AllLife Bank Customer Segmentation

Nibu Kuriakose

Jul 2021

Content

1. Introduction and Proposed Approach

Background

- AllLife Bank’s marketing department thinks that the bank can increase its credit card service market penetration. Also, the customer service is rated poorly by the customers.
- The bank aims to do the following:
  - ✓ Run personalised campaign targeting new customers and upsell to existing customers.
  - ✓ Reduce time to resolve customer queries by upgrading the service delivery model.

Purpose and Benefits

- Perform exploratory data analysis to understand about the customers.
- Find customer segments based on their spending pattern and past interaction with the bank
- Provide insights and recommendation based on the data.

Proposed Approach

1. Perform exploratory data analysis.
2. Perform clustering using K-means clustering and hierarchical clustering to identify customer segments.
2. Explain the differences between segments.

2. Dataset Information and Feature Engineering

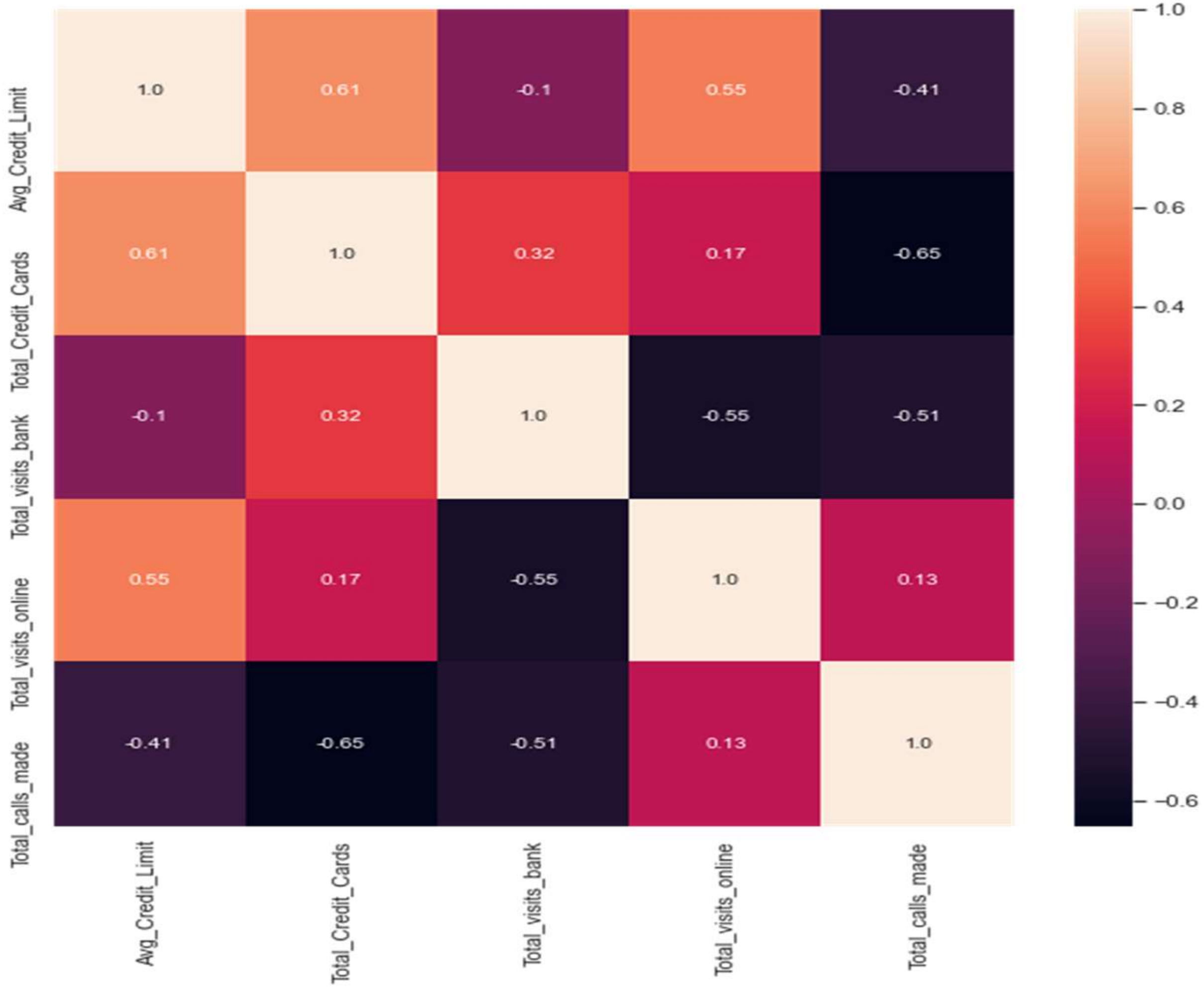
- There are 660 samples. Each sample has 7 attributes
- Once the serial number and customer key were removed, there were 11 duplicate records. Duplicate records were removed before clustering.

3.1 Univariate Analysis

Attribute	Description
Avg_Credit_Limit	• Right skewed data. 50% of customers have credit limit of $18k or less, but there are number of outliers above $100k.
Total_Credit_Cards	• Left skewed data. 50% of customers have 5 or less credit cards. 25% of customers have 6 credit cards or more.
Total_visit_banks	• Right skewed data. 50% of customers visited the bank 2 times or lesser. 25% of customers visited the bank 4 or 5 times.
Total_visit_online	• Right skewed data. 50% of customers used online 2 times or lesser. 25% of customers used online 4 times or more.
Total_calls_made	• Right skewed data. 50% of customers made 3 calls or lesser. 25% of customers made 5 calls or more.

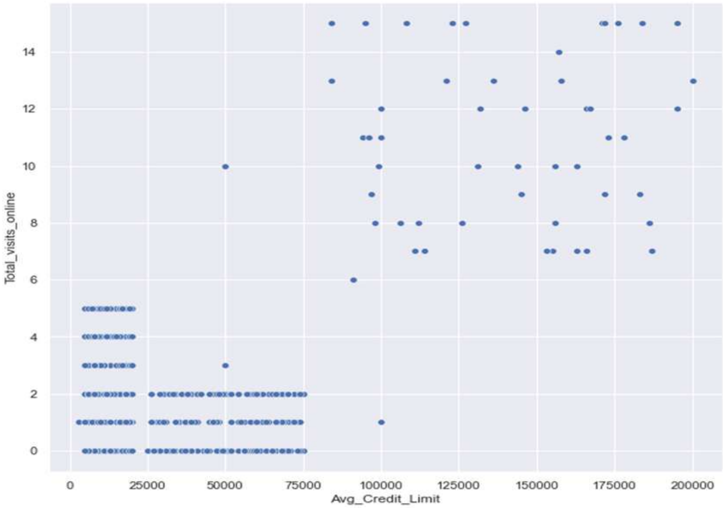
	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made
count	660.0	660.0	660.0	660.0	660.0
mean	34,574.2	4.7	2.4	2.6	3.6
std	37,625.5	2.2	1.6	2.9	2.9
min	3,000.0	1.0	0.0	0.0	0.0
25%	10,000.0	3.0	1.0	1.0	1.0
50%	18,000.0	5.0	2.0	2.0	3.0
75%	48,000.0	6.0	4.0	4.0	5.0
max	200,000.0	10.0	5.0	15.0	10.0

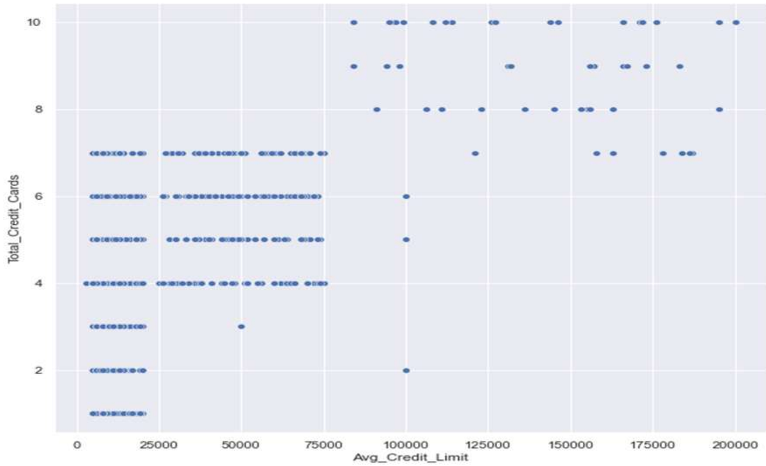
3.2 Multivariate Analysis

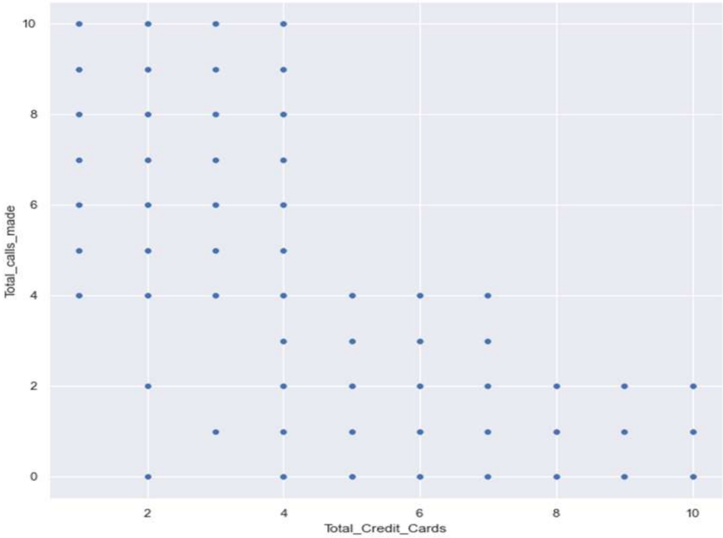


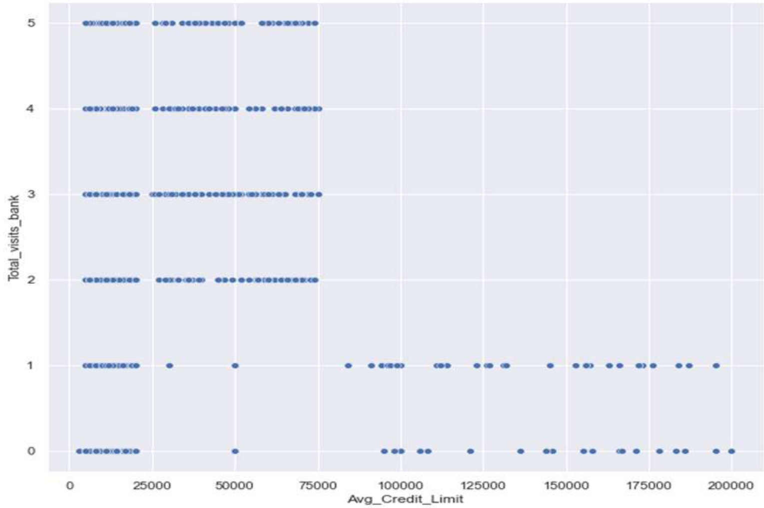
- The credit card limit is positively correlated to the number of credit cards and number of online visits done by the customer.
- Number of calls made by the customer is negatively correlated to the credit card limit, number of credit cards and number of visits to the bank.
- Number of visits to the bank is negatively correlated to the number of customers using online and calls made to the bank.

3.2 Multivariate Analysis



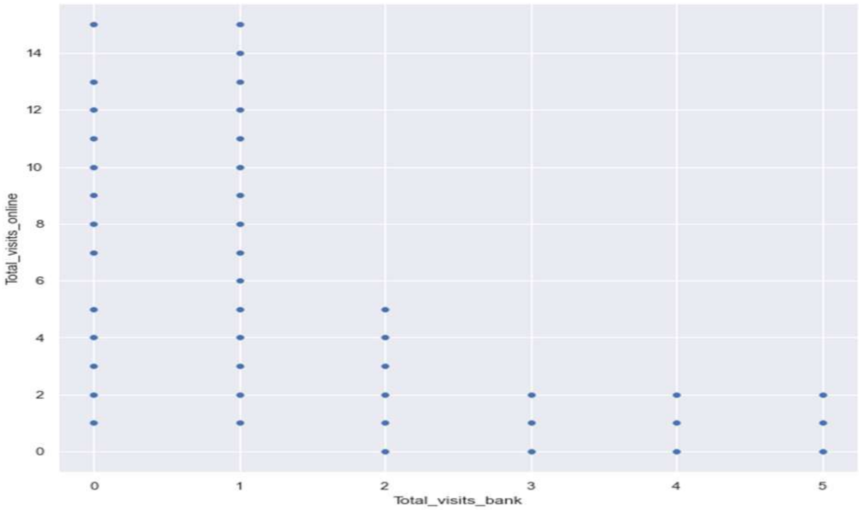


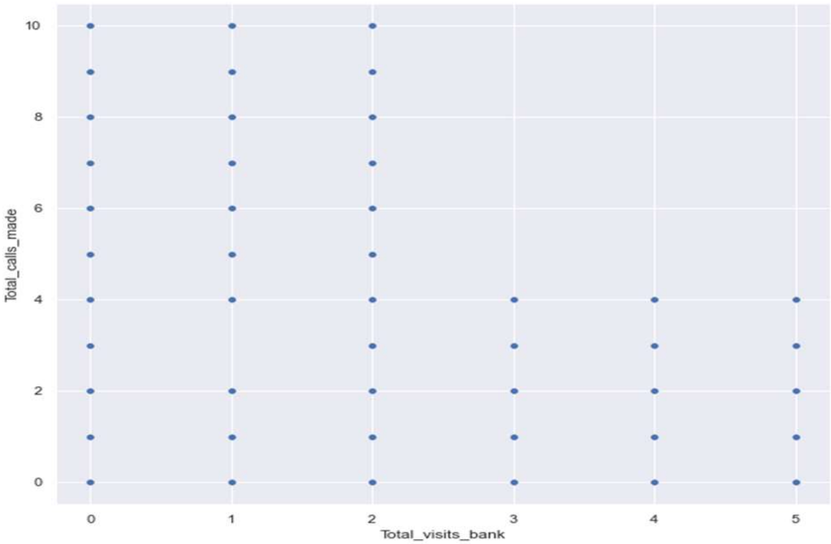


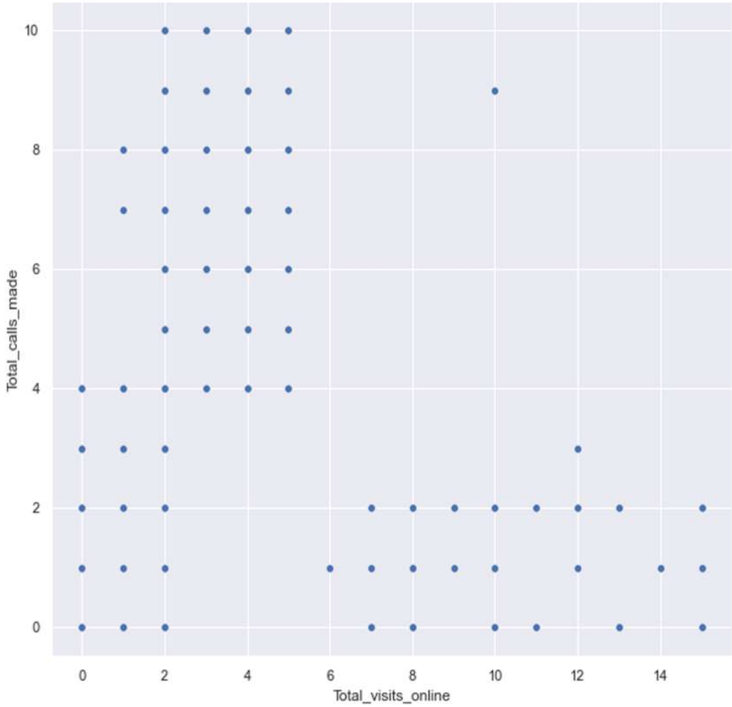


- Customers with high credit limit has mostly used online service. They make less number of visits to banks and calls compared to customers with lower credit limit.
- Customers with lower credit limit uses less online service, but makes more visits to the bank and makes more calls compared to customers with higher credit limit.
- Customers with higher credit limit has more number of credit cards.

3.2 Multivariate Analysis



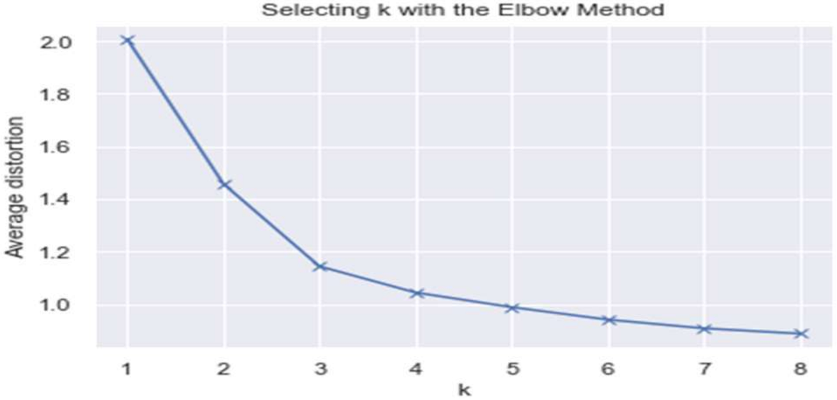




- Customer who does more visits to the bank are customers who has lower usage of online services and makes less number of calls to the bank.
- Customers who either have very low online usage or high online usage, does not make high number of calls to the bank.

4. K-Means Clustering

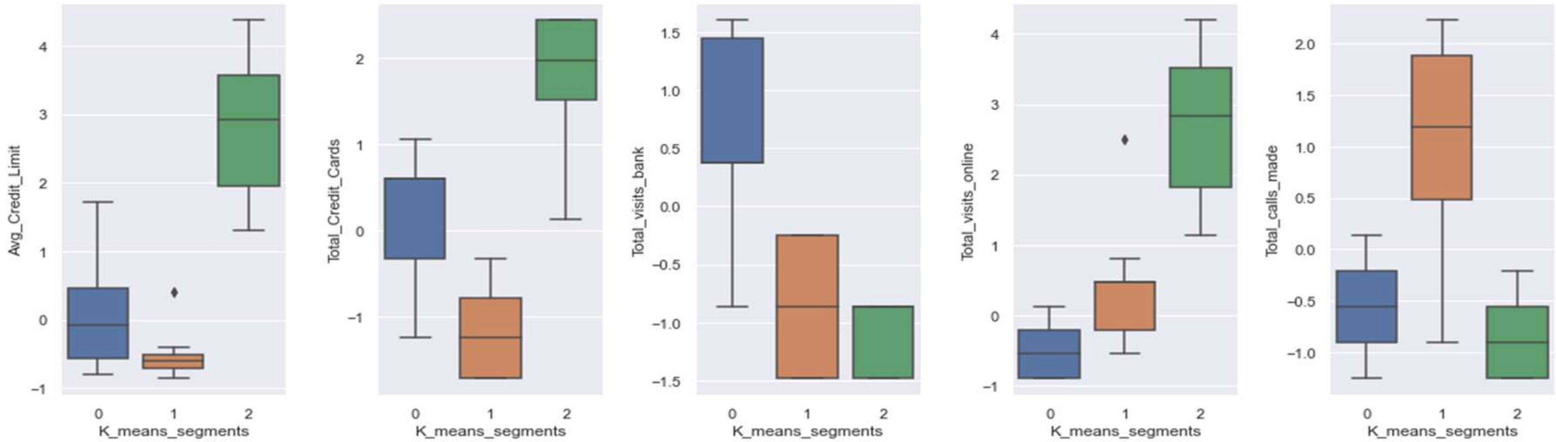
- K-Means clustering gave the below average distortion and silhouette score when we use 1 to 8 clusters.
- We will do further analysis with 3 and 4 clusters due to the following reasons:
  - the drop of distortion is low when we move from 3 to 4 clusters, and 4 to 5 clusters
  - Silhouette score is highest for 3 clusters



Clusters	Average Distortion	Silhouette Score
2	1.46	0.42
3	1.14	0.52
4	1.04	0.36
5	0.99	0.27
6	0.94	0.26
7	0.91	0.25
8	0.89	0.22

4. K-Means Clustering- 3 Clusters

Boxplot of numerical variables for each cluster

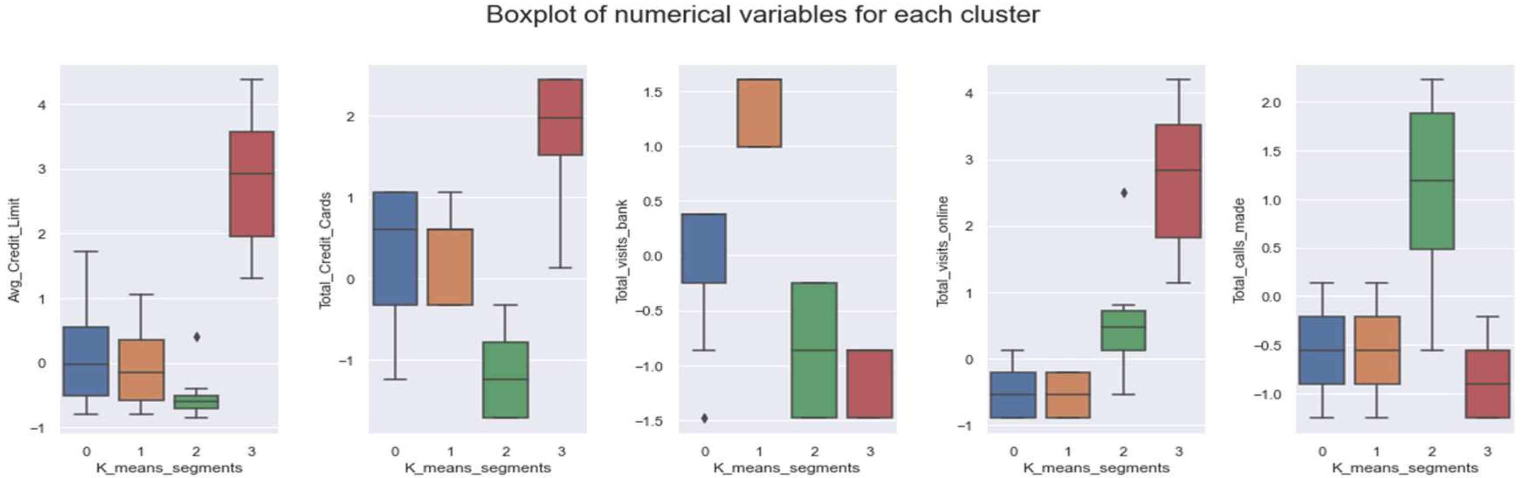


K_means_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	HC_Clusters	count_in_each_segments
0	34071.428571	5.518519	3.484127	0.981481	1.992063	2.489418	378
1	12239.819005	2.411765	0.945701	3.561086	6.891403	0.000000	221
2	141040.000000	8.740000	0.600000	10.900000	1.080000	1.000000	50

K_means_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	HC_Clusters	count_in_each_segments
0	34071.428571	5.518519	3.484127	0.981481	1.992063	2.489418	378
1	12239.819005	2.411765	0.945701	3.561086	6.891403	0.000000	221
2	141040.000000	8.740000	0.600000	10.900000	1.080000	1.000000	50

- 3 clusters gives below segments:
  - ✓ High credit limit customers
  - ✓ Average credit limit customers
  - ✓ Low credit limit customers
- Depending on the customer segment, their mode of contacting the bank (online, call, bank visit) changes.

4. K-Means Clustering- 4 Clusters



K_means_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	35857.142857	5.525510	2.479592	0.974490	2.056122	196
1	31832.432432	5.481081	4.513514	1.005405	1.940541	185
2	12233.944954	2.394495	0.940367	3.582569	6.944954	218
3	141040.000000	8.740000	0.600000	10.900000	1.080000	50

K_means_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	35857.142857	5.525510	2.479592	0.974490	2.056122	196
1	31832.432432	5.481081	4.513514	1.005405	1.940541	185
2	12233.944954	2.394495	0.940367	3.582569	6.944954	218
3	141040.000000	8.740000	0.600000	10.900000	1.080000	50

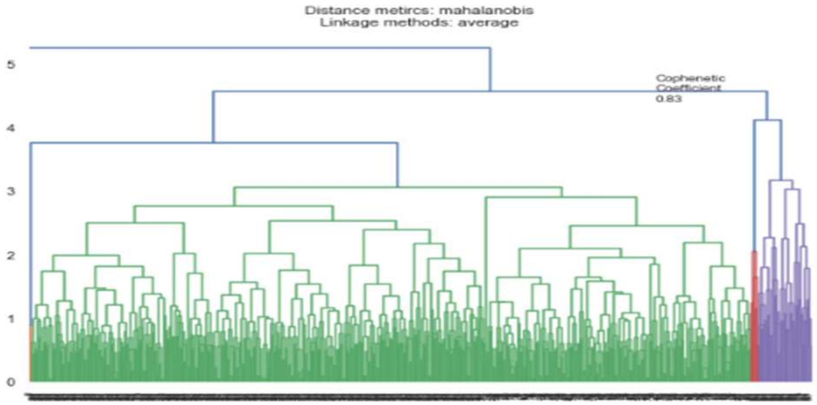
- The segmentation using four clusters has provided an additional segment by splitting the average credit limit customer into segments which are:
  - ✓ (category-1) Average credit limit customers who make high number of visits to the bank compared to other segments.
  - ✓ (category-0) Average credit limit customers who has the least online usage compared to other segments, but do not make high number of bank visits compared to the above category.
- As four clusters provides more separation of customer types with unique features, **4 clusters is preferred over 3 clusters.**

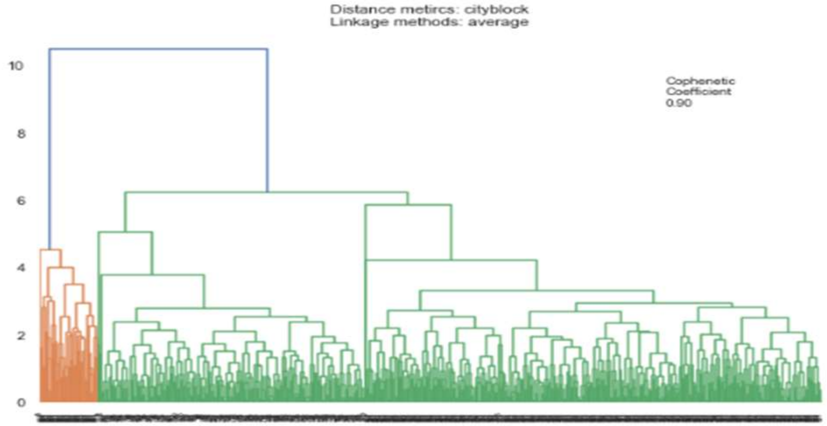
5. Hierarchical Clustering

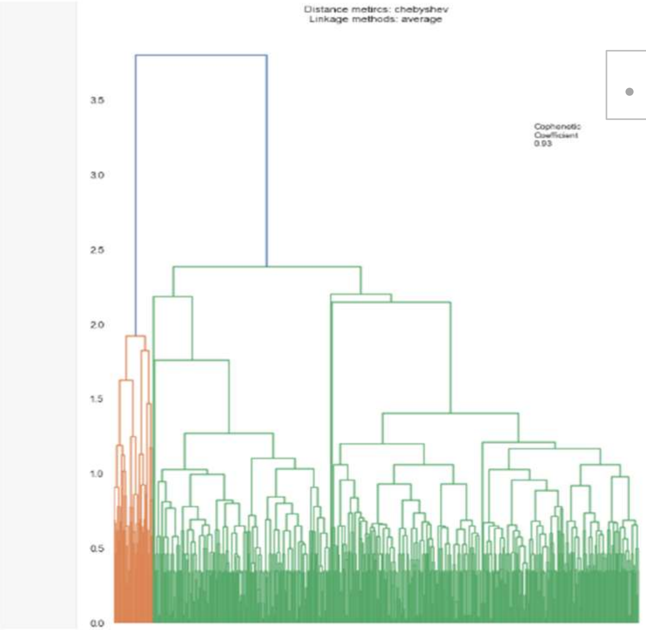
- Various linkage methods using different distance measuring types were used to get the Cophenetic coefficient.
- The combinations marked in red boxes gave high cophenetic coefficient. The dendograms for these will be plotted to select which one should be selected for creating the clusters.

Distance	Linkage Method	Cophenetic Coefficient
Euclidean	Single	0.74
Euclidean	Complete	0.88
Euclidean	Average	0.90
Euclidean	Median	0.88
Euclidean	Ward	0.74
Euclidean	Weighted	0.86
Euclidean	Centroid	0.89
chebyshev	Single	0.74
chebyshev	Complete	0.86
chebyshev	Average	0.90
chebyshev	Weighted	0.89
mahalanobis	Single	0.71
mahalanobis	Complete	0.62
mahalanobis	Average	0.83
mahalanobis	Weighted	0.78
cityblock	Single	0.73
cityblock	Complete	0.87
cityblock	Average	0.90
cityblock	Weighted	0.88

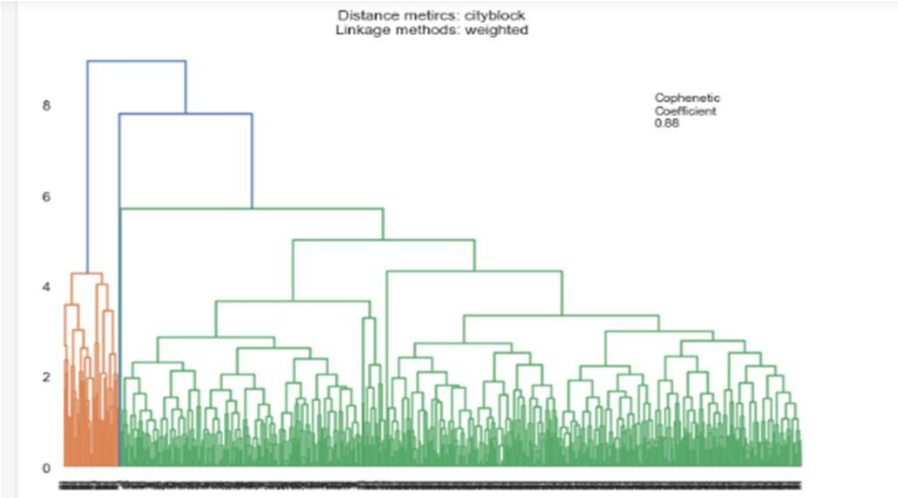
5. Hierarchical Clustering- Dendograms





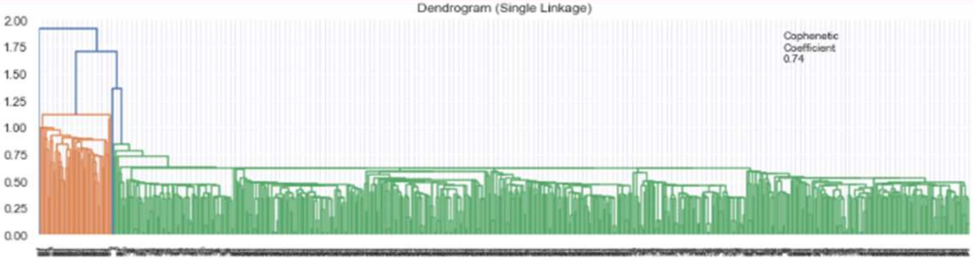


- None of these dendograms show distinct clusters.

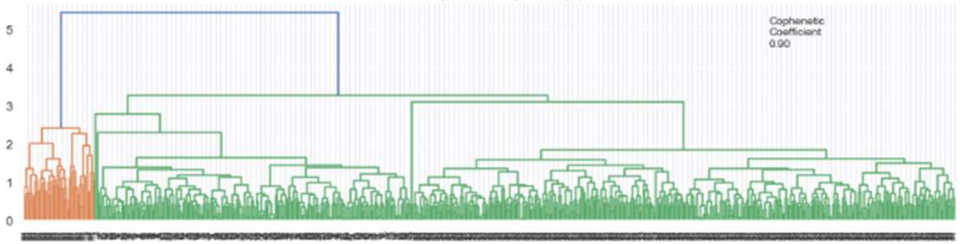


5. Hierarchical Clustering- Dendograms

Dendrogram (Single Linkage)

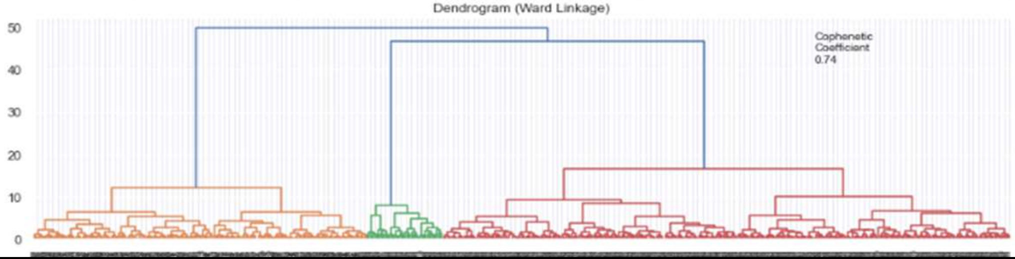


Dendrogram (Average Linkage)

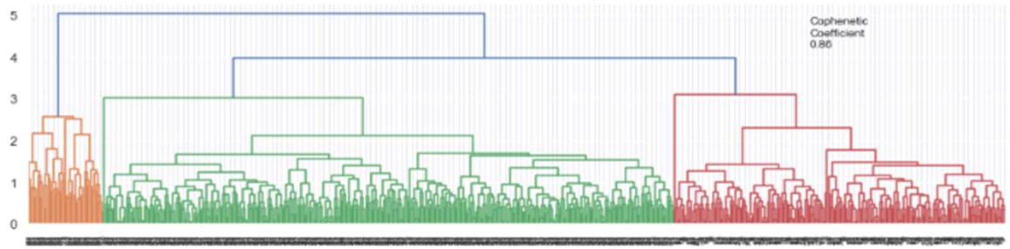


Dendrogram (Complete Linkage)

Dendrogram (Ward Linkage)

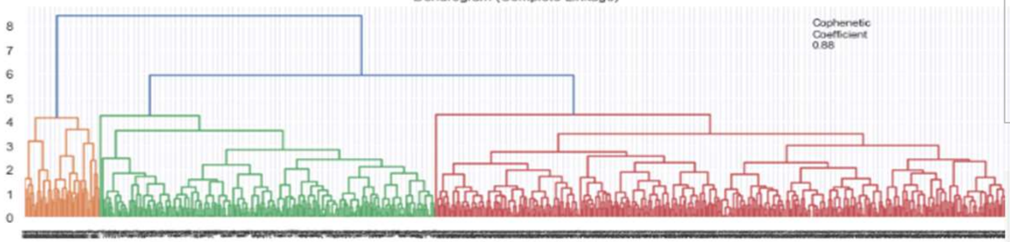


Dendrogram (Weighted Linkage)

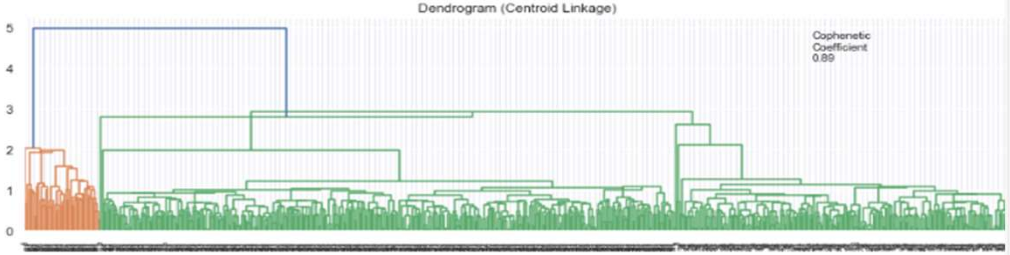


- The dendogram with Euclidean distance and Ward (in black box above) method gives four distinct clusters. Hence, this will be used for creating the clusters.

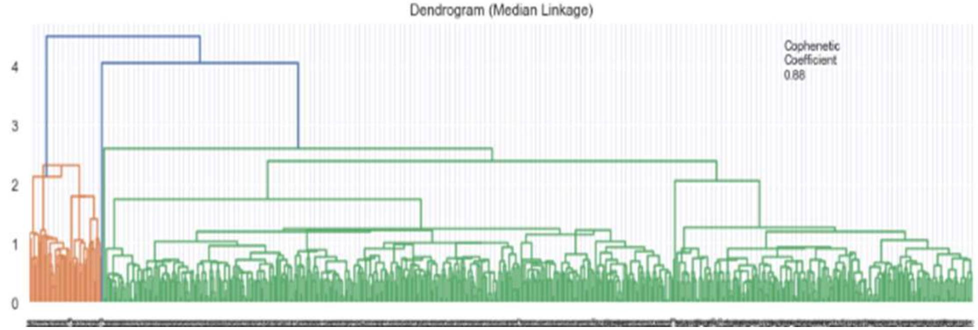
Dendrogram (Complete Linkage)



Dendrogram (Centroid Linkage)

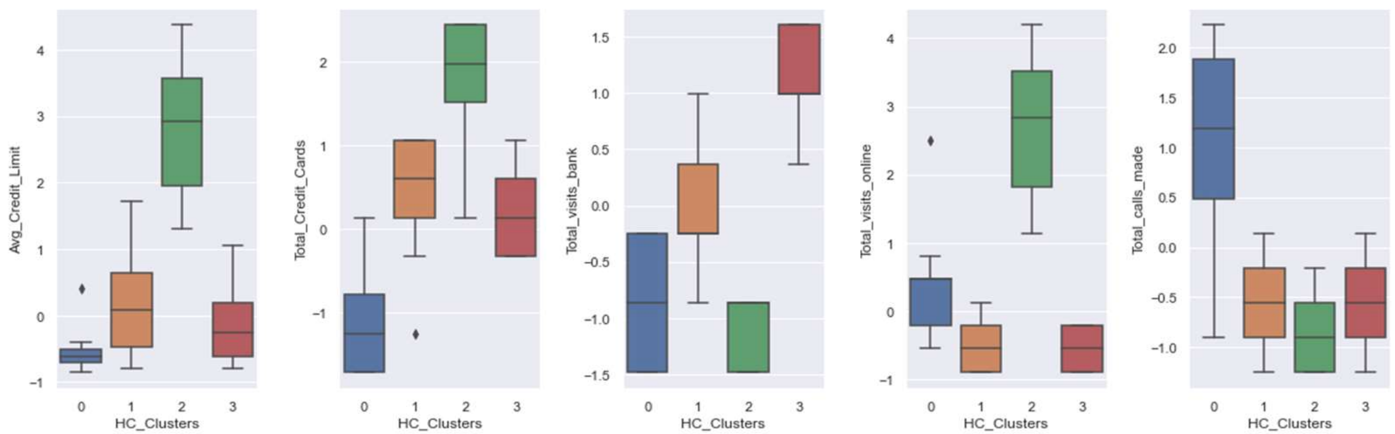


Dendrogram (Median Linkage)



5. Hierarchical Clustering- 4 Clusters

Boxplot of numerical variables for each cluster



HC_Clusters	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	12216.216216	2.423423	0.950450	3.554054	6.878378	222
1	38906.593407	5.719780	2.510989	0.972527	2.043956	182
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50
3	29697.435897	5.333333	4.400000	0.984615	1.933333	195

HC_Clusters	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	12216.216216	2.423423	0.950450	3.554054	6.878378	222
1	38906.593407	5.719780	2.510989	0.972527	2.043956	182
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50
3	29697.435897	5.333333	4.400000	0.984615	1.933333	195

- 4 clusters gives below segments:
  - ✓ Cluster-2: High credit limit customers with high online usage.
  - ✓ Cluster-1: Average credit limit customers, with average visit to bank and lowest online usage
  - ✓ Cluster-3: Average credit limit customers, with high visit to bank and low online usage
  - ✓ Cluster-0: Low credit limit customers with high calls made to the bank.
- Depending on the customer segment, their mode of contacting the bank (online, call, bank visit) changes.
- A detailed analysis of each of the clusters is given in the following slide.

6. Cluster Analysis

Cluster	# of customers	Credit Limit	Credit Cards	Visit to bank	Phone Calls	Online	Summary
0	222	**Low**	Low	Low	**High**	Average	Low credit limit customers who uses phone calls to contact the bank, and have relatively low visit to banks and use of online services.
1	182	**Average**	Average	Average	Low	Low	Customers who have average credit limit, and average visit to banks. They make relatively less number of visits to bank and has low usage of online services.
3	195	**Average**	Average	**High**	Low	Low	Customers who has average credit limit, and high number of visit to banks. They make relatively less phone calls to the bank and low use of online services.
2	50	**High**	**High**	Low	Low	**High**	Customers who have high credit limit, and high usage of online services. They make relatively less number of visits to bank and calls to the bank.

HC_Clusters	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	12216.216216	2.423423	0.950450	3.554054	6.878378	222
1	38906.593407	5.719780	2.510989	0.972527	2.043956	182
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50
3	29697.435897	5.333333	4.400000	0.984615	1.933333	195

HC_Clusters	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	12216.216216	2.423423	0.950450	3.554054	6.878378	222
1	38906.593407	5.719780	2.510989	0.972527	2.043956	182
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50
3	29697.435897	5.333333	4.400000	0.984615	1.933333	195

7. Comparison- K-Means and Hierarchical Clustering

K-Means Clusters

K_means_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	35857.142857	5.525510	2.479592	0.974490	2.056122	196
1	31832.432432	5.481081	4.513514	1.005405	1.940541	185
2	12233.944954	2.394495	0.940367	3.582569	6.944954	218
3	141040.000000	8.740000	0.600000	10.900000	1.080000	50

Hierarchical Clusters

HC_Clusters	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	12216.216216	2.423423	0.950450	3.554054	6.878378	222
1	38906.593407	5.719780	2.510989	0.972527	2.043956	182
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50
3	29697.435897	5.333333	4.400000	0.984615	1.933333	195

- Clusters formed using k-means and hierarchical clustering are similar, with some differences.
- Key similarities between the clusters using the two clustering methods are:
  - Four clusters- one cluster each for low and high credit limit customers, and two clusters for average credit limit customers
  - Preferred method of contacting the bank for each cluster
- Key differences between the clusters using the two clustering methods are:
  - Credit limit, number of customers and credit cards for the low and average credit limit clusters.
  - Average count of various methods of contacting the bank is also different.

8. Key Insights

Generic

1. 50% of customers have credit limit between $10k and $48k indicating majority of the customers have average credit limit, with smaller number of customers with low and high credit limit. There are only 7% of customers in the high credit limit customer segment.
2. Customers with high credit card limit have relatively high number of credit cards. This might be due to their high spending power and good credit history.
3. Highest customer contact has been through phone calls (42%), followed by online (30%) and visiting the bank (28%). Similarly, the average number of phone calls made to the bank is much higher than average number of online visits and bank visits. The high usage of phone calls might be a reason for poor perception of customer service as the resolution time is higher over phone compared to online or visit to the bank.

Cluster Analysis

The way of contacting the bank differs based on customer segment as given below:

Online

- Used mostly by high credit limit customers and these customers makes low number of visits to the bank and phone calls.
- Online usage is low by average and low credit limit customers

Bank visit

- Mostly used by average credit limit customers- The customers at the lower end of this segment have high number of bank visits compared to the customers in the upper end of this segment
- Visit to banks is relatively low for high and low credit limit customers.

Phone calls

- Mostly used by low credit limit customers which forms the biggest customer segment
- Phone calls is low for high and average credit customers who prefers online and bank visits respectively.

9. Recommendations

1. Target high income customers or business community due to the following:
   - to increase number of customers with high credit limit.
   - high credit limit customers have more credit cards and is more likely to take up a new credit card

2. Target lower income customers to increase the customer base as there are only few customers with lower credit limit. This might provide a higher market penetration and provide lower risk customers due to lower credit limit.

3. Measure service levels and resolution time for services provided through phone to understand if there is an underlying issue.

4. Identify why the customers at the lower end of the average credit limit visits the bank very often compared to the average credit limit customers at the higher end of the average credit limit segment.

5. Conduct a survey for low and average credit limit customers to understand why they prefer services through phones calls or bank visits, and what is preventing them from using online services.

6. Incentivise customers to use more online service by the following as it allows to provide better service at lower cost:
   - educating users to use online services
   - advertising about online service in the bank branches and during the phone calls
   - lower service charges for using online services- this will be effective considering the low and average credit limit will be encouraged to use the online service if it lowers their service charge
   - provide incentive to buy laptops and hand-held devices using the credit card