**MALL CUSTOMERS SEGMENTATION**

**Introduction / Problem Statement**

"We aim to segment customers of a supermarket mall based on spending behavior to enable targeted marketing."

**Data Description**

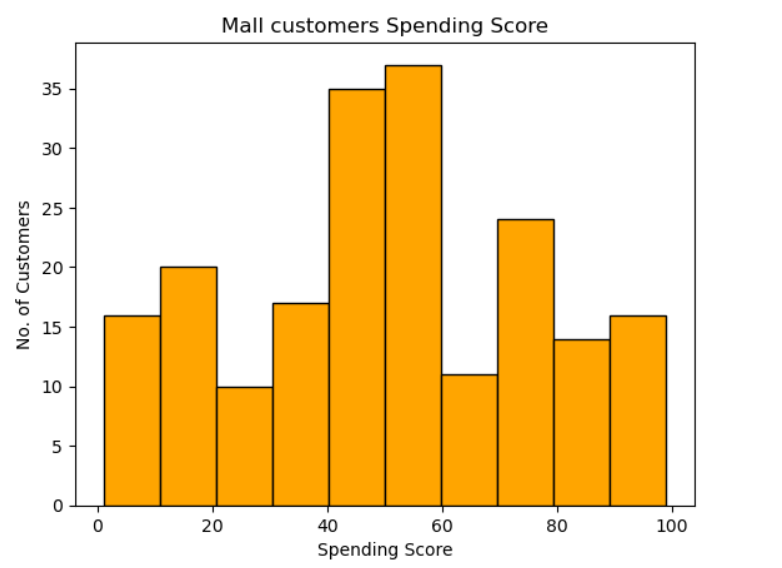
Number of customers - 200

Features used – Age, Gender, Annual Income, Spending Score

Target variable – Spending Score

A graph of age distribution

AI-generated content may be incorrect.A graph of a number of people

AI-generated content may be incorrect.

**Methodology**

1. Data cleaning

* Checked for missing values, duplicated values and possible outliers.
* Data integrity check using summary statistics
* Rows-to-columns eligibility check

1. Feature selection

* Transformed categorical variable, gender to dummy variable.
* Ensured non-collinearity of dependent variables.

A chart with yellow dots

AI-generated content may be incorrect.

* Selected features: Age, Annual Income, Gender (dummy)

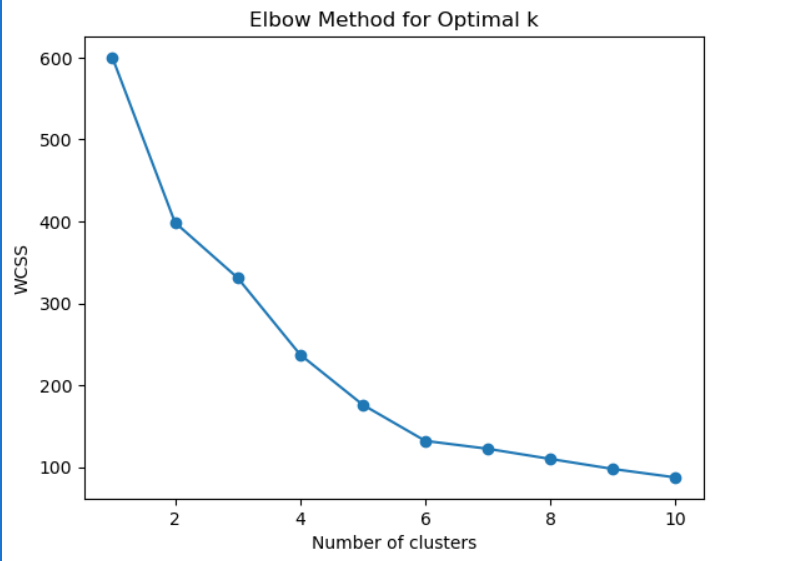
A screenshot of a graph

AI-generated content may be incorrect.

1. Scaling

* Standardization of numerical variables to avoid over-influence of high magnitude variables on the model, Eg: Annual Income.

1. Elbow Method for Optimal k



Using elbow method, we found the optimal value for hyperparameter, k is 6.

1. K-Means model fitting

* Trained k-means algorithm on the scaled dataset with optimal, k= 6, no, of clusters.

**Evaluation**

1. Silhouette Score – 0.42
2. Cluster sizes –

|  |  |
| --- | --- |
| **Cluster** | **Size** |
| 0 | 29 |
| 1 | 49 |
| 2 | 30 |
| 3 | 35 |
| 4 | 33 |
| 5 | 24 |

1. Inertia - 131.95
2. Summary table of average values for each cluster

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Average** | **Age** | **Annual Income** | **Spending Score** | **Gender** |
| **Cluster** |  |  |  |  |
| **0** | 58.41 | 52.34 | 37.03 | Male |
| **1** | 33.82 | 81.86 | 54.47 | Female |
| **2** | 28.13 | 32.23 | 57.20 | Female |
| **3** | 34.09 | 86.14 | 50.74 | Male |
| **4** | 53.52 | 50.24 | 42.00 | Female |
| **5** | 25.67 | 39.29 | 59.13 | Male |

**Visualization**

1. Power BI/ Tableau dashboard link:

**Insights**

* The spending score is high among young customers, with an average age of 36 years.

Insights:

* **Most Valuable Segment**: Cluster 3 has the highest spending score of 72.14, comprising of customers with age ranging from 27 to 45 and annual income from 70 to 140K.
* **Old customers** (clusters 0 & 5) are more **budget conscious** and spend less compared to their younger counterparts, irrespective of their annual incomes.
* **Potential value shoppers**: Cluster 1 has a significantly high income, but the least spending score.
* **Impulsive buyers**: Cluster 2 and 4 comprising of young customers (with combined avg. age, 27 years) are among the top spenders.
* Gender has no significant impact on the spending score (confirmed with a hypothesis test)

*Hint:*

1. *Analyse summary statistics of clusters,*
2. *Compare cluster sizes and balance,*
3. *Interpret cluster centroids,*
4. *Use visualizations to spots relationships,*
5. *Segment behavioral archetypes,*
6. *Assess model quality, and*
7. *Strategies based on insights.*

**Recommendations**

*Guidelines/ Tips:*

1. *Anchor recommendations for business goals.*
2. *Use the Insight-Impact-Action Framework.*
3. *Make recommendations SMART (Specific, Measurable, Achievable, Relevant, Time-bound)*
4. *Prioritize recommendations.*
5. *Tailor the language based on stakeholder type.*
6. *Include visuals to back up recommendations.*
7. *Create an executive summary slide*

*Final Checklist Before Presenting:*

1. *Does each recommendation tie back to a real, measurable insight?*
2. *Are the business impacts clearly stated?*
3. *Can the suggested actions be taken forward by the relevant team?*

Your Recommendations:

**Conclusion**