# Customer Segmentation Report

## 1. Approach Used

### Exploratory Data Analysis (EDA):

- The dataset was loaded and inspected for missing values, duplicates, and inconsistencies.

- Summary statistics were computed to understand the distribution of features.

- Histograms were plotted to analyze the distribution of Age, Annual Income, and Spending Score.

### Feature Engineering & Preprocessing:

- The dataset was filtered to include relevant features: Annual Income and Spending Score.

- Standardization was applied using StandardScaler to normalize feature values for clustering.

### K-Means Clustering:

- The Elbow Method was used to determine the optimal number of clusters.

- K-Means was applied with k=5, and clusters were assigned to customers.

- A scatter plot was generated to visualize customer segments.

### Hierarchical Clustering:

- A Dendrogram was plotted to determine the number of clusters.

- Agglomerative Clustering was applied with Ward’s method.

- A scatter plot was generated to compare hierarchical clustering results with K-Means.

## 2. Challenges Faced

- Feature Selection: Initially, Age was included in clustering, but it introduced noise and was removed.

- Cluster Number Selection: The Elbow Method suggested an optimal cluster count, but fine-tuning was needed.

- Data Scaling: Without standardization, clustering results were biased due to income scale dominance.

- Hierarchical Clustering Performance: Computationally expensive for large datasets; K-Means was faster and scalable.

## 3. Model Performance & Improvements

### K-Means Clustering:

- Provided distinct customer segments, capturing differences in spending behavior.

- Faster and more efficient for large datasets.

- Future improvement: Experiment with other distance metrics and clustering algorithms (e.g., DBSCAN).

### Hierarchical Clustering:

- Useful for understanding relationships between clusters.

- More interpretable due to dendrogram visualization.

- Future improvement: Apply PCA for dimensionality reduction to improve efficiency.

## 4. Conclusion

This segmentation system successfully identified different customer groups based on spending patterns and income. Businesses can use these insights for targeted marketing, personalized offers, and better customer retention strategies. Future work may include integrating additional features like transaction history and online behavior to refine segmentation further.