**Project Report**

* **Project Title:** Customer Segmentation for a Retail Store
* **Project Manager:** Satyam Maheshwari
* **Start Date:** 13-07-24
* **End Date:** 17-07-24
* **Objectives:** To segment customers into distinct groups based on their purchasing behavior.
* **Scope:**
  + Data cleaning
  + EDA
  + Customer Segmentation using K-Means
  + Visualization using Matplotlib and Power BI
* **Technical Requirements Document (TRD)**
* **Data Sources:** Mall Customers dataset
* **Technologies:**
  + Python
  + Jupyter Notebook
  + Matplotlib
  + Seaborn
  + Scikit-learn
  + Power BI
* **Architecture:**
  + Data preprocessing
  + EDA
  + Clustering
  + Visualization
* **Data Flow:** Import data → Clean data → Analyze data → Segment customers → Visualize results
* **Deliverables:**
  + - Conclusions

{Insights & Recommendations}

* **Tasks:**

**1. Data Loading and Initial Inspection**

The dataset 'Mall\_Customers.csv' is loaded into a pandas Data Frame, and the first 10 rows are displayed to get an initial look at the data structure and values.

**2. Handling Missing Values**

* **Age Column:** Missing values in the 'Age' column are filled with the mean age.
* **Gender Column:** Missing values in the 'Gender' column are filled with the mode (most frequent value) of the column.

**3. Data Cleaning and Transformation**

* Columns are renamed for better readability: CustomerID, Gender, Age, Annual Income, Spending Score.
* The 'Gender' column is encoded numerically: 0 for Male and 1 for Female.
* All remaining missing values are confirmed to be handled.

**4. Exploratory Data Analysis (EDA)**

A statistical summary of the dataset is provided, giving insights into measures such as mean, median, standard deviation, etc.

**5. Visualizing Distributions**

* **Age Distribution:** A histogram with a kernel density estimate (KDE) overlay is created to show the age distribution of customers.
* **Annual Income Distribution:** A histogram with KDE shows how customers' annual incomes are distributed.
* **Spending Score Distribution:** A histogram with KDE illustrates the distribution of customers' spending scores.

**6. Visualizing Relationships**

A scatter plot is created to show the relationship between Annual Income and Spending Score, with points colored by Gender, revealing any patterns based on gender differences.

**7. Clustering with K-Means**

* **Feature Selection:** Relevant features (Age, Annual Income, Spending Score) are selected for clustering.
* **Standardization:** Features are standardized to ensure equal weighting during clustering.
* **K-Means Clustering:** The K-Means algorithm is applied with 5 clusters to segment customers into distinct groups.
* **Cluster Visualization:** A scatter plot visualizes the resulting customer segments, colored by cluster, highlighting distinct groupings based on spending and income.
* **Conclusion:**

Through this analysis, we have cleaned and preprocessed the data, visualized key distributions and relationships, and applied K-Means clustering to identify distinct customer segments.

The visualizations indicate clear patterns in customer spending behavior and income levels.

These insights can help tailor marketing strategies and improve customer relationship management by targeting specific segments based on their characteristics and spending habits.

Overall, the approach effectively segments customers and provides actionable insights for business decisions.

* **Charts & Code Snippets:**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

Description automatically generated**

**A screenshot of a computer

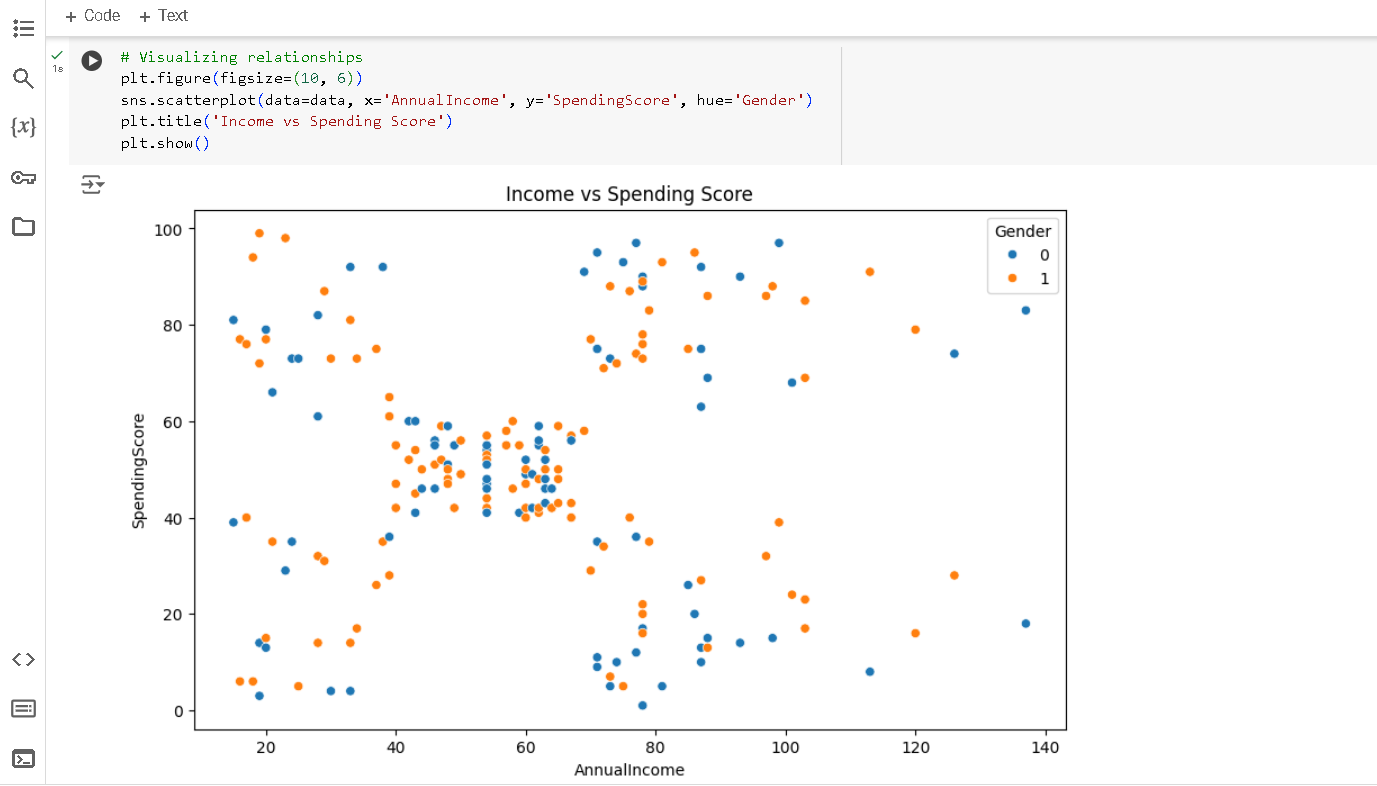
Description automatically generated**

**A screenshot of a computer program

Description automatically generated**

**A screenshot of a graph

Description automatically generated**

****

**A screenshot of a computer screen

Description automatically generated**

**A chart with colorful dots

Description automatically generated**

A blue and red circle

Description automatically generated