**Customer Segmentation Project Report**

**1. Context**

The dataset used in this project focuses on customer segmentation, a critical aspect of market basket analysis. The goal is to segment customers based on their demographic data (age, gender, annual income, spending score) using an unsupervised machine learning algorithm, specifically KMeans clustering. This helps to identify target customers who can be more easily converted through focused marketing strategies.

**2. Objective**

As the owner of a supermarket, the objective is to understand customer behaviors to aid the marketing team in identifying the most likely customers to convert. The segmentation is based on basic customer information such as:

* Customer ID
* Age
* Gender
* Annual income
* Spending score (a metric based on customer purchasing behavior)

**3. Data Overview**

* The dataset consists of 200 rows and 5 columns: Customer ID, Gender, Age, Annual Income (in k$), and Spending Score (1-100).
* No missing values were found during the data exploration.

**4. Data Exploration and Visualization**

* **Distribution Plots**:
  + Histograms were created for variables like Age, Annual Income, and Spending Score, showing the distribution of these variables.
  + A gender count plot showed a nearly equal distribution of male and female customers.
* **Relationship Analysis**:
  + Scatter plots revealed relationships between age, annual income, and spending score.
  + The comparison of "Age vs. Annual Income" and "Annual Income vs. Spending Score" by gender highlighted possible trends for each gender in terms of spending behavior.
* **Violin and Swarm Plots**: These were used to further explore the relationship between gender and variables like age, annual income, and spending score.

**5. Clustering with KMeans**

**5.1 Age and Spending Score**

* **KMeans Model**: Clustering was performed using the combination of Age and Spending Score.
* **Elbow Method**: To determine the optimal number of clusters, inertia was plotted against the number of clusters. The optimal number of clusters was found to be 4.
* **Cluster Visualization**: The clusters and centroids were visualized, showing clear segmentation of customers based on age and spending score.

**5.2 Annual Income and Spending Score**

* **KMeans Model**: Clustering was then applied to Annual Income and Spending Score.
* **Elbow Method**: The optimal number of clusters was determined to be 5, based on inertia calculations.
* **Cluster Visualization**: Clusters were plotted, displaying customer segments based on these two variables.

**5.3 Age, Annual Income, and Spending Score (3D Clustering)**

* **KMeans Model**: A final clustering was performed using all three variables: Age, Annual Income, and Spending Score.
* **Optimal Clusters**: The elbow method suggested 6 clusters for this case.
* **3D Visualization**: A 3D scatter plot was created to visualize the clusters across all three dimensions, providing a comprehensive view of customer segmentation.

**6. Insights and Marketing Implications**

* **Targeted Marketing**: Based on the clustering results, specific customer groups can be targeted for marketing strategies. For example, clusters with high spending scores and moderate incomes can be prioritized for premium services.
* **Customized Offers**: Each cluster reveals a different pattern in customer spending behavior, allowing the mall to design tailored promotions and offers based on income levels, spending habits, or age demographics.

**7. Conclusion**

The customer segmentation analysis successfully identified distinct customer groups using KMeans clustering. This segmentation can help the marketing team devise more effective strategies by focusing on the most promising customer groups. This analysis also demonstrates the value of applying unsupervised learning techniques to gain insights into customer behaviors and preferences.