



Credit Cardholders Segmentations for Marketing Strategies

Background Information

Statement of marketing question

how to use data analytics to make a customer segmentation for defining the marketing strategy based on the Credit Card Dataset

Description of the dataset

- usage behavior of about 9000 active credit card holders
- 18 behavioral variables

- Basic data exploration – 2 columns have NAs

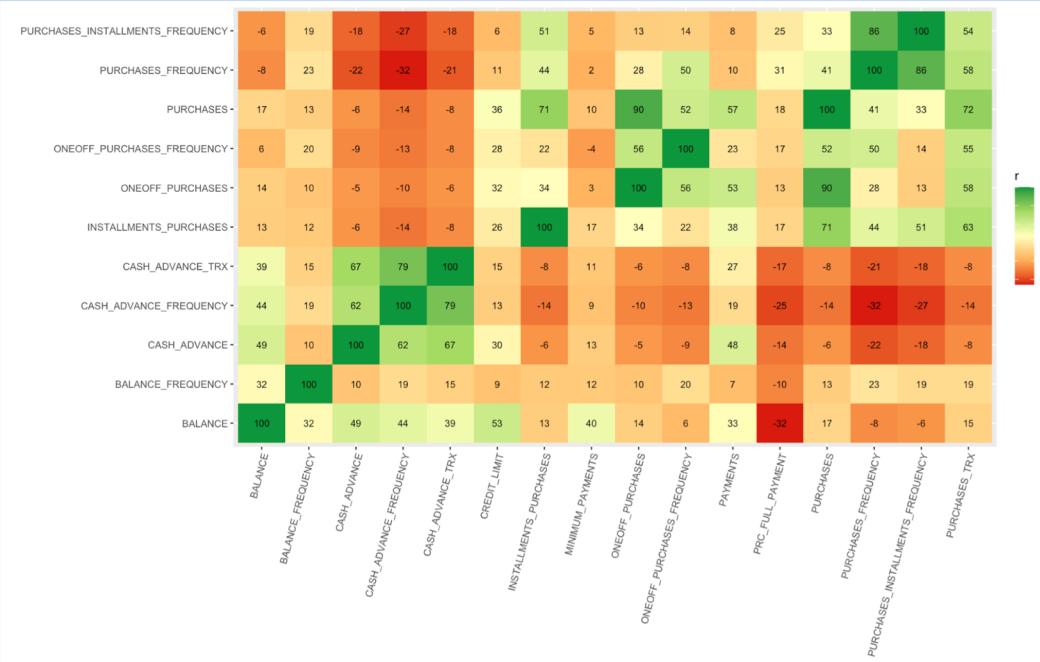
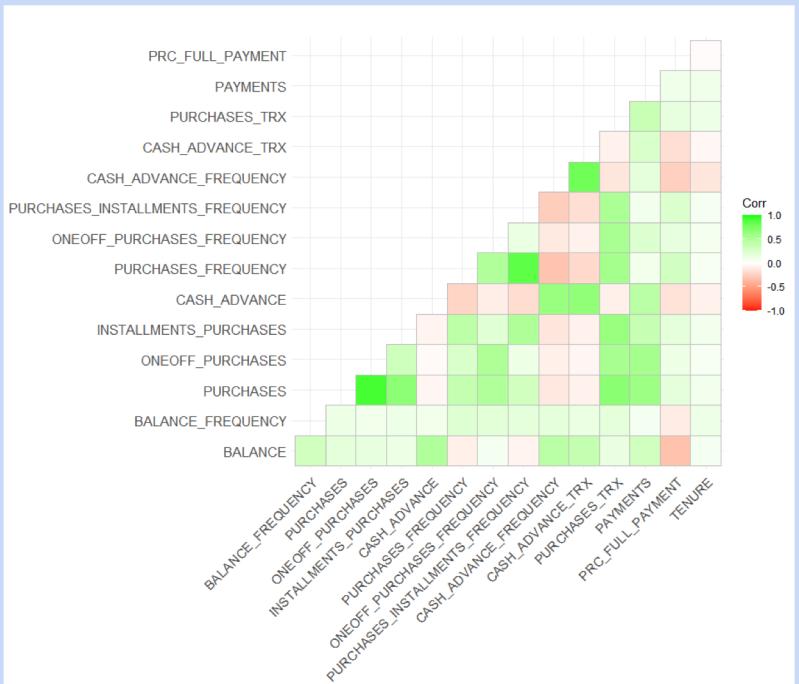
```
> summary(ccdata)
   CUST_ID          BALANCE        BALANCE_FREQUENCY PURCHASES      ONEOFF_PURCHASES
Length:8950      Min. :  0.0  Min. :0.0000  Min. :  0.00  Min. :  0.0
Class :character  1st Qu.:128.3 1st Qu.:0.8889  1st Qu.: 39.63  1st Qu.:  0.0
Mode :character   Median : 873.4  Median :1.0000  Median :361.28  Median : 38.0
                           Mean :1564.5  Mean :0.8773  Mean :1003.20  Mean : 592.4
                           3rd Qu.:2054.1 3rd Qu.:1.0000  3rd Qu.:1110.13 3rd Qu.: 577.4
                           Max. :19043.1  Max. :1.0000  Max. :49039.57  Max. :40761.2

  INSTALLMENTS_PURCHASES  CASH_ADVANCE    PURCHASES_FREQUENCY ONEOFF_PURCHASES_FREQUENCY
Min. :  0.0      Min. :  0.0      Min. :0.00000  Min. :0.00000
1st Qu.:  0.0      1st Qu.:  0.0      1st Qu.:0.08333  1st Qu.:0.00000
Median : 89.0      Median :  0.0      Median :0.50000  Median :0.08333
Mean : 411.1      Mean : 978.9      Mean :0.49035  Mean :0.20246
3rd Qu.: 468.6      3rd Qu.:1113.8      3rd Qu.:0.91667  3rd Qu.:0.30000
Max. :22500.0      Max. :47137.2      Max. :1.00000  Max. :1.00000

  PURCHASES_INSTALLMENTS_FREQUENCY CASH_ADVANCE_FREQUENCY CASH_ADVANCE_TRX PURCHASES_TRX
Min. :0.0000  Min. :0.0000  Min. :  0.000  Min. :  0.00
1st Qu.:0.0000 1st Qu.:0.0000  1st Qu.:  0.000  1st Qu.: 1.00
Median :0.1667  Median :0.0000  Median :  0.000  Median : 7.00
Mean : 0.3644  Mean : 0.1351  Mean :  3.249  Mean :14.71
3rd Qu.:0.7500 3rd Qu.:0.2222  3rd Qu.:  4.000  3rd Qu.:17.00
Max. :1.0000  Max. :1.5000  Max. :123.000  Max. :358.00

  CREDIT_LIMIT      PAYMENTS    MINIMUM_PAYMENTS PRC_FULL_PAYMENT      TENURE
Min. : 50  Min. :  0.0  Min. :  0.02  Min. :0.0000  Min. : 6.00
1st Qu.:1600 1st Qu.: 383.3 1st Qu.:169.12  1st Qu.:0.0000  1st Qu.:12.00
Median :3000  Median : 856.9  Median :312.34  Median :0.0000  Median :12.00
Mean : 4494  Mean :1733.1  Mean : 864.21  Mean :0.1537  Mean :11.52
3rd Qu.:6500 3rd Qu.:1901.1 3rd Qu.:825.49  3rd Qu.:0.1429  3rd Qu.:12.00
Max. :30000  Max. :50721.5  Max. :76406.21  Max. :1.0000  Max. :12.00
NA's : 1       NA's :313            NA's :313            NA's :313            NA's :313
```

Correlation

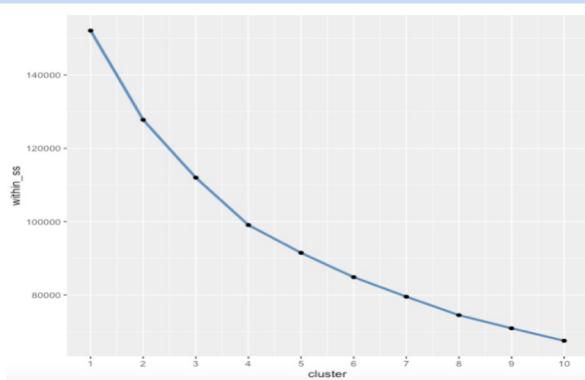


No obvious relationship

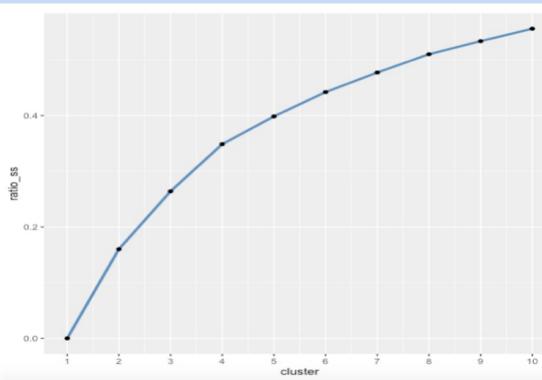
Most people buy thing within affordability and the purchasing behavior doesn't change

Determining the number of clusters

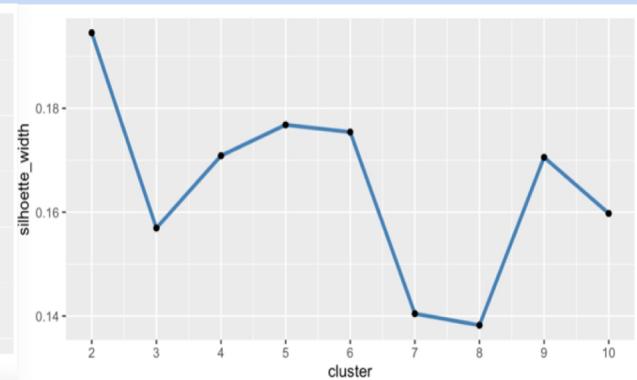
Total Within Sum of Square Plot



Ratio Plot



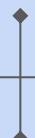
Silhouette Plot



3,4



4

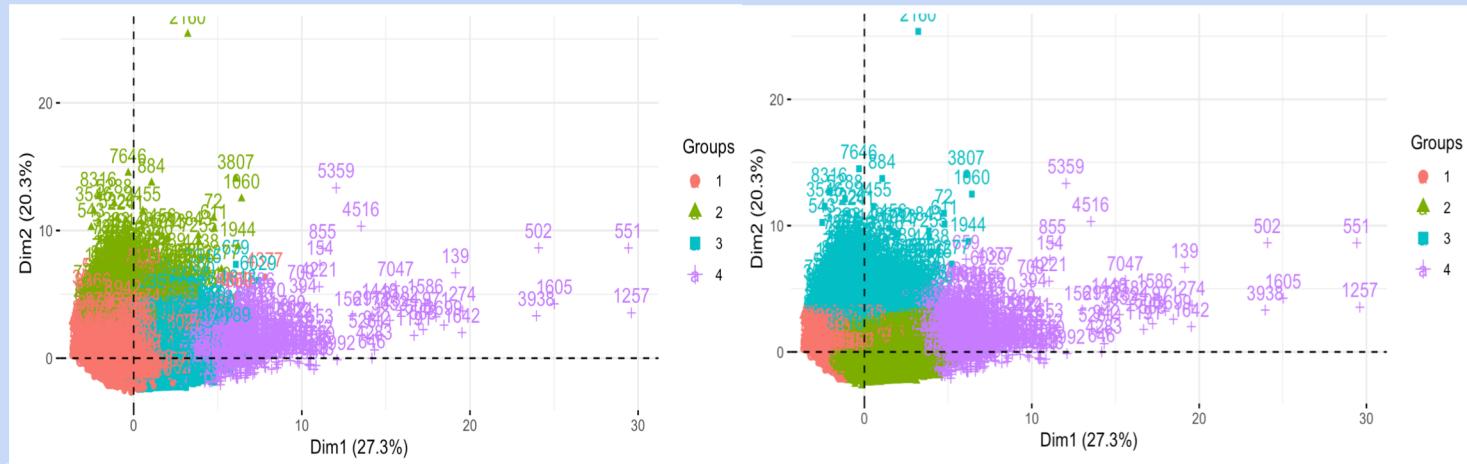


2



4

Hierarchical & K-means PCA Comparison

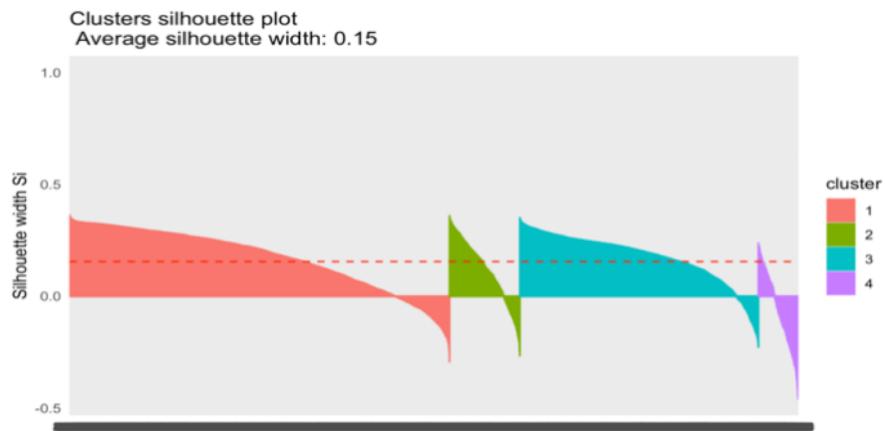


	Cluster 1	Cluster 2	Cluster 3	Cluster 4
hierarchical clustering	4668	864	2931	487
k-means	3977	3366	1198	409

Hierarchical Clustering v.s. K-means

- Fail to scale well for large dataset
- Comparatively LOW avg Silhouette width
 - many observations in that method are probably in the wrong clusters

Hierarchical Clustering Silhouette Plot

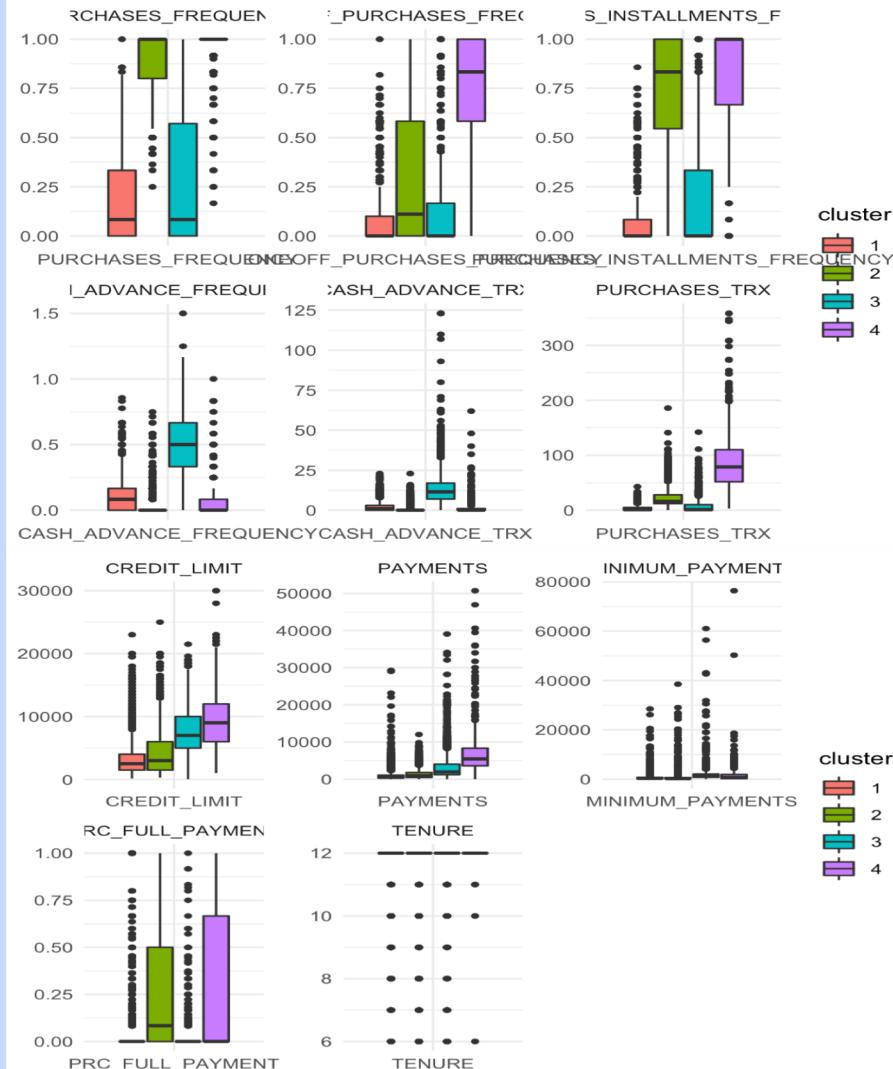
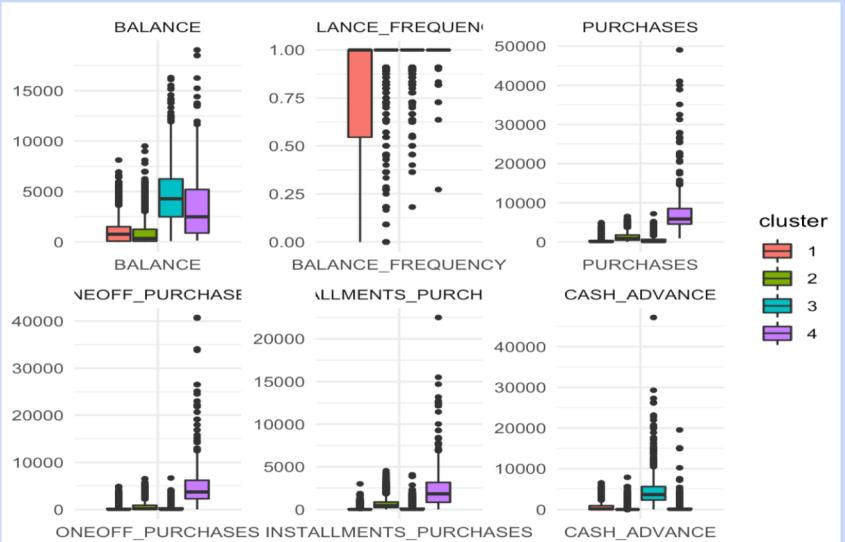


- Suitable for large dataset
- Comparatively HIGH avg Silhouette width 

K-means Silhouette Plot



Analyzing Box Plots & Conclusion



Customer Profile – interesting results

- Cluster 1:** Rare user, with (probably) low to mid income
- Cluster 2:** Frequent user, with (probably) low income
- Cluster 3:** Mid to rare users, with (probably) high income
- Cluster 4:** Super frequent user, with (probably) mid to high income, frequently purchasing with installment

Limitation: K-means is sensitive to outliers but we did not eliminate outliers, because there are too many of them in the dataset.

Improvements:

1. Address outliers problem
2. Dimension Reduction to deal with sparse observations

Dimension Reduction

- low income (cluster1&2): more one-off purchases
- rich group (cluster3): more purchase installment
- middle income (cluster4): more cash advance and purchase installment

