YENEPOYA INSTITUTE OF ARTS, SCIENCE, COMMERCE AND MANAGEMENT
A CONSTITUENT UNIT OF YENEPOYA (DEEMED TO BE UNIVERSITY)
BALMATTA, MANGALORE

INTERNSHIP REPORT
ON
DATA ANALYTICS

SUBMITTED BY
MOHAMMED AAQIL RAYYAN
III BCA
(ROBOTICS , MACHINE LEARNING AND ARTIFICIAL INTELLIGENCE) WITH MICROSOFT

22BCARMA051

UNDER THE GUIDANCE OF

MS. FATHIMA THOUFEERA

DEPARTMENT OF COMPUTER SCIENCE

IN PARTIAL FULFILLMENT OF THE REQUIREMENT FOR THE AWARD OF THE DEGREE OF

BACHELORS IN COMPUTER APPLICATION

May - 2025

INTERNSHIP REPORT

ON

DATA ANALYTICS

Internship organization:

Aydeed Complex, YMCA Cross Road,
Kozhikode - 673001

Under the guidance of
Ms. Fathima Thoufeera
Lecturer, Department of Computer Science
Submitted by
Student name: Mohammed Aaqil Rayyan
Register No: 22BCARMA051

BACHELOR OF COMPUTER APPLICATION

(Robotics, Machine Learning And Artificial
Intelligence) with Microsoft

SUBMITTED TO

YENEPOYA INSTITUTE OF ARTS, SCIENCE, COMMERCE & MANAGEMENT
(A constituent unit of Yenepoya Deemed to be University) Deralakatte,
Mangaluru – 575018, Karnataka, India

CERTIFICATE

This is to certify that the internship entitled "**Data Analytics** " has been successfully carried out in the Graduate Studies and Research in Computer Science by **Mohammed Aaqil Rayyan (RegNo:22BCARMA051)**, student of **III BCA (Robotics , Machine Learning and Artificial Intelligence) with Microsoft** , under the supervision and guidance of **Ms. Fathima Thoufeera.**

Internal Guide: Ms. Fathima Thoufeera   **Chairperson**

Internal Examiner: **External Examiner :**

PRINCIPAL Prof (Dr.) Arun Bhagwath The Yenepoya Institute of Arts, Science, Commerce and Management (Deemed to be University)

Submitted for the viva- voice examination held on: 31/05/2025   **Place: Mangalore**

DECLARATION

I, **Mohammed Aaqil Rayyan**, a student of **BCA (in Robotics, Machine Learning and Artificial Intelligence) with Microsoft** at **Yenepoya Institute of Arts, Science, Commerce, and Management, Balmatta, Mangalore**, affiliated with **Yenepoya (Deemed to be University)** hereby declare that this internship report titled **Data Analytics** is a genuine and original record of the work undertaken by me as part of my academic curriculum.

This report documents the **knowledge, skills, and practical experience** acquired during my internship at **Quest Innovative Solutions Pvt. Ltd.** It includes methodologies, analytical processes, and investigative approaches aligned with **recognized industry standards in focus in data-driven decision making, predictive modeling,** and **the communication of findings.**

I extend my sincere gratitude to **Ms. Thoyyiba PK , Internship Program Head**, for her valuable guidance, mentorship, and support throughout the internship period. I also express my appreciation to **Quest Innovative Solutions Pvt. Ltd.** for providing an enriching environment and exposure to real-world Data analytics operations.

Furthermore, I acknowledge the support and encouragement received from my institutional guide, **Ms. Fathima Thoufeera** and the faculty of **Yenepoya Institute of Arts, Science, Commerce, and Management**. Their insights and assistance have been instrumental in successfully completing this internship.

Intern Signature:

Institution Guide Signature:

Name: Mohammed Aaqil Rayyan

Name: Ms. Fathima Thoufeera
Institute: Yenepoya Institute of Arts, Science, Commerce, and Management

Date: 31/05/2025

Date: 31/05/2025

Internship Supervisor Signature:

Head of the Department (HOD) Signature:

Name: Thoyyiba PK Designation: Internship Program Head
Organization: Quest Innovative Solutions Pvt. Ltd
Date: 24/05/2025

Name: Dr. Rathnakar Shetty P
Head of Department of Computer Science
Institute: Yenepoya Institute of Arts, Science,Commerce, and Management
Date:

ACKNOWLEDGEMENT

I sincerely express my gratitude to **Yenepoya Institute of Arts, Science, Commerce, and Management, Balmatta, Mangalore** affiliated with **Yenepoya (Deemed to be University), Mangalore**, for providing me with the opportunity to undertake this internship as part of my academic curriculum.

I extend my heartfelt thanks to **Dr. Arun Bhagwath, Principal**, for his continuous support and guidance in facilitating this learning opportunity. I also express my sincere appreciation to **Dr. Rathnakara Shetty P**, Head of Department (Computer Science), for his encouragement and academic support throughout my internship journey.

I am deeply grateful to **Ms Thoyyiba PK and Ms Jasheena** for their visionary leadership and for fostering a dynamic learning environment in **Quest Innovative Solutions Pvt. Ltd.** A special note of appreciation goes to **Ms. jasheena** for her invaluable mentorship, expert guidance, and for sharing his vast experience in the field. I also extend my sincere gratitude to the **Ms Thoyyiba PK** , for their technical insights, hands-on training, and unwavering support during my internship. Their guidance has been instrumental in enhancing my practical understanding of data manipulation, statistical inference, and machine learning algorithms and data visualization techniques.

I am also thankful to my internal guide, **Ms. Fathima Thoufeera** , for her academic mentorship and for providing valuable insights that helped bridge the gap between theoretical learning and practical application.

Lastly, I am grateful to my faculty mentors, family, and peers for their encouragement and motivation throughout this journey.

Place: Mangalore

Date: 31/05/2025

Mohammed Aaqil Rayyan
III BCA ROBOTICS
22BCARMA051

Table of Contents

INTERNSHIP OFFER LETTER

QUEST
INNOVATIVE SOLUTIONS

Date: 13-02-2025

To,

The Head of the Department

Bachelor of Computer Applications,

Yenepoya Institute of Arts Science

Commerce and Management, Mangalore

Sub: Internship Opportunity- For BCA 3rd Year

Ref to: Letter from Mohammed Aaqil Rayyan, by 08.02.2025

Respected Officer,

With reference to the request letter received from **Mr. Mohammed Aaqil Rayyan (Reg ID-22BCARMA051)**, on 08.02.2025, I would like to officially notify the Head of the Department that Quest Innovative Solution Pvt. Ltd. has approved his internship request as part of the **6th Semester, Bachelor of Computer Applications (Robotics, Artificial Intelligence, and Machine Learning with Microsoft)** curriculum. The internship is slated to commence from 03.03.2025 and conclude on 03.05.2025, and will be conducted at Quest Innovative Solution Calicut Office.

Domain: Data Analytics

For Quest Innovative Solutions Pvt. Ltd

Manoj Jacob
Managing Director

....redefining innovation

1

INTERNSHIP COMPLETION CERTIFICATE

QUEST
INNOVATIVE SOLUTIONS

QIS-HR/2025/06-05
May 06, 2025

TO WHOMSOEVER IT MAY CONCERN

This is to certify that Mr. Mohammed Aaqil Rayyan, BCA (Robotics, Machine Learning and Artificial Intelligence with Microsoft) student of Yenepoya Institute of Arts Science Commerce and Management, Mangalore has successfully completed an internship in Data Analytics at Quest Innovative Solutions Pvt. Ltd. from 03/03/2025 to 03/05/2025.

During the internship period, his dedication, adaptability and eagerness to learn have been commendable. We are confident that his skills, knowledge and experiences gained during this internship will serve as a strong foundation for his future endeavours.

We wish Mr. Mohammed Aaqil Rayyan continued success in his academic and professional journey.

For Quest Innovative Solutions Pvt. Ltd

Manoj Jacob
Managing Director

....redefining innovation

COMPANY PROFILE: Quest Innovative Solutions Pvt. Ltd.

Quest Innovative Solutions (QIS) Pvt. Ltd. Kozhikode, Kerala, is a versatile IT company engaged in both software development and comprehensive IT training. Established around the early 2000s, with some records pointing to a business foundation in 2017, QIS emphasizes a holistic approach integrating skilled personnel, streamlined processes, and cutting-edge technology to ensure high levels of customer satisfaction. Their software development arm offers tailored solutions in areas such as embedded systems, systems software, and web applications, complemented by their own suite of software products designed for diverse business needs, including UPS management with Q-Safe and various business productivity tools. A cornerstone of QIS's operations is their training division, QIS Academy, which delivers industry-focused programs across a spectrum of in-demand IT domains. These include full- stack development, embedded systems, data science, artificial intelligence, and more. QIS Academy distinguishes itself by its emphasis on practical, hands-on learning through well- equipped labs, real-world projects, and experienced instructors. Furthermore, they provide robust placement assistance and ongoing support, aiming to equip graduates with the necessary skills and confidence to excel in the IT industry, thereby bridging the divide between academic learning and the practical demands of the professional world.

[Company Seal]

OBJECTIVES OF THE INTERNSHIP

The primary aim of this internship was to gain practical experience in the field of data analytics by applying theoretical knowledge to real-world scenarios and developing proficiency in essential tools and techniques.

Apply What I've Learned: * To put into practice the fundamental concepts of data analytics and data science learned during the internship, including the data analysis lifecycle and various analytical methodologies. * To utilize Python programming skills and its relevant libraries (NumPy, Pandas, Matplotlib, Seaborn) to execute data analysis tasks effectively.

Learn More About Data Analytics Fundamentals: * To deepen the understanding of core data analytics concepts, their significance across industries, and the diverse roles within the data analytics field. * To gain a strong grasp of fundamental Python programming concepts and Object-Oriented Programming (OOP) principles relevant to data manipulation and analysis.

Get Hands-On with Key Data Analysis Libraries: * To acquire practical skills in using Jupyter Notebook for interactive data exploration and analysis. * To develop proficiency in leveraging NumPy for numerical computations and efficient handling of datasets. * To gain hands-on experience in utilizing Pandas for data manipulation, cleaning, and preprocessing, including loading data from different sources. * To learn and apply Matplotlib and Seaborn for creating insightful data visualizations to identify patterns and trends.

 Understand Data Preprocessing, EDA, and Statistical Concepts: * To master data preprocessing techniques, including data inspection, cleaning, handling missing values, and feature engineering. * To conduct thorough Exploratory Data Analysis (EDA) to understand data characteristics and identify potential insights. * To grasp fundamental statistical and probability concepts and their application in data analysis and decision-making. By focusing on these objectives, this internship aimed to provide a comprehensive and practical introduction to the field of data analytics, building a strong foundation for future growth and application of these skills.

Learn How to Work in a Team: I wanted to develop my teamwork and communication skills by collaborating with other interns and mentors. I learned how to divide tasks, track progress, and share updates effectively
.
Build Industry Connections: I was also hoping to meet and learn from professionals in the field, gaining insights into how things work in the tech industry and making valuable connections for the future.

ROLES AND RESPONSIBILITIES

During my internship, I gained a strong foundation in data analytics and applied these skills to support various projects.

 * **Data Collection and Preprocessing:** I gathered data from various sources (CSV, Excel) and prepared it for analysis. This involved inspecting data using functions like head(), tail(), and info(), handling missing values, removing duplicates, correcting data types, and dropping irrelevant columns. I used Pandas for efficient data manipulation.

* **Exploratory Data Analysis (EDA):** I performed EDA to understand the characteristics of the datasets. This included calculating summary statistics (mean, median, mode, quartiles, etc.), exploring data distributions, and identifying potential issues.

* **Feature Engineering:** I created new features from existing ones to improve the performance of analytical models. This involved techniques like label encoding to handle categorical variables.

* **Data Visualization:** I used data visualization techniques (line charts, bar charts, histograms, scatter plots, heatmaps) to identify patterns, trends, and insights within the data. I utilized libraries like Matplotlib and Seaborn for creating effective visualizations.

* **Statistical Analysis:** I applied statistical concepts (mean, median, mode, standard deviation, variance, probability distributions) to analyze data and draw meaningful conclusions.

* **Python Programming:** I used Python for various data analytics tasks, leveraging its simplicity, readability, and extensive library support (NumPy, Pandas, Matplotlib, Seaborn). I applied my knowledge of Python fundamentals (variables, data types, operators, conditional statements, loops, functions, OOP concepts) to write efficient and well-structured code.

* **Jupyter Notebook:** I used Jupyter Notebook as an interactive environment for data analysis and visualization, writing and executing Python code, documenting my work using Markdown, and presenting my findings.

TOOLS & TECHNOLOGIES USED

During my data analytics internship, I gained hands-on experience with a variety of tools and technologies essential for data analysis, manipulation, visualization, and machine learning. My work involved data preprocessing, exploratory data analysis (EDA), and applying fundamental statistical and mathematical concepts.

Programming Languages & Environments

Python: The primary language used throughout the internship. I utilized it for: * Scripting and automating tasks. * Data manipulation and cleaning. * Implementing statistical calculations. * Developing and evaluating machine learning models. * Creating data visualizations.

Jupyter Notebook: An interactive environment for writing and executing Python code, creating visualizations, and documenting the data analysis process. I used it for: * Developing and testing code snippets. * Performing exploratory data analysis. * Documenting my workflow and findings using Markdown.

Core Python Libraries

NumPy: A fundamental library for numerical computing, used for:
* Efficiently storing and manipulating large, multi-dimensional arrays. * Performing mathematical operations on arrays.

Pandas: A powerful library for data manipulation and analysis, used for: * Creating and manipulating DataFrames and Series. * Loading and cleaning data from various sources (CSV, Excel). * Performing data inspection and summary statistics. * Feature engineering and handling categorical variables.

Matplotlib: A widely used library for creating static, interactive, and animated visualizations in Python.

Seaborn: A library built on top of Matplotlib, providing a high-level interface for creating informative and attractive statistical graphics. Statistical Concepts & Techniques

Descriptive Statistics: Learned and applied measures such as mean, median, mode, standard deviation, variance, and range to summarize and describe datasets.

Probability: Understood and applied concepts like permutations, combinations, probability distributions (including normal distribution), and Z-scores.

Data Visualization : Created various types of charts and plots to visualize data and communicate insights.

Exploratory Data Analysis (EDA): Employed techniques to explore and understand data, identify

TOOLS & TECHNOLOGIES USED

Preprocessing And Visualization Of Datasets

```python
import pandas as pd
```

```python
data=pd.read_csv('iris .csv')
```

```python
data
```

	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Virginica
146	6.3	2.5	5.0	1.9	Virginica
147	6.5	3.0	5.2	2.0	Virginica
148	6.2	3.4	5.4	2.3	Virginica
149	5.9	3.0	5.1	1.8	Virginica

150 rows × 5 columns

```python
data.shape
```

```
(150, 5)
```

```python
data.size
```

```
750
```

```python
data.head(10)
```

	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa
5	5.4	3.9	1.7	0.4	Setosa
6	4.6	3.4	1.4	0.3	Setosa
7	5.0	3.4	1.5	0.2	Setosa
8	4.4	2.9	1.4	0.2	Setosa
9	4.9	3.1	1.5	0.1	Setosa

```python
data.tail(10)
```

	sepal.length	sepal.width	petal.length	petal.width	variety
140	6.7	3.1	5.6	2.4	Virginica
141	6.9	3.1	5.1	2.3	Virginica
142	5.8	2.7	5.1	1.9	Virginica
143	6.8	3.2	5.9	2.3	Virginica
144	6.7	3.3	5.7	2.5	Virginica
145	6.7	3.0	5.2	2.3	Virginica
146	6.3	2.5	5.0	1.9	Virginica
147	6.5	3.0	5.2	2.0	Virginica
148	6.2	3.4	5.4	2.3	Virginica
149	5.9	3.0	5.1	1.8	Virginica

```python
[ ]: data.columns
```

```
[ ]: Index(['sepal.length', 'sepal.width', 'petal.length', 'petal.width',
             'variety'],
            dtype='object')
```

```python
[ ]: data.dtypes
```

```
[ ]: sepal.length    float64
     sepal.width     float64
     petal.length    float64
     petal.width     float64
     variety          object
     dtype: object
```

```python
[ ]: data.sample(10,random_state=25)
```

	sepal.length	sepal.width	petal.length	petal.width	variety
139	6.9	3.1	5.4	2.1	Virginica
125	7.2	3.2	6.0	1.8	Virginica
67	5.8	2.7	4.1	1.0	Versicolor
3	4.6	3.1	1.5	0.2	Setosa
113	5.7	2.5	5.0	2.0	Virginica
52	6.9	3.1	4.9	1.5	Versicolor
18	5.7	3.8	1.7	0.3	Setosa
137	6.4	3.1	5.5	1.8	Virginica
20	5.4	3.4	1.7	0.2	Setosa
64	5.6	2.9	3.6	1.3	Versicolor

```python
[ ]: data.isnull().sum()
```

```
[ ]: sepal.length    0
     sepal.width     0
     petal.length    0
     petal.width     0
     variety         0
     dtype: int64
```

```python
[ ]: data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 150 entries, 0 to 149
Data columns (total 5 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   sepal.length  150 non-null    float64
 1   sepal.width   150 non-null    float64
 2   petal.length  150 non-null    float64
 3   petal.width   150 non-null    float64
 4   variety       150 non-null    object
dtypes: float64(4), object(1)
memory usage: 6.0+ KB
```

```python
[ ]: data['sepal.length']
```

```
[ ]: 0      5.1
     1      4.9
     2      4.7
     3      4.6
     4      5.0
            ...
     145    6.7
     146    6.3
     147    6.5
     148    6.2
     149    5.9
     Name: sepal.length, Length: 150, dtype: float64
```

```python
[ ]: data.iloc[[24,10]]
```

	sepal.length	sepal.width	petal.length	petal.width	variety
24	4.8	3.4	1.9	0.2	Setosa
10	5.4	3.7	1.5	0.2	Setosa

```python
[ ]: data['variety'].nunique()
```

```
[ ]: 3
```
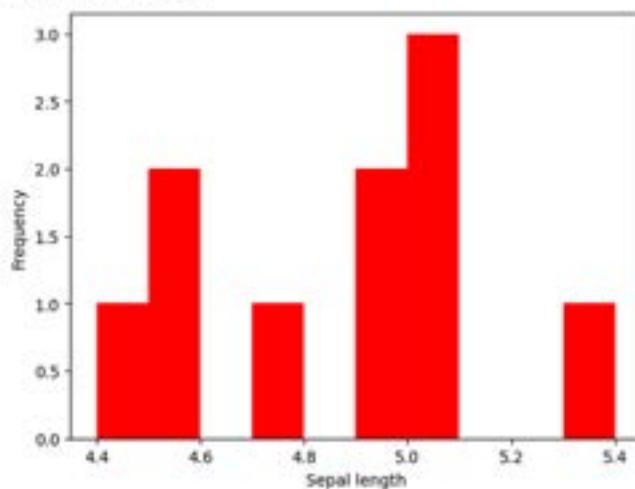
```python
[ ]: data['variety'].value_counts()
```

```
[ ]: variety
     Setosa        50
     Versicolor    50
     Virginica     50
     Name: count, dtype: int64
```

TOOLS & TECHNOLOGIES USED
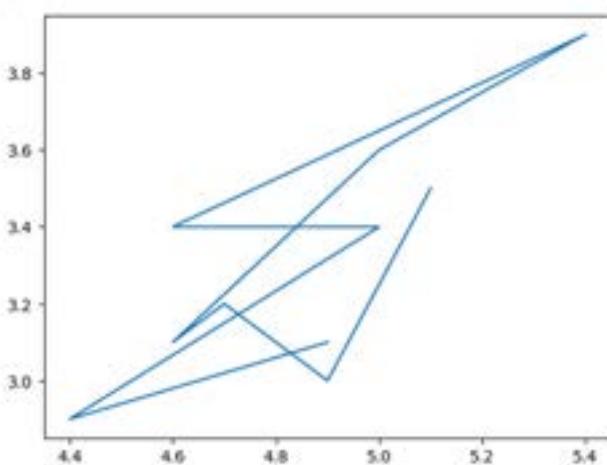
```
[130]: plt.hist(a, color ="red")
       plt.xlabel('Sepal length')
       plt.ylabel('Frequency')

[130]: Text(0, 0.5, 'Frequency')
```



```
[131]: a=data1['sepal.length'][:10]
       b=data1['sepal.width'][:10]
       plt.plot(a,b)

[131]: [<matplotlib.lines.line2D at 0x1e38cb860d0>]
```

Machine Learning Algorithms On Datasets

logistic Regression

```python
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score,confusion_matrix
from sklearn.linear_model import LogisticRegression
```

```python
data=pd.read_csv('iris .csv')
data
```

	sepal.length	sepal.width	petal.length	petal.width	variety
0	5.1	3.5	1.4	0.2	Setosa
1	4.9	3.0	1.4	0.2	Setosa
2	4.7	3.2	1.3	0.2	Setosa
3	4.6	3.1	1.5	0.2	Setosa
4	5.0	3.6	1.4	0.2	Setosa
...	...	...	...	...	...
145	6.7	3.0	5.2	2.3	Virginica
146	6.3	2.5	5.0	1.9	Virginica
147	6.5	3.0	5.2	2.0	Virginica
148	6.2	3.4	5.4	2.3	Virginica
149	5.9	3.0	5.1	1.8	Virginica

150 rows × 5 columns

```python
x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=20)
```

```python
x_train.shape,x_test.shape
```
```
((120, 4), (30, 4))
```

```python
y_train.shape,y_test.shape
```
```
((120,), (30,))
```

```python
model=LogisticRegression()
model.fit(x_train,y_train)
```
```
▸ LogisticRegression
LogisticRegression()
```

```python
model.score(x_train,y_train)
```
```
0.9666666666666667
```
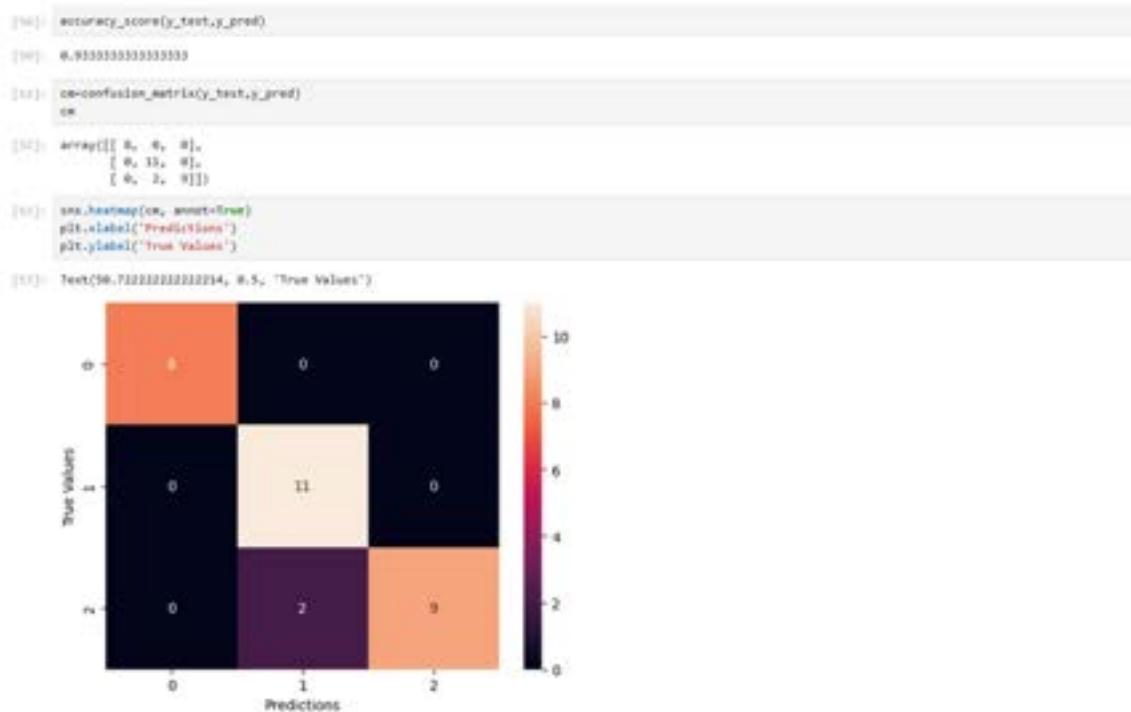
```python
model.score(x_test,y_test)
```
```
0.9333333333333333
```

```python
y_pred=model.predict(x_test)
```

```python
pd.DataFrame({'True Values':y_test,'Predicted Values':y_pred})
```

	True Values	Predicted Values
47	Setosa	Setosa
73	Versicolor	Versicolor
74	Versicolor	Versicolor
129	Virginica	Virginica

# TOOLS & TECHNOLOGIES USED

```
[24]: accuracy_score(y_test,y_pred)
```
```
[24]: 0.9333333333333333
```
```
[25]: cm=confusion_matrix(y_test,y_pred)
      cm
```
```
[25]: array([[ 8,  0,  0],
             [ 0, 11,  0],
             [ 0,  2,  9]])
```
```
[27]: sns.heatmap(cm, annot=True)
      plt.xlabel('Predictions')
      plt.ylabel('True Values')
```
```
[27]: Text(50.722222222222214, 0.5, 'True Values')
```

Dataset-2

```
[72]: data1=pd.read_csv('healthcare-dataset-stroke-data.csv')
```
```
[74]: data1
```

	id	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	9046	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	51676	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	31112	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	60182	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	1665	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
...	...	...	...	...	...	...	...	...	...	...	...	...
5105	18234	Female	80.0	1	0	Yes	Private	Urban	83.75	NaN	never smoked	0
5106	44873	Female	81.0	0	0	Yes	Self-employed	Urban	125.20	40.0	never smoked	0
5107	19723	Female	35.0	0	0	Yes	Self-employed	Rural	82.99	30.6	never smoked	0
5108	37544	Male	51.0	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
5109	44679	Female	44.0	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

5110 rows × 12 columns

```
[75]: data1=data1.drop("id",axis=1)
```
```
[76]: data1
```

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
1	Female	61.0	0	0	Yes	Self-employed	Rural	202.21	NaN	never smoked	1
2	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
...	...	...	...	...	...	...	...	...	...	...	...
5105	Female	80.0	1	0	Yes	Private	Urban	83.75	NaN	never smoked	0
5106	Female	81.0	0	0	Yes	Self-employed	Urban	125.20	40.0	never smoked	0

TOOLS & TECHNOLOGIES USED

```python
data1['bmi']=data1['bmi'].fillna(data1['bmi'].mean())
```

```python
from sklearn.preprocessing import LabelEncoder
```

```python
le=LabelEncoder()
```

```python
data1.dtypes
```

```
gender               object
age                  float64
hypertension         int64
heart_disease        int64
ever_married         object
work_type            object
Residence_type       object
avg_glucose_level    float64
bmi                  float64
smoking_status       object
stroke               int64
dtype: object
```

```python
data1['gender']=le.fit_transform(data1['gender'])
```

```python
data1['gender'].unique()
```

```
array([1, 0, 2])
```

```python
data1['gender'].info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 5110 entries, 0 to 5109
Series name: gender
Non-Null Count  Dtype
--------------  -----
5110 non-null   int64
dtypes: int64(1)
memory usage: 40.1 KB
```

```python
data1['ever_married']=le.fit_transform(data1['ever_married'])
```

```python
data1['ever_married'].unique()
```

```
array([1, 0])
```

```python
data1.dropna()
```

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	Male	67.0	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
2	Male	80.0	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
3	Female	49.0	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
4	Female	79.0	1	0	Yes	Self-employed	Rural	174.12	24.0	never smoked	1
5	Male	81.0	0	0	Yes	Private	Urban	186.21	29.0	formerly smoked	1
...	...	...	...	...	...	...	...	...	...	...	...
5104	Female	13.0	0	0	No	children	Rural	103.08	18.6	Unknown	0
5106	Female	81.0	0	0	Yes	Self-employed	Urban	125.20	40.0	never smoked	0
5107	Female	35.0	0	0	Yes	Self-employed	Rural	82.99	30.6	never smoked	0
5108	Male	51.0	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
5109	Female	44.0	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

4909 rows × 11 columns

```python
data1.columns
```

```
Index(['gender', 'age', 'hypertension', 'heart_disease', 'ever_married',
       'work_type', 'Residence_type', 'avg_glucose_level', 'bmi',
       'smoking_status', 'stroke'],
      dtype='object')
```

```python
data1.isnull().sum()
```

```
gender               0
age                  0
hypertension         0
heart_disease        0
ever_married         0
work_type            0
Residence_type       0
avg_glucose_level    0
bmi                  201
smoking_status       0
stroke               0
dtype: int64
```

TOOLS & TECHNOLOGIES USED

```python
[111]: data1['work_type']=le.fit_transform(data1['work_type'])
```

```python
[112]: data1['work_type'].unique()
```

```
[112]: array([2, 3, 0, 4, 1])
```

```python
[113]: data1["Residence_type"]=le.fit_transform(data1['Residence_type'])
```

```python
[114]: data1["Residence_type"].unique()
```

```
[114]: array([1, 0])
```

```python
[115]: data1['smoking_status']=le.fit_transform(data1['smoking_status'])
```

```python
[116]: data1['smoking_status'].unique()
```

```
[116]: array([1, 2, 3, 0])
```

```python
[117]: from sklearn.preprocessing import StandardScaler,MinMaxScaler
```

```python
[118]: ss=StandardScaler()
       new_data=ss.fit_transform(data1)
       pd.DataFrame(new_data, columns=data1.columns)
```

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	1.188073	1.051434	-0.328602	4.185032	0.723884	-0.153836	0.984080	2.706375	1.001234e+00	-0.351701	4.418381
1	-0.840344	0.790070	-0.328602	-0.238947	0.723884	0.763438	-1.016178	2.121559	4.615554e-16	0.581552	4.418381
2	1.188073	1.626390	-0.328602	4.185032	0.723884	-0.153836	-1.016178	-0.005028	4.685773e-01	0.581552	4.418381
3	-0.840344	0.255342	-0.328602	-0.238947	0.723884	-0.153836	0.984080	1.437158	7.154182e-01	1.514885	4.418381
4	-0.840344	1.582163	3.043196	-0.238947	0.723884	0.763438	-1.016178	1.501184	-6.357112e-01	0.581552	4.418381
...	...	...	...	...	...	...	...	...	...	...	...
5105	-0.840344	1.626390	3.043196	-0.238947	0.723884	-0.153836	0.984080	-0.494658	4.615554e-16	0.581552	-0.226327
5106	-0.840344	1.670617	-0.328602	-0.238947	0.723884	0.763438	0.984080	0.420775	1.442949e+00	0.581552	-0.226327
5107	-0.840344	-0.363842	-0.328602	-0.238947	0.723884	0.763438	-1.016178	-0.511443	2.217363e-01	0.581552	-0.226327
5108	1.188073	0.343796	-0.328602	-0.238947	0.723884	-0.153836	-1.016178	1.328257	-4.278451e-01	-0.351701	-0.226327
5109	-0.840344	0.034205	-0.328602	-0.238947	0.723884	-1.986384	0.984080	-0.460867	-3.498953e-01	-1.285114	-0.226327

5110 rows × 11 columns

```python
[119]: mm=MinMaxScaler()
       new_data=mm.fit_transform(data1)
       pd.DataFrame(new_data, columns=data1.columns)
```

	gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
0	0.5	0.816895	0.0	1.0	1.0	0.50	1.0	0.801265	0.301260	0.333333	1.0
1	0.0	0.743652	0.0	0.0	1.0	0.75	0.0	0.679023	0.212981	0.666667	1.0
2	0.5	0.975586	0.0	1.0	1.0	0.50	0.0	0.234512	0.254296	0.666667	1.0
3	0.0	0.597168	0.0	0.0	1.0	0.50	1.0	0.536008	0.276060	1.000000	1.0
4	0.0	0.963379	1.0	0.0	1.0	0.75	0.0	0.549349	0.158830	0.666667	1.0
...	...	...	...	...	...	...	...	...	...	...	...
5105	0.0	0.975586	1.0	0.0	1.0	0.50	1.0	0.132167	0.212981	0.666667	0.0
5106	0.0	0.987793	0.0	0.0	1.0	0.75	1.0	0.323516	0.340206	0.666667	0.0
5107	0.0	0.426270	0.0	0.0	1.0	0.75	0.0	0.128658	0.232532	0.666667	0.0
5108	0.5	0.621582	0.0	0.0	1.0	0.50	0.0	0.513203	0.175258	0.333333	0.0
5109	0.0	0.536133	0.0	0.0	1.0	0.00	1.0	0.139230	0.182131	0.000000	0.0

5110 rows × 11 columns

13
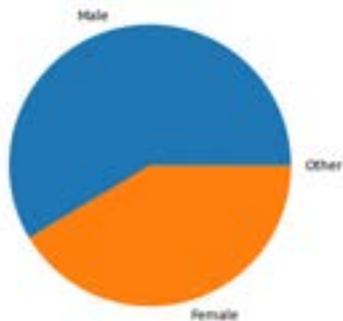
```
[120]: count= data['gender'].value_counts()
       lab=data['gender'].unique()
```

```
[121]: plt.pie(count, labels=lab)
```

```
[121]: ([<matplotlib.patches.Wedge at 0x1e30cb425d0>,
        <matplotlib.patches.Wedge at 0x1e30cb42950>,
        <matplotlib.patches.Wedge at 0x1e30cb41ad0>],
        [Text(-0.19326237984960356, 1.0681790320141637, 'Male'),
        Text(0.29264065906052367, -1.0603581513181293, 'Female'),
        Text(1.0999997920134026, -0.00007641955547407801, 'Other')])
```
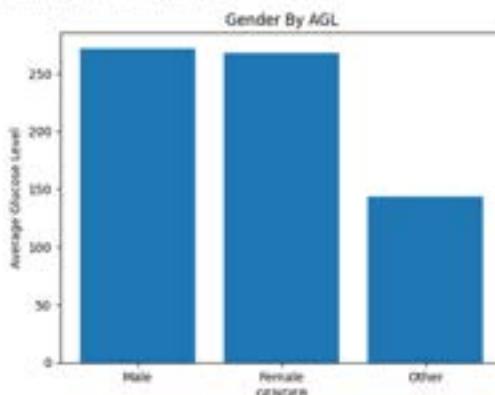


```
[119]: plt.bar(data['gender'], data['avg_glucose_level'])
       plt.xlabel("GENDER")
       plt.ylabel("Average Glucose Level")
       plt.title("Gender By AGL")
```
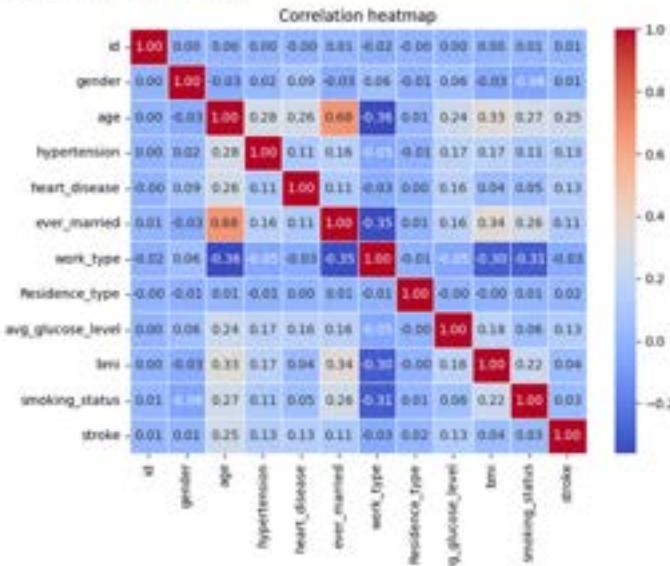
```
[119]: Text(0.5, 1.0, 'Gender By AGL')
```



```
[115]: import seaborn as sns
       import matplotlib.pyplot as plt
       corr = data.corr()
```

```
[116]: plt.figure(figsize=(9,8))
       sns.heatmap(corr, annot = True, cmap='coolwarm', fmt='.2f', linewidth=0.5)
       plt.title("Correlation heatmap")
```

```
[116]: Text(0.5, 1.0, 'Correlation heatmap')
```

TOOLS & TECHNOLOGIES USED

Dataset-3

```python
import pandas as pd
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_absolute_error, mean_squared_error
from sklearn.model_selection import train_test_split
```

```python
data=pd.read_csv('Admission_predict.csv')
data
```

	Serial No.	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	1	337	118	4	4.5	4.5	9.65	1	0.92
1	2	324	107	4	4.0	4.5	8.87	1	0.76
2	3	316	104	3	3.0	3.5	8.00	1	0.72
3	4	322	110	3	3.5	2.5	8.67	1	0.80
4	5	314	103	2	2.0	3.0	8.21	0	0.65
...	...	...	...	...	...	...	...	...	...
395	396	324	110	3	3.5	3.5	9.04	1	0.82
396	397	325	107	3	3.0	3.5	9.11	1	0.84
397	398	330	116	4	5.0	4.5	9.45	1	0.91

```python
data.isnull().sum()***
```

```
Serial No.            0
GRE Score             0
TOEFL Score           0
University Rating     0
SOP                   0
LOR                   0
CGPA                  0
Research              0
Chance of Admit       0
dtype: int64
```

```python
data.columns
```

```
Index(['Serial No.', 'GRE Score', 'TOEFL Score', 'University Rating', 'SOP',
       'LOR ', 'CGPA', 'Research', 'Chance of Admit '],
      dtype='object')
```

```python
data.dtypes
```

```
Serial No.            int64
GRE Score             int64
TOEFL Score           int64
University Rating     int64
SOP                 float64
LOR                 float64
CGPA                float64
Research              int64
Chance of Admit     float64
dtype: object
```

```python
data.drop('Serial No.',axis=1,inplace=True)
data
```

	GRE Score	TOEFL Score	University Rating	SOP	LOR	CGPA	Research	Chance of Admit
0	337	118	4	4.5	4.5	9.65	1	0.92
1	324	107	4	4.0	4.5	8.87	1	0.76
2	316	104	3	3.0	3.5	8.00	1	0.72
3	322	110	3	3.5	2.5	8.67	1	0.80
4	314	103	2	2.0	3.0	8.21	0	0.65
...	...	...	...	...	...	...	...	...
395	324	110	3	3.5	3.5	9.04	1	0.82
396	325	107	3	3.0	3.5	9.11	1	0.84
397	330	116	4	5.0	4.5	9.45	1	0.91
398	312	103	3	3.5	4.0	8.78	0	0.67
399	333	117	4	5.0	4.0	9.66	1	0.95

400 rows × 8 columns

```python
#Model Creation And Training
x=data.drop('Chance of Admit ',axis=1)
y=data['Chance of Admit ']
```

```
[10]: x
```

```
[10]:     GRE Score  TOEFL Score  University Rating  SOP  LOR  CGPA  Research
    0      337        118                4        4.5  4.5  9.65     1
    1      324        107                4        4.0  4.5  8.87     1
    2      316        104                3        3.0  3.5  8.00     1
    3      322        110                3        3.5  2.5  8.67     1
    4      314        103                2        2.0  3.0  8.21     0
    ...    ...        ...              ...        ...  ...  ...     ...
    395    324        110                3        3.5  3.5  9.04     1
    396    325        107                3        3.0  3.5  9.11     1
    397    330        116                4        5.0  4.5  9.45     1
    398    312        103                3        3.5  4.0  8.78     0
    399    333        117                4        5.0  4.0  9.66     1

    400 rows × 7 columns
```

```
[11]: y
```

```
[11]: 0      0.92
    1      0.76
    2      0.72
    3      0.80
    4      0.65
           ...
    395    0.82
    396    0.84
    397    0.91
    398    0.67
    399    0.95
    Name: Chance of Admit , Length: 400, dtype: float64
```

```
[12]: x_train,x_test,y_train,y_test=train_test_split(x,y,test_size=0.2,random_state=23)
```

```
[13]: x_train.shape,x_test.shape
```

```
[13]: ((320, 7), (80, 7))
```

```
[14]: y_train.shape,y_test.shape
```

```
[14]: ((320,), (80,))
```

```
[15]: model=LinearRegression()
    model.fit(x_train,y_train)
```

```
[15]:  ▾ LinearRegression  ⊙ ⊙
      LinearRegression()
```

```
[16]: model.coef_ # slope of each columns
```

```
[16]: array([[0.00195942, 0.00261344, 0.00040984, 0.00118067, 0.01662844,
            0.11889927, 0.01008301]])
```

```
[17]: model.intercept_
```

```
[17]: np.float64(-1.29790150235465)
```

```
[18]: #Model Performance
    model.score(x_test,y_test)
```

```
[18]: 0.7393817181673338
```

```
[19]: model.score(x_train,y_train)
```

```
[19]: 0.8144427567157716
```

```
[20]: y_pred=model.predict(x_test)
```

TOOLS & TECHNOLOGIES USED

```
[11]: pd.DataFrame({"True values": y_test, "Predicted values": y_pred})
```

	True values	Predicted values
133	0.79	0.793264
331	0.73	0.623077
167	0.64	0.642757
335	0.83	0.843480
239	0.59	0.530874
...	...	...
155	0.77	0.710218
74	0.74	0.764523
270	0.72	0.641204

```
[12]: mean_absolute_error(y_test,y_pred)
```
```
[12]: 0.04930378418949355
```
```
mean_squared_error(y_test,y_pred) ***
```
```
[13]: 0.004393350036548289
```
```
[14]: model.predict([[300,113,4,2,1,7.9,0]])
```
```
C:\Users\user\AppData\Local\Programs\Python\Python313\Lib\site-packages\sklearn\utils\validation.py:2739: UserWarning: X does not have valid featur
e, but LinearRegression was fitted with feature names
  warnings.warn(
```
```
[14]: array([0.61044973])
```

Dataset-4
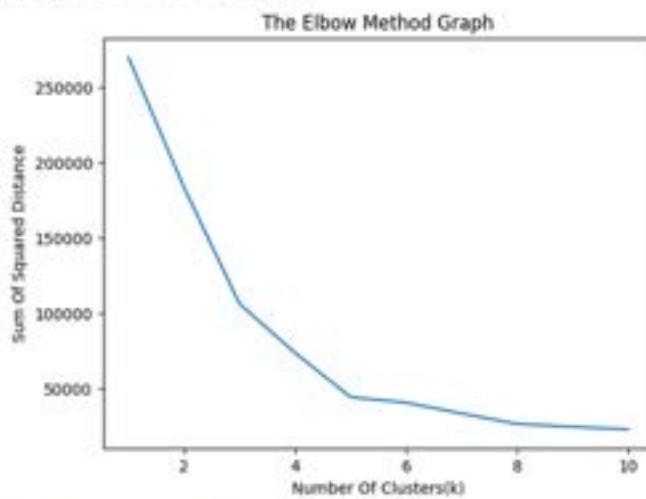
KMeans Clustering

```
[6]: import pandas as pd
import matplotlib.pyplot as plt
from sklearn.cluster import KMeans
```
```
[7]: data=pd.read_csv('Mall_Customers.csv')
data
```

	CustomerID	Gender	Age	Annual Income (k$)	Spending Score (1-100)
0	1	Male	19	15	39
1	2	Male	21	15	81
2	3	Female	20	16	6
3	4	Female	23	16	77
4	5	Female	31	17	40
...	...	...	...	...	...
195	196	Female	35	120	79
196	197	Female	45	126	28
197	198	Male	32	126	74
198	199	Male	32	137	18
199	200	Male	30	137	83

200 rows × 5 columns

```
[10]: x=data[['Annual Income (k$)','Spending Score (1-100)']].values
x
```
```
array([[15, 39], ***
```
```
[11]: sse=[]
for i in range(1,11):
    km=KMeans(n_clusters=i, random_state=42)
    km.fit(x)
    sse.append(km.inertia_)
```

# TOOLS & TECHNOLOGIES USED

```
[11]: y=range(1,11)
      z=sse
      plt.plot(y, z)
      plt.title('The Elbow Method Graph')
      plt.xlabel('Number Of Clusters(k)')
      plt.ylabel('Sum Of Squared Distance')
```

```
[11]: Text(0, 0.5, 'Sum Of Squared Distance')
```



```
[10]: plt.scatter(x[y_predict == 0,0], x[y_predict == 0,1],s=100,color='skyblue',label='Cluster 1')
      plt.scatter(x[y_predict == 1,0], x[y_predict == 1,1],s=100,color='lightgreen',label='Cluster 2')
      plt.scatter(x[y_predict == 2,0], x[y_predict == 2,1],s=100,color='pink',label='Cluster 3')
      plt.scatter(kmeans.cluster_centers_[:,0], kmeans.cluster_centers_[:,1],s=100,color='red',label='Cenroid',marker='s')
      plt.title('Clusters Of Customers')
      plt.xlabel('Annual Income (k$)')
      plt.ylabel('Spending Score (1-100)')
      plt.legend()
```

```
[10]: <matplotlib.legend.Legend at 0x1ca5f80cf50>
```
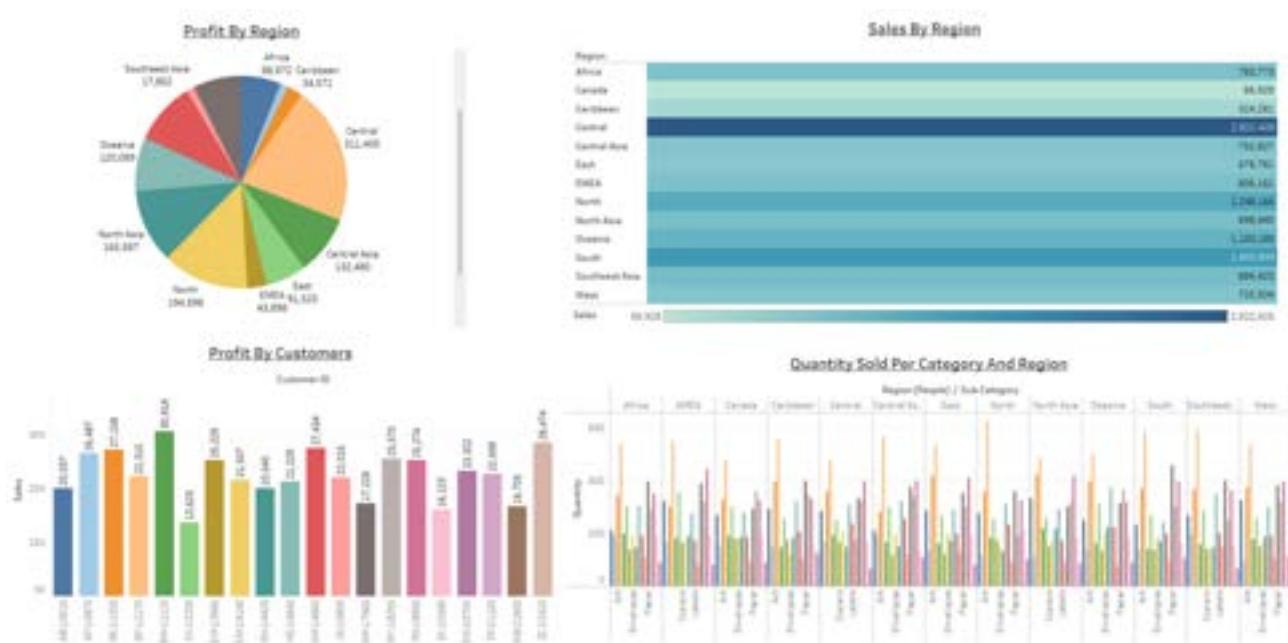
TOOLS & TECHNOLOGIES USED

Tableau Dashboards

Dashboard-1

Dashboard-2

LEARNING OUTCOMES

During this intensive data analytics internship, I gained practical experience and a comprehensive understanding of the data analysis pipeline, from foundational programming to advanced analytical techniques. Key areas of learning include:

Python and SQL Skills Improvement: I became proficient in utilizing Python for data manipulation, cleaning, building, and visualization. I also developed the ability to write efficient SQL queries to interact with databases and retrieve relevant data for analysis.

Data Preprocessing and Cleaning: I understood the critical importance of clean and well-prepared data for accurate analysis. I practiced identifying and handling missing values, correcting data types, removing duplicates and irrelevant information, and applying data transformation techniques using libraries like Pandas.

Exploratory Data Analysis (EDA): I learned how to conduct thorough exploratory data analysis to understand the characteristics of datasets. This involved using various techniques to inspect data (e.g., head, tail, info), calculate descriptive statistics, identify patterns and anomalies, and gain initial insights.

Data Visualization: I developed skills in creating effective data visualizations using libraries such as Matplotlib and Seaborn. I learned to choose appropriate chart types (e.g., line charts, bar charts, histograms, scatter plots, heatmaps) to communicate data insights clearly and identify trends.

Foundational Statistics and Probability: I grasped fundamental statistical and probability concepts essential for data analysis. This included understanding different types of data, sampling methods, descriptive and inferential statistics, probability rules, and common probability distributions like the normal distribution. I also gained practical experience in calculating key statistical measures.

NumPy Proficiency: I gained hands-on experience with NumPy, a powerful library for numerical computing in Python. I learned how to create and manipulate multi-dimensional arrays and perform efficient mathematical operations, which are crucial for data manipulation and analysis.

Pandas Mastery: I developed strong skills in using the Pandas library for data manipulation and analysis. I learned to create and work with DataFrames and Series, load data from various sources, perform data cleaning and transformation, and conduct exploratory data analysis.

Object-Oriented Programming (OOP) in Python: I gained a solid understanding of OOP principles in Python, including classes, objects, inheritance, polymorphism, and encapsulation. This knowledge enables me to write more organized, reusable, and maintainable code for data analysis tasks.

Foundational Knowledge of Data Analytics and Data Science: I acquired a strong foundational understanding of the fields of Data Analytics and Data Science, including their real-world applications across various industries and the diverse job opportunities available.

Jupyter Notebook Proficiency: I became comfortable using Jupyter Notebook as an interactive development environment for data analysis, including writing and executing Python code, documenting my work using Markdown, and organizing my analysis in a reproducible manner.

INTERN WEEKLY REPORT

Intern: **MOHAMMED AAQIL RAYYAN**   Week Beginning: **03/03/2025** To: **09/032025**

Position: **DATA ANALYTICS INTERN**

Employer: **QUEST INNOVATIVE SOLUTIONS**

Internship Supervisor's Name: **THOYYIBA PK**   Phone: **9778160684**

Internship Coordinator: **JASHEENA**

DAY	TIME IN	TIME OUT	TOTAL HOURS	DUTIES PERFORMED
MONDAY	9:00 AM	12:00 PM	3 HR	FUNDAMENTAL CONCEPTS OF DATA ANALYTICS ABOUT PYTHON AND ITS ROLE IN DATA ANALYTICS AND ITS FEATURES
TUESDAY	9:00 AM	12:00 PM	3 HR	INSTALLATION AND CONFIGURATION OF PYTHON,SETUP AND WORKING OF JUPYTER NOTEBOOK
WEDNESDAY	9:00 AM	12:00 PM	3 HR	LEARNED ABOUT BASIC PYTHON FEATURES AND THEIR SIGNIFICANCE OPERATORS,DATATYPES ETC.
THURSDAY	9:00 AM	12:00 PM	3 HR	CONTINUED LEARNING ABOUT IDENTITY OPERATORS CONDITIONAL STATEMENTS
FRIDAY	OFF	OFF	OFF	OFF
SATURDAY	9:00 AM	1:00 PM	4 HR	LOOPS IN PYTHON FOR AUTOMATING REPETITIVE TASK
SUNDAY				
TOTAL HOURS			16 HR	

Signatures:

_____
Intern

_____
Internship Supervisor

PLEASE RETURN TO: Internship Guide

INTERN WEEKLY REPORT

Intern: **MOHAMMED AAQIL RAYYAN**   Week Beginning: **10/03/2025** To: **16/03/2025**

Position: **DATA ANALYTICS INTERN**

Employer: **QUEST INNOVATIVE SOLUTIONS**

Internship Supervisor's Name: **THOYYIBA PK**   Phone: **9778160684**

Internship Coordinator: **JASHEENA**

DAY	TIME IN	TIME OUT	TOTAL HOURS	DUTIES PERFORMED
MONDAY	9:00 AM	12:00 PM	3 HR	FUNCTIONS IN PYTHON AND THEIR SIGNIFICANC,DIFFERENT FUCNTION ARGUMENTS
TUESDAY	9:00 AM	11:00 AM	2 HR	VARIABLE SCOPE IN PYTHON,CORE OOP CONCEPTS
WEDNESDAY	9:00 AM	12:00 PM	3 HR	CONSTRUCTORS AND DESTRUCTORS IN PYTHON,INHERITANCE,DIFFERENT TYPES OF INHERITANCE
THURSDAY	9:00 AM	11:00 AM	2 HR	POLYMORPHISM,ABOUT BUILD IN POLYMORPHISM IN PYTHON METHOD OVERLOADING AND METHOD OVERRIDING
FRIDAY	OFF	OFF	OFF	OFF
SATURDAY	OFF	OFF	OFF	OFF
SUNDAY				
TOTAL HOURS			10 HR	

Signatures:

_____
Intern

_____
Internship Supervisor

PLEASE RETURN TO: Internship Guide

22

WEEKLY INTERNSHIP LOG

INTERN WEEKLY REPORT

Intern: **MOHAMMED AAQIL RAYYAN**   Week Beginning: **17/03/2025** To: **23/032025**

Position: **DATA ANALYTICS INTERN**

Employer: **QUEST INNOVATIVE SOLUTIONS**

Internship Supervisor's Name: **THOYYIBA PK**   Phone: **9778160684**

Internship Coordinator: **JASHEENA**

DAY	TIME IN	TIME OUT	TOTAL HOURS	DUTIES PERFORMED
MONDAY	9:00 AM	12:00 PM	3 HR	ENCAPSULATION,ACCESS MODIFIERS IN PYTHON AND EXPLORED DATA ABSTRACTION
TUESDAY	9:00 AM	11:00 AM	2 HR	UNDERSTANDING OF NUMPY,NUMPY ARRAYS,VARIOUS MANIPULATION TECHNIQUES AND ARRAY OPERATIONS
WEDNESDAY	9:00 AM	12:00 PM	3 HR	FUNDAMENTALS OF STATISTICS,UNDERSTANDING OF PERMUTATION AND COMBINATION
THURSDAY	9:00 AM	12:00 PM	3 HR	PROBABILITY, ITS KEY TERMINOLOGIES AND TYPES INTRODUCTION TO PANDAS
FRIDAY	OFF	OFF	OFF	
SATURDAY	9:00 AM	1:00 PM	4 HR	UTILIZED ESSENTIAL LIBRARIES,PANDAS,MATPLOTLIB SEABORN ETC LOADED DATASETS PERFORMED DATA INSPECTION FEATURE ENGINEERING
SUNDAY				
TOTAL HOURS			15 HR	

Signatures:

_____
Intern

_____
Internship Supervisor

PLEASE RETURN TO: Internship Guide

INTERN WEEKLY REPORT

Intern: **MOHAMMED AAQIL RAYYAN**   Week Beginning: **24/03/2025** To: **30/032025**

Position: **DATA ANALYTICS INTERN**

Employer: **QUEST INNOVATIVE SOLUTIONS**

Internship Supervisor's Name: **THOYYIBA PK**   Phone: **9778160684**

Internship Coordinator: **JASHEENA**

DAY	TIME IN	TIME OUT	TOTAL HOURS	DUTIES PERFORMED
MONDAY	OFF	OFF	OFF	
TUESDAY	9:00 AM	12:00 PM	3 HR	UNDERSTANDING OF DATA VISUALIZATION,VARIOUS VISUALIZATION TECHNIQUIES AND THEIR APPLICATIONS
WEDNESDAY	9:00 AM	12:00 PM	3 HR	GAINED KNOWLEDGE ON TABLEAU SETUP OFF TABLEAU AND DATA TYPES IN TABLEAU
THURSDAY	9:00 AM	12:00 PM	3 HR	WITH THE HELP OF TABLEAU CREATED VARIOUS VISUALIZATIONS,DASHBOARDS AND STORYBOARD
FRIDAY	OFF	OFF	OFF	
SATURDAY	9:00 AM	12:00 PM	3 HR	CREATING DIFFERENT CHARTS USING MULTIPLE DATASETS MODIFYING AND CUSTOMIZING CHARTS
SUNDAY				
TOTAL HOURS			12 HR	

Signatures:

_____
Intern

_____
Internship Supervisor

PLEASE RETURN TO: Internship Guide

WEEKLY INTERNSHIP LOG

INTERN WEEKLY REPORT

Intern: **MOHAMMED AAQIL RAYYAN**   Week Beginning: **31/03/2025** To: **06/042025**

Position: **DATA ANALYTICS INTERN**
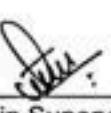
Employer: **QUEST INNOVATIVE SOLUTIONS**

Internship Supervisor's Name: **THOYYIBA PK**   Phone: **9778160684**

Internship Coordinator: **JASHEENA**

DAY	TIME IN	TIME OUT	TOTAL HOURS	DUTIES PERFORMED
MONDAY	OFF	OFF	OFF	OFF
TUESDAY	OFF	OFF	OFF	OFF
WEDNESDAY	9:00 AM	12:00 PM	3 HR	CHARTS IN TABLEAU,HOW TO MODIFY VISUAL ELEMENTS AND BINNING COLUMNS
THURSDAY	9:00 AM	11:00 AM	2 HR	ABOUT MACHINE LEARNING AND ITS TYPES, SUBTYPES,ADVANTAGES, LIMITATIONS
FRIDAY	OFF	OFF	OFF	OFF
SATURDAY	9:00 AM	12:00 PM	3 HR	CONTINUED WITH SUPERVISIED LEARNING,LINEAR REGRESSION PROBLEMS,LOGISTIC REGRESSION MATHEMATICAL FOUNDATION USE CASES AND EVALUATION METRICS SUCH AS ACCURACY,CONFUSION MATRIX
SUNDAY				
TOTAL HOURS			8 HR	

Signatures:

_____
Intern

_____
Internship Supervisor

PLEASE RETURN TO: Internship Guide

WEEKLY INTERNSHIP LOG

INTERN WEEKLY REPORT

Intern: **MOHAMMED AAQIL RAYYAN**   Week Beginning: **07/04/2025** To: **13/042025**

Position: **DATA ANALYTICS INTERN**

Employer: **QUEST INNOVATIVE SOLUTIONS**

Internship Supervisor's Name: **THOYYIBA PK**   Phone: **9778160684**

Internship Coordinator: **JASHEENA**

DAY	TIME IN	TIME OUT	TOTAL HOURS	DUTIES PERFORMED
MONDAY	9:00 AM	11:00 AM	2 HR	UNSUPERVISIED LEARNING,FOCUSING ON CLUSTERING TECHNIQUES,UNDERSTANDING OF K-MEANS CLUSTERING ALGORITHM
TUESDAY	9:00 AM	11:00 AM	2 HR	THEORETICAL KNOWLEDGE OF GENERATIVE AI,EXPLORED TYPES OF ARTIFICIAL INTELLIGENCE.INDEPTH UNDERSTANDING OF LARGE LANGUAGE MODELS
WEDNESDAY	9:00 AM	12:00 PM	3 HR	APPLIED THE CONCEPTS OF GENERATIVE AI AND NLP BY DEVELOPING A SENTIMENT ANALYSIS APPLICATION,IMPLEMENTED THE APPLICATION IN PYTHON,DEPLOYED THE MODEL USING GRADIO
THURSDAY	9:00 AM	12:00 PM	3 HR	GAINED PRACTICAL EXPERIENCE IN DEVELOPING A CHATBOT USING OPEN AI's API.UNDERSTOOD THE WORKFLOW OF SENDING USER INPUTS TO THE API's AND RECEIVING INTELLIGENT,CONTEXT-AWARE RESPONSES
FRIDAY	OFF	OFF	OFF	
SATURDAY	9:00 AM	11:00 AM	2 HR	RECEIVED A DETAILED PROJECT BRIEFING,INCLUDING GUIDELINES ON HOW TO APPROACH AND EXECUTE A DATA ANALYTICS PROJECT
SUNDAY				
TOTAL HOURS			12 HR	

Signatures:

_____   _____
Intern   Internship Supervisor

PLEASE RETURN TO: Internship Guide

CERTIFICATIONS & OTHER ACTIVITIES

Project Title: AI Sentiment Analyzer

Project Description:

The AI Sentiment Analyzer is a web-based tool developed using Python that analyzes the sentiment of user-input text. It uses the TextBlob library to determine whether the sentiment is Positive, Negative, or Neutral based on polarity scores.

The user interface is built with Gradio, allowing real-time interaction where users can input sentences and instantly see the sentiment result. This project showcases my understanding of Natural Language Processing (NLP) and integrating ML models into user-friendly applications.

AI SENTIMENT ANALYSER

```
[2]: from textblob import TextBlob
     import gradio as gr # Used to quickly create web-based user interfaces (UIs) for machine learning models or Python functions.
     from openai import OpenAI # Used for: Accessing OpenAI's language models (like GPT-4), embeddings, etc., via API calls.

     C:\Users\user\AppData\Local\Programs\Python\Python313\Lib\site-packages\tqdm\auto.py:21: TqdmWarning: IProgress not found. Please update jupyter and
     idgets. See https://ipywidgets.readthedocs.io/en/stable/user_install.html
       from .autonotebook import tqdm as notebook_tqdm

[6]: def analyze_sentiment(text):
         sentiment_score = TextBlob(text).sentiment.polarity
         if sentiment_score > 0 :
             sentiment = "Positive"
         elif sentiment_score < 0 :
             sentiment = "Negative"
         else:
             sentiment = "Neutral"
         return sentiment

[11]: # Gradio UI
      iface = gr. Interface(fn=analyze_sentiment
                          ,inputs="text"
                          ,outputs="text"
                          ,title="AI Sentiment Analyzer"
                          ,description="Enter a sentence, and the AI will analyze its sentiment!")
      iface.launch()

      * Running on local URL: http://127.0.0.1:7860 ***
```

Output

```
* Running on local URL:  http://127.0.0.1:7860
* To create a public link, set `share=True` in `launch()`.
```

AI Sentiment Analyzer

Enter a sentence, and the AI will analyze its sentiment!

text	output
It was a Great day	Positive

Clear	Submit	Flag

27

CERTIFICATIONS & OTHER ACTIVITIES

Project Title: AI Chatbot

Project Description:

This project involves the development of an AI-powered chatbot using OpenAI's GPT-4o-mini model. The chatbot interacts with users in a conversational format, answering queries and maintaining context through a stored conversation history.

Built with Python, the system takes user input, generates responses using OpenAI's API, and continues the dialogue until the user exits. This chatbot demonstrates practical skills in API integration, prompt engineering, and building conversational AI systems.

```
AI CHATBOT

[1]: #Initialize OpenAI Client
     Client = OpenAI(
         api_key="sk-proj-21p4PW4cuos6PKN72wpdIeFmUtxkYy5w1-1ue4YEkBXea4ZuSKKHTJ6OcKK2fram43wn1sWm5oT3B1b4FJ_F5YYwBrYbwB8lO1t1Y_9V6Ub2mulngS7%6NQu6pG-Mp3qL
     ◀ ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━

[4]: #Conversation History
     conversation_history = []
     def chatbot_response(user_input):
         #Add user input to history
         conversation_history.append({"role":"user","content": user_input})

         #Get AI Response
         completion = Client.chat.completions.create(
             model="gpt-4o-mini",
             store=True,
             messages=conversation_history
         )

         #Extract Reply
         reply = completion.choices[0].message.content

         #Add AI Response To History
         conversation_history.append({"role":"assistant","content":reply})

         return reply

[5]: #Chat Loop
     print("AI ChatBot: Hello! How Can I assist You Today?")
     while True:
         user_input=input("You:")
         if user_input.lower() in ["exit","quit","bye"]:
             print("AI ChatBot: Goodbye!")
             break
         response = chatbot_response(user_input)
         print("AI ChatBot:",response)
```

Output

```
AI ChatBot: Hello! How Can I assist You Today?

You: heyy

AI ChatBot: Hey! How can I assist you today?

You: what are the top five Cars in the world?

AI ChatBot: The "top" cars in the world can mean different things depending on the criteria-performance, luxury, sales, or innovation. Here are five st
andout cars from various categories based on performance and acclaim as of my last update in October 2023:

1. **Bugatti Chiron Super Sport 300+** - Known for its extreme performance and luxury, the Chiron has a top speed that exceeds 300 mph and represents o
ne of the pinnacle achievements in automotive engineering.

2. **Tesla Model S Plaid** - Renowned for its electric performance, the Model S Plaid combines rapid acceleration (0-60 mph in under 2 seconds), long-r
ange capabilities, and advanced technology, making it a significant player in the electric vehicle segment.

3. **Porsche 911 (992)** - A classic sports car that combines timeless design with modern technology, the latest 911 is praised for its driving dynamic
s, performance, and everyday usability.

4. **Ferrari SF90 Stradale** - A hybrid supercar that balances incredible speed, advanced hybrid technology, and stunning design. It features a powerfu
```

Project Title: UEFA Champions League Data Visualization Dashboard
Project Overview:
This project focuses on the creation of an interactive and visually compelling data visualization dashboard that analyzes historical performance in the UEFA Champions League—Europe's premier club football tournament. The dashboard provides a comprehensive overview of top-performing players, clubs, and coaches across multiple performance metrics including appearances, goals scored, and titles won.
Built using powerful data visualization tools (such as Power BI or Tableau), this project involved collecting, cleaning, and transforming raw football data into structured insights that fans, analysts, and stakeholders can understand at a glance.

Dashboard Features:
1. Player Statistics:
- Top 10 Players by Appearances: Lists players like Cristiano Ronaldo, Iker Casillas, and Lionel Messi based on total Champions League appearances.
- Top 10 Goal Scorers: Displays top scorers such as Cristiano Ronaldo, Messi, and Robert Lewandowski along with their total goals.
- Seasonal Records:
  - Most goals in one season: Cristiano Ronaldo (17)
  - Most appearances in a single season: Raúl González (15)
2. Club Statistics:
- Top Clubs by Matches Played: Visualizes clubs with over 200 games, including Real Madrid, Bayern Munich, and Barcelona.
- Top Clubs by Goals Scored: Displays clubs with 300+ goals, with Real Madrid leading at 1,047 goals.
- Goals For vs Goals Against: Compares offensive and defensive strengths of leading clubs.
3. Coach Statistics:
- Top 10 Coaches by Total Appearances: Includes legendary managers like Carlo Ancelotti, Sir Alex Ferguson, and Arsène Wenger, reflecting their long-term impact on the competition.
4. Title Rankings:
- All-Time Winners Ranking: Shows number of titles won by clubs, with Real Madrid far ahead at 14 titles, followed by AC Milan (7) and Liverpool (6).
- Total Titles by Country: A map visualization that shows national dominance, highlighting countries like Spain, Italy, England, and Germany.

Tools and Technologies Used:
- Data Visualization Tool: Power BI / Tableau
- Data Sources: UEFA official statistics, football databases (e.g., Transfermarkt, FBref)
- Data Processing: Microsoft Excel / Python (for data wrangling)
- Design Elements: Custom color palette, dynamic charts (bar, radial, map), filters

Skills Demonstrated:
- Data Collection & Cleaning
- Dashboard Design Principles
- Analytical Thinking & Metric Selection
- Storytelling with Data
- Sports Data Interpretation

Conclusion:

This project illustrates the power of data visualization in sports analytics. It helps uncover patterns and comparisons among elite football players, managers, and clubs. Whether used by sports journalists, analysts, or football fans, this dashboard offers a data-driven view of the Champions League's rich history.

CONCLUSION

My two-month internship at Quest Innovative Solutions Pvt. Ltd. was an incredibly valuable experience that allowed me to apply the concepts I learned in my BCA program—especially in the areas of Python programming and data analytics—to real-world problems. This hands-on exposure helped me grow both technically and professionally.In the beginning, I focused on the basics of Python . I worked on industry-relevant tools, enhanced my problem-solving abilities, and developed a deeper understanding of machine learning, data handling, and AI integration. These skills were essential for preparing the data for machine learning tasks.Apart from technical skills, I also improved my problem-solving, time management, and teamwork abilities. Attending meetings, sharing updates, and working with experienced mentors gave me a sense of what it's like to be part of a professional team. I became more confident in explaining my ideas and asking questions when I needed help.Overall, this internship has helped me connect classroom learning with industry practices. I now feel more confident in working with data. I am very thankful to Quest Innovative Solutions ,my mentors **Ms. Thoyyiba PK and Ms. Jasheena** ,and my academic guide **Ms Fathima Thoufeera** for their support and guidance.This experience has strengthened my passion for data analytics and artificial intelligence and prepared me to pursue advanced opportunities in this field. I'm truly grateful for this journey and excited about what lies ahead

Mohammed Aaqil Rayyan
Robotics , Machine Learning And Artificial Intelligence
Yenepoya Institute of Arts, Science, Commerce, and Management

REFERENCES

- **NumPy. Retrieved from :** https://numpy.org
- **pandas. Retrieved from** : https://pandas.pydata.org
- **Matplotlib. Retrieved from :**https://matplotlib.org
- **Seaborn. Retrieved from :** https://seaborn.pydata.org
- **OpenAI. Retrieved from :** https://openai.com
- **Gradio. Retrieved from** : https://gradio.app
- **Tableau. Retrieved from :** https://www.tableau.com

REMARKS

I have successfully completed a two-month internship that focused on practical applications of Python programming and data analytics with active involvement in solving real-world problems . This experience allowed me to strengthen my technical foundation, improve my coding practices, and gain deeper insights . During the internship, I worked extensively on data preprocessing, including data cleaning, handling missing values, feature engineering, and exploratory data analysis (EDA) using libraries such as Pandas, NumPy, Matplotlib, and Seaborn. These steps were crucial in understanding the data and preparing it for meaningful analysis.