

Project Title: Customer Segmentation Project

Objective:

The primary goal of this project is to analyze customer behavior and segment customers into different groups using clustering techniques. This segmentation helps the mall management understand the shopping patterns and preferences of their customers. By identifying key customer segments, the mall can target marketing strategies more effectively and improve customer satisfaction.

Dataset Overview:

The dataset consists of the following variables:

- **CustomerID:** A unique identifier for each customer.
- **Gender:** The gender of the customer (Male/Female).
- **Age:** The age of the customer.
- **Annual Income:** The customer's annual income in thousands of dollars.
- **Spending Score:** A score assigned to customers based on their spending habits.

Analysis Process:

1. Exploratory Data Analysis (EDA):

- **Distribution Analysis:**
 - A frequency count of customer gender was performed. This revealed an approximately equal distribution of male and female customers.
 - Age and annual income distributions were also analyzed using count plots. The age distribution shows a younger customer base, with the majority being between 20-40 years old. Annual income values were spread relatively evenly between \$15,000 and \$135,000.
- **Spending Score Analysis:**
 - Spending score analysis revealed customer tendencies toward higher or lower spending behavior, useful for differentiating between big spenders and more frugal customers.

2. Clustering Techniques: Two clustering techniques were applied in the project

a. Hierarchical Clustering:

- **Dendrogram:** A dendrogram was created to visualize the hierarchical relationships between customers. Various linkage methods were applied (Ward, Complete, and Average) to evaluate the structure of the clusters.
- **Agglomerative Clustering:** This technique was used to group customers based on their annual income and spending score. The optimal number of clusters was identified using the dendrogram and silhouette analysis, which measures how well each customer fits within its cluster.
- **Cluster Analysis:** Based on hierarchical clustering, customers were grouped into distinct segments. For example, one group could represent high-income individuals with low spending scores, while another group could include low-income customers with high spending scores.

b. KMeans Clustering:

- **Feature Selection:** KMeans clustering was applied to a subset of the data, specifically focusing on the customers' age and spending score.
- **Cluster Identification:** The KMeans algorithm divided customers into five distinct clusters, each representing a specific combination of age and spending habits. For instance, one cluster could represent younger customers with a high spending score, while another could indicate older customers with more moderate spending behavior.
- **Cluster Visualization:** The clusters were visualized using scatter plots, clearly showing the distinct groups of customers. Each cluster displayed different characteristics in terms of spending patterns, age, and income.

3. **Cluster Interpretation:** After segmenting the customers, the analysis focused on understanding the key characteristics of each cluster:
- **High Spenders:** A segment of customers was identified as high spenders, often in the younger age group with middle to high income.
 - **Low Spenders:** Another segment showed low spending scores despite varying income levels. This group may be more conservative with their purchases.
 - **Targeted Marketing Insights:** Based on the clustering results, mall management can focus marketing campaigns on specific clusters. For example, high-spending customers might receive personalized offers, while younger customers could be targeted with special promotions.

Conclusion:

The clustering analysis successfully segmented the customers into distinct groups, each with unique shopping behaviors. By analyzing factors such as age, income, and spending score, the mall can better understand its customers and tailor its marketing strategies. Hierarchical and KMeans clustering were both effective in identifying key patterns in customer behavior, with each method providing valuable insights.

Future Strategies for Enhancing Customer Insights

Based on the customer segmentation analysis in this project, several strategic actions and enhancements can be applied moving forward to optimize the use of these insights and further refine the analysis.

1. Personalized Marketing Campaigns

- **Target High-Spending Customers:** The high-spending cluster should be the focus of premium and personalized marketing campaigns. Offering exclusive discounts, loyalty programs, and special events for this segment could further enhance their shopping experience, potentially increasing revenue.
- **Engage Low-Spending Customers:** For customers with low spending scores but moderate or high incomes, incentives such as limited-time offers or bundled deals could help convert them into more frequent or higher-value shoppers.
- **Age-Based Promotions:** Younger customers with high spending scores could be targeted with trendy products and social media campaigns, while older, lower-spending customers might respond better to more traditional loyalty programs or in-store promotions.

2. Enhance Data Collection and Enrich Segmentation

- **Incorporate More Customer Attributes:** While age, income, and spending score provide valuable insights, adding more dimensions such as purchasing history, preferred shopping times, product categories, and online behavior would enable more precise segmentation.
- **Behavioral Segmentation:** Track customer behavior over time, such as the frequency of visits, type of purchases (luxury vs. necessity), and responsiveness to sales promotions. This can uncover new customer segments that could benefit from specific engagement strategies.
- **Customer Feedback and Satisfaction Data:** Incorporating customer feedback, surveys, or ratings can help identify customer satisfaction levels and areas where the shopping experience could be improved, particularly within different segments.