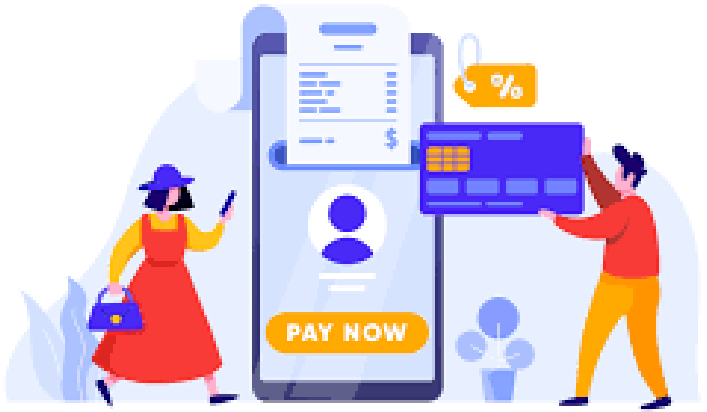

Customer
Segmentation
ALL LiFE Bank

Contents

1. Business Problem Overview
2. Data Overview
3. Exploratory Data Analysis (EDA)
4. Model Performance
5. Business Recommendations



ALL LIFE Bank

Business Problem Overview

AllLife Bank wants to focus on its credit card customer base in the next financial year.

They have been advised by their marketing research team, that the penetration in the market can be improved.

Based on this input, the Marketing team proposes to run personalized campaigns to target new customers as well as upsell to existing customers.

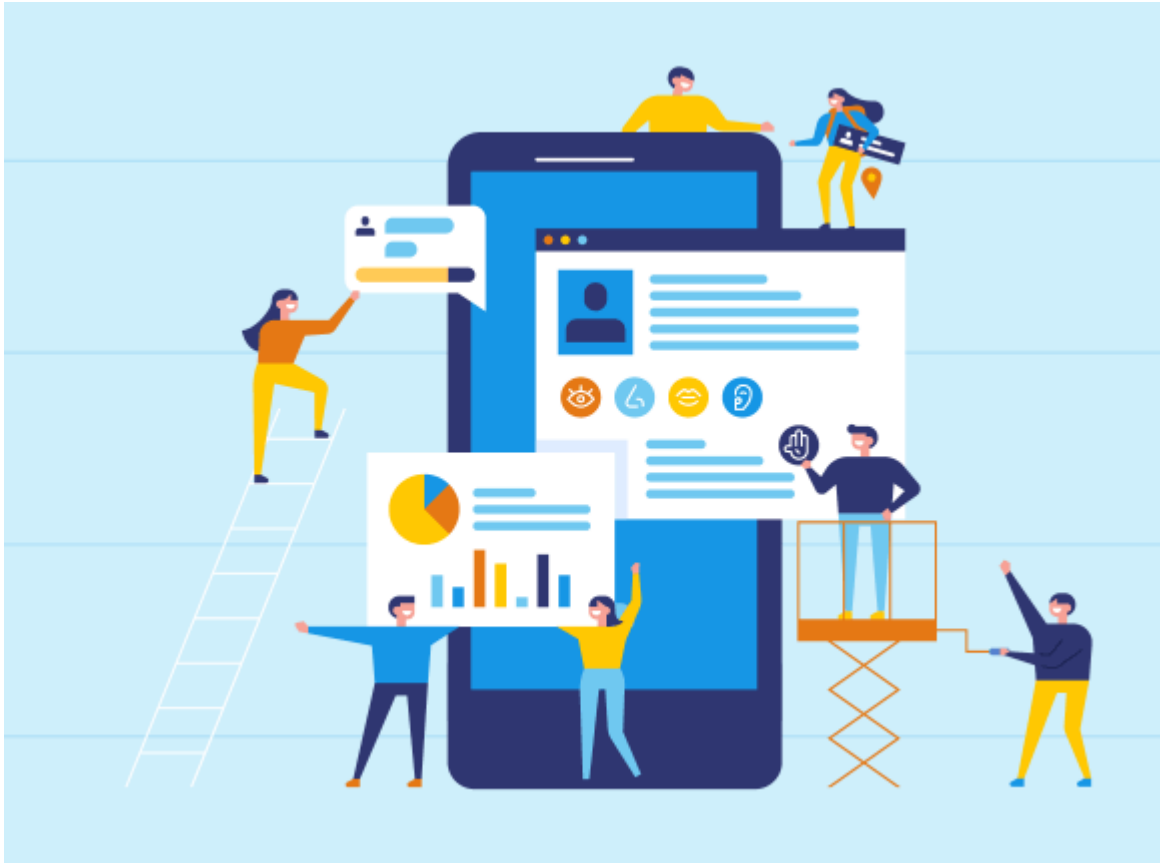
Another insight from the market research was that the customers perceive the support services of the back poorly. Based on this, the Operations team wants to upgrade the service delivery model, to ensure that customer queries are resolved faster.

Data Overview

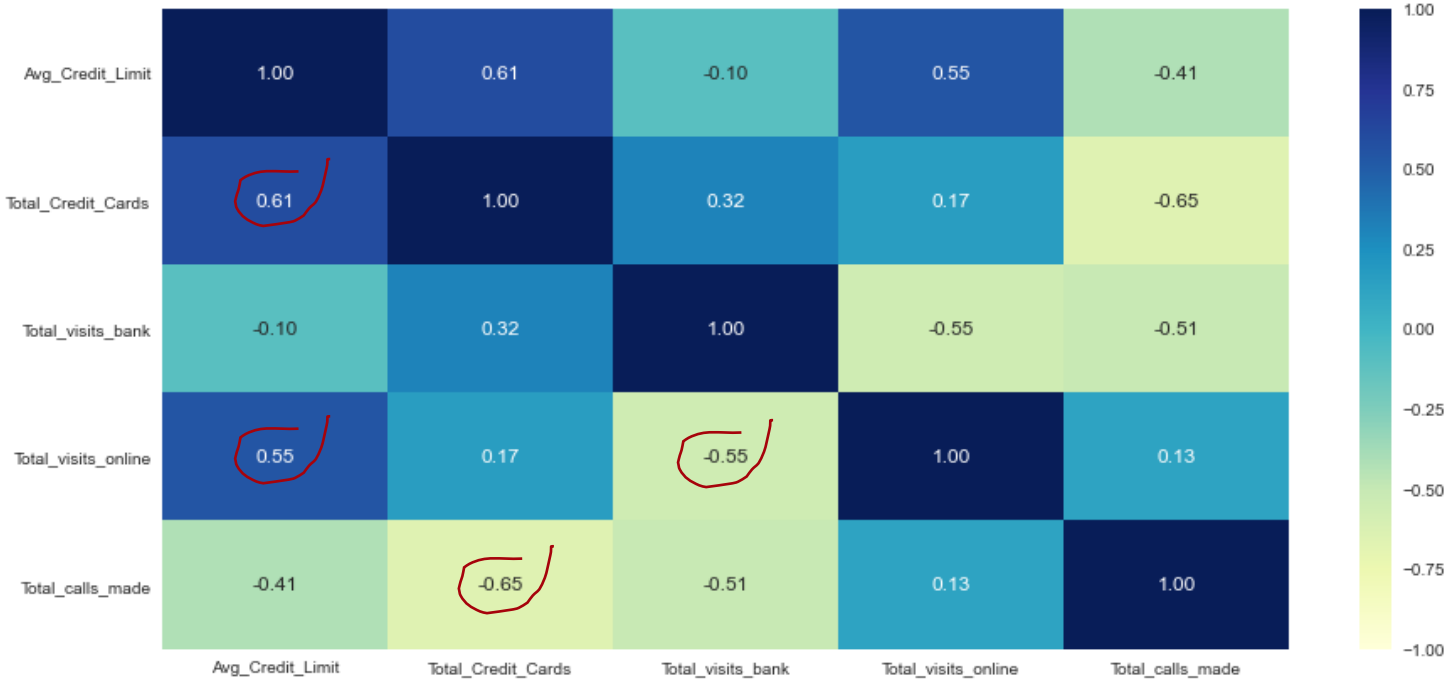
- Dataset is not labelled, we will proceed with Unsupervised Learning techniques

- The data provided is of 660 customers and their 5 financial attributes like:
  - Credit limit;
  - The total number of credit cards the customer has;
  - And different channels through which customers have contacted the bank for any queries (including visiting the bank, online and through a call center).

- There is no Missing Values or duplicated rows, but there are 5 Customers duplicated and values might correspond to customer profile changes, and as such, there is no need to delete these records as these are actual occurrences at some point in the time.



Exploratory Data Analysis



i. As expected, Total_Credit_Cards and Avg_Credit_Limit have a High Positive Correlation, since more credit cards, more like hood to have a higher Credit Limit.

ii. Total_Visit_Online is Positive Correlated with Avg_Credit_Limit suggesting that customers with higher Credit Limit tend to use more the Online tools and Negative Correlated with Total_visits_bank which indicates that online customers tend not to visit the bank.

iii. Total_calls_made is Negative Correlated with Avg_Credit_Limit, Total_Credit_Cards and Total_visitis_bank, which indicate that customers who use more calls channels, are more like hood to visit less the bank and also have a small Avg_Credit_Limit.

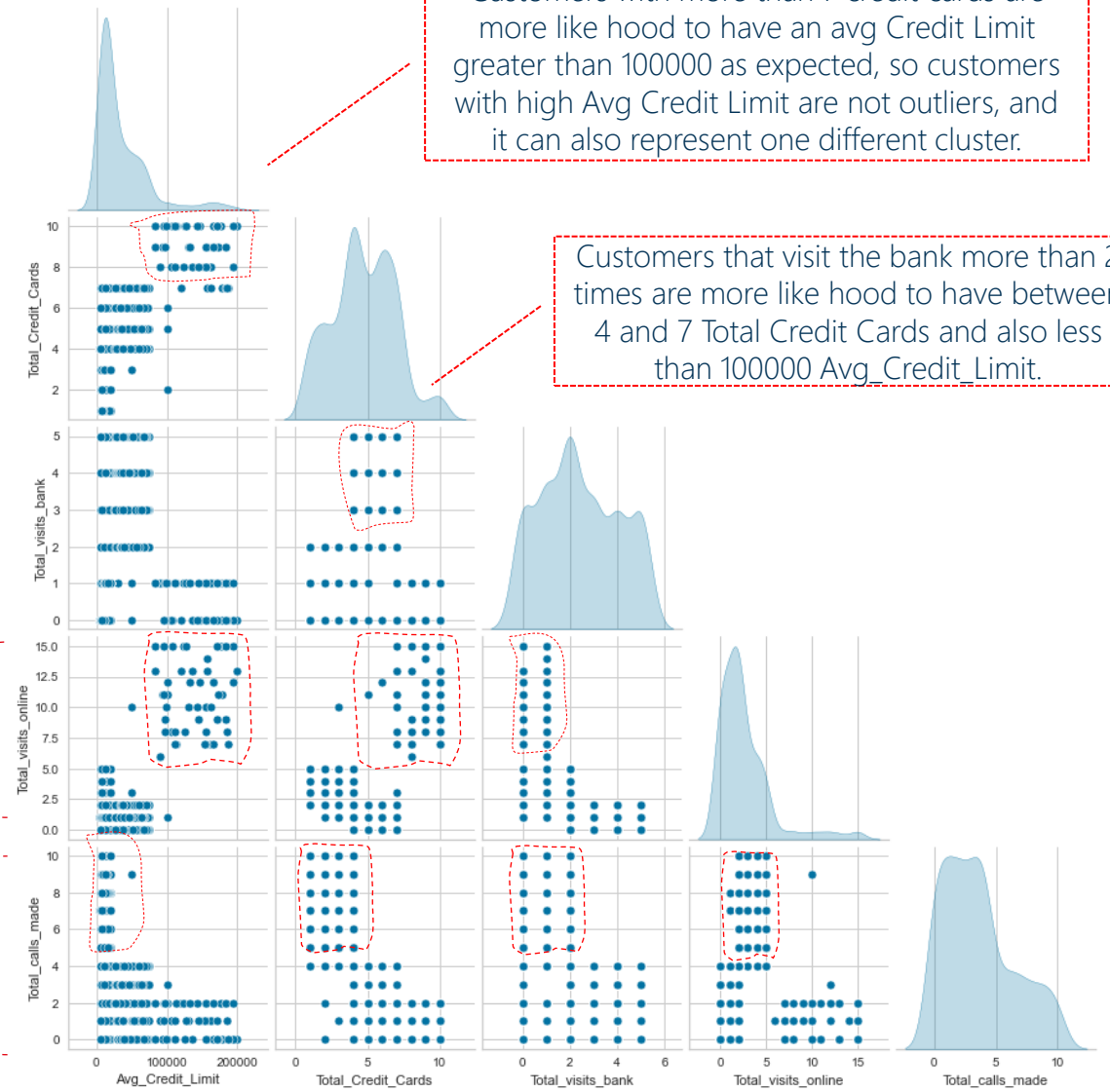
Exploratory Data Analysis

Customers with more than 7 credit cards are more like hood to have an avg Credit Limit greater than 100000 as expected, so customers with high Avg Credit Limit are not outliers, and it can also represent one different cluster.

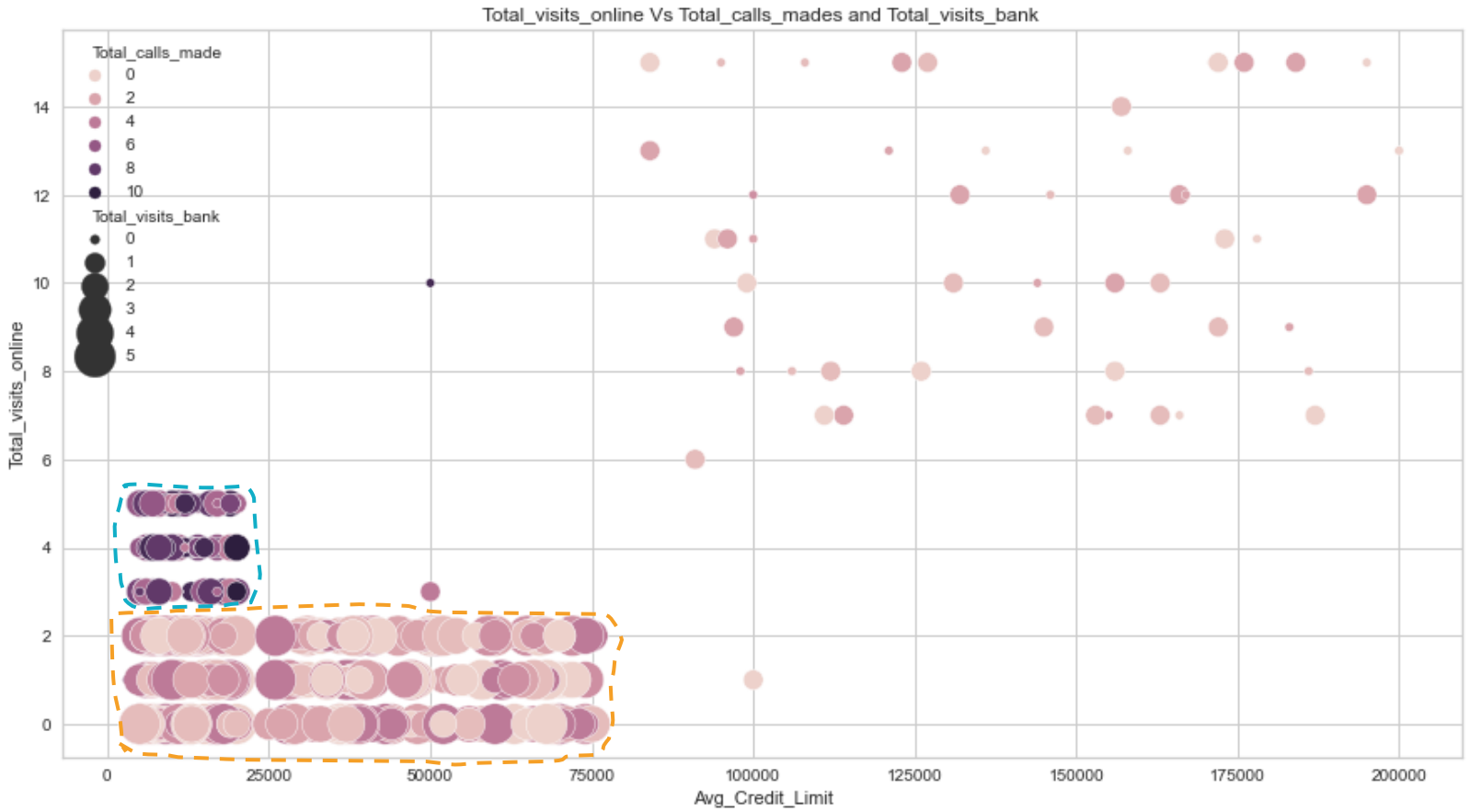
Customers that visit the bank more than 2 times are more like hood to have between 4 and 7 Total Credit Cards and also less than 100000 Avg_Credit_Limit.

Customers that have high Avg Credit Card and high total credit card are more like hood to do more online visits and less bank visit.

Customers with low Avg Credit Limit, 1 to 4 total credit card, prefers to make calls, visiting less the bank and 1 to 5 visit on the online channels.



Exploratory Data Analysis



Customers with less than 78000 Credit Limit, 2 or less visits online, less than 5 calls made are more like hood to use the In Person channel, visiting the bank more than 3 times yearly.

Customers that are starting to use our online channels (3 to 5 access) they still calling between 4 to 10 times, moderate visits to the bank and they have a low avg_credit_limit.

Customers with high Credit Limit (grater than 78000) prefers visits online, doing less than 2 visits to the bank and calls

Model Performance Summary

I. We want to find a best customer segmentation.

II. We will use Unsupervised Techniques as K-mean and Hierarchical clustering for that.

III. First we will standardize our data set to avoid biased towards the variables with a higher magnitude.

IV. K-means: We will check Elbow Method and Silhouette scores to define the best k.

V. We will check K – means Cluster Profiling.

VI. Hierarchical: We will check distance metrics and linkage methods.

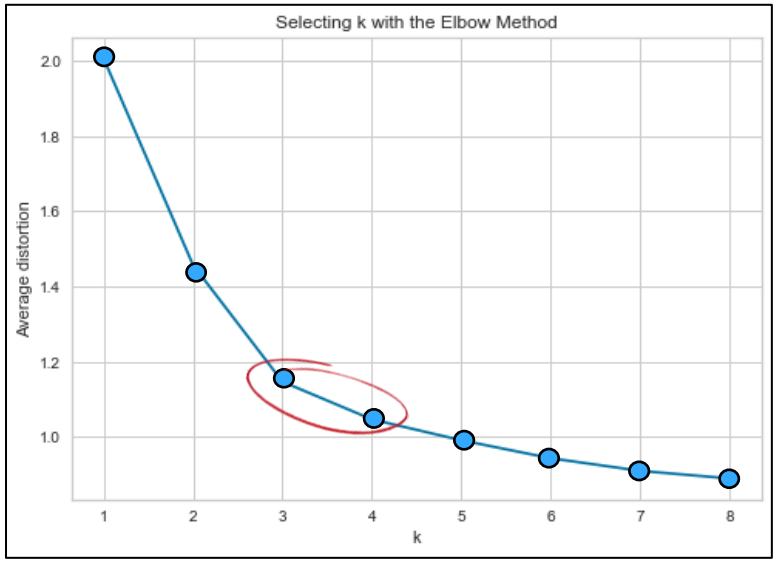
VII. The best distance metrics we will applied with different linkage and see the dendrograms.

VIII. With best distance metrics and linkage defined we will do silhouette scores to help choose the best k.

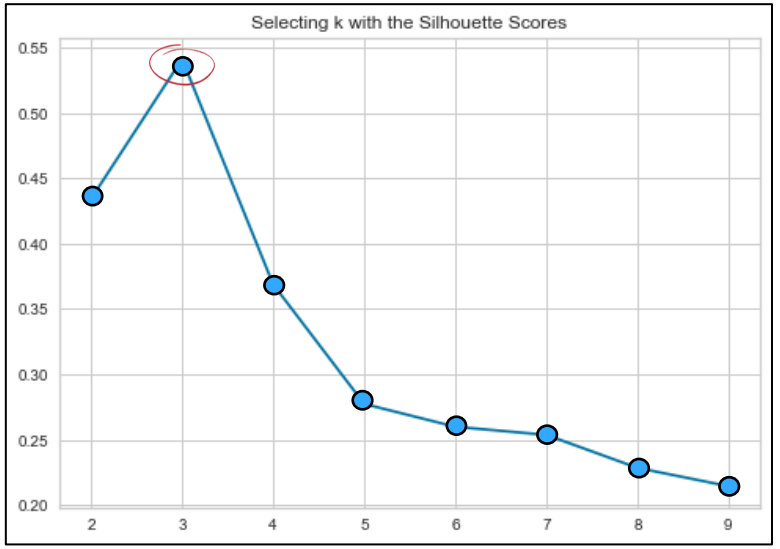
IX. We will check Hierarchical Cluster Profiling

K-means Performance

Choosing K:



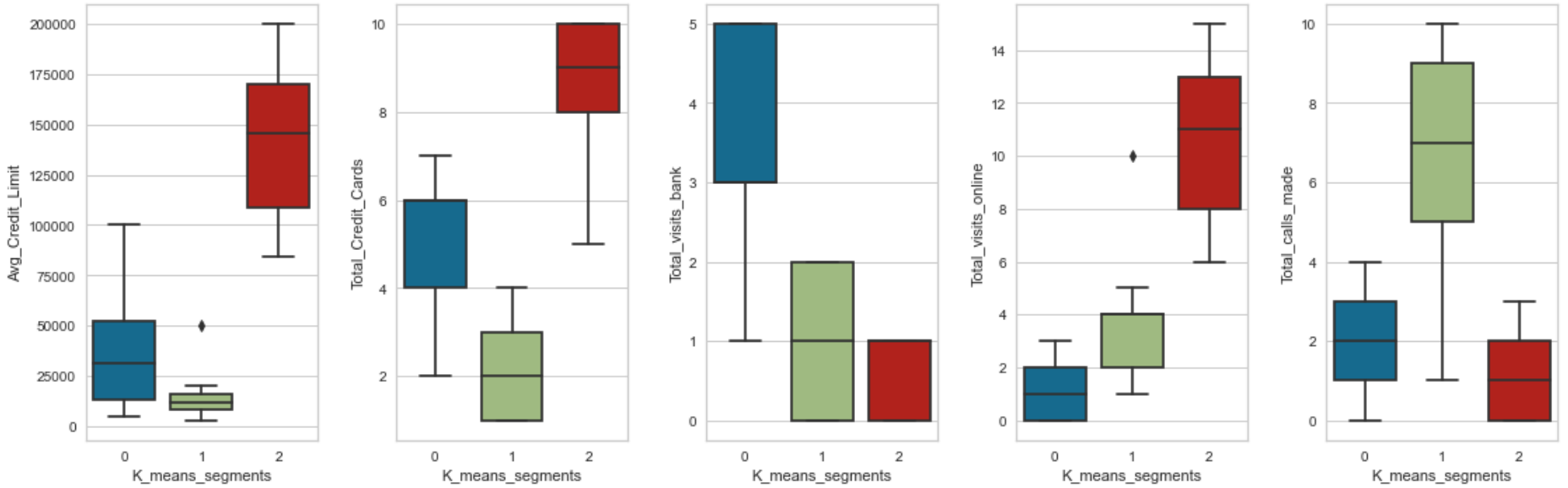
The distortion declines most at 3 and 4. Hence the optimal value for k seems to be 3 or 4.



Silhouette score for k=3 is higher than for 4. So, we will choose 3 as value of k.

K-means Cluster Profiling

Boxplot of original numerical variables for each cluster



Cluster 0 : GOLD CUSTOMERS:

- ✓ Moderate Average Credit Limit (34000);
- ✓ Total of 6 Credit Cards.
- ✓ Visit the bank the most (4 times yearly);
- ✓ Almost do not use our online channels;
- ✓ Mean of 2 calls made per customer yearly.

Therefore, this cluster is composed of Moderate Average Limit and in person visit preferable.

Cluster 1 : SILVER CUSTOMERS:

- ✓ Low Average Credit Limit (12174.11);
- ✓ Total of 3 Credit Cards.
- ✓ Visit the bank only 1 times yearly;
- ✓ Moderate use our online channels (4 times);
- ✓ High calls made 7 per customer yearly.

Accordingly, this cluster is composed of low Average Limit and the preferable contact is to call customer service department.

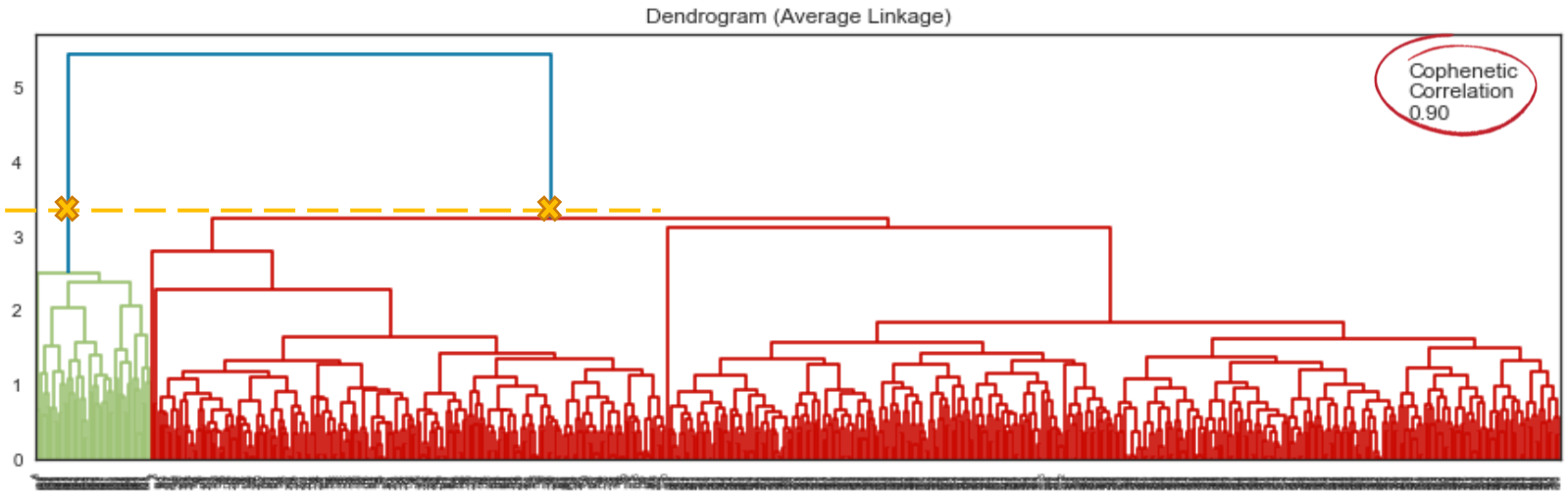
Cluster 2 : PLATINUM CUSTOMERS:

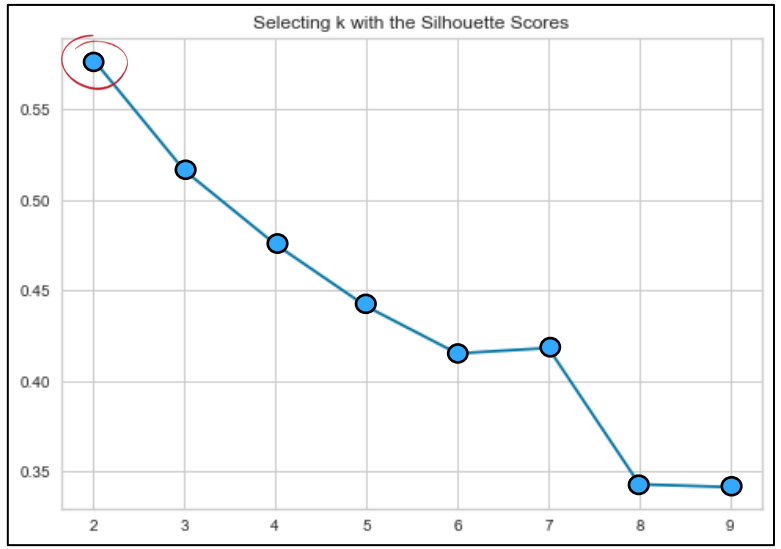
- ✓ Highest Average Credit Limit (141040);
- ✓ Mean Total of 9 Credit Cards.
- ✓ Almost never visit the bank (1 times yearly);
- ✓ Prefers the online channels using it almost 11 times per year.
- ✓ Mean of 2 calls made per customer yearly.

Hence, customers on this cluster prefers the Online contact and have the highest Average Credit Limit.

Hierarchical Performance

Choosing K:





	Linkage	Cophenetic Coefficient
0	single	0.739122
1	complete	0.859973
2	average	0.897708
3	centroid	0.893939
4	ward	0.741516
5	weighted	0.886175

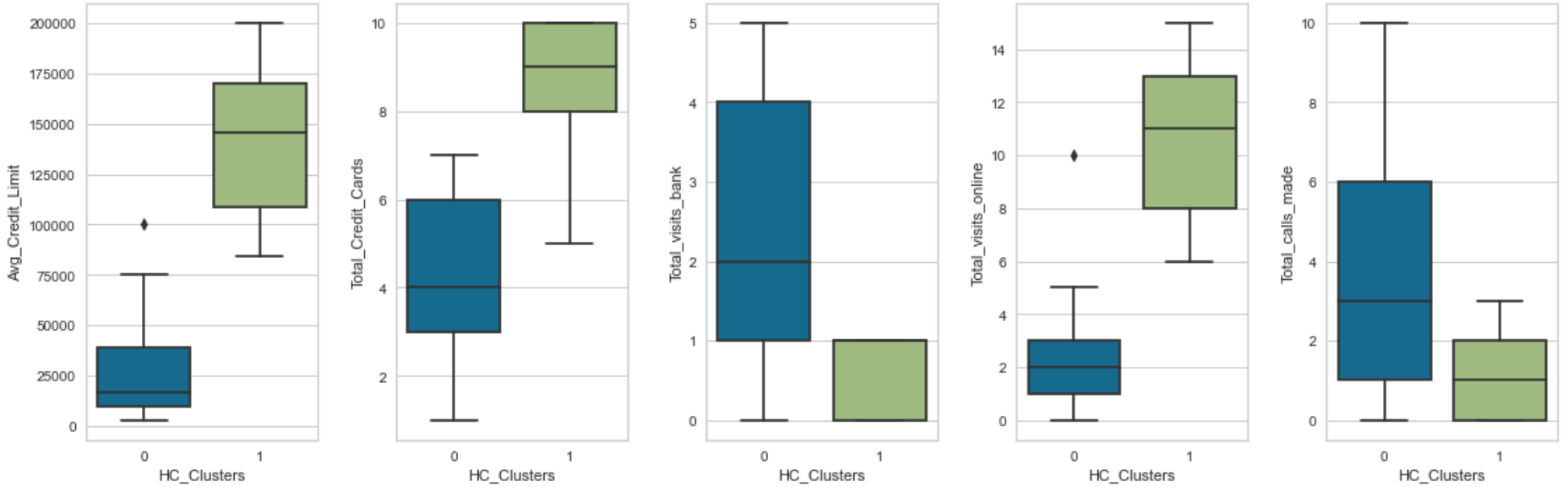
i. The cophenetic correlation is highest for average linkage methods and also is giving us distinct and separate clusters.

ii. 2 would be the appropriate number of clusters considering that the dendrogram shows a long distance (the y axis) for the two final groups and each of the subsequent groups has a drastically shorter distance.

Silhouette score for k=2 is higher (0.57). So, we will choose 2 as value of k.

Hierarchical Cluster Profiling

Boxplot of original numerical variables for each cluster



Cluster 0 : SILVER CUSTOMERS:

- ✓ Moderate Average Credit Limit (25847.5);
- ✓ Total of 5 Credit Cards.
- ✓ Visit the bank the most (3 times yearly);
- ✓ Almost do not use our online channels;
- ✓ Mean of 4 calls made per customer yearly.

Therefore, this cluster is composed of Moderate Average Limit and preferable contact is to call customer service department.

Cluster 1 : PLATINUM CUSTOMERS:

- ✓ Highest Average Credit Limit (141040);
- ✓ Mean Total of 9 Credit Cards.
- ✓ Almost never visit the bank (1 times yearly);
- ✓ Prefers the online channels using it almost 11 times per year.
- ✓ Mean of 2 calls made per customer yearly.

Hence, customers on this cluster prefers the Online contact and have the highest Average Credit Limit.

Comparing Clusters and Profiling

	K- means Cluster Profile			Hierarchical Cluster Profile	
# of Clusters	0	1	2	0	1
Avg_Credit_Limit	33782.4	12174.1	141040.0	25847.5	141040.0
Total_Credit_Cards	6	3	9	5	9
Total_visits_bank	5	1	1	3	1
Total_visits_online	1	4	11	2	11
Total_calls_made	2	7	2	4	2
Count in each segments	386	224	50	610	50

Both Technique took almost same time for execution, considering is a small dataset, but Hierarchical clustering use a high space as we need to store the similarity matrix in the RAM.

K-means gave us more distinct clusters, grouping the customers in 3 distinct clusters:

a) Cluster 0 : 386 customers with moderate Average Limit and in person visit preferable behavior.

b) Cluster 1 : This cluster is composed of 224 customers with low Average Limit and the preferable contact is to call customer service department.

c) Cluster 2 : Online channel are preferable and have the highest Average Credit Limit, composed of 50 customers.

Insights and Recommendations

We have been able to build a clustering model that:

a) The Bank can deploy this model to do customers segmentations.

b) The bank can use it to find the key behaviors that differs one cluster from another.

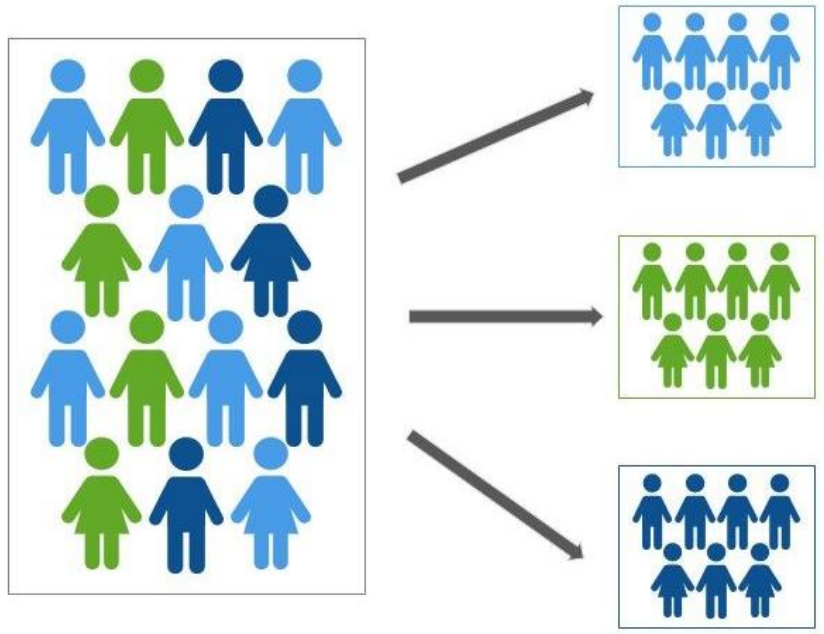
c) Based on it, the bank can take appropriate actions to build personalized campaigns and optimize support service for customers.

Behaviors that differ clusters:

a) Low, Moderate or High Average Credit Limit;

b) How customers prefers to handle bank transactions (In person, By call or Online Channels).

Armed with this segmentation, All Bank Life can not only craft better value propositions but also identify groups that are not well served by current offers.



Insights and Recommendations

Cluster 2: Our Platinum customers have high total credit cards; they look for easy of use through online channels and for better benefits and fees.

- We can use more Behavioral Segmentation like Spending habits, Usage of Rewards and track what they look on the online channels to understand more they needs.

- Considering their high Credit Limited, they are looking for more sophisticate tools in our online channel like Card usage (day, week, month), Spending by category, Rewards points, Promotions etc. That will help them track they spending and manage their credit cards.

- Also we can set online as the preferable communication channel, using e-mail text and app pop up to talk with our customers offering new products, campaigns and for satisfaction survey.

Cluster 1: These are our customers that use online channel moderate and a high call contact and have a low Credit Limit.

- We need to tabulate the calls to understand what they are looking for, are the calls related to how to use the online channel?

- We should use our call center to tabulate the data capturing customers concerns, monitor complains and develop dashboards to detect addressable patterns that can help us to improve our support service, personalize campaigns , keep our customers and increase our customer base.

- As phone is the preferable communication channel, we can text and call this customers for personalized campaigns and satisfaction survey.

Cluster 0: 59% of our customer base falls into this cluster, with moderate Credit Card Limit and in person contact preferable followed by calls.

- We should use Demographic Segmentation as age, to understand their preferable contact model, Are they young or middle age customers?

- We also should map their visits, what are they looking for? Deep understand our customers needs will help us to improve on our service and better target customers with our campaigns.

- These are the king of customers we should reach by mail.

- We can do a campaign to incentive them use our online channels, start with a tour through the online channel and interactive menus that helps them understand easily how it works.

Insights and Recommendations

i. Therefore, redesign our online channels with more intuition bottoms, adding the most used functions on a more visibly way, like balance, due date and payments methods.

ii. Adding more sophisticate tools like Card usage (day, week, month), Spending by category, Rewards points

iii. Run periodic satisfaction survey in all different contact channels to understand what customers wants and what others Credit Card are offering can help us develop the best marketing campaign.

iv. Understand how satisfied are the customers and what we can improve to provide the best experience for them is the most important takeaway to differs our company.

Thank You

Amanda Mendonca