ALL LIFE BANK CUSTOMER SEGMENTATION

JAMEEL KHAN

Objective

- 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 back 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

- To identify different segments in the existing customer based on their spending patterns as well as past interaction with the bank

- Provide Key insights and Recommendations from Data Analysis

Dataset Characteristics

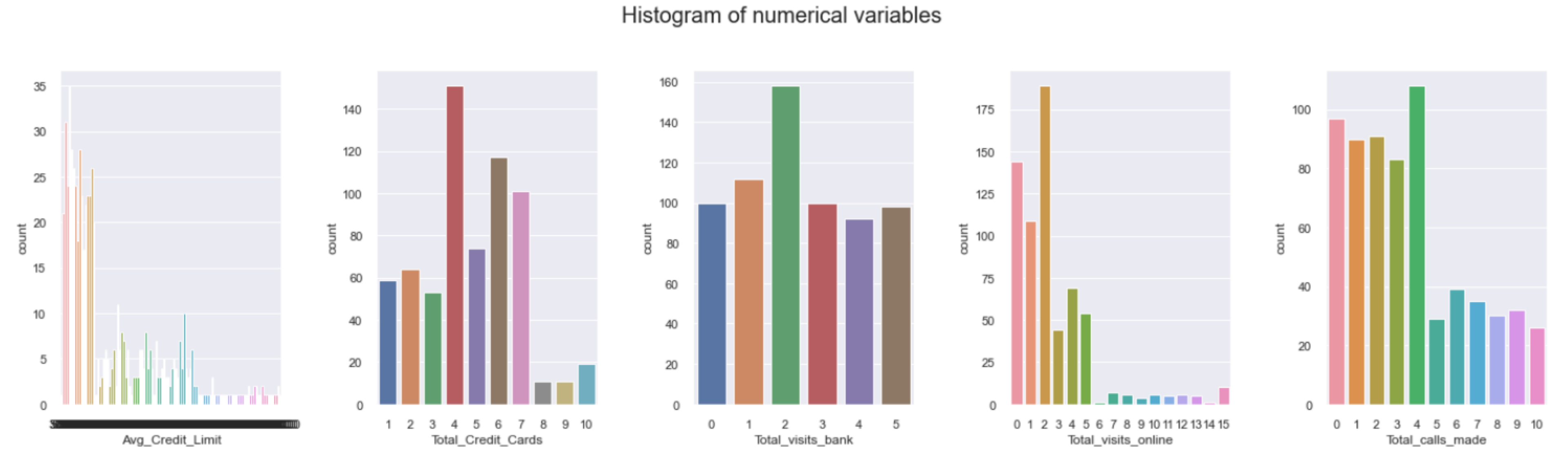
Dataset Column	Description
Sl_No	Serial Number
Customer Key	Customer Unique ID
Avg_Credit_Limit	Average Credit Limit of all cards in possession
Total_Credit_Cards	Number of credit cards in possession
Total_visits_online	Total visits of online portal used by customer for contact
Total_calls_made	Total phone calls made by customer for contact
Total_visits_bank	Total visits to brick and mortar banks made by customer for contact

The dataset contains customer centric and other metrics of Bank customers

Observations on Data Set:

1. There are 7 columns of data for each customer, with a total of 660 rows of data
2. The dataset has no missing data.

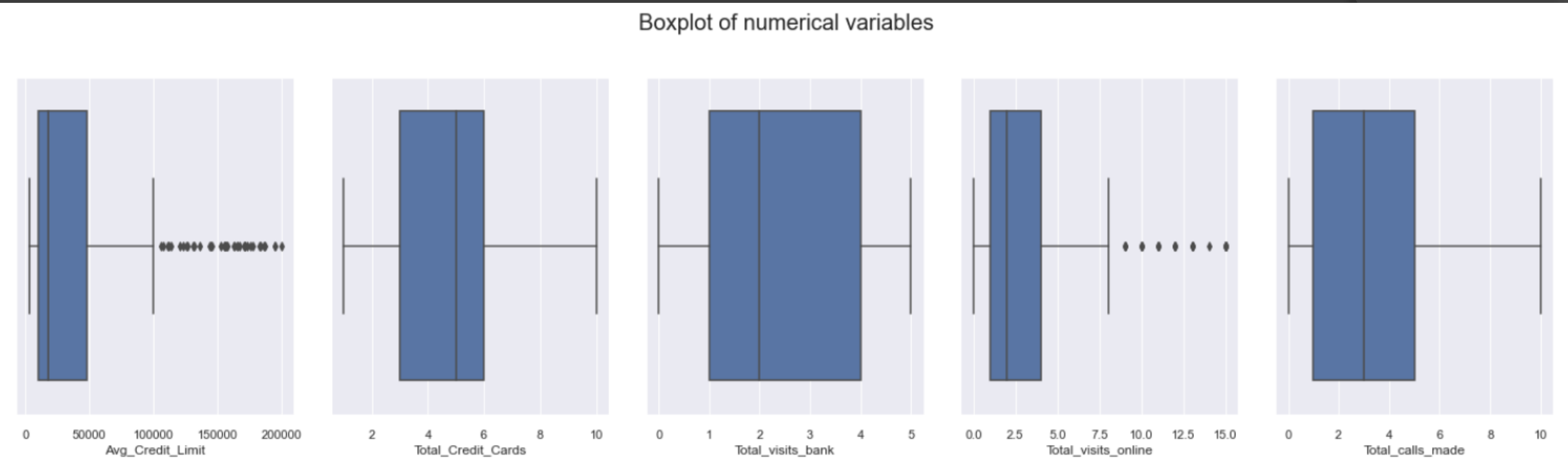
Exploratory Data Analyses – Distribution



Observations on Numerical Data Distribution in the dataset:

- Most of the features appear to be right skewed, as expected.
- Most customers have 4-7 credit cards. This section could be targetted better for retention
- Banks visits are evenly distributed.
- Customers with low Credit Limit prefer to call-in with queries
- Whereas, Customers with Medium to High Credit limit prefer to use the online bank portal
- Sl_No and Customer Key is just serialized information and will have no significance to modelling. So, we will drop them. All other features are significant for clustering.

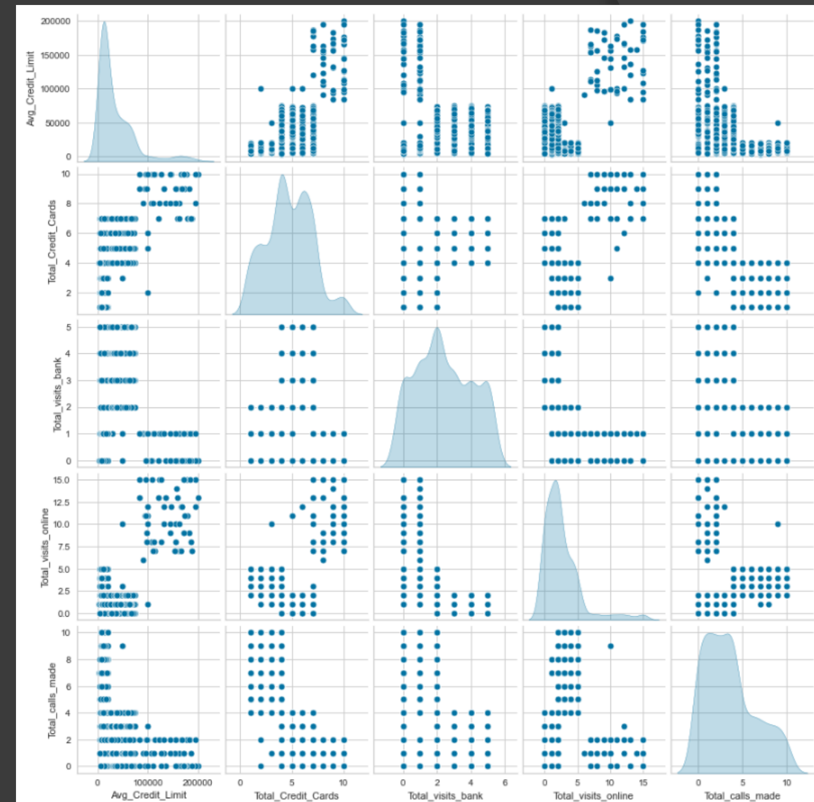
Exploratory Data Analyses – Outliers



Observations on Numerical Data Distribution in the dataset:

- There are outliers in Avg_Credit_Limit and Total_visits_online. However, they are uniformly distributed so we will leave them untreated and they could be form a cluster.

Exploratory Data Analyses – Correlation



Observation on pair plots and Correlation matrix

- Interestingly, there is a negative correlation between the three query approaches made by the customers. So the segment of customers making phone queries are different from online bank customer segment and from brick-and-mortar customers. These segments are mutually exclusive. This implies that the customers can be clustered into different segments
- Total Credit cards and Average Credit Limit are highly correlated.
- Lower the credit limit, higher the phone calls and higher the visits to the bank
- Higher credit limit, higher the online bank visits
- Most distributions are rightly skewed

ML Model Building – K Means Cluster

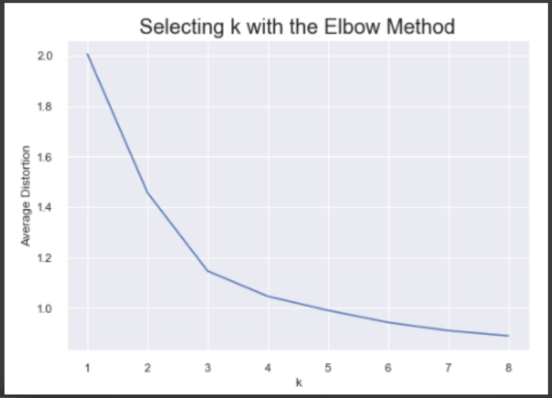
Post Data Preprocessing,

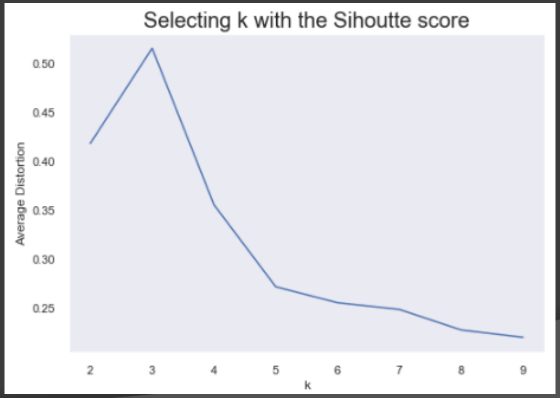
1. We standardized the data with StandScaler that normalizes all the data to a zero mean and 1 standard deviation.

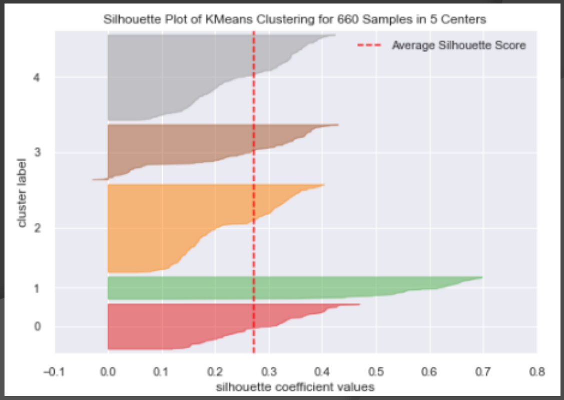
2. Next, we will find the clusters by segmenting the customers using K means

3. Next, we used the elbow method to find the number of clusters. The appropriate value of k from elbow curve seems to be 4 or 5

4. Next, we will check the Silhouette score. Let us take 5 as appropriate no. of clusters as silhouette score is high enough for all features.



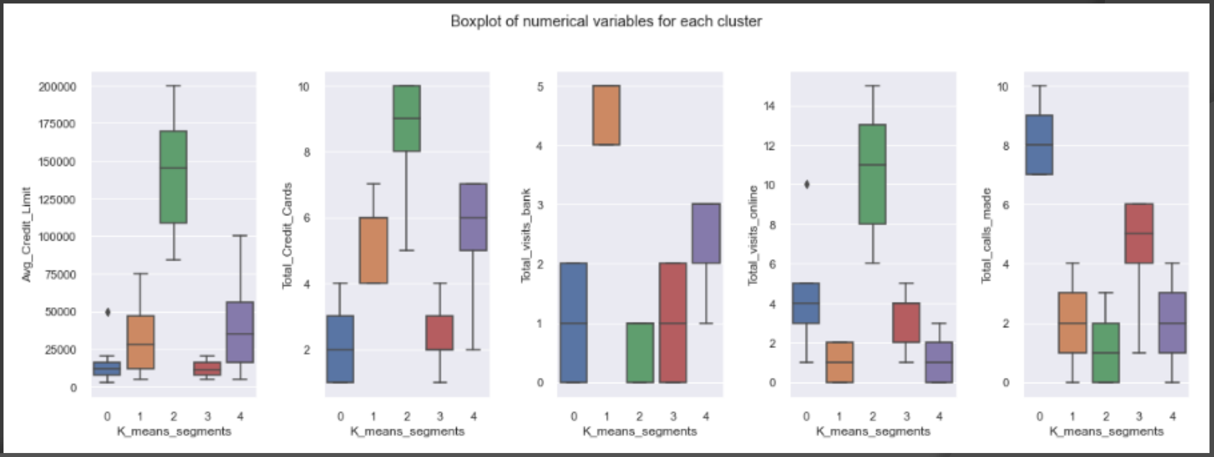




ML Model Building – K Means Cluster

Insights from the clusters found by K means,

- Clusters 0 :
  - Credit Limit is low
  - Total credit cards is medium
  - Total bank visits is low
  - Total visits online is medium
  - Total calls made is highest
- Cluster 1 :
  - Credit Limit is medium
  - Total credit cards is medium
  - Total bank visits is high
  - Total visits online is low
  - Total calls made is low
- Cluster 2 :
  - Credit Limit is highest
  - Total credit cards is highest
  - Total bank visits is low
  - Total visits online is highest
  - Total calls made is low
- Cluster 3:
  - Credit Limit is low
  - Total credit card is low
  - Total bank visits is low
  - Total visits online is low
  - Total calls made is medium
- Cluster 4:
  - Credit Limit is medium
  - Total credit card is medium
  - Total bank visits is medium
  - Total visits online is low
  - Total calls made is low



ML Model Building – Hierarchical Clustering

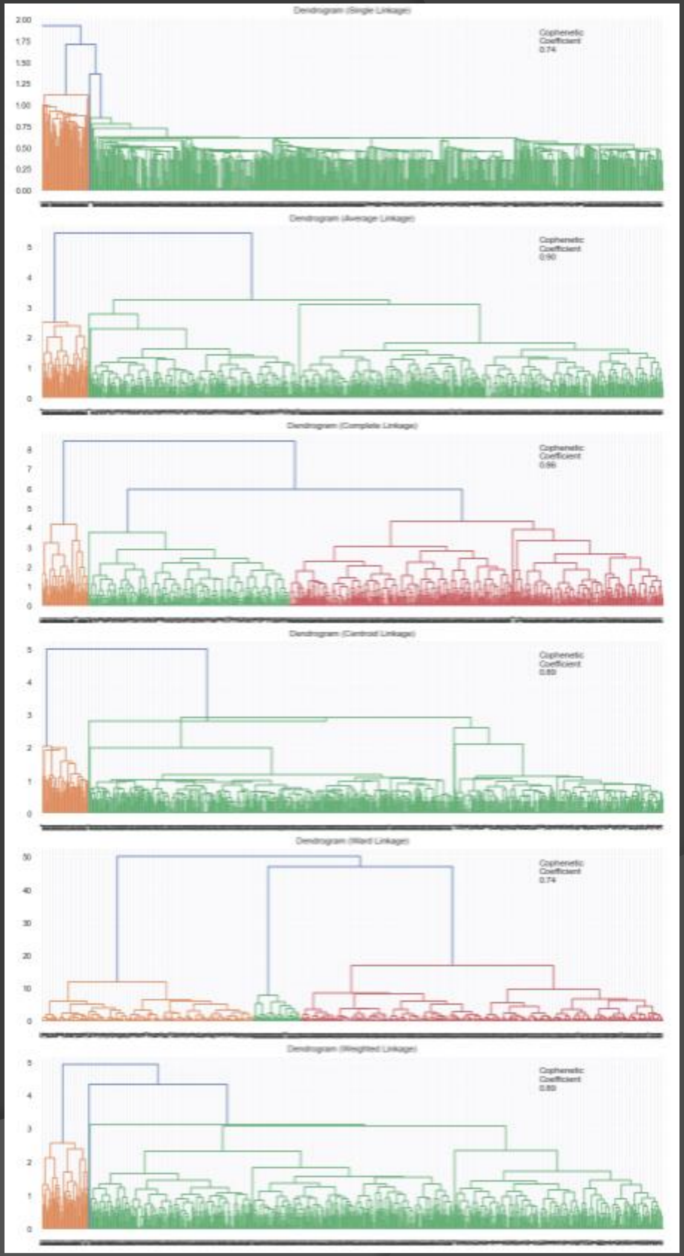
1. We will segment customers into clusters using Hierarchical clustering and compare with K means

2. For model building, we used various linkage methods such as 'single', 'average', 'complete', 'centroid', 'ward', 'weighted’. We saved results with Linkage and Cophenetic Coefficient

3. Next, we plot the Dendrograms using Euclidean and Cophenetic coefficients

4. Out of all the dendrogram we see, it is clear that dendrogram with ward linkage method gave us separate and distinct clusters

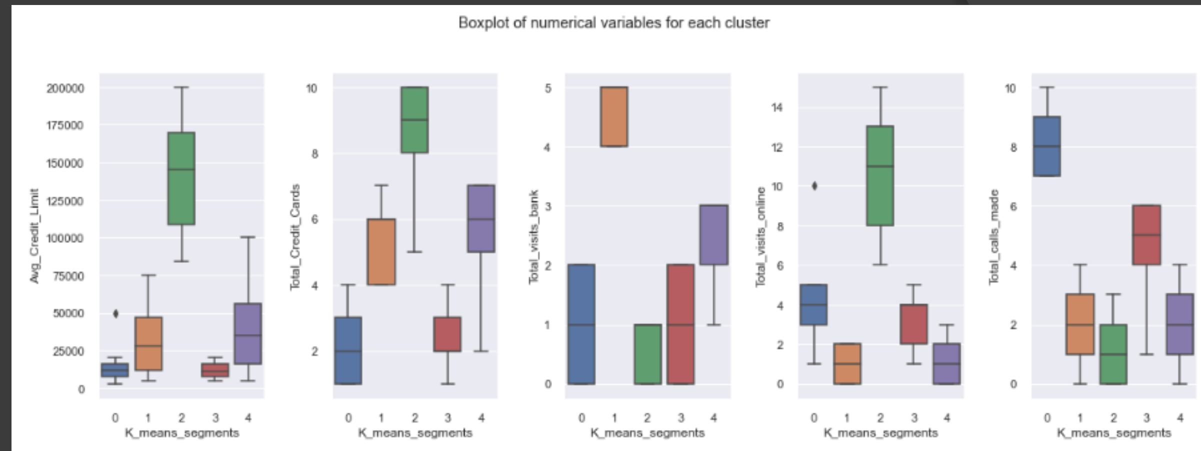
5. 5 cluster would be appropriate number of cluster from dendrogram with ward linkage method

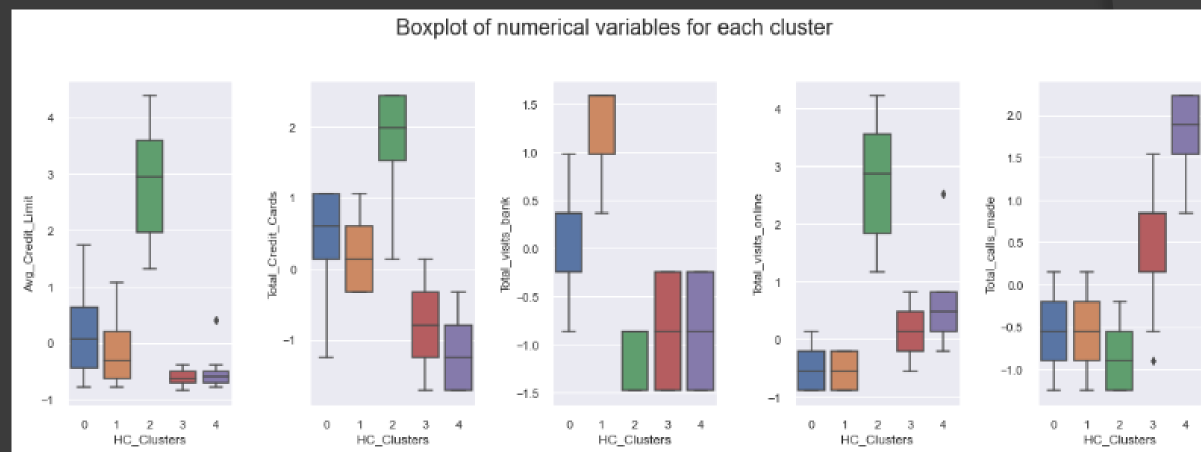


K Means Cluster v/s Hierarchical Clustering

Observation on comparison between these ML models:

- As seen in the table below, the clusters found in K-Means and Hierarchical Clustering are quite similar.
- The clusters found between K means and Hierarchical Clustering have matching high and lows
- The numbers of clusters found are the same in K-Means and Hierarchical Clustering. The counts in each cluster are quite similar





Clusters	Credit Limit		Credit cards		Bank visits		Online visits		Calls made		Count in Clusters	
	K means	HCA	K means	HCA	K means	HCA	K means	HCA	K means	HCA	K means	HCA
0	Low	Low	Low	Low	Low	High	Medium	Low	Highest	Low	123	191
1	Medium	Medium	Medium	Medium	High	Highest	Low	Low	Low	Low	190	194
2	Highest	High	Highest	Highest	Low	Low	Highest	High	Low	Medium	50	50
3	Low	Low	Low	Low	Low	High	Low	Low	Medium	Medium	101	133
4	Medium	Low	Medium	Low	Medium	Low	Low	Medium	Low	Highest	196	92

Actionable Insights & Recommendations

1. Cluster 0 - have customers with low credit limit and medium credit card utilization and they prefer to call in with queries, based on K-means

   ** Recommendation - Business can offer better phone service to these customers as they are a sizeable percentage of population. Feedback on service delivery can be requested from customers to improve perception of delivery. New customers could be added to this category by offering cards with low credit limit. They can be offered more credit cards as their credit card utilization is low.

2. Cluster 1 - have customers with medium credit limit and medium credit card utilization and they prefer to goto Bank with queries.

   ** Recommendation - Business can offer better in person service to these customers as they are a sizeable percentage of population. They can be offered more information about online banking for some services, keeping their convenience in mind.

3. Cluster 2 - have customers with high credit limit and high credit card utilization and they prefer to use Online Bank for queries.

   ** Recommendation - Business can look to add more customers in this category, as this cluster is a small percentage of the customers population.

4. Cluster 3 - have customers with low credit limit and low credit card utilization and they seldom make queries.

   ** Recommendation - Business can contact these customers to offer credit cards or to get feedback about service.

5. Cluster 4 - have customers with low credit limit and low credit card utilization and they prefer to call in with queries, based on HCA clustering

   ** Recommendation - Business can offer better phone service to these customers as they are a sizeable percentage of population. Feedback on service delivery can be requested from customers to improve perception of delivery They can be offered more credit cards as their credit card utilization is low