

Customer Segmentation Using Mall Customers Dataset

Approach

The project applied unsupervised machine learning to segment mall customers based on demographic and spending attributes. The dataset consisted of 200 entries with four key variables: CustomerID, Gender, Age, Annual Income (k\$), and Spending Score (1–100). Preprocessing steps included data cleaning (minimal, since no missing values were found), and feature selection focusing on Annual Income and Spending Score. K-Means clustering was applied, and the Elbow Method determined the optimal number of clusters to be 5. Scatter plots were used for visualization.

Challenges

1. Feature selection dilemma: including all variables did not yield clear clusters. 2. Optimal cluster identification required subjective judgment using the elbow method. 3. Interpretability: translating raw clusters into business insights required domain reasoning. 4. Limited dataset size (200 records) may limit generalizability.

Outcomes

The K-Means model identified 5 distinct customer segments: 1. Low Income – Low Spending → Price-sensitive customers. 2. High Income – Low Spending → Wealthy but disengaged customers. 3. Average Income – Average Spending → Moderate shoppers. 4. High Income – High Spending → Premium customers, highly valuable. 5. Low Income – High Spending → Aspirational shoppers, often younger. Business value: Enables targeted marketing, product placement, and customer retention strategies.

Conclusion

The project successfully applied unsupervised machine learning for customer segmentation. The results provide meaningful insights to inform data-driven marketing strategies despite challenges in interpretability and dataset limitations. Future improvements could include integrating purchase history, testing other clustering methods, and validating results with larger datasets.