Customer segmentation was performed using K-Means clustering on aggregated data from Customers.csv and Transactions.csv. The dataset was pre-processed to combine customer profiles and transaction histories, with features such as total spending, average transaction value, and transaction count engineered to represent customer behaviour. After analysing clustering performance for 2 to 10 clusters using the Elbow Method and Davies-Bouldin Index (DB Index), the optimal number of clusters was determined to be 5. This choice balanced cluster separation and compactness, achieving a DB Index of 1.33, indicating well-defined clusters. Additionally, the average Silhouette Score was 0.26, suggesting moderate cluster consistency. Each cluster represents a distinct customer group, such as high spenders, frequent buyers, and low engagement users. These insights can guide personalized marketing and improve customer retention strategies.

Visualization using PCA (Principal Component Analysis) confirmed the separation of clusters in a two-dimensional space, showing clear distinctions among the customer groups. These findings can be leveraged for targeted marketing campaigns, product recommendations, and resource allocation to improve customer satisfaction and retention. This analysis demonstrates the value of clustering in identifying actionable customer segments for data-driven decision-making.