Reporting No:1 Week No:1 From :15/05/2025 To :22/05/2025 College ID**: 23AIML010**

Project Title: Exploratory Data Analysis and Customer Segmentation using Python

# WEEKLY REPORT

**Work done in last week ( Attach supporting Documents):**

1.**Studied the fundamentals of Exploratory Data Analysis (EDA)** and **Inferential Statistics**, with focus on their application in solving real-world problems such as financial market analysis and consumer behavior modeling.

2. **Watched curated educational videos related to sentiment analysis and customer data analytics, including:**

* [What is EDA? | Exploratory Data Analysis with Python](https://www.youtube.com/watch?v=Za0C2Zxxb0g)
* [Inferential Statistics: Introduction and Concepts](https://www.youtube.com/watch?v=0zZYBALbZgg)
* [Customer Segmentation in Python using KMeans](https://www.youtube.com/watch?v=6Ejga4kJUts)
* [Sentiment Analysis using Python | NLP Tutorial](https://www.youtube.com/watch?v=Yk9IEpRjcLQ)

3. **Explored and worked with industry-standard tools and techniques for EDA and segmentation:**

* Used **pandas** and **matplotlib** for cleaning and visualizing datasets.
* Performed **statistical analysis** including mean, variance, correlation, and outlier detection.
* Used **K-Means Clustering** from scikit-learn for unsupervised segmentation of customer data.
* Created visual outputs using **seaborn** and **plotly** for intuitive representation of clusters and insights.

**Reason for incomplete work:**  **N/A – All planned tasks completed for Week 1**.

**Plans for next week:**

**1. Project: Predicting House Prices (Linear Regression)**

* **Goal: Predict house prices using multiple numerical features.**
* **Key Learnings: Multiple Linear Regression, model evaluation.**
* **Workflow:**
  + **Collect and clean data**
  + **Select relevant features**
  + **Train model using Scikit-Learn**
  + **Evaluate with MSE, RMSE, R²**
  + **Visualize actual vs. predicted prices**

**2. Project: Wine Quality Prediction**

* **Goal: Predict wine quality based on chemical properties.**
* **Dataset: Includes features like acidity, alcohol, and density.**
* **Approach:**
  + **Preprocess and explore data**
  + **Train 3 classifiers (e.g., Logistic Regression, Random Forest, SVM)**
  + **Evaluate using accuracy, F1-score, and confusion matrix**
  + **Visualize feature impact and model performance**

**References:**

## Oasis Infobyte : <https://oasisinfobyte.com/>

## Clustering in Machine learning : <https://www.geeksforgeeks.org/clustering-in-machine-learning/>

## Signature of External Guide Signature of Internal Guide

A black text on a white background

AI-generated content may be incorrect.

**Student Id: 23AIML010**

**Student Name: OM CHOKSI**