

Report

- After merging the 3 datasets provided i have created some more features for clustering like Total Revenue, Total Quantity Sold,Purchase Frequency,Recency,Most Expensive ,Least Expensive Product Bought and Number of days since Signup Date
- After the features are produced i have created dummy features for Region and Category (Categorical) Variables.
- I have used MinMax scaler for the remaining Numerical Variables
- Then i have used Principal Component Analysis for Dimensionality reduction and found out the number of components most suitable is 6 using the Elbow Method
- I have used the K-means Clustering Algorithm
- I have found out that for clusters of 9, the DB Score is the minimum(0.84) and visualized the Clusters. However, As there are 6 features ,Visualisation in a 2d space is not clearly seen. The silhouette score for this instance is 0.555
- For proper Visualisation i have created 2 features by taking mean of odd and even PCA's respectively and put the synthetically designed features in the K-Means Algorithm.Where I have found that for 9 clusters the DB score is the minimum(0.40) with corresponding silhouette Score to be same 0.555.I have then visualised the clusters in the 2d plane