

Final Report

Executive Summary

1.1 Summary of the Project

The Mall Customers Segmentation project was designed to understand the various profiles of customers visiting the mall, allowing for more effective and personalized marketing strategies. By segmenting customers based on their demographic and purchasing behavior data, the project aimed to enhance customer satisfaction and increase sales.

1.2 Objectives

- Segment customers into distinct groups based on demographics and purchasing behaviour.
- Improve targeted marketing strategies for each customer segment.
- Enhance overall customer satisfaction and drive sales growth.

1.3 Outcomes

- Successfully segmented customers into actionable groups.
- Developed visualizations and dashboards to represent customer segments.
- Provided actionable recommendations for targeted marketing strategies.

Introduction

2.1 Project Background

The mall management struggled to understand the diverse profiles of their customers, leading to generalized and less effective marketing strategies. The Mall Customers Segmentation project was initiated to gain deeper insights into customer behavior and preferences.

2.2 Objectives

- Collect and analyze customer data to identify distinct segments.
- Use clustering techniques to categorize customers.
- Visualize the segments and provide recommendations for targeted marketing.

Methodology

Steps Taken to Achieve the Objectives 3.1 Data Collection

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

3.2 Data Cleaning

- Addressed missing values, removed duplicates, scaled data using Min-Max Scaler, and normalized data for consistency and accuracy.

3.3 Exploratory Data Analysis (EDA)

- Conducted 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 using Matplotlib and Seaborn to represent customer segments.

- Developed interactive dashboards in Power BI for 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.

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 behaviour.

4.2 Clustering Results

- Applied the K-means algorithm with the 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 for 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.

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.

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 of Dataset

```
[3] import pandas as pd
data = pd.read_csv('/content/Mall_Customers.csv')
print(data.head(6))
```

| | CustomerID | Genre | 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 |
| 5 | 6 | Female | 22 | 17 | 76 |

✓ Brief Overview of the Dataset

```
[ ] print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 200 entries, 0 to 199
Data columns (total 5 columns):
#   Column                Non-Null Count  Dtype
---  -
0   CustomerID            200 non-null   int64
1   Genre                 200 non-null   object
2   Age                   200 non-null   int64
3   Annual Income (k$)    200 non-null   int64
```

7. Scatterplot of Annual Income vs Spending Score

```
# Scatter plot of Annual Income vs Spending Score
plt.figure(figsize=(10, 6))
plt.scatter(data['Annual Income (k$)'], data['Spending Score (1-100)'], alpha=0.7)
plt.title('Annual Income vs Spending Score')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score')
plt.show()
```

6. Scatterplot of Age vs Spending Score

```
[ ] # Scatter plot of Age vs Spending Score
plt.figure(figsize=(10, 6))
plt.scatter(data['Age'], data['Spending Score (1-100)'], alpha=0.7)
plt.title('Age vs Spending Score')
plt.xlabel('Age')
plt.ylabel('Spending Score')
plt.show()
```

8. 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()
```

✓ K-means clustering

Applying K-Means Clustering to segment customers into distinct groups.

```
[ ] from sklearn.cluster import KMeans
kmeans = KMeans(n_clusters=5, random_state=0)
X = data[['Annual Income (k$)', 'Spending Score (1-100)']]
kmeans.fit(X)
data['Cluster'] = kmeans.labels_
```

 /usr/local/lib/python3.10/dist-packages/sklearn/cluster/_kmeans.py:870: FutureWarning: The default value of `n_init` will change from 10 to 1 in the future. This will also affect the results of the fit. Please set `n_init` to a number greater than 1 to silence this warning. If you wish to keep the old behavior, use `n_init=1`.

```
import matplotlib.pyplot as plt
plt.scatter(data['Annual Income (k$)'], data['Spending Score (1-100)'], c=data['Cluster'], cmap='viridis')
plt.title('Customer Segmentation')
plt.xlabel('Annual Income (k$)')
plt.ylabel('Spending Score (1-100)')
plt.show()
```