Task 4 Customer Segmentation Report

Introduction

This report outlines the approach, challenges faced, and model performance for the customer segmentation project. The project leveraged data visualization techniques, clustering algorithms, and interactive visual tools to segment mall customers based on spending habits and demographics.

Approach Used

1. **Data Exploration and Cleaning:**
    - Imported and inspected the dataset.
    - Identified missing values and confirmed data integrity.
    - Visualized gender distribution and calculated mean income for both genders.
1. **Exploratory Data Analysis (EDA):**
    - Visualized Annual Income, Age, and Spending Score distributions.
    - Created heatmaps to identify correlations between key features.
    - Used boxplots to analyze gender-based spending patterns.
2. **Clustering Algorithms:**
    - Employed the K-Means Algorithm to segment customers using the Elbow Method for optimal cluster selection.
    - Applied Agglomerative Clustering and plotted a dendrogram to identify hierarchical clusters.
3. **Model Evaluation:**
    - Evaluated K-Means clusters using the Silhouette Score to assess model quality.
    - Visualized customer segments in 2D and 3D plots for better interpretation.
4. **Deployment:**
    - Integrated the solution into a Streamlit Web App for interactive data visualization and cluster exploration.

Challenges Faced

- Optimal Cluster Selection: Determining the ideal number of clusters required multiple visual assessments using the Elbow Method and Silhouette Score.
- Data Scaling Issues: To ensure accurate clustering, scaling features like income and spending score was crucial.
- Cluster Visualization Complexity: Displaying high-dimensional clusters effectively was challenging, but 3D visualization improved interpretability.

Model Performance and Improvements

- The K-Means Model with 5 clusters provided well-defined customer segments with distinct spending behaviors.
- The Silhouette Score confirmed cluster quality, indicating optimal separation.
- Future improvements may involve integrating additional customer attributes for enhanced segmentation insights.

Conclusion

The implemented solution effectively segments mall customers using clustering algorithms. The interactive web app allows businesses to analyze customer groups for targeted marketing strategies. Future enhancements could involve advanced clustering methods like DBSCAN or Gaussian Mixture Models (GMM) for improved performance.