

Final Report:

1. Executive Summary:

1.1 Summary of the Project

The Mall Customers Segmentation project aimed to understand the different profiles of customers visiting the mall to enable more targeted marketing strategies. By segmenting customers based on demographic and purchasing behaviour data, the project sought to improve customer satisfaction and boost sales.

1.2 Objectives

- To segment customers into distinct groups based on their demographics and purchasing behaviour.
- To enhance targeted marketing strategies for each customer segment.
- To improve overall customer satisfaction and increase sales.

1.3 Outcomes

- Successful segmentation of customers into actionable groups.
- Development of visualizations and dashboards to represent customer segments.
- Provision of actionable recommendations for targeted marketing strategies.

2. Introduction:

2.1 Project Background

The mall management faced challenges in understanding the diverse profiles of their customers, leading to generalized and less effective marketing strategies. To address this issue, the Mall Customers Segmentation project was initiated to gain deeper insights into customer behavior and preferences.

2.2 Objectives

- To collect and analyze customer data to identify distinct customer segments.
- To use clustering techniques to categorize customers.
- To visualize the segments and provide recommendations for targeted marketing.

3. Methodology:

Steps Taken to Achieve the Objectives

3.1 Data Collection

- Collected the mall customers dataset, which included demographic information (age, gender, income) and purchasing behaviour (spending scores).

3.2 Data Cleaning

- Handled missing values, removed duplicates, and normalized data to ensure consistency and accuracy.

3.3 Exploratory Data Analysis (EDA)

- Performed descriptive statistics and visualizations to understand data distributions and relationships.

3.4 Clustering

- Applied the K-means clustering algorithm to segment customers into distinct groups based on their demographics and purchasing behaviour.

3.5 Visualization

- Created visualizations to represent the customer segments using Matplotlib and Seaborn.
- Developed interactive dashboards in Power BI to provide a comprehensive view of the customer segments.

3.6 Documentation

- Documented the data analysis and clustering process.
- Prepared user guides and technical documentation to ensure the solution's usability.

4. Results:

Findings from Data Analysis and Segmentation

4.1 Descriptive Statistics and EDA

- The dataset consisted of 200 customers with attributes including age, gender, annual income, and spending score.
- Initial EDA revealed key trends and patterns in the data, such as age distribution and spending behavior.

4.2 Clustering Results

- Applied the K-means algorithm with an optimal number of clusters determined through the elbow method.
- Segmented customers into five distinct groups, each with unique characteristics in terms of age, income, and spending score.

4.3 Visualization of Segments

- Created scatter plots and bar charts to visualize the customer segments.
- Developed Power BI dashboards to allow interactive exploration of the segments.

4.4 Key Customer Segments

- **Segment 1:** Young, high spenders with moderate income.

- **Segment 2:** Middle-aged, high-income customers with high spending scores.
- **Segment 3:** Older customers with low spending scores and moderate income.
- **Segment 4:** Young customers with low income and low spending scores.
- **Segment 5:** Middle-aged customers with moderate income and spending scores.

5. Conclusion:

5.1 Key Insights

- The segmentation revealed distinct customer profiles that can be targeted with specific marketing strategies.
- Younger, high-spending customers could be targeted with promotions for trendy products and events.
- High-income, high-spending middle-aged customers may respond well to premium offerings and loyalty programs.

5.2 Recommendations

- Implement targeted marketing campaigns for each customer segment to increase engagement and sales.
- Use the Power BI dashboards to monitor segment performance and adjust strategies as needed.
- Regularly update the customer data and refine the segmentation model to maintain accuracy and relevance.

6. Appendices:

6.1 Additional Charts

- Histograms, scatter plots, and other visualizations used in the EDA and clustering process.

6.2 Code Snippets

- Python code for data cleaning, EDA, clustering, and visualization

✓ Loading the Dataset

```
import pandas as pd
```

```
data = pd.read_csv("Mall_Customers.csv")  
print(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 x 5 columns]
```

✓ 1. Bar Plot of Total Spending Scores by Gender

```
import matplotlib.pyplot as plt  
import seaborn as sns
```

```
# Bar plot of total spending scores by gender  
plt.figure(figsize=(8, 6))  
sns.barplot(x=total_spending_by_gender.index, y=total_spending_by_gender.values, palette='viridis')  
plt.xlabel('Gender')  
plt.ylabel('Total Spending Score')  
plt.title('Total Spending Score by Gender')  
plt.show()
```

▼ Applying KMeans Algorithm

```
▶ from sklearn.cluster import KMeans
   from sklearn.preprocessing import StandardScaler

   # Feature selection
   features = data[['Age', 'Annual Income', 'Spending Score']]

   # Standardizing the features
   scaler = StandardScaler()
   scaled_features = scaler.fit_transform(features)

   # Applying K-Means clustering
   kmeans = KMeans(n_clusters=5, random_state=42)
   data['Cluster'] = kmeans.fit_predict(scaled_features)

   # Evaluating cluster quality
   plt.figure(figsize=(10, 6))
   sns.scatterplot(data=data, x='Annual Income', y='Spending Score', hue='Cluster', palette='viridis')
   plt.title('Customer Segments')
   plt.show()
```