Clustering Report

Clustering is an unsupervised learning problem that is a type of machine learning that uses algorithms to learn from unlabeled data. Here we did clustering with customers and transaction history. I first extracted features and did feature engineering to extract new features such as tenure, recency, etc. As well I linked customers with their history and items bought.

I found the optimal numbers of clusters using the minimum DB score as the minimum DB score is better. We also calculated silhouette scores and plotted clusters using PCA.

Key Points:
- Number of clusters: 6
- Davies-Bouldin Index: 1.1596
- Silhouette Score: 0.3101

Customer Clusters (PCA Visualization)
DB Index: 1.16, Silhouette Score: 0.31