Final Report

Executive Summary :

This report details a customer segmentation project aimed at identifying distinct groups within a customer base to enhance targeted marketing strategies. The project involved collecting and cleaning customer data, performing exploratory data analysis (EDA), applying clustering algorithms, and interpreting the results to define actionable segments.

The key outcomes include the identification of six unique customer segments

Introduction:

Customer behaviour is crucial for marketing strategies and boost revenue. This project is aimed at identifying different customer groups and their behaviour.

Objectives:

Identify distinct customer segments.

Understand the characteristics and behaviors of each segment.

Provide recommendations for targeted marketing strategies.

Methodology:

Data Collection:

- Gathered customer data, including ID, gender, age, income, and spending score.

Data Cleaning:

- Removed duplicates and handled missing values.
- Standardized data formats and normalized numerical features.
- Addressed outliers appropriately.

Exploratory Data Analysis (EDA):

- Performed descriptive statistics to summarize data.
- Visualized distributions and correlations between features.

- Identified key patterns and insights.

Clustering:

- Chose the K-means clustering algorithm.
- Determined the optimal number of clusters using the Elbow method and Silhouette score.
- Trained the K-means model and assigned cluster labels to data points.
- Analyzed and interpreted the characteristics of each cluster.

Visualization:

- Created scatter plots, histograms, and stacked charts to represent cluster distributions and characteristics.
- Creating Dashboard for interactivity
- Ensured visualizations were clear and informative for stakeholders.

Documentation:

- Documented all steps, methodologies, and insights.
- Prepared a comprehensive final report summarizing the findings and recommendations.

Results:

Cluster 0: *Low to moderate spending middle class males*
Gender: Predominantly Male
Age Range: 35 - 71
Income Range: $19k - $71k
Spending Score Range: 4 – 60

Description: This cluster consists of mature males with diverse ages and incomes. Their spending habits range from conservative to moderate.

Cluster 1: *Low spending rich males*
Gender: Predominantly Male
Age Range: 19 - 59
Income Range: $71k - $137k
Spending Score Range: 1 – 36

Description: This cluster includes males with high incomes but exhibit low to moderate spending. They prioritize value and long-term investments.

Cluster 2: *High spending young males*
Gender: Predominantly Male
Age Range: 18 - 40
Income Range: $15k - $101k
Spending Score Range: 39 – 97

Description: This cluster consists of young males with varying incomes but consistently high spending scores, indicating a propensity for discretionary spending.

Cluster 3: *Moderate to high spending rich adults*
Gender: Predominantly Female
Age Range: 27 - 45
Income Range: $69k - $137k
Spending Score Range: 23 – 95

Description: This cluster comprises females with high incomes and moderate to high spending scores. They are inclined towards luxury and premium products.

Cluster 4: *Young middle-class females with moderate to high spending*
Gender: Predominantly Female
Age Range: 18 - 35
Income Range: $16k - $72k
Spending Score Range: 29 – 99

Description: This cluster includes young females with moderate to high incomes and spending scores, indicating a preference for trendy and lifestyle products.

Cluster 5: *low to moderate spending old females*
Gender: Predominantly Female
Age Range: 34 - 68
Income Range: $18k - $101k
Spending Score Range: 5 – 59

Description: This cluster consists of older females with varied incomes and cautious spending habits, preferring essential and value-driven purchases.

Conclusion:

- **Considerable Insights:**
  Majority of customer base:
  - Are females (slightly more)
  - Have low to moderate income
  - Spend moderately

- **Recommendation based on customer groups:**

  Cluster 0: *Low to moderate spending middle class males*
  Tailor marketing strategies based on age and income segments. Offer a variety of products and services to accommodate different spending behaviors. Implement targeted loyalty programs and consider financial advisory services.

  Cluster 1: *Low spending rich males*
  Focus on savings and investment products. Offer premium products emphasizing value for money. Develop exclusive memberships and educational content on financial planning.

  Cluster 2: *High spending young males*
  Market trendy and lifestyle products. Offer subscription services and exclusive experiences. Utilize influencer marketing and engage through social media for effective outreach.

  Cluster 3: *Moderate to high spending rich adults*
  Highlight luxury and premium offerings. Develop exclusive memberships and VIP programs. Focus on personalized marketing and health and wellness products.

  Cluster 4: *Young middle-class females with moderate to high spending*
  Market affordable luxury and subscription services. Offer student and young professional discounts. Engage through social media and influencers for targeted outreach.

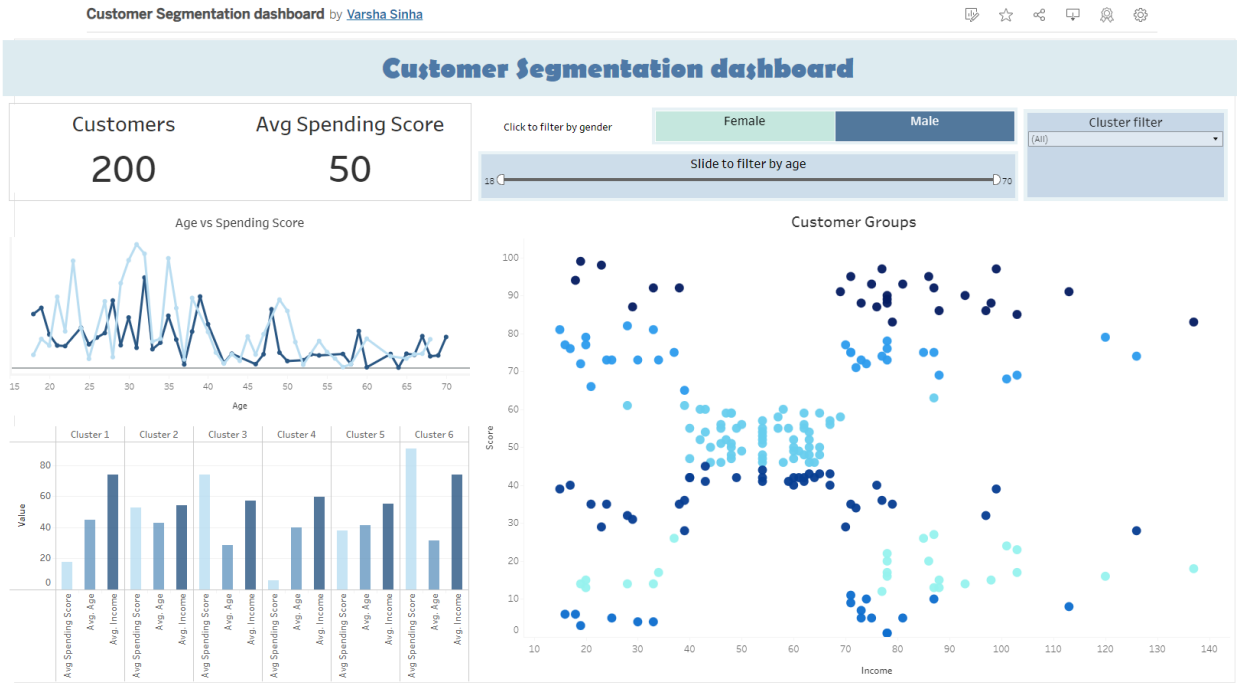
  Cluster 5: *low to moderate spending old females*
  Focus on essential products and services. Offer value-for-money options and senior discounts. Provide financial planning and health-related products and services.

Appendices :

Data : Mall Customers.csv

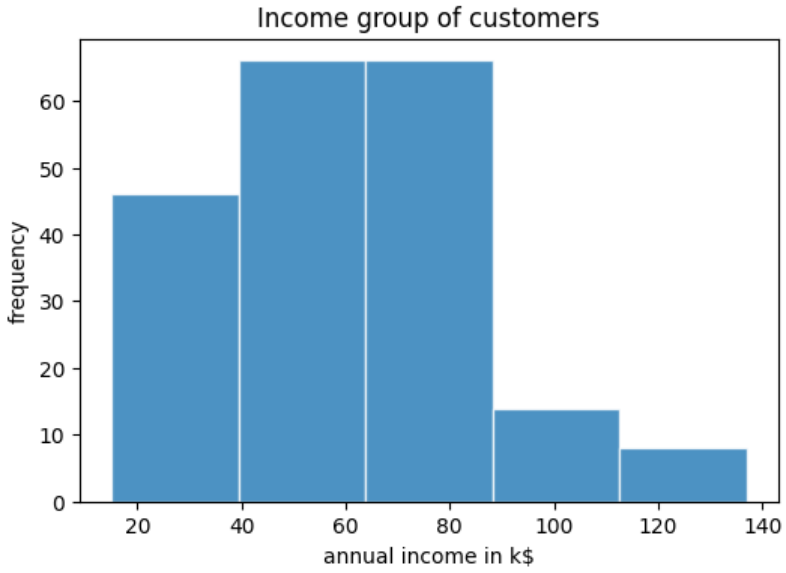
Insights Visualizations:

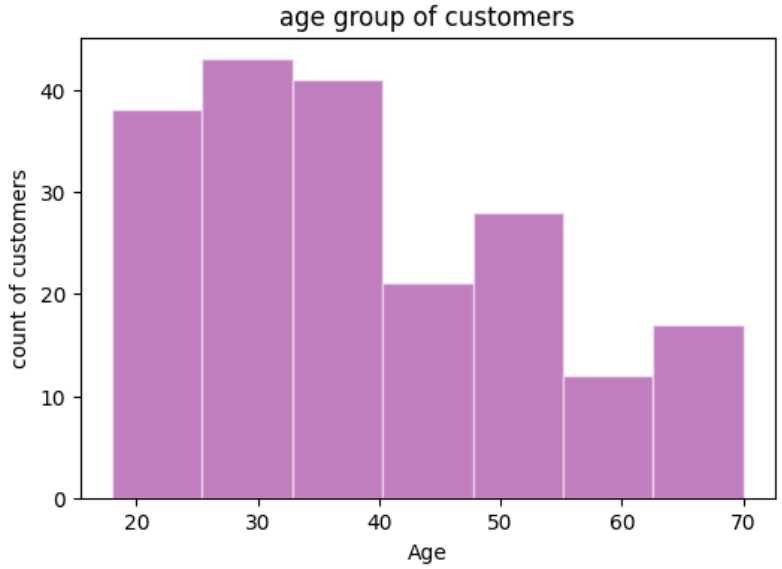
1. Dashboard for interactive interpretation

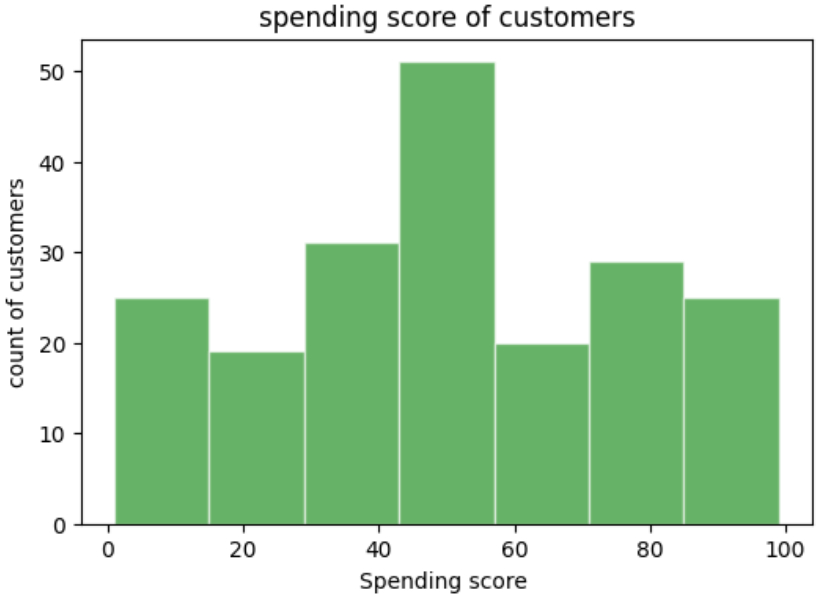


2. Summarizing customers purchasing behavior by age and gender

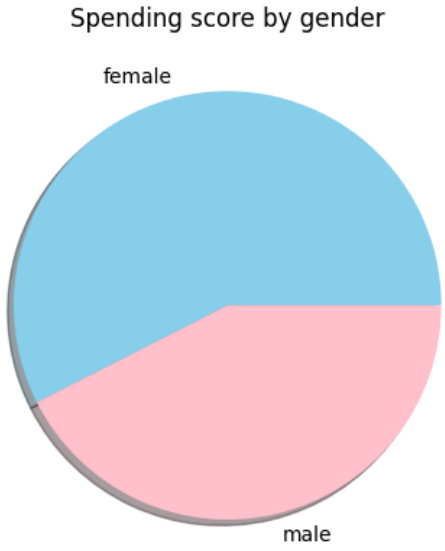
3. Data Distribution



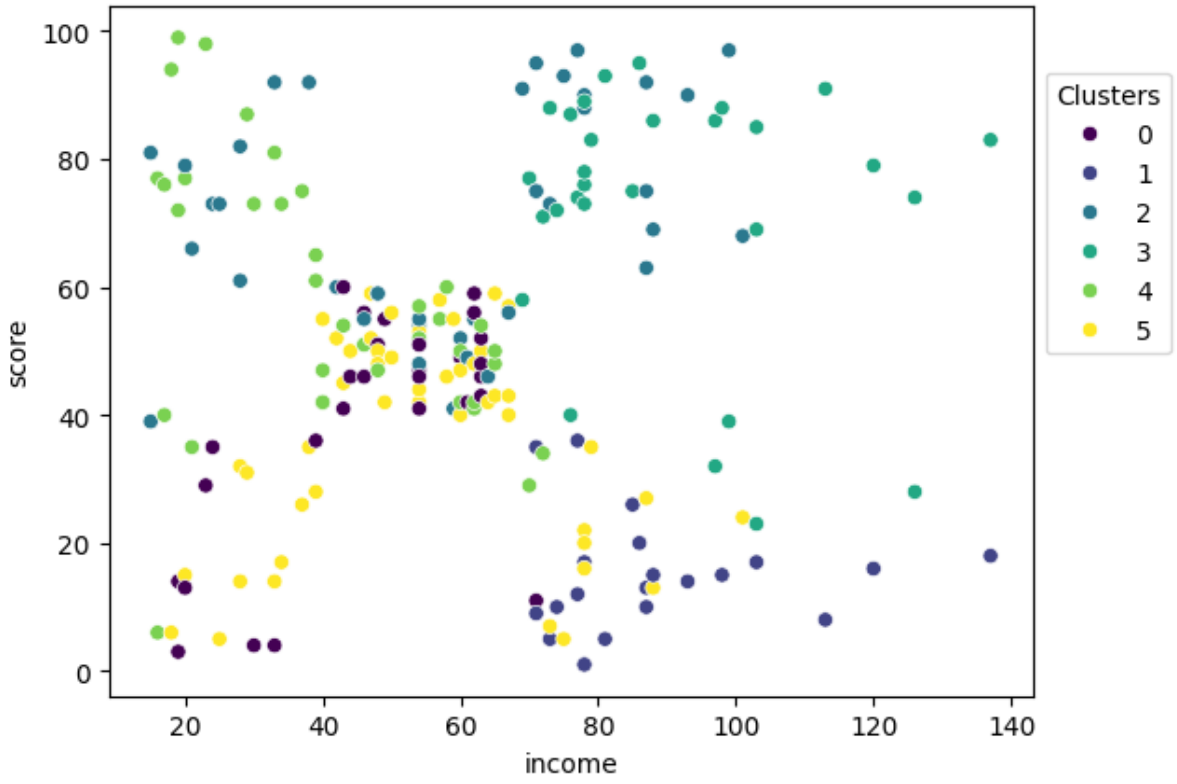

age group of customers
count of customers
Age
0
10
20
30
40
20
30
40
50
60
70


spending score of customers
count of customers
Spending score
0
10
20
30
40
50
0
20
40
60
80
100

4. Customer's gender analysis



5. Cluster

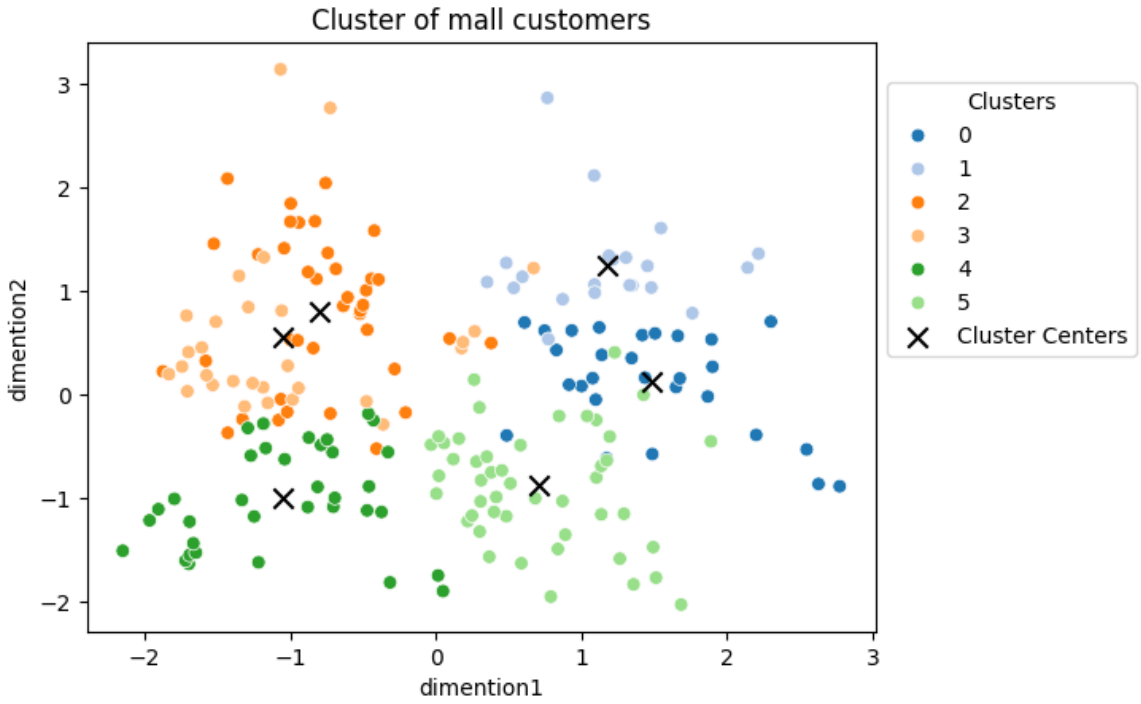


6. Evaluation of clusters by both, visualization and evaluation metrics

Cluster Evaluation

In [11]:
```
labels= Kmeans_model.predict(X)
print(f"inertia : {Kmeans_model.inertia_}")
print(f"silhouette score : {silhouette_score(X, Kmeans_model.labels_)}")
```

```
inertia : 277.3897401985191
silhouette score : 0.3347543475669217
```



References :

- Data cleaning :

https://github.com/varshas-08/Mall-Customer-Segmentation/blob/main/Data%20cleaning.ipynb

- Eda :

https://github.com/varshas-08/Mall-Customer-Segmentation/blob/main/EDA.ipynb

- Clustering :

https://github.com/varshas-08/Mall-Customer-Segmentation/blob/main/Clustering.ipynb

- Dashboard :

https://public.tableau.com/app/profile/varsha.sinha/viz/CustomerSegmentationdashboard_17212998480490/Dashboard1