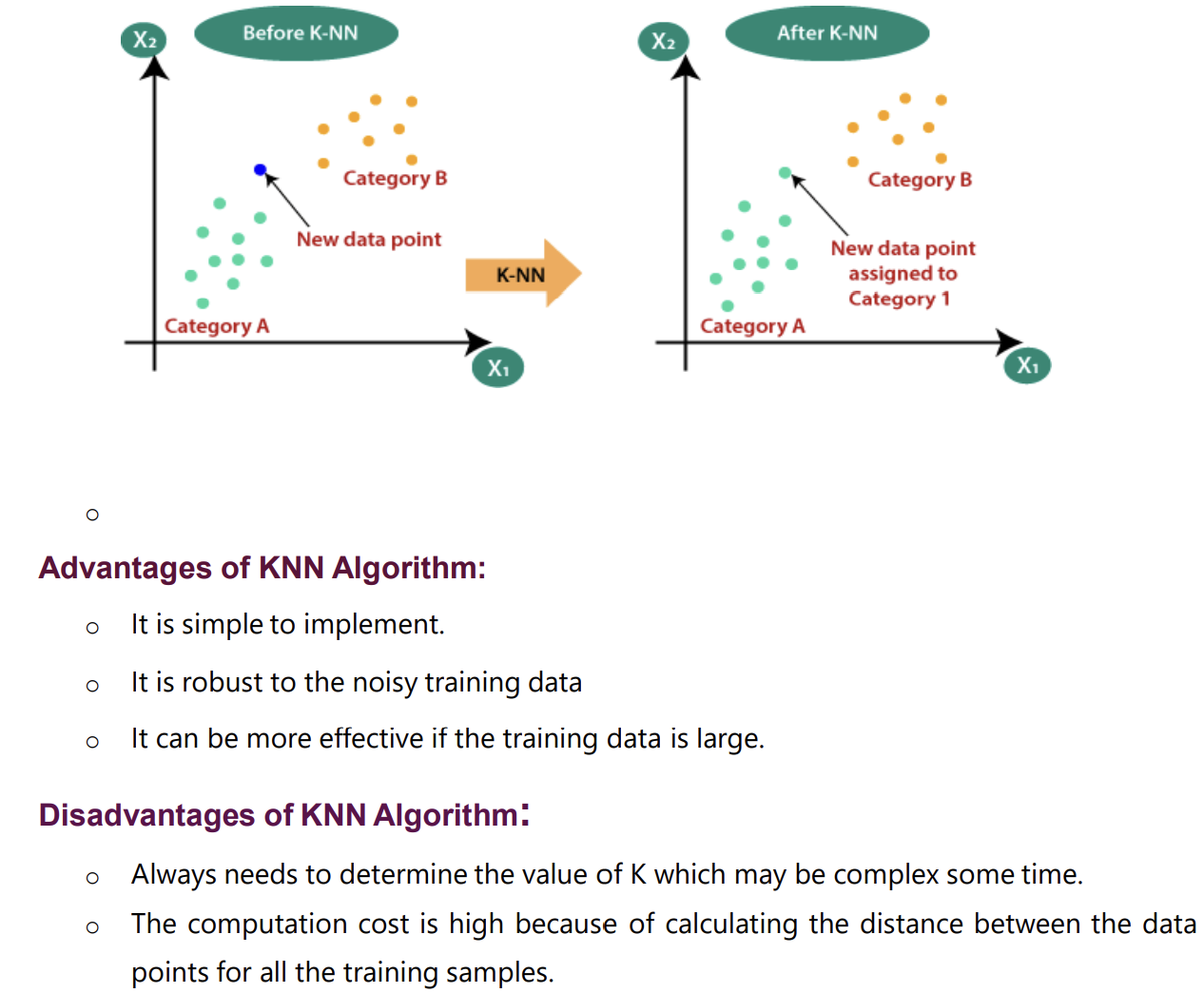
**K-Nearest Neighbor (KNN) Algorithm**

> K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.

> K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories.

> K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm.

> K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems.



**Different Libarires used in Data Science and Machine Learning**

**Matplotlib:**

**Description:**

**Purpose:** Matplotlib is a powerful 2D plotting library for Python. It enables the creation of a wide variety of static, animated, and interactive plots.

* **Features:**
  + Supports line plots, scatter plots, bar plots, histograms, and more.
  + Highly customizable plot elements.
  + Integrates well with Jupiter notebooks.
* **Use Cases:** Matplotlib is widely used in the Python ecosystem for creating static visualizations in various domains, including data science and scientific research.

**Tableau:**

**Description:**

**Purpose:** Tableau is a leading data visualization tool that provides a user-friendly interface for creating interactive and shareable dashboards. It allows users to connect to diverse data sources and explore insights visually.

* **Features:**
  + Drag-and-drop interface for creating dashboards without coding.
  + Offers a variety of pre-built visualizations.
  + Connects to various data sources, facilitating easy data integration.
* **Use Cases:** Tableau is employed for business intelligence, data analysis, and creating interactive dashboards for decision-making.

**Seaborn:**

**Description:**

* **Purpose:** Seaborn is a statistical data visualization library built on top of Matplotlib. It provides a high-level interface for creating visually appealing statistical graphics with less code.
* **Features:**
  + Attractive default styles and colour palettes.
  + Specialized functions for visualizing statistical relationships in data.
  + Integration with Pandas data structures.

**Database:**

A database is an organized collection of structured information or data that is stored electronically in a computer system. Databases are designed to efficiently manage, store, and retrieve data. They are a critical component in various applications, enabling the storage and manipulation of vast amounts of information. Key concepts related to databases include:

1. **Database Management System (DBMS):**
   * A software system that provides an interface to interact with the database.
   * Manages the storage, retrieval, and organization of data.
2. **Relational Database:**
   * Organizes data into tables with rows and columns.
   * Uses relationships between tables to establish connections and maintain data integrity.
3. **NoSQL Database:**
   * A type of database that does not strictly adhere to the traditional relational model.
   * Suited for handling unstructured or semi-structured data and provides flexible schema design.
4. **Key-Value Store:**
   * A NoSQL database model where each data item is stored as a key-value pair.
5. **Document Store:**
   * A NoSQL database model that stores data in semi-structured documents (e.g., JSON or XML).
6. **Graph Database:**
   * A database designed for handling data with complex relationships, using a graph structure.
7. **ACID Properties:**
   * A set of properties (Atomicity, Consistency, Isolation, Durability) that guarantee the reliability of database transactions.

**SQL (Structured Query Language):**

SQL is a specialized programming language designed for managing and manipulating relational databases. It is used to perform various operations on databases, such as querying, updating, inserting, and deleting data. Key concepts and statements in SQL include:

1. **SELECT Statement:**
   * Retrieves data from one or more tables.

sqlCopy code

SELECT column1, column2 FROM table WHERE condition;

1. **INSERT Statement:**
   * Adds new records to a table.

sqlCopy code

INSERT INTO table (column1, column2) VALUES (value1, value2);

1. **UPDATE Statement:**
   * Modifies existing records in a table.

sqlCopy code

UPDATE table SET column1 = value1 WHERE condition;

1. **DELETE Statement:**
   * Removes records from a table.

sqlCopy code

DELETE FROM table WHERE condition;

1. **CREATE TABLE Statement:**
   * Defines a new table with its structure.

sqlCopy code

CREATE TABLE table ( column1 datatype1, column2 datatype2, ... );

1. **JOIN Operation:**
   * Combines rows from two or more tables based on a related column.

sqlCopy code

SELECT \* FROM table1 INNER JOIN table2 ON table1.column = table2.column;

1. **INDEX:**
   * Improves the speed of data retrieval operations on a database table.

sqlCopy code

CREATE INDEX index\_name ON table (column);

1. **Normalization:**
   * The process of organizing data in a database to reduce redundancy and improve data integrity.

**PROJECT WORK**

During the period of (6 week) Internship they have assigned two projects to do:

1. Minor Project
2. Major Project

**Minor Project:**

In Minor project, I have done “**live data analytics**” and “**Real Time object Detection”.**

Live data analytics refers to the real-time analysis and interpretation of data as it is generated or received, allowing organizations to make informed decisions and take immediate actions based on current information.

**Machine Learning Libraries:**

**Scikit-learn:** A popular machine learning library in Python that provides simple and efficient tools for data analysis and modelling.

**TensorFlow:** An open-source machine learning library developed by the Google Brain team that is widely used for building and training deep learning models.

**Py Torch:** Another popular deep learning library that is known for its dynamic computational graph, making it easier to work with dynamic and variable-sized inputs.

**Time Series Analysis:**

**Stats models:** A library for estimating and testing statistical models, including time series analysis.

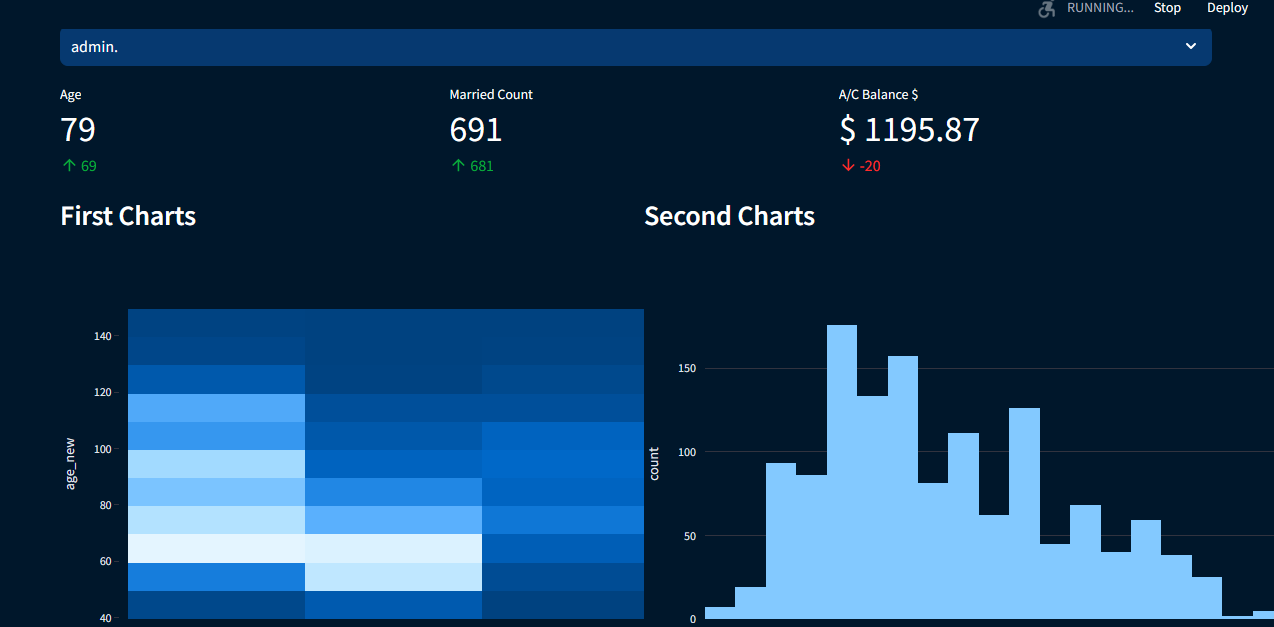
**Real-Time Data Processing:**

**Apache Kafka:** A distributed streaming platform that is often used for building real-time data pipelines and streaming applications.

**Apache Flink:** A stream processing framework for big data processing and analytics.

**Web Frameworks for Real-Time Predictions:**

* + **Flask:** A lightweight web framework for Python that is commonly used for building APIs and serving machine learning models.

fig: Output od live data analytics

**“Real Time Object Detection”**

**Real-time object detection** is a computer vision task that involves identifying and locating objects in a video or image stream as it occurs, typically at a rate that provides instantaneous results. This technology has diverse applications, ranging from video surveillance and autonomous vehicles to augmented reality and interactive systems

**TensorFlow Object Detection API:**

**Description:** TensorFlow provides a dedicated Object Detection API that includes pre-trained models and tools for building custom models. Models like SSD (Single Shot MultiBox Detector) and Faster R-CNN can be used for real-time object detection. import cv2

import utlis

###################################

webcam = True

path = '1.jpg'

cap = cv2.VideoCapture(0)

cap.set(10,160)

cap.set(3,1920)

cap.set(4,1080)

scale = 3

wP = 210 \*scale

hP= 297 \*scale

###################################

while True:

    if webcam:success,img = cap.read()

    else: img = cv2.imread(path)

    imgContours , conts = utlis.getContours(img,minArea=50000,filter=4)

    if len(conts) != 0:

        biggest = conts[0][2]

        #print(biggest)

        imgWarp = utlis.warpImg(img, biggest, wP,hP)

        imgContours2, conts2 = utlis.getContours(imgWarp,

                                                 minArea=2000, filter=4,

                                                 cThr=[50,50],draw = False)

        if len(conts) != 0:

            for obj in conts2:

                cv2.polylines(imgContours2,[obj[2]],True,(0,255,0),2)

                nPoints = utlis.reorder(obj[2])

                nW = round((utlis.findDis(nPoints[0][0]//scale,nPoints[1][0]//scale)/10),1)

                nH = round((utlis.findDis(nPoints[0][0]//scale,nPoints[2][0]//scale)/10),1)

                cv2.arrowedLine(imgContours2, (nPoints[0][0][0], nPoints[0][0][1]), (nPoints[1][0][0], nPoints[1][0][1]),

                                (255, 0, 255), 3, 8, 0, 0.05)

                cv2.arrowedLine(imgContours2, (nPoints[0][0][0], nPoints[0][0][1]), (nPoints[2][0][0], nPoints[2][0][1]),

                                (255, 0, 255), 3, 8, 0, 0.05)

                x, y, w, h = obj[3]

                cv2.putText(imgContours2, '{}cm'.format(nW), (x + 30, y - 10), cv2.FONT\_HERSHEY\_COMPLEX\_SMALL, 1.5,

                            (255, 0, 255), 2)

                cv2.putText(imgContours2, '{}cm'.format(nH), (x - 70, y + h // 2), cv2.FONT\_HERSHEY\_COMPLEX\_SMALL, 1.5,

                            (255, 0, 255), 2)

        cv2.imshow('A4', imgContours2)

    img = cv2.resize(img,(0,0),None,0.5,0.5)

    cv2.imshow('Original',img)

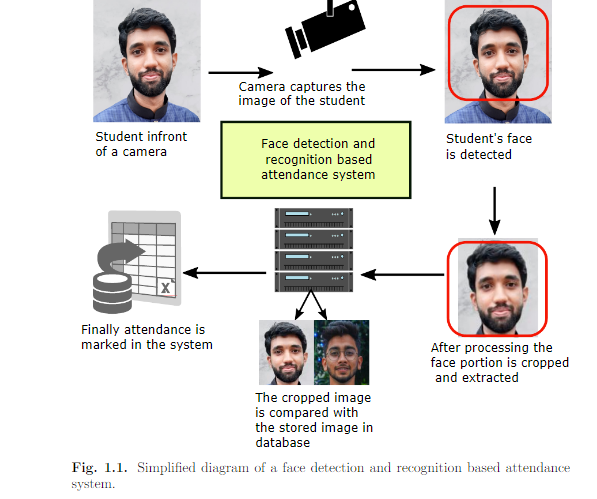
    cv2.waitKey(1)

**2.Major Project**

**On major project, I have done on topic “Online Attendance Using Face Recognition In Python”**

Online attendance is time saving and effective.Online attendance management system enables school administrators to record, manage & compile daily student attendance data. Along with student attendance, this software also allows teachers to generate 100% accurate student attendance reports.

* Face Recognition is capable of detecting tracking, identifying or verifying human faces from an image or video captured using digital camera
* This proposed system aims to develop an automated system that records the student’s attendance by using facial recoginition technology.

****It works like given below:

We used following libraries,

* Cv2
* Os
* Flask
* Datetime
* Numpy
* Sklearn.neighbors
* Pandas
* Joblib

I have used visual studio code to perform program. It capture photos of students and take a attendance.

**Input output dataset/screenshots**

import cv2

import os

from flask import Flask,request,render\_template

from datetime import date

from datetime import datetime

import numpy as np

from sklearn.neighbors import KNeighborsClassifier

import pandas as pd

import joblib

#### Defining Flask App

app = Flask(\_\_name\_\_)

#### Saving Date today in 2 different formatspip i

datetoday = date.today().strftime("%m\_%d\_%y")

datetoday2 = date.today().strftime("%d-%B-%Y")

#### Initializing VideoCapture object to access WebCam

face\_detector = cv2.CascadeClassifier('static/haarcascade\_frontalface\_default.xml')

cap = cv2.VideoCapture(0)

#### If these directories don't exist, create them

if not os.path.isdir('Attendance'):

    os.makedirs('Attendance')

if not os.path.isdir('static/faces'):

    os.makedirs('static/faces')

if f'Attendance-{datetoday}.csv' not in os.listdir('Attendance'):

    with open(f'Attendance/Attendance-{datetoday}.csv','w') as f:

        f.write('Name,Roll,Time')

#### get a number of total registered users

def totalreg():

    return len(os.listdir('static/faces'))

#### extract the face from an image

def extract\_faces(img):

    gray = cv2.cvtColor(img, cv2.COLOR\_BGR2GRAY)

    face\_points = face\_detector.detectMultiScale(gray, 1.3, 5)

    return face\_points

#### Identify face using ML model

def identify\_face(facearray):

    model = joblib.load('static/face\_recognition\_model.pkl')

    return model.predict(facearray)

#### A function which trains the model on all the faces available in faces folder

def train\_model():

    faces = []

    labels = []

    userlist = os.listdir('static/faces')

    for user in userlist:

        for imgname in os.listdir(f'static/faces/{user}'):

            img = cv2.imread(f'static/faces/{user}/{imgname}')

            resized\_face = cv2.resize(img, (50, 50))

            faces.append(resized\_face.ravel())

            labels.append(user)

    faces = np.array(faces)

    knn = KNeighborsClassifier(n\_neighbors=5)

    knn.fit(faces,labels)

    joblib.dump(knn,'static/face\_recognition\_model.pkl')

#### Extract info from today's attendance file in attendance folder

def extract\_attendance():

    df = pd.read\_csv(f'Attendance/Attendance-{datetoday}.csv')

    names = df['Name']

    rolls = df['Roll']

    times = df['Time']

    l = len(df)

    return names,rolls,times,l

#### Add Attendance of a specific user

def add\_attendance(name):

    username = name.split('\_')[0]

    userid = name.split('\_')[1]

    current\_time = datetime.now().strftime("%H:%M:%S")

    df = pd.read\_csv(f'Attendance/Attendance-{datetoday}.csv')

    if int(userid) not in list(df['Roll']):

        with open(f'Attendance/Attendance-{datetoday}.csv','a') as f:

            f.write(f'\n{username},{userid},{current\_time}')

################## ROUTING FUNCTIONS #########################

#### Our main page

@app.route('/')

def home():

    names,rolls,times,l = extract\_attendance()

    return render\_template('home.html',names=names,rolls=rolls,times=times,l=l,totalreg=totalreg(),datetoday2=datetoday2)

#### This function will run when we click on Take Attendance Button

@app.route('/start',methods=['GET'])

def start():

    if 'face\_recognition\_model.pkl' not in os.listdir('static'):

        return render\_template('home.html',totalreg=totalreg(),datetoday2=datetoday2,mess='There is no trained model in the static folder. Please add a new face to continue.')

    cap = cv2.VideoCapture(0)

    ret = True

    while ret:

        ret,frame = cap.read()

        if extract\_faces(frame)!=():

            (x,y,w,h) = extract\_faces(frame)[0]

            cv2.rectangle(frame,(x, y), (x+w, y+h), (255, 0, 20), 2)

            face = cv2.resize(frame[y:y+h,x:x+w], (50, 50))

            identified\_person = identify\_face(face.reshape(1,-1))[0]

            add\_attendance(identified\_person)

            cv2.putText(frame,f'{identified\_person}',(30,30),cv2.FONT\_HERSHEY\_SIMPLEX,1,(255, 0, 20),2,cv2.LINE\_AA)

        cv2.imshow('Attendance',frame)

        if cv2.waitKey(1)==27:

            break

    cap.release()

    cv2.destroyAllWindows()

    names,rolls,times,l = extract\_attendance()

    return render\_template('home.html',names=names,rolls=rolls,times=times,l=l,totalreg=totalreg(),datetoday2=datetoday2)

#### This function will run when we add a new user

@app.route('/add',methods=['GET','POST'])

def add():

    newusername = request.form['newusername']

    newuserid = request.form['newuserid']

    userimagefolder = 'static/faces/'+newusername+'\_'+str(newuserid)

    if not os.path.isdir(userimagefolder):

        os.makedirs(userimagefolder)

    cap = cv2.VideoCapture(0)

    i,j = 0,0

    while 1:

        \_,frame = cap.read()

        faces = extract\_faces(frame)

        for (x,y,w,h) in faces:

            cv2.rectangle(frame,(x, y), (x+w, y+h), (255, 0, 20), 2)

            cv2.putText(frame,f'Images Captured: {i}/50',(30,30),cv2.FONT\_HERSHEY\_SIMPLEX,1,(255, 0, 20),2,cv2.LINE\_AA)

            if j%10==0:

                name = newusername+'\_'+str(i)+'.jpg'

                cv2.imwrite(userimagefolder+'/'+name,frame[y:y+h,x:x+w])

                i+=1

            j+=1

        if j==500:

            break

        cv2.imshow('Adding new User',frame)

        if cv2.waitKey(1)==27:

            break

    cap.release()

    cv2.destroyAllWindows()

    print('Training Model')

    train\_model()

    names,rolls,times,l = extract\_attendance()

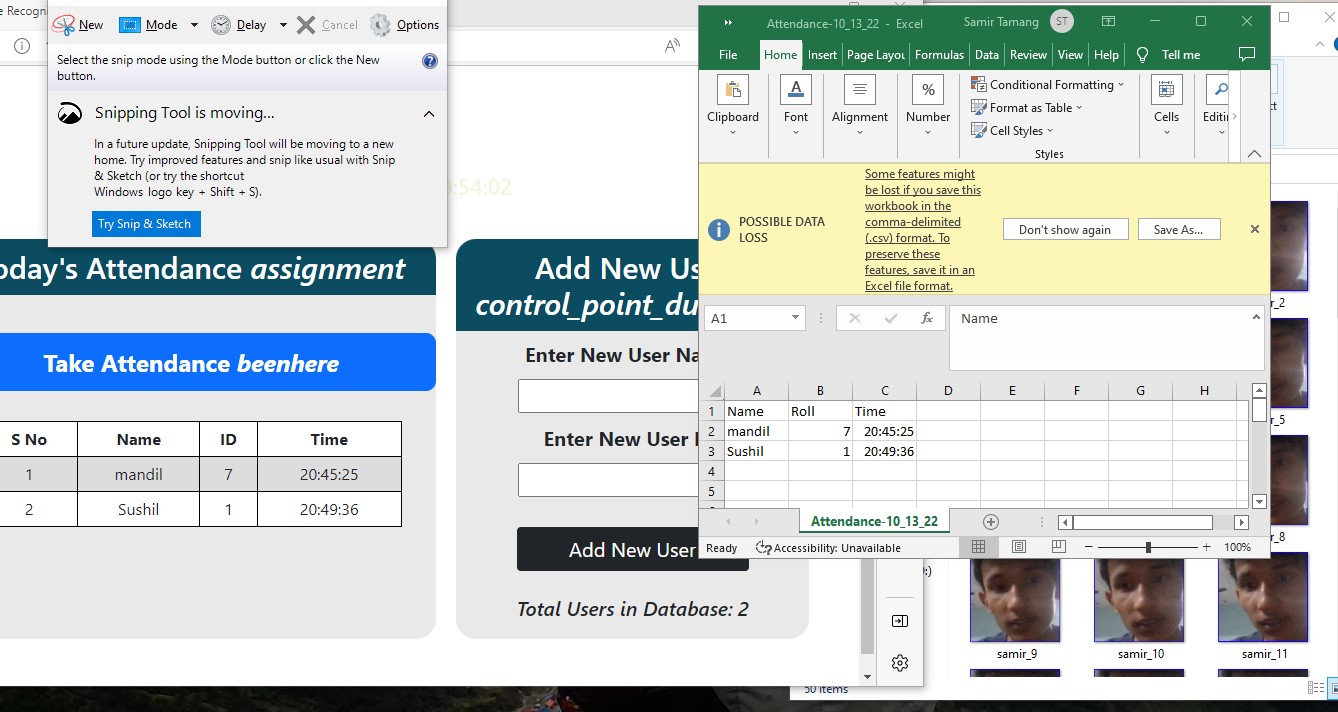
    return render\_template('home.html',names=names,rolls=rolls,times=times,l=l,totalreg=totalreg(),datetoday2=datetoday2)

#### Our main function which runs the Flask App

if \_\_name\_\_ == '\_\_main\_\_':

    app.run(debug=True)

**Output:**

****

1. **CASE STUDY**

In this 21st century the usage of smartphones has increased exponentially and use of social media too. Concerning future prospects, learning Machine learning is strongly recommended for all individuals. And when we come about Machine learning, nowadays it become part of all organization. It is part of Artificial Intelligence and whole world is going towards AI. Why machine learning and Data Science are important because of

* **Explainable AI (XAI):** As AI systems become more integral to decision-making, there is a growing emphasis on making AI models interpretable and explainable.
* **Advanced AI and Deep Learning**: Continued advancements in deep learning and neural networks are expected.
* **Quantum Machine Learning**: The intersection of quantum computing and machine learning is an emerging area.
* **Natural Language Processing (NLP) Advancements**: NLP will continue to advance, enabling more sophisticated language understanding, sentiment analysis, and language generation
* **Natural Language Processing (NLP) Advancements:** NLP will continue to advance, enabling more sophisticated language understanding, sentiment analysis, and language generation

1. **CONCLUSION**

The Machine learning and Data Science are more useful nowadays as increasing number of people spending time on social media.

The project is about Online Attendance through face recognition.

In this system attendance will take automatically by capturing photos and it stored in proper location. It is time Consuming and effective ways.

The symbiotic relationship between data science and machine learning has given rise to unprecedented capabilities, shaping industries, influencing policy decisions, and enhancing various aspects of our daily lives. In navigating the future landscape of data science and machine learning, it is essential to embrace a commitment to ethical practices, inclusivity, and ongoing education. As we harness the power of data and AI, the responsible application of these technologies will determine their positive impact on individuals, organizations, and society as a whole. As stewards of these transformative forces, the collaboration between data scientists, machine learning practitioners,

From this internship I have grabbed many skills and knowledge. This internship is real life useful. It makes me so much knowledgeable and helps to gather new skills.

From Corner of heart, I would like to thank for “YBI Foundation” for made me more skill able, guiding me and giving good instructions.

**REFERENCES**

<https://www.eisystems.in/onlinetraining/datamachine.php>

<https://yourstory.com/companies/ybi-foundation>