

SDSC3002

Credit Card Customer segmentation

PRESENTATION

A solid green horizontal bar spanning the width of the slide, located at the bottom.

Introduction

Introduction

1. Analysis of how credit card products should be developed
2. Based on current customer segments
3. Targets for existing products or new products.
4. Segmentation is critical because
 - A. Limited resources
 - B. Identify and serve customers

Description of the dataset

About the dataset

1. Usage behavior of about 9000 active credit card holders
2. In last 6 months
3. 18 behavioral variables such as:
 1. Balance
 2. Purchase behavior
 3. Credit limit

Key statistics of the dataset

1. Number of data: 8950
2. Number of column: 18
3. Number of NaN data: 313

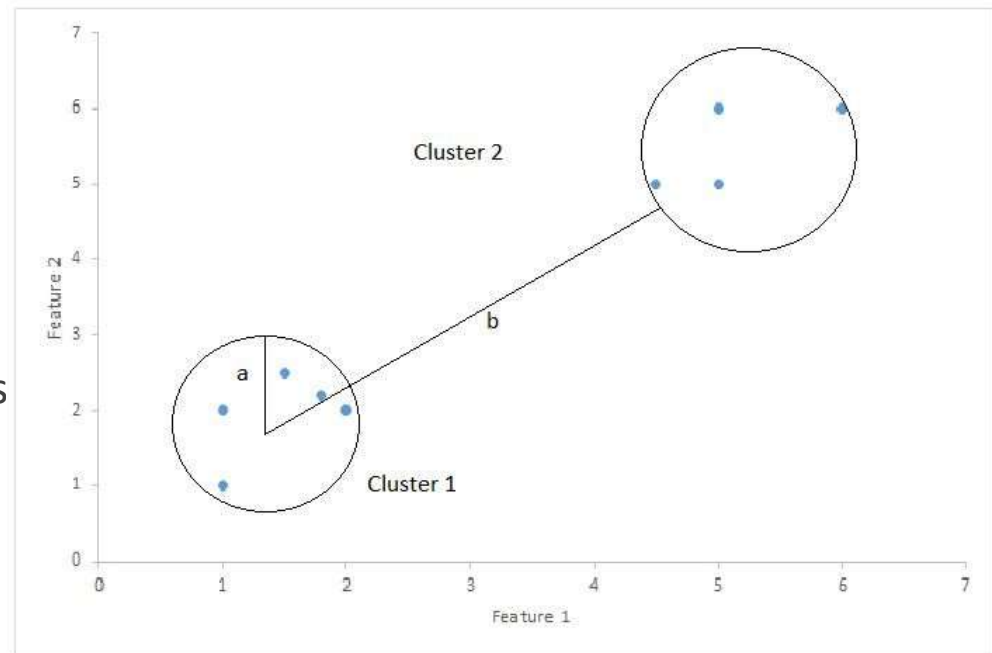
Description of data mining tasks

Data mining techniques used


1. For finding the number of clusters:
 1. Elbow Method
 2. Silhouette scores
2. Hierarchical Clustering
3. K-mean Clustering
4. PCA for visualization

Silhouette Score

- Silhouette Score = $(b-a)/\max(a,b)$,
- where
- a = average intra-cluster distance
- b = average inter-cluster distance
- 1: clusters clearly distinguished.
- 0: no clear distinction between clusters
- -1: clusters mis-assigned



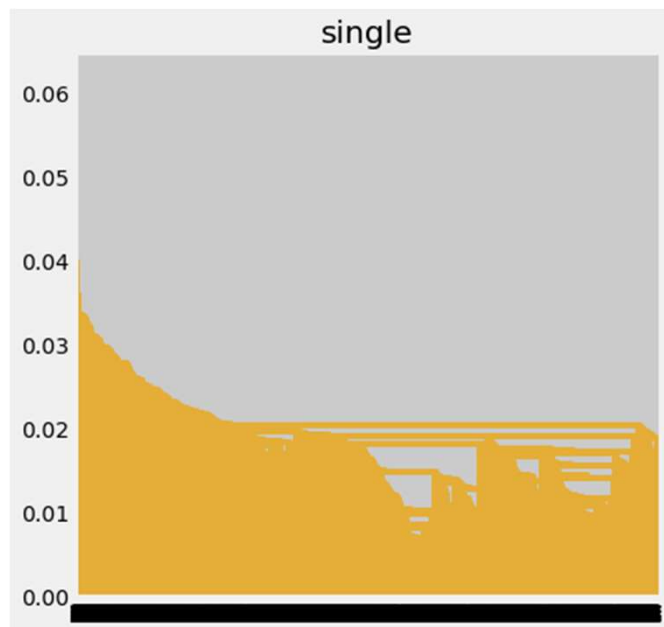
Linkage Criteria

- Single linkage: shortest distance between a pair of observations in two clusters
 - Complete linkage: longest distance between a pair of observations in two clusters
 - Average linkage: average distance between each pair of observations in each cluster
 - Ward linkage: sum of intra-cluster variance of two clusters
- 
- A solid green horizontal bar spanning the width of the slide, located at the bottom.

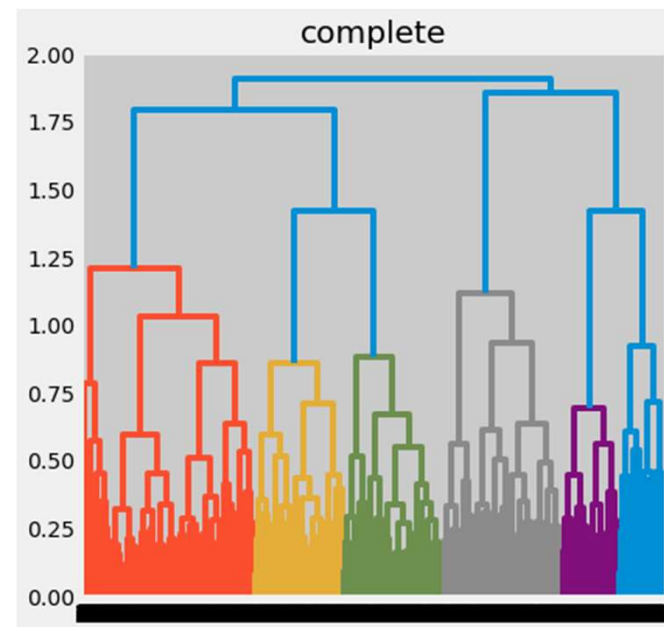
Results for Hierarchical Clustering

Dendrogram

SINGLE

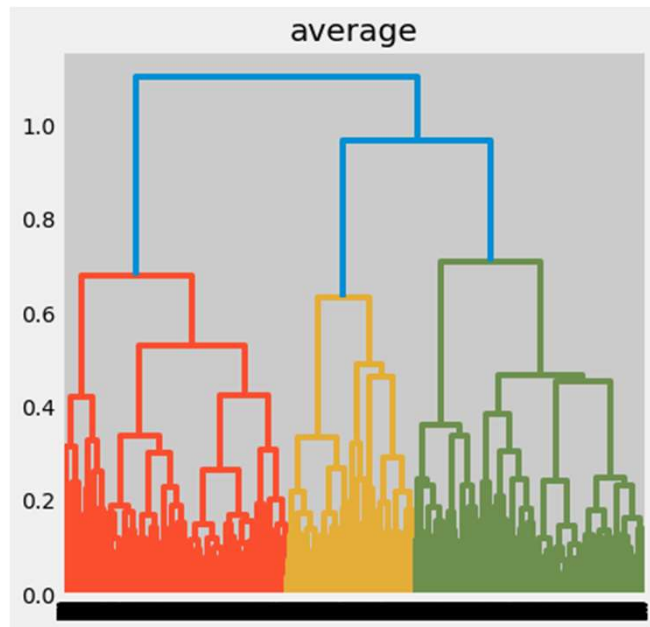


COMPLETE

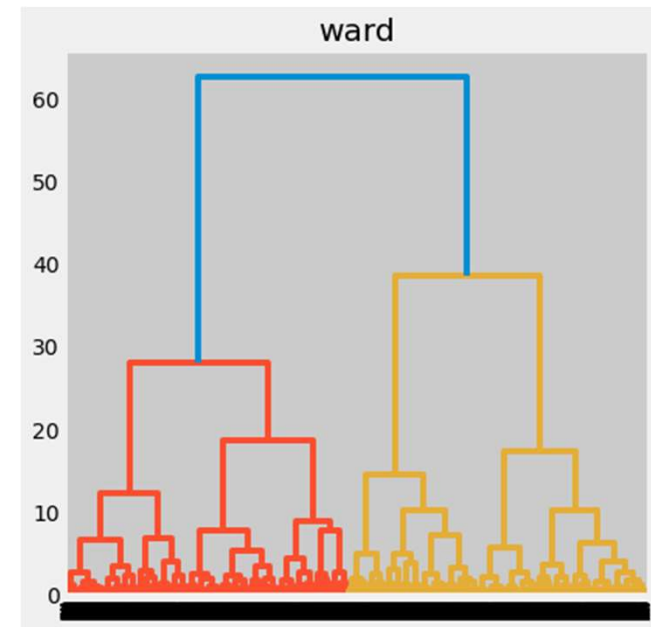


Dendrogram

AVERAGE

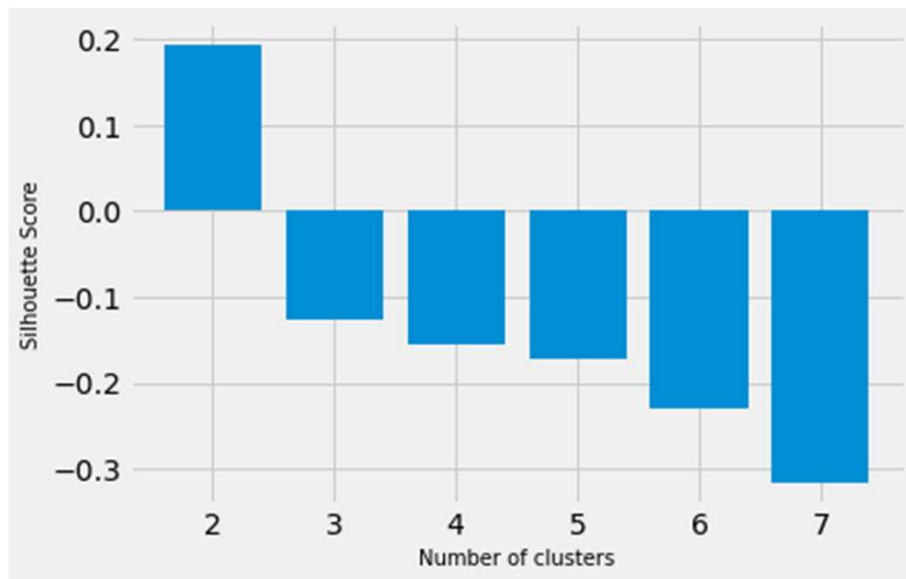


WARD

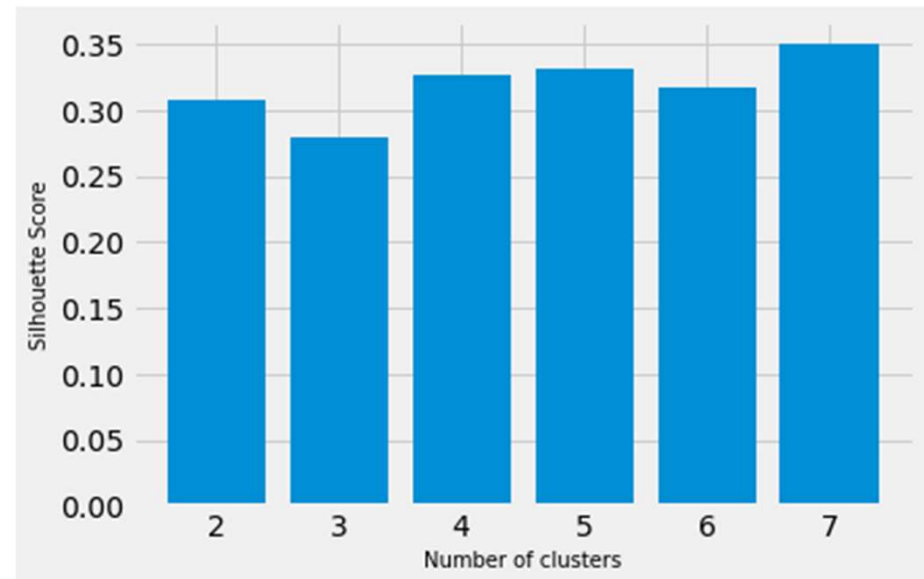


Silhouette scores for different linkage setting

SINGLE

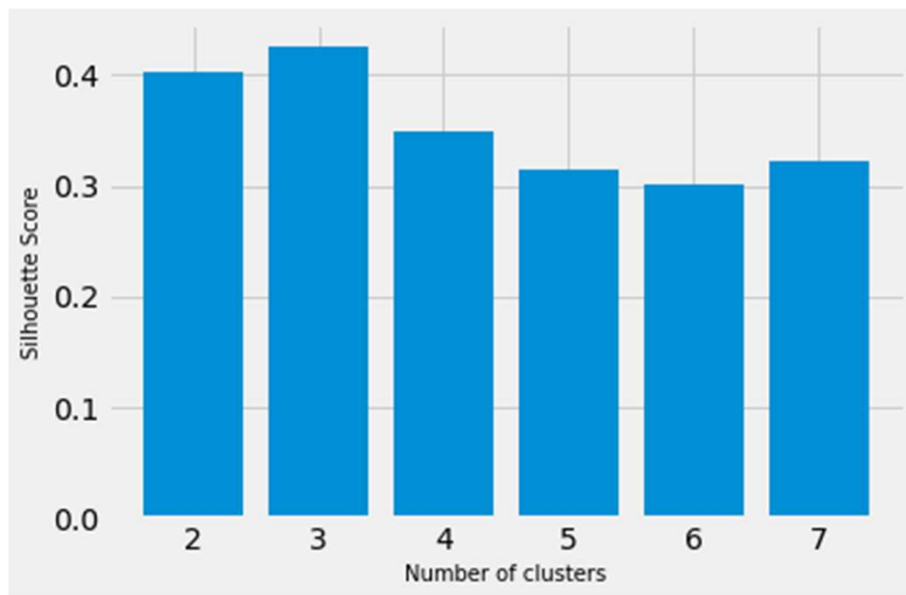


COMPLETE

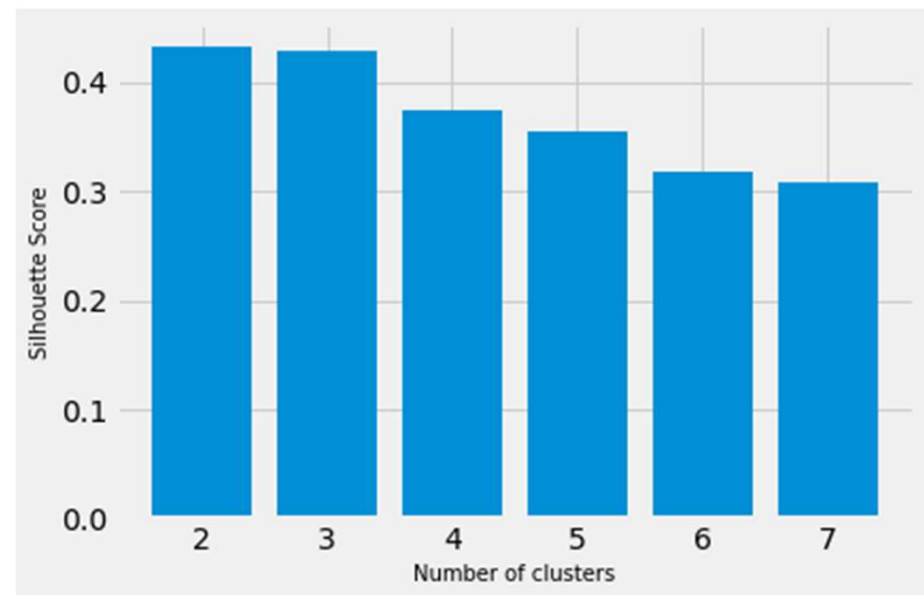


Silhouette scores for different linkage setting

AVERAGE

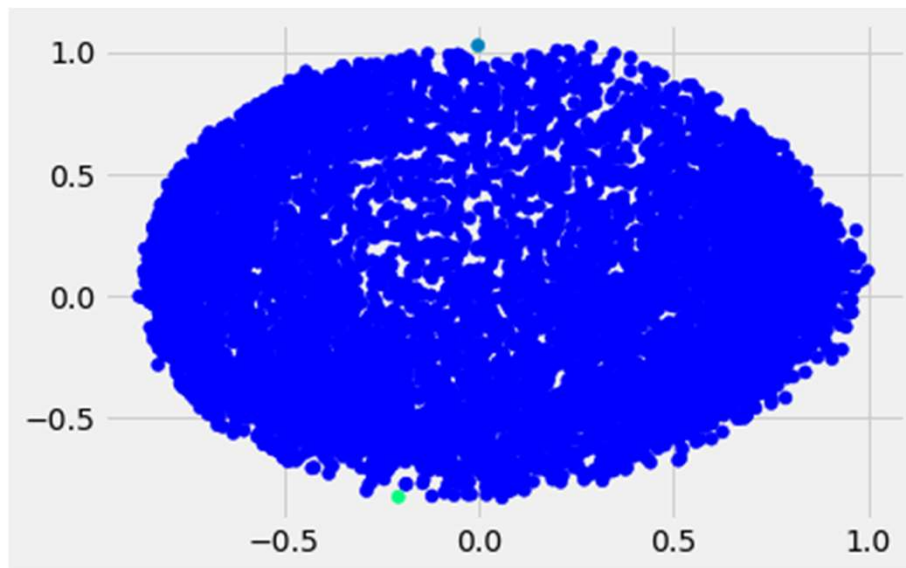


WARD

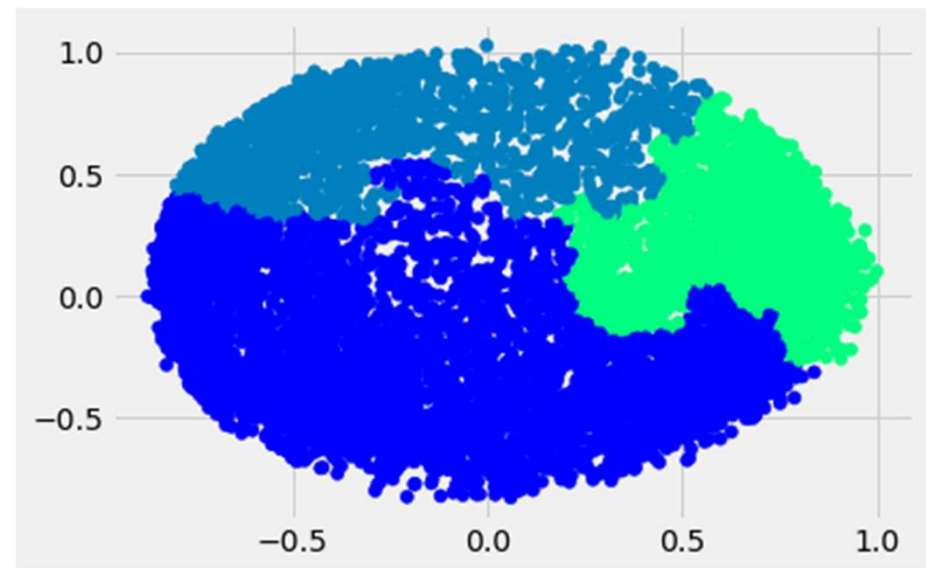


Clustering visualization in different linkage setting with $n=3$

SINGLE

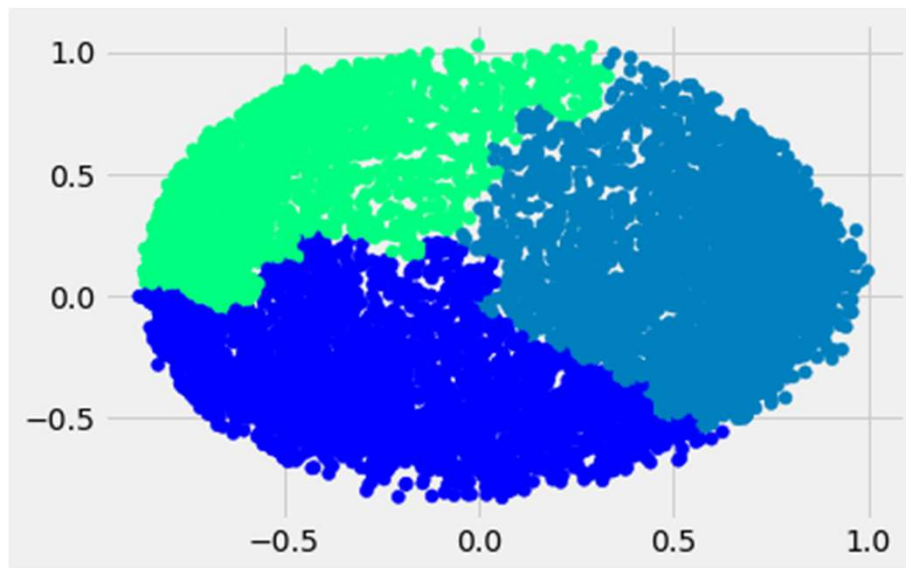


COMPLETE

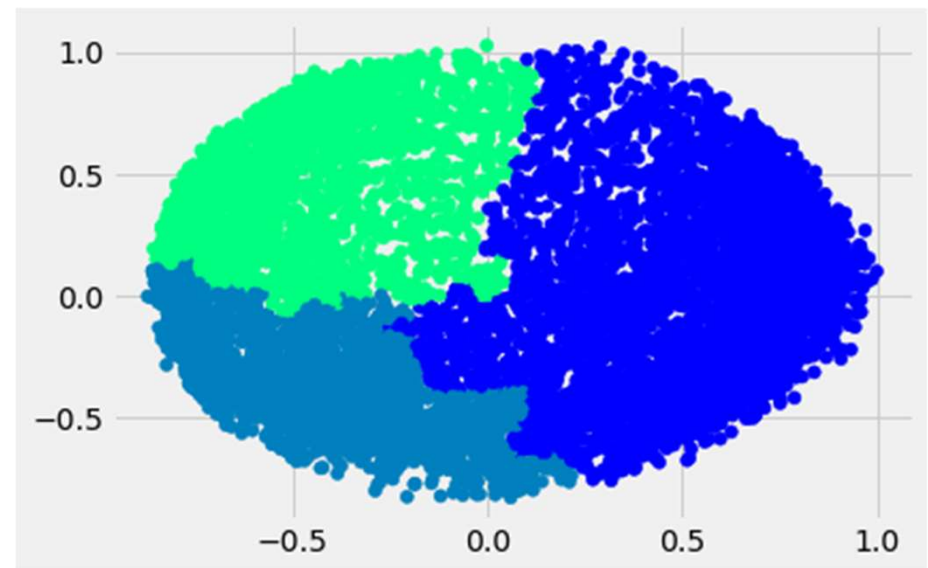


Clustering visualization in different linkage setting with $n=3$

AVERAGE

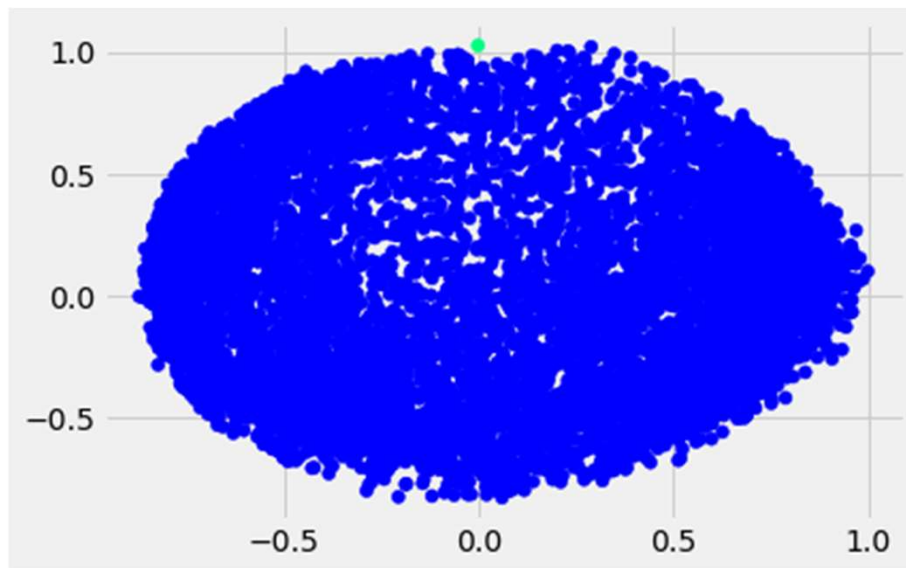


WARD

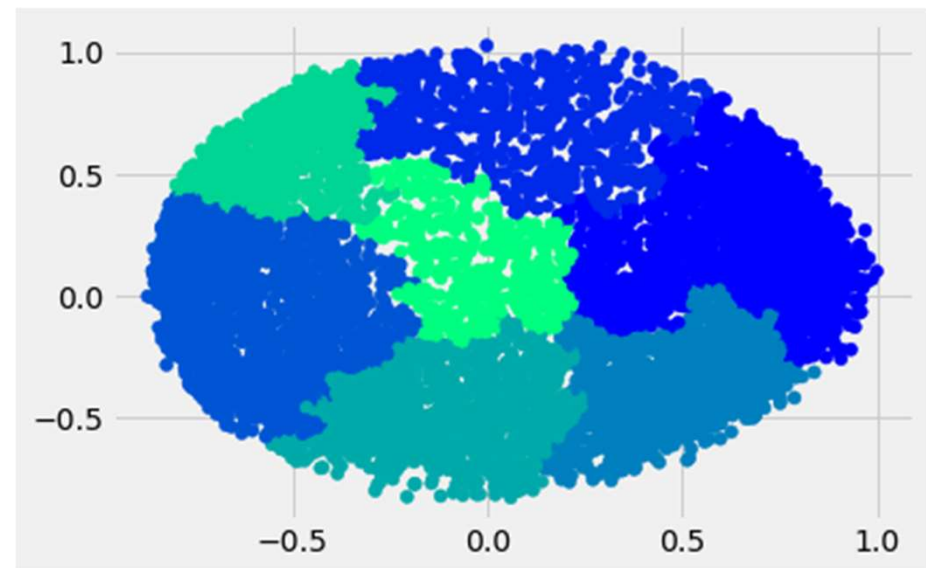


Clustering visualization in different linkage setting with best Silhouette scores

SINGLE

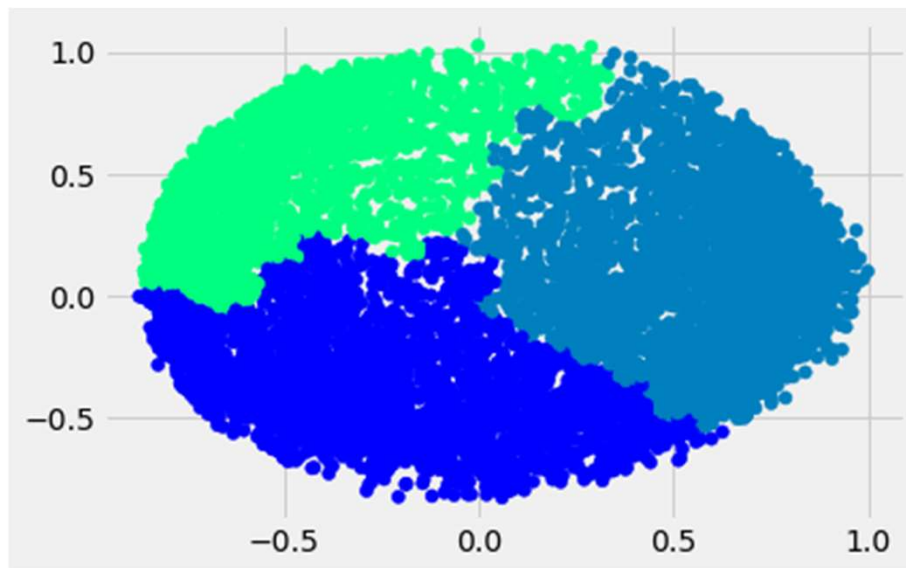


COMPLETE

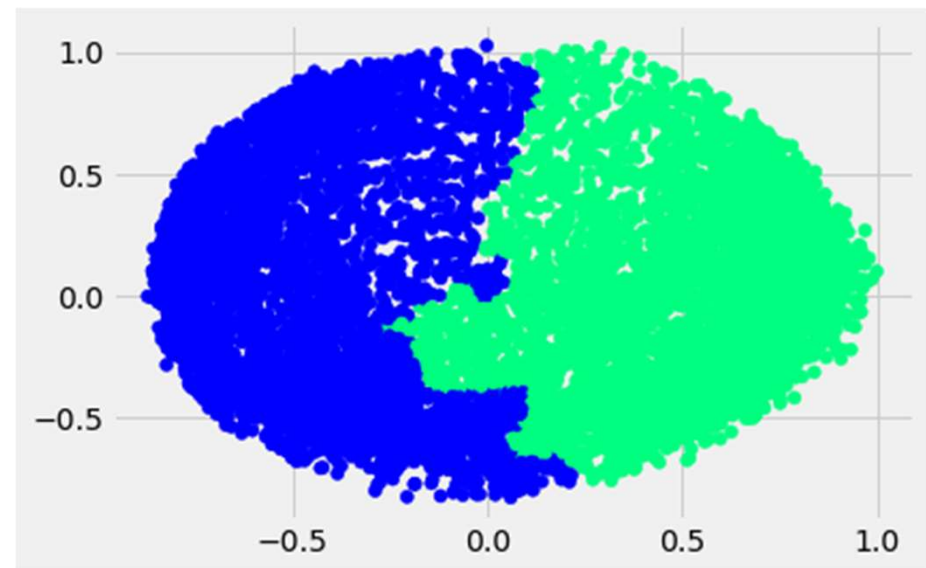


Clustering visualization in different linkage setting with best Silhouette scores

AVERAGE

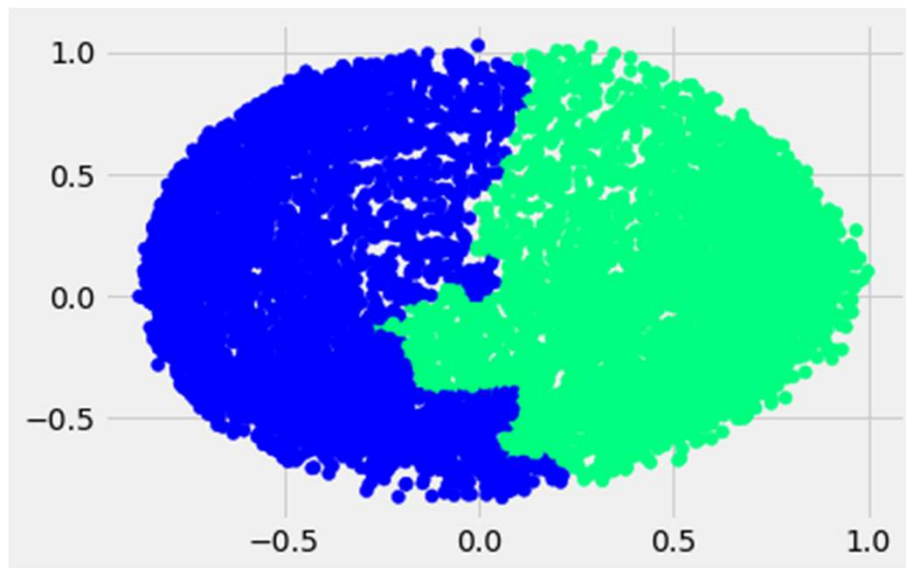


WARD

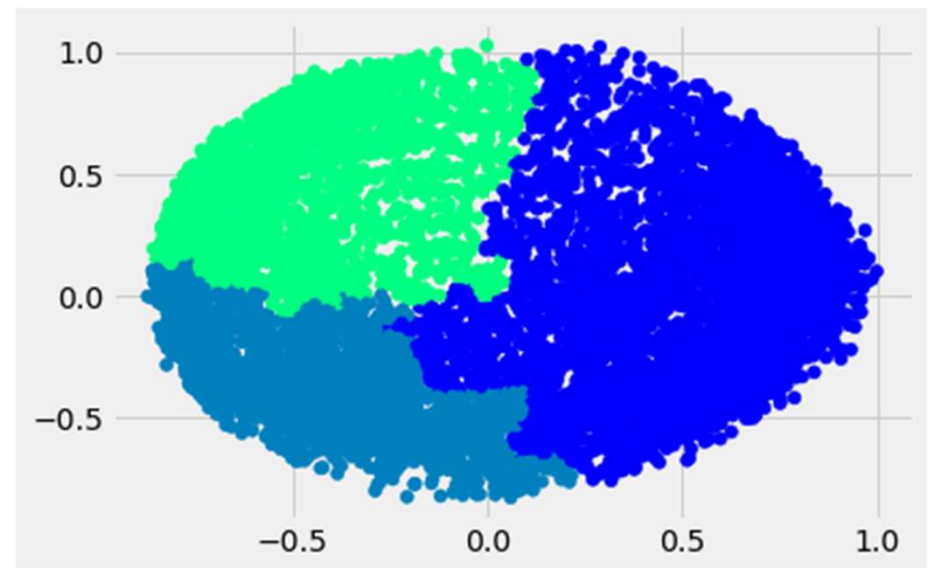


Clustering visualization with Ward

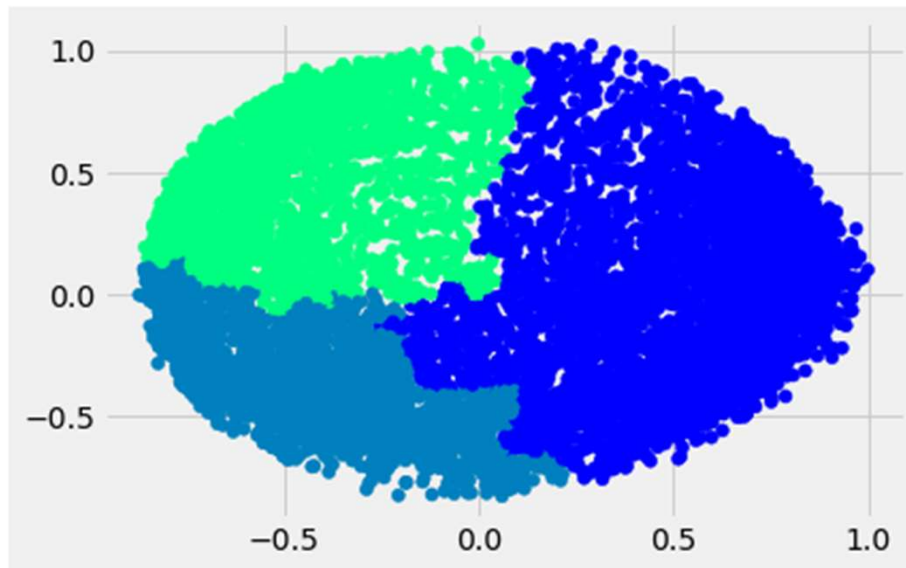
2 CLUSTER



3 CLUSTER



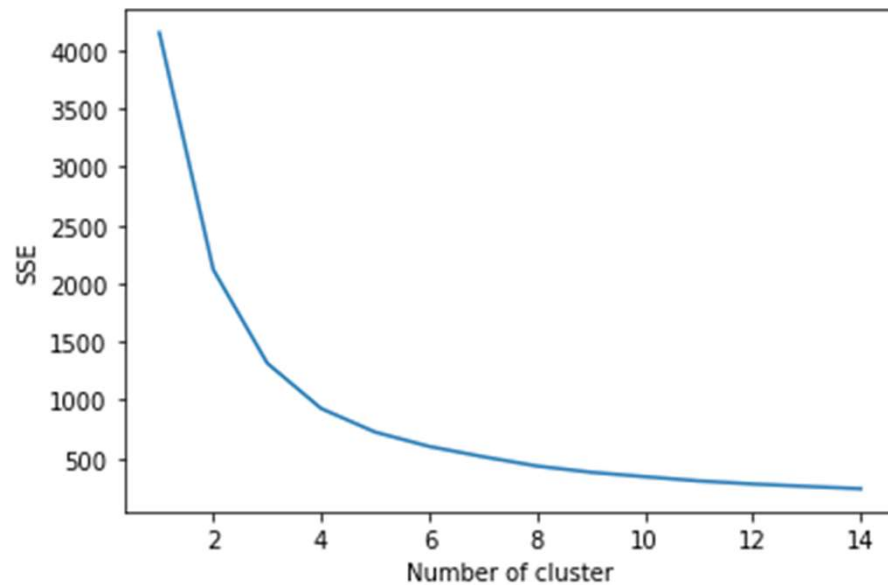
Results for Hierarchical Clustering



1. Method "Ward"
 1. Highest Silhouette scores amount all method
2. Number of cluster: 3
 1. 2 and 3 has similar score
 2. More cluster, more flexible
 3. Provide more precise strategy to different group of customers

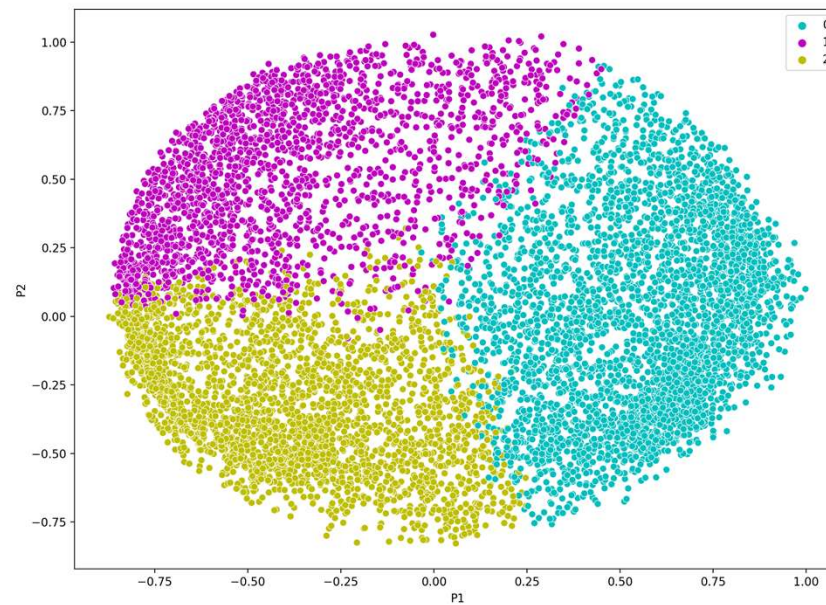
Results for K-mean Clustering

Elbow Method



1. 2 and 3 both are good choice
2. Validated the results for Hierarchical Clustering and choose 3

Visualization by PCA for $n=2$



Recommendation for 3 clusters

Recommendation for 3 cluster: Cluster 1

1. Low balance
2. Low purchase frequency
3. Lowest credit limit
4. Not the main target customers

	Cluster 1	Cluster 2	Cluster 3
Account balance	Low	Medium	High
Purchase frequency	Low	High	Medium
Credit limit	Low	Medium	High
Cash advance	Medium	Low	High

Recommendation for 3 cluster: Cluster 2

1. High purchase frequency
2. Medium balance
3. Low cash advance
4. Main target because these customer frequently use the card to purchase

	Cluster 1	Cluster 2	Cluster 3
Account balance	Low	Medium	High
Purchase frequency	Low	High	Medium
Credit limit	Low	Medium	High
Cash advance	Medium	Low	High

Recommendation for 3 cluster: Cluster 3

1. High balance
2. High cash advance
3. High credit limit
4. Use credit cards as loans

	Cluster 1	Cluster 2	Cluster 3
Account balance	Low	Medium	High
Purchase frequency	Low	High	Medium
Credit limit	Low	Medium	High
Cash advance	Medium	Low	High

Thank you
