Business Presentation

AllLife Bank Customer Segmentation Case Study

Background & Context

- **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.
- **Head of Marketing** and **Head of Delivery** both decide to **reach out to the Data Science team** for help

Objective

- **Explore and visualize data**
- **Build a clustering model**
- **Model should be able to**
    - To identify different segments in the existing customer base, based on their spending patterns as well as past interaction with the bank and come up with customer/cluster profiling.
- **Draw conclusions and Business Recommendations**

Data Overview

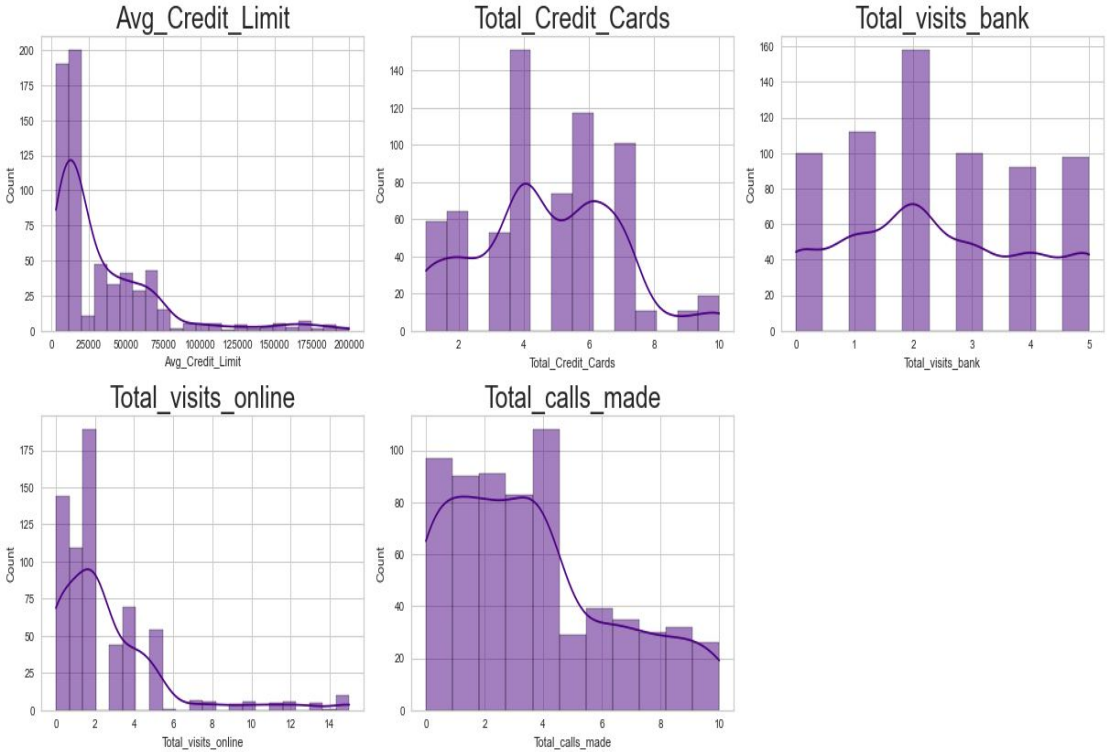
- Data contains **660 (rows)** customer data with **7 (columns)** characteristics.
- There are **no missing values** in **any of the columns.**
- **No duplicates** in dataset.
- The **Sl.No and Customer Key** column is of no significance and will be **removed** for analysis.
- **Customer Key** was **renamed** to **Customer_Key** for data analysis.

Variable	Description
Sl_No	Primary key of the records
Customer Key	Customer identification number
Avg_Credit_Limit	Average credit limit of each customer for all credit cards
Total_Credit_Cards	Total number of credit cards possessed by the customer
Total_visits_bank	Total number of Visits that customer made (yearly) personally to the bank
Total_visits_online	Total number of visits or online logins made by the customer (yearly)
Total_calls_made	Total number of calls made by the customer to the bank or its customer service department (yearly)

Exploratory Data Analysis - Univariate

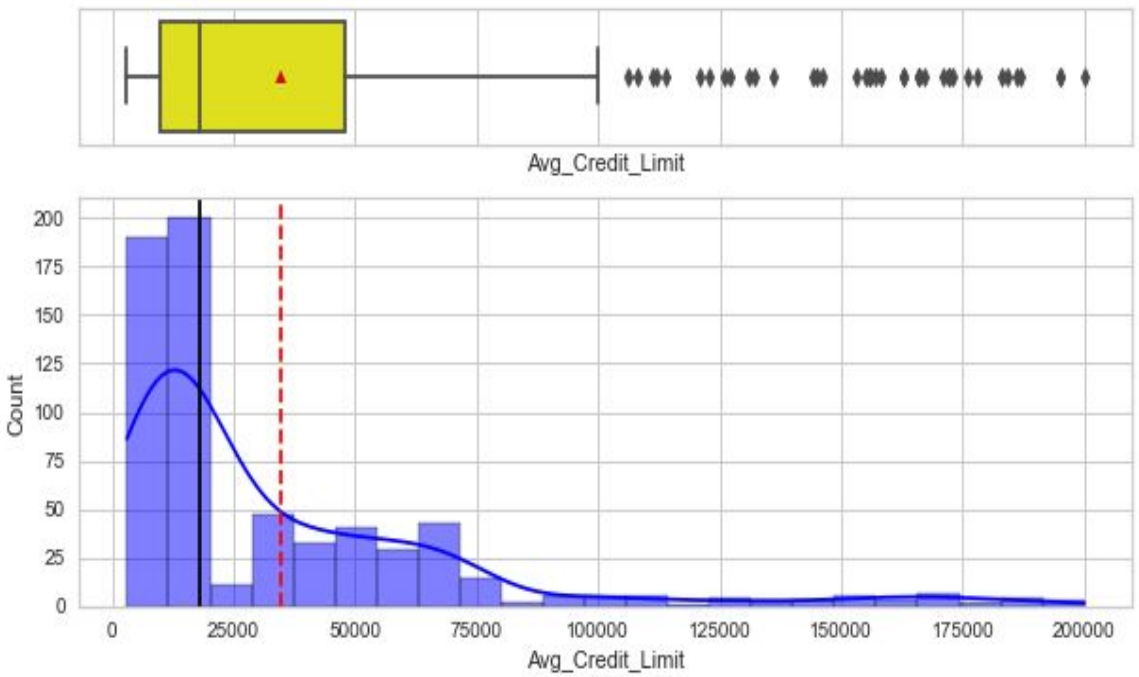
Observations:

- ***Avg_credit_limit:*** The average credit limit is heavily **right skewed** indicating **presence of outliers** on the higher end on the data.
- ***Total_Credit_Cards:*** We can see a kind of **bimodal distribution** in data with two peaks and not many outliers.
- ***Total_visits_bank:*** This is kind of **normally distributed** with no outliers.
- ***Total_visits_online:*** This variable is **right skewed** indicating the **presence of outliers**.The average peak here is around 2.
- ***Total_calls_made:*** This distribution tells us around 50% of the customers has made around 3 to 4 calls and the remaining over 5 times.The average peak is around 3.



Exploratory Data Analysis - Univariate

Avg_Credit_Limit

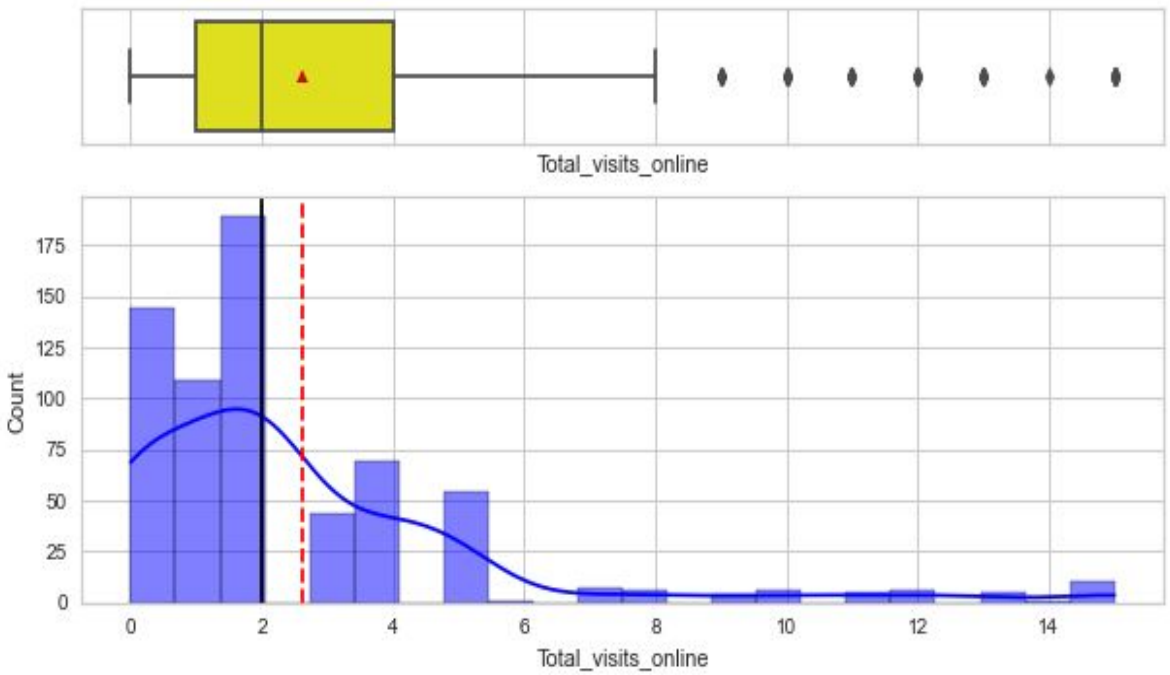


Observations:

- Average **credit limit** is **37k**.
- Around 50% of the customers credit limit is below 20k and the **higher end outliers** have skewed the mean to be greater than the median value.
- There are some **outliers beyond 75k**.

Exploratory Data Analysis - Univariate

Total_online_visits



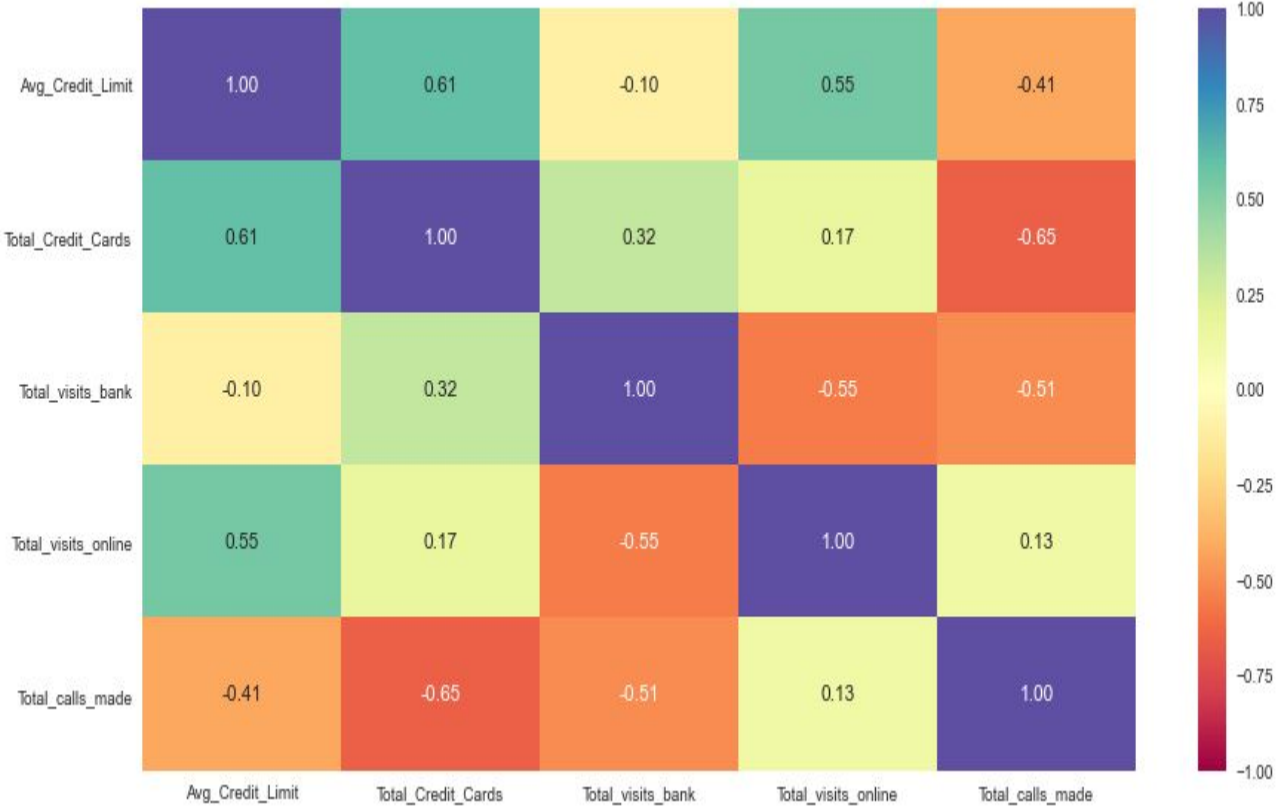
Observations:

- The **average count** of online visits made by the customer is **3**.
- We can see that the data is **right skewed** with **outliers** beyond the value of **8 up until 15**.

Exploratory Data Analysis - Bivariate Heat Map

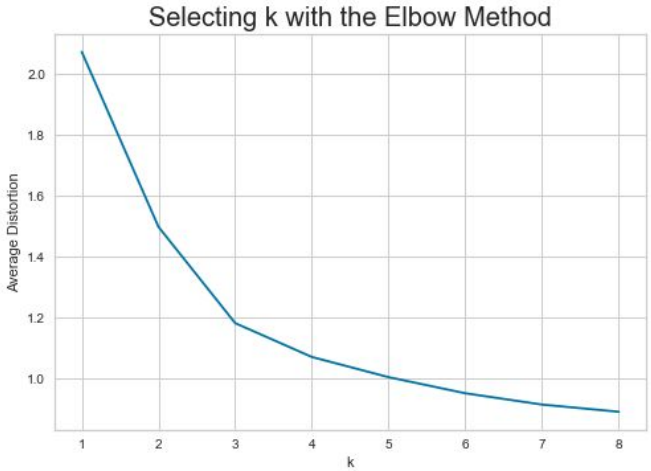
Observations:

- We **cannot see any strong correlations** here.
- The **top positive correlation** we can see is between Total_Credit_cards and Avg_Credit_Limit is **0.61,** the more number of cards the higher the credit limit.
- The **top negative correlation** is between Total_Credit_cards and Total_call_made is **-0.65**, the more number of cards a customer has the lesser calls he has made to the bank(happy customers)as well as the lesser number of credit cards the more calls made to the bank(unhappy customers).

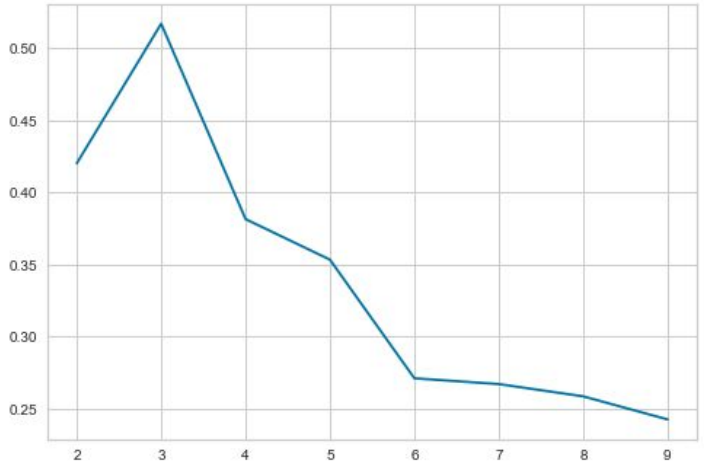


K Means Clustering - Elbow curve vs Silhouette Score

Elbow Curve



Silhouette Score Plot



Observations:

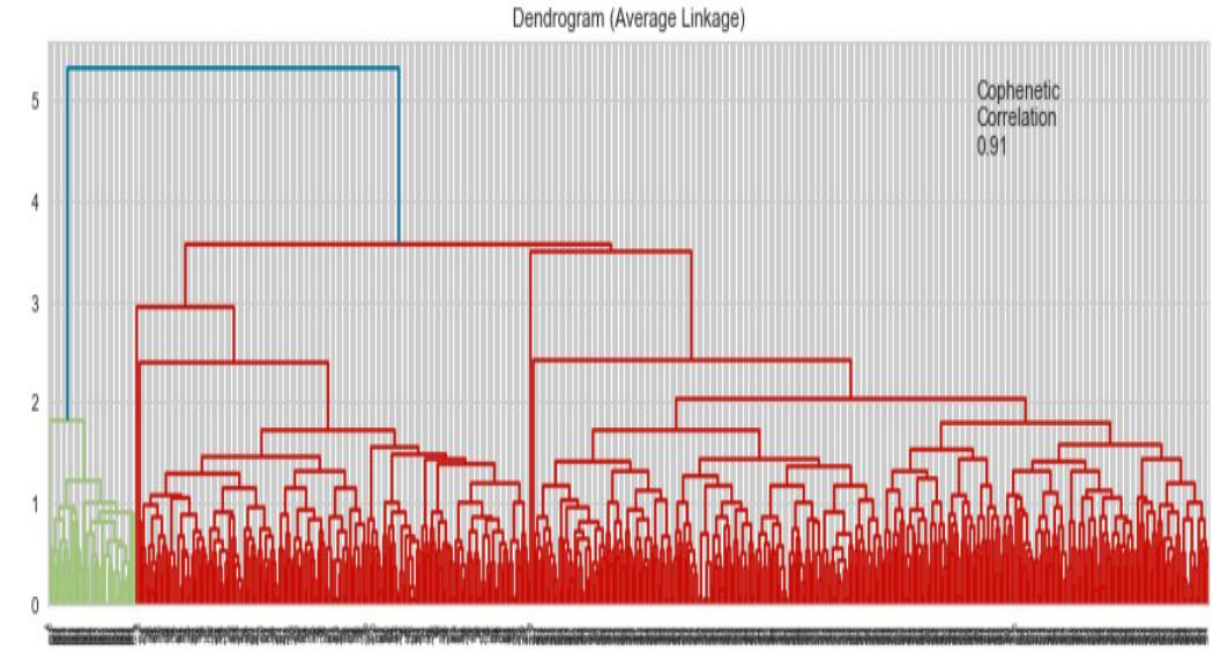
- We can see the **knick at 3** in the elbow curve post which the curve tends to flatten gradually.

Observations:

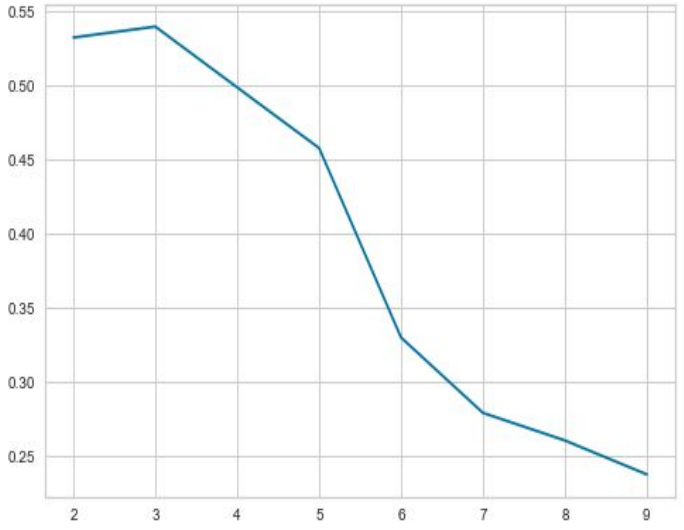
- The **maximum rise** is at **3** post which it follows a decreasing trend.
- The **Silhouette coefficient**(how close is the distance between the cluster points) is **highest** at the value **3** which **exactly matches** to the **knick** at the **Elbow curve**.

Hierarchical Clustering - Dendrogram vs Silhouette Score

Dendrogram - Average Linkage - Euclidean Distance



Silhouette Score Plot



Observations:

- From the **Dendrogram** and **Silhouette score plot** we can see that **3** is the **optimal number of clusters** for this data set.
- **Silhouette score**(distance between the cluster points) **is** the **maximum** at **3** and the **Dendrogram height** keeps **reducing** gradually **below 3**.

K Means Clustering - Cluster Profiling

Cluster 0:

- The **average credit limit** is below **30k**.
- Around **50% of customers** in this category **have around 4 credit cards** and the minimum here is 2 maximum is around 7.
- **Customers** in this cluster group have a **minimum of 1 visit** and **maximum of 5 visits to the bank**.This cluster have the more **traditional/conservative customers who would visit the bank** for queries and activities rather than using other modes.
- **Customers** in this group have **least online usage**.50% of the customers do not do online visits at all.
- **Customer**s in this cluster group **has made 2 calls** to the bank on average and other 50 to 75% of customers in this group has made 3 - 4 calls

Cluster 1:

- The **average credit limit** is around **15k**.All the customers credit limit is less than 20k in this cluster.
- Around **50% of customers** in this category have around **3 credit cards** and the **minimum here is 1 maximum is around 4**.
- This group has the **lowest number of credit cards**.
- **Customers** in this cluster group have a **minimum of 0 visit and maximum of 2 visits to the bank**.
- **Customers** in this group have minimum of 1 to **maximum of 5 online visits**.50% of the customers are between 2-4 online visits.
- **Customers** in this cluster group **has made maximum calls**.25% of customers has made between 1-5 calls,25% has made 5-7 calls,25% has made **7-9 calls** to the bank on average.

Cluster 2:

- The **average credit limit** is below **110k**.This seems to be the **group capturing all the outliers**.
- **More than 25% of customers** in this category **have around 8-10 credit cards** and the minimum here is 5 maximum is around 10.
- **Customers** in this cluster group do a **minimum number of bank visits 0 or 1**.They are mostly online customers.
- **Customers** in this group have **maximum online usage**.Most customers has a minimum of 8 online visits to 9 online being the maximum.
- **Customers** in this cluster group **has made least number of calls to the bank** on average with a 25% making no calls and maximum of 3 calls.

Hierarchical Clustering - Cluster Profiling

Cluster 0:

- The **average credit limit** is below **30k**.
- Around **50% of customers** in this category **have around 4 credit cards** and the minimum here is 2 maximum is around 7.
- **Customers** in this cluster group have a minimum of 1 visit and **maximum of 5 visits to the bank**.This cluster have the more t**raditional/conservative customers who would visit the bank for queries** and activities rather than using other modes.
- **Customers** in this group **have least online usage**.50% of the customers do not do online visits at all.
- **Customers** in this cluster group **has made 2 calls to the bank** on average and other 50 to 75% of customers in this group has made 3 - 4 calls

Cluster 1:

- The **average credit limit** is below **110k**.This seems to be the group capturing all the outliers.
- **More than 25% of customers** in this category **have around 8-10 credit cards** and the minimum here is 5 maximum is around 10.
- **Customers** in this cluster group do a **minimum number of bank visits 0 or 1**.They are mostly online customers.
- **Customers** in this group have **maximum online usage**.Most customers has a minimum of 8 to 9 online being the maximum.
- **Customers** in this cluster group **has made least number of calls to the bank** on average with a 25% making no calls and maximum of 3 calls.

Cluster 2:

- The **average credit limit** is around **15k**.All the customers credit limit is less than 20k in this cluster.
- Around **50% of customers** in this category have around **3 credit cards** and the minimum here is 1 maximum is around 4.
- This group has the **lowest number of credit cards**.
- **Customers** in this cluster group have a **minimum of 0 visit and maximum of 2 visits to the bank.**
- **Customers** in this group have minimum of 1 to **maximum of 5 online visits**.50% of the customers are between 2-4 online visits.
- **Customers** in this cluster group **has made maximum calls**.25% of customers has made between 1-5 calls,25% has made 5-7 calls,25% has made **7-9 calls to the bank** on average.

Cluster Profiling - Insights

Cluster Group	K Means Clustering - Customer Characteristics	Hierarchical Clustering - Customer Characteristics
0	**Traditional/Conservative Customers** ● Visiting bank is the preferred way of communication. ● No or least online banking usage. ● Mid range credit Limit of 37k . ● Moderate Credit card users with average of 5 cards. ● Low number of telephonic calls made to the bank.	**Traditional/Conservative Customers** ● Visiting bank is the preferred way of communication. ● No or least online banking usage. ● Mid range credit Limit of 37k . ● Moderate Credit card users with average of 5 cards. ● Low number of telephonic calls made to the bank.
1	**Target Customers for Upselling and Services** ● Calling the bank is the preferred way of communication. ● Moderate online banking usage. ● Low credit card users holding 3 cards on average. ● Least bank visitors. ● Lower credit limit of 15k	**High Profile Customer/Premium Customers** ● Online banking is the preferred way of communication. ● High range credit limit of 110k. ● Heavy credit card users holding 9 cards on average. ● Least or no telephonic calls/visits made to the bank.
2	**High Profile Customer/Premium Customers** ● Online banking is the preferred way of communication. ● High range credit limit of 110k. ● Heavy credit card users holding 9 cards on average. ● Least or no telephonic calls/visits made to the bank.	**Target Customers for Upselling and Services** ● Calling the bank is the preferred way of communication. ● Moderate online banking usage. ● Low credit card users holding 3 cards on average. ● Least bank visitors. ● Lower credit limit of 15k

Conclusion

- Both **K Means and Hierarchical clustering** has given us **exactly the same results** with same number of cluster groups.
- **K Means and Hierarchical cluster groups 1 and 2** has cluster **labels swapped** with same data grouping.
- **Data** in all the **three clusters** across both the methods are **identical**.

Recommendations

- **Cluster 1 in K Means** and **Cluster 2 in Hierarchical** are the **target group for Business** to **upsell their credit cards** and to **improve** the **customer/delivery services** to make the customers happy as this group hold the lowest number of credit cards and makes the maximum number of phone calls.They also use online banking moderately, **online banking channel for advertisements** can also be a **marketing strategy**.
- **Cluster 2 in K Means and Cluster 1 in Hierarchical clustering** have **captured** all the **high end outliers**.**Online marketing strategies** and **online services can be targeted/focussed** for this group.**Premium Services** can also be **targeted**.
- **Cluster 0** has more **traditional /conservative customers** and **in person banking services** with attractive flyers on the bank promotions and friendly personal banker services can be the **areas of focus for this group**.