

AllLife Credit Card Customer Segmentation

By:
Sukhmani Kaur Bedi

Background

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 personalised 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 bank poorly. Based on this, the Operations team wants to upgrade the service delivery model, to ensure that customers queries are resolved faster. Head of Marketing and Head of Delivery both decide to reach out to the Data Science team for help.

Business Problem and Key Questions

- To identify different segments in the existing customer based on their spending patterns as well as past interaction with the bank.
- How many different segments of customers are there?
- How are these segments different from each other?
- What are your recommendations to the bank on how to better market to and service these customers?

Solution to approach

- Perform univariate analysis on the data to better understand the variables and to get an idea about the no of clusters.
- Execute K-means clustering to create clusters
- Execute hierarchical clustering (with different linkages) with the help of dendrogram and cophenetic coeff.
Analyse clusters formed

Data Overview

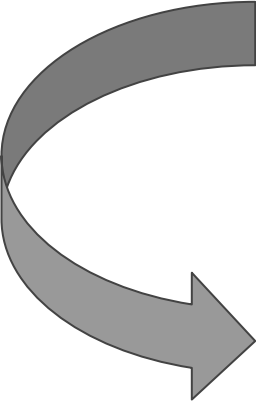
- There are 660 rows and 6 attributes in the data.

Variable	Datatype	Description
Customer Key	int64	Identifier for the customer
Avg_Creit_Limit	int64	Average credit limit across all the credit cards
Total_Creit_Cards	int64	Total number of credit cards
Total_visits_bank	int64	Total number of bank visits
Total_visits_online	int64	Total number of online visits
Total_calls_made	int64	Total number of calls made by the customer

- The data had no standard and non-standard missing values.
- There were no duplicate entries in the data set.

Hypothesis Generated

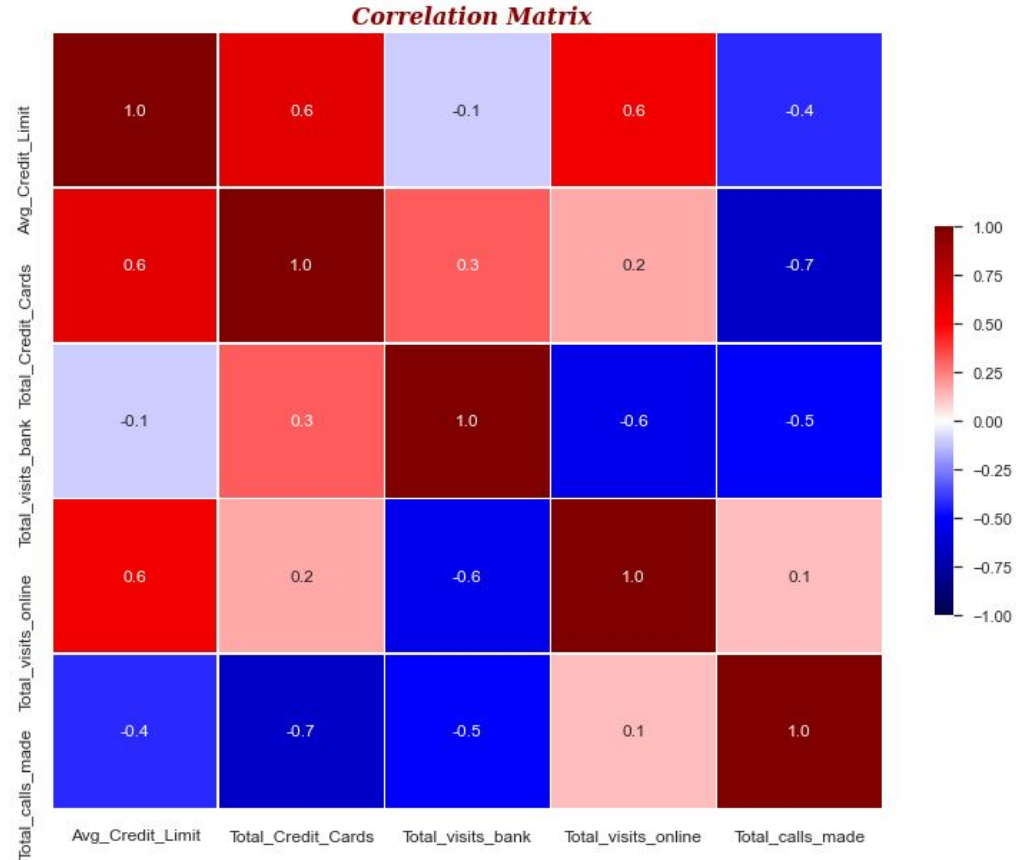
CONTACT METHOD



I assume that there would be three clusters for contact method, where customers would stick to their preferred method for interacting with their bank (online, in person, and through the phone). Shall proceed with further exploratory data analysis and Clustering technique to prove the hypothesis statement.

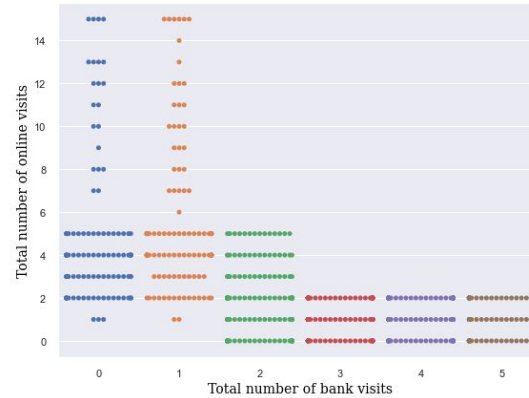
Exploratory Data Analysis

- Total number of credit cards is strongly negatively correlated with Total number of calls made
- Total number online visits`is negatively correlated with Total number of bank visits, which is obvious because customers resolving their issues online do not need to visit the bank.
- Average credit limit is positively correlated with Total number of online visits and Total number of credit cards



Exploratory Data Analysis contd....

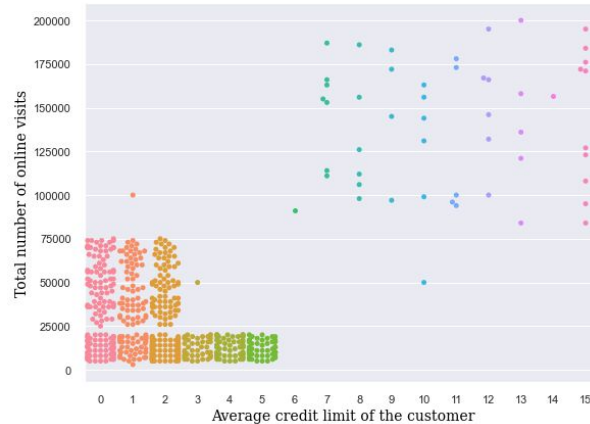
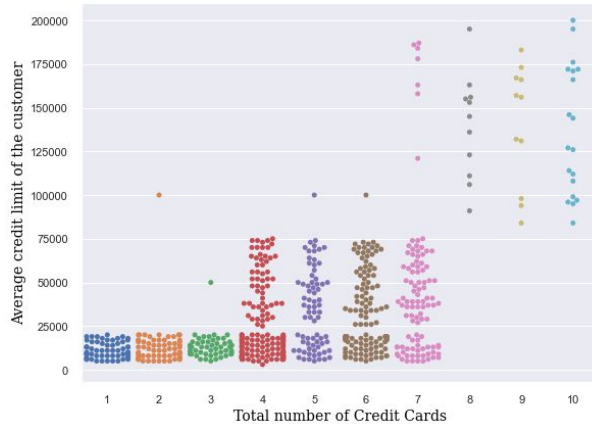
NEGATIVE CORRELATION



- The higher the credit limit, less support calls are made by the customer
- As the number increases for online visits, a decline can be seen for the number of bank visits.
- It can be assumed that customers with less credit limit requires more support service from the bank. Also, we can see that there could be 3 possible clusters for customer segmentation.

Exploratory Data Analysis contd....

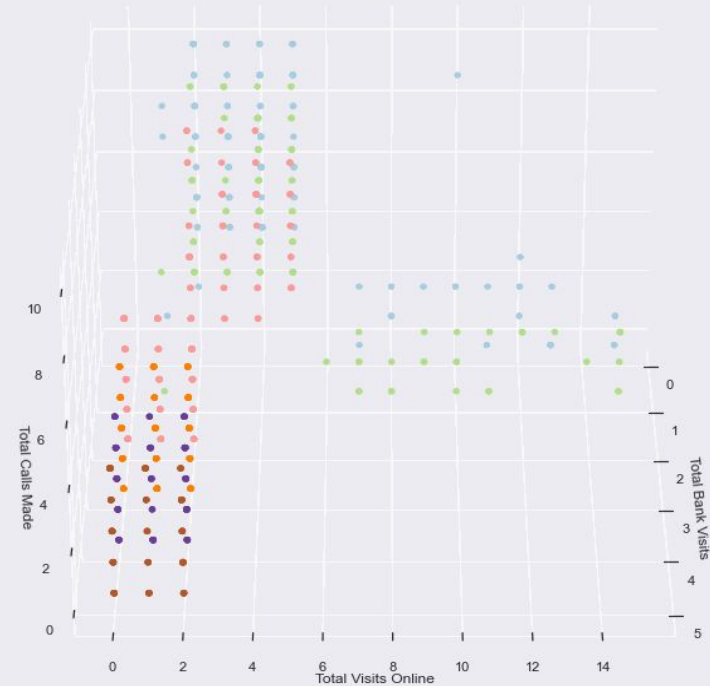
POSITIVE CORRELATION



- It is obvious that customers with more number of credit cards will have a higher average credit limit
- Also, customers with higher credit limit prefers to visit bank online rather than being physically present there

Exploratory Data Analysis contd....

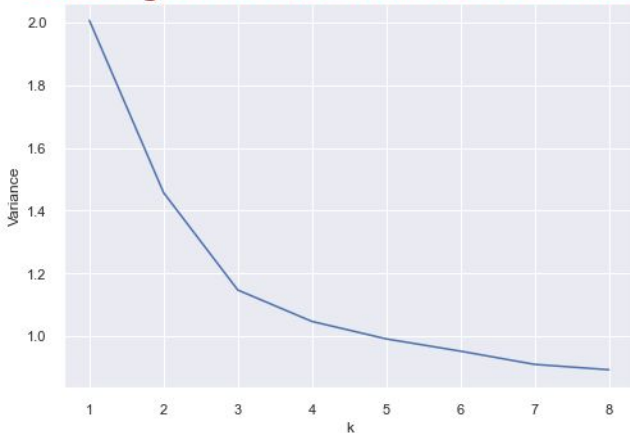
This 3D rotating scatter plot shows the hypothesis formed is correct



K-Means Clustering

First, we want to iterate through and view the performance of each value of K for K-means, then using a line graph to find the elbow of the plot we can select the optimal number of cluster and then finally check the silhouette score for the optimal number of clusters.

Selecting k with the Elbow Method- Variance



Silhouette Score for each value of k

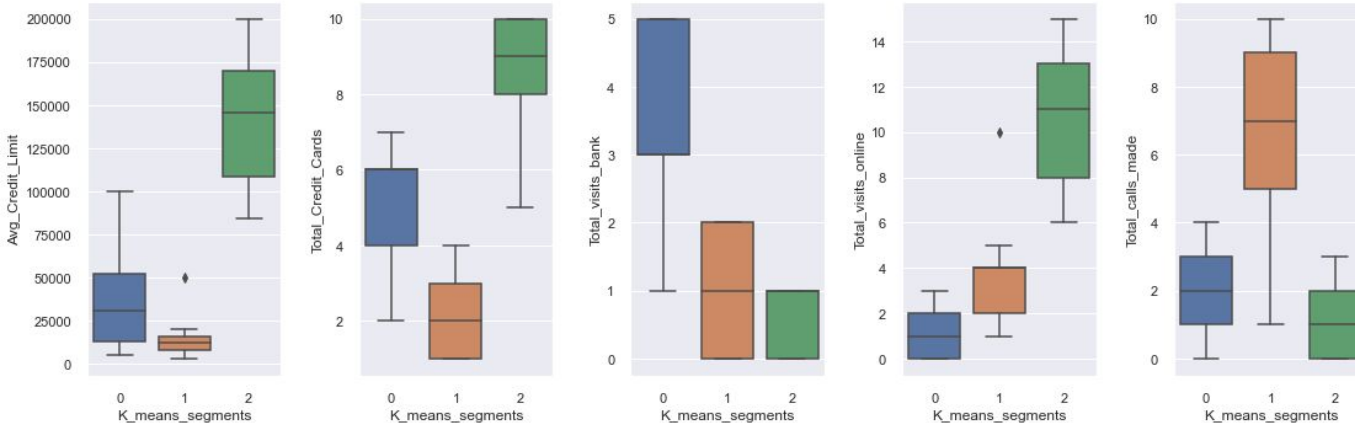


Appropriate number of clusters is 3

K-Means Clustering contd...

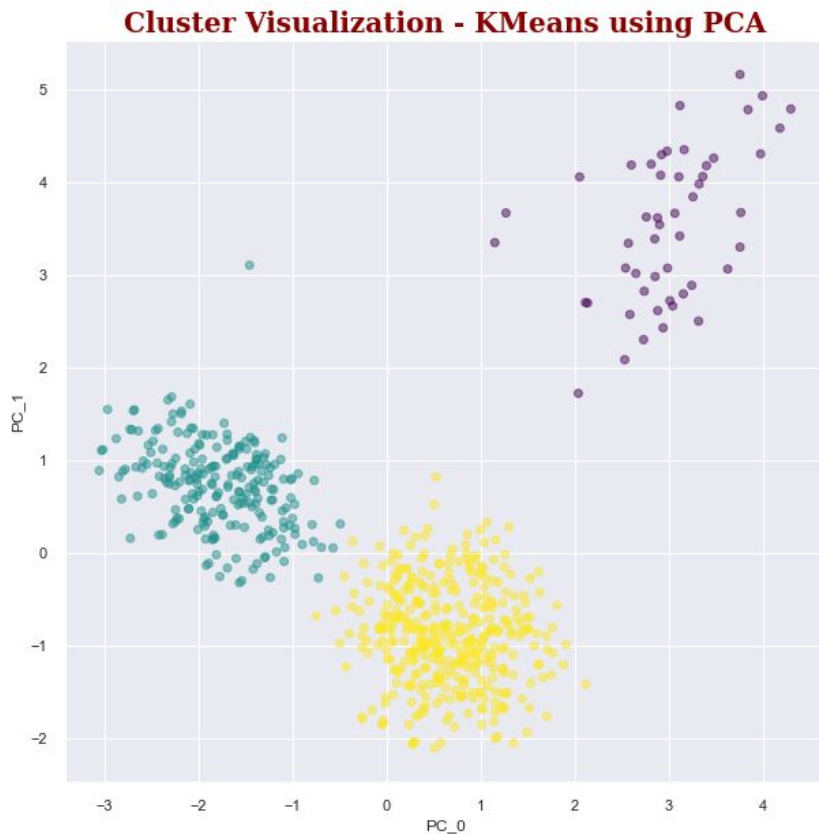
	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
K_means_segments						
0	33782.383420	5.515544	3.489637	0.981865	2.000000	386
1	12174.107143	2.410714	0.933036	3.553571	6.870536	224
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50

Boxplot of numerical variables for each cluster- KMeans Clustering



- Highest number of customer base(386) can be seen for cluster 0, which has the highest mean value of total bank visits
- Second highest customer base(224) can be seen for cluster 1, which has the highest mean value of total calls made
- Finally, the least customer base(50) can be seen for cluster 2, which has the highest mean value of total online visits

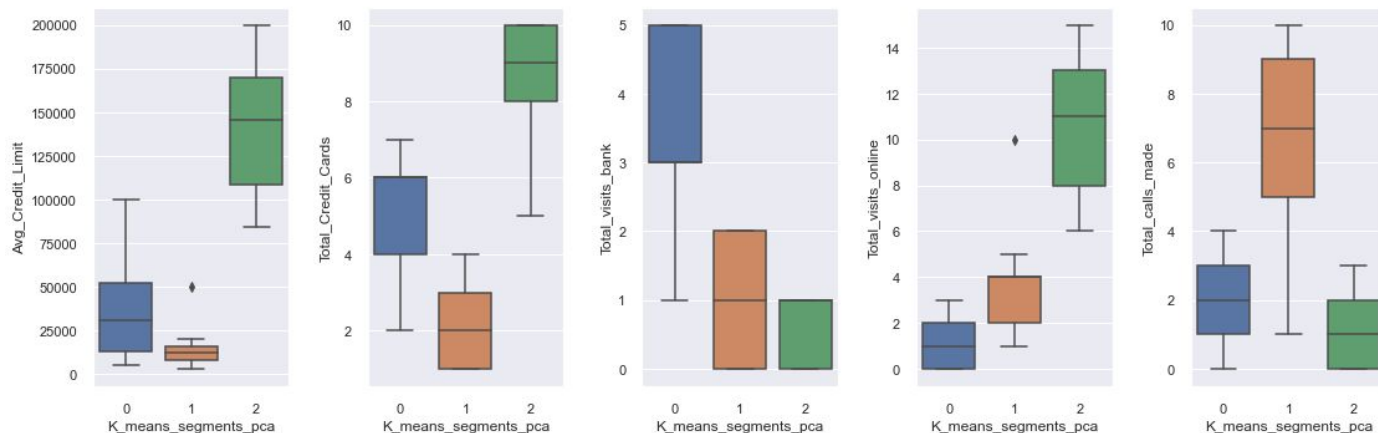
K-Means Cluster Visualization using PCA



Clusters are segregated well

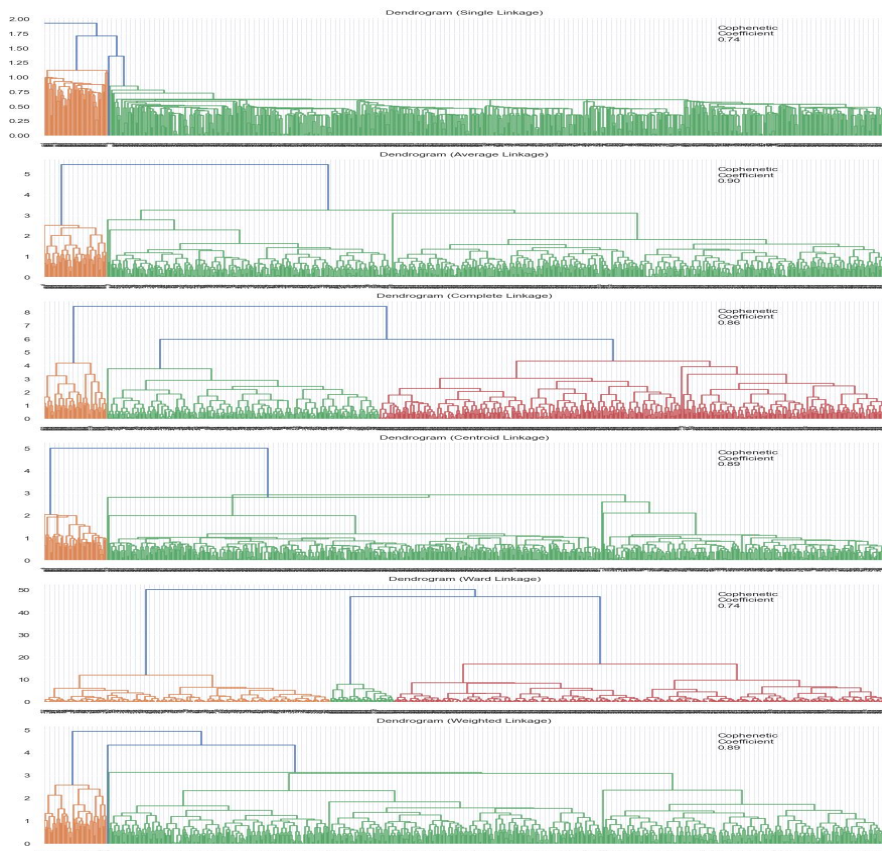
K-Means Clustering using PCA

Boxplot of numerical variables for each cluster - KMeans Clustering - PCA



No difference in composition of clusters can be seen after dimensionality reduction

Hierarchical Clustering

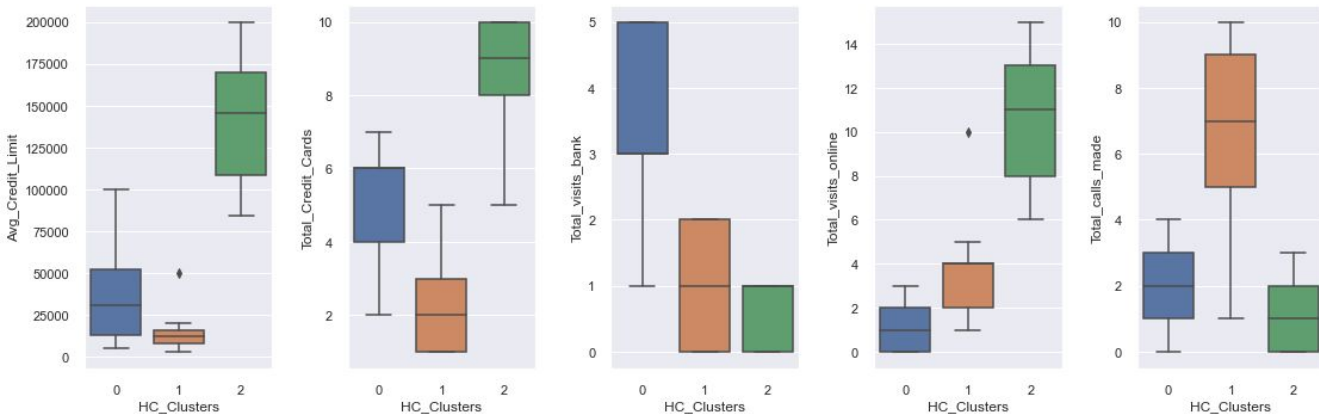


According to the Dendrogram and Cophenetic Coefficient and Silhouette score, Ward Linkage seems to be the best one with 3 as appropriate number of clusters

Hierarchical Clustering contd...

	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
HC_Clusters						
0	33851.948052	5.516883	3.493506	0.979221	1.994805	385
1	12151.111111	2.422222	0.937778	3.546667	6.857778	225
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50

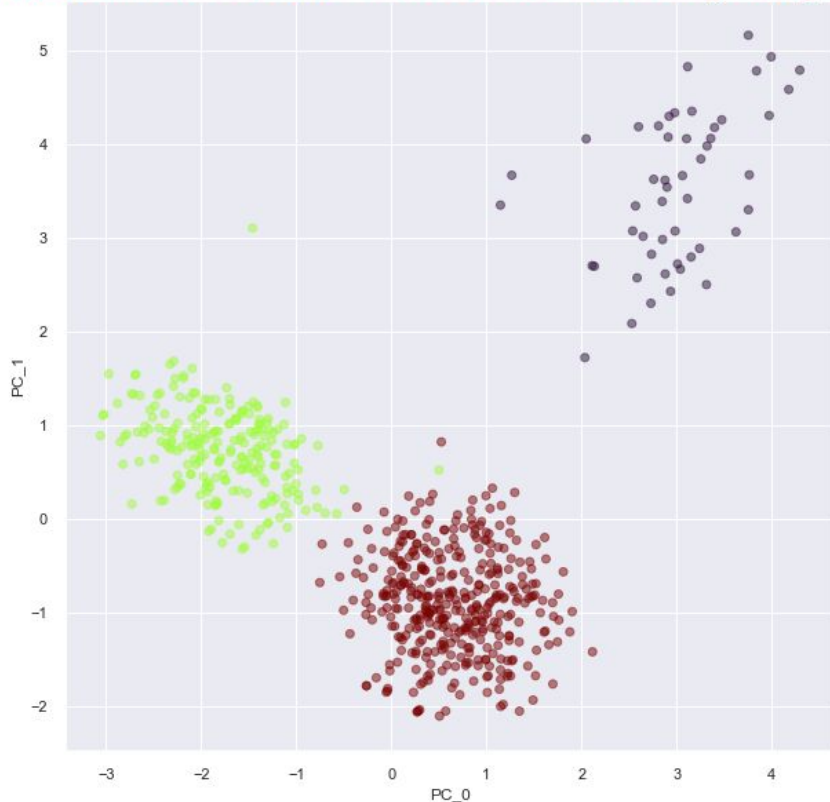
Boxplot of numerical variables for each cluster - Hierarchical Clustering



- Highest number of customer base(386) can be seen for cluster 0, which has the highest mean value of total bank visits
- Second highest customer base(224) can be seen for cluster 1, which has the highest mean value of total calls made
- Finally, the least customer base(50) can be seen for cluster 2, which has the highest mean value of total online visits

Hierarchical Cluster Visualization using PCA

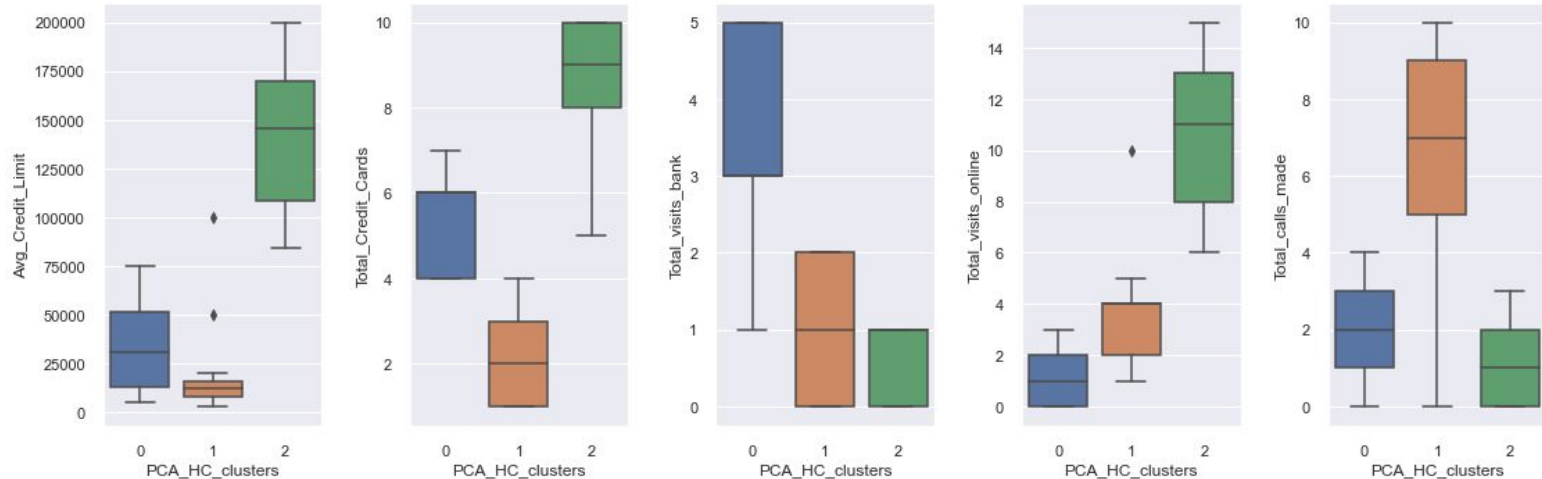
Cluster Visualization - Hierarchical clustering using PCA



Clusters are segregated well,
but some overlapping of one
or two data points can be seen

Hierarchical Clustering using PCA

Boxplot of numerical variables for each cluster - Hierarchical Clustering using PCA



- The range of average credit limit has for cluster 0 has decreased and increased for cluster.
- Similarly, the range of total number of credit cards has been changed.

Cluster Analysis

Cluster 0

Contact Method : IN PERSON

- This group is about 58.5% of the total customer base
- Average Credit Limit is between Rs. 5000 and Rs. 10,000, which is a medium range
- Customers have total number of credit cards between 2 and 7. 50% of the customers have total number of credit cards between 4 and 6, which is a medium range according to the boxplots
- Highest number of bank visits can be seen for this cluster. 50% of the customers visited bank 3 to 5 times.
- Lowest number of online visit customers. 50% of the customers visited the bank via online 0 to 2 times.
- 50% of the customers made support calls between 1 and 3, which is a medium range.

Cluster Analysis

Cluster 1

Contact Method : PHONE

- This group is about 33.9% of the total customer base
- Customers in this cluster have the lowest average credit limit, ranging between Rs. 8000 and Rs. 16,000
- Customers with lowest number of credit cards are grouped in this cluster, ranging between 1 and 4
- Number of times customers visited the bank was between 0 and 2, which can be called a medium range
- Number of times customers visited the bank online was between 1 and 5. An extreme values is also seen, where the customer visited the bank online 10 times.
- Highest number of support calls made by the customers are in this cluster. 75% of the customers had made support calls at most 9 times.

Cluster Analysis

Cluster 2

Contact Method : ONLINE

- This group is about 7.6% of the total customer base
- Customers with maximum average credit limit are grouped in this cluster, ranging between Rs. 85,000 and Rs. 170,000
- Customers with maximum number of credit cards are grouped in this cluster, which is obvious because there is a positive correlation between average credit limit and total number of credit cards. The range is between 5 and 10.
- Lowest number of total bank visits, which is again obvious as these two attributes have a negative correlation. The range is between 0 and 1.
- Customers with maximum number of online visits are grouped in this cluster, ranging between 6 and 15.
- Lowest number of support calls made by the customers are in this cluster, ranging between 0 and 3

Cluster Analysis Summary

- It would appear that my hypothesis of the clusters forming along query method has proven correct.
- If we look at the data we see that there is a group which prefers **cluster 2 online** interactions with their bank, they have a much higher credit limit and also have more credit cards.
- The customers who prefer **cluster 1 telephonic** interactions tend to have the least number of credit cards and the lowest credit limit.
- The customers who contact **in_person cluster 0** are in the middle

One additional observation is that if we tally up the number of interactions per group (how many times they have used online, phone, or in-person services) we see the in-person customers appear to be the most active. This was the opposite of my initial expectations as visiting a bank in person has the highest friction (effort required) to complete.

Business Recommendations :-

- We can tailor contact methods to these customer preferences.
- Online/phone users will probably prefer email/text notifications, while in-person users prefer mail notifications and upselling (when at the bank location)
- Online users tend to have (and presumably use) the most credit, these may be the demographic we want to target with our next ad campaign, focusing on digit recruiting
- Telephonic users own less credit card than others, bank should target them to upsell credit cards services
- Since cluster 0, that is users visiting the bank in person, uses the online banking the least, bank should promote more to group 0 in order for them to use it.
- Assuming cluster 1 who make most phone calls are the customers perceive the support services of the bank poorly. Bank should target group 1 and provide better customers service by conducting feedback survey through phone.