

# Project Credit Card Customer Segmentation Unsupervised Learning Business Report

Prepared by: Smriti Shukla

Great learning PGP DSBA

September 2025



## Table of Contents

List of Figures and plots .....	3
List of tables .....	4
<b>Customer Segmentation Models in Banking</b> .....	5
<b>Exploratory Data Analysis</b> .....	8
Problem Statement .....	8
Data background and contents.....	9
Statistical summary of the dataset .....	9
Univariate analysis .....	10
Bivariate Analysis .....	14
Insights based on EDA.....	15
<b>Data Preprocessing</b> .....	19
Duplicate value and Missing value check .....	19
Outlier treatment.....	20
Feature engineering and data preparation for modelling .....	25
Data Scaling.....	25
<b>Model Building</b> .....	26
<b>K-means Clustering</b> .....	26
Checking Elbow Plot.....	26
Checking Silhouette Scores .....	27
Figure out the appropriate number of clusters .....	36
Cluster Profiling: K-means Clustering .....	37
<b>Hierarchical Clustering</b> .....	47
Computing Cophenetic Correlation .....	47
Checking Dendrograms.....	48
Figure out the appropriate number of clusters .....	49
Cluster Profiling: Hierarchical Clustering .....	49
<b>K-means vs Hierarchical Clustering</b> .....	58
<b>Principal component analysis (PCA)for Visualization</b> .....	62
<b>Actionable Insights Business Recommendations</b> .....	63

## List of Figures and plots

No	Name	Page
1	Histogram and box plot figure no.1 for distribution of Avg_Credit_Limit	10
2	Figure no.2 for distribution of Total_Credit_Cards	11
3	Histogram and box plot figure no. 3 for distribution of Total_Credit_Cards	11
4	Histogram and box plot figure no.4 for distribution of Total_visits_bank	12
5	Labelled bar plot figure no.5 for distribution of Total_visits_bank	12
6	Labelled bar plot figure no.6 for Total_visits_online column	13
7	Labelled bar plot figure no.7 for distribution of Total_calls_made	13
8	Correlation plot figure no.8 correlation matrix for numerical features	14
9	Pair plot figure no. 9 to visualize correlation as pairplot for numerical features	15
10	Boxplots figure no. 10 to show distribution of data to check outliers	20
11	Boxplot figure no. 11 to show distribution of Avg_credit_Limit to check outliers	21
12	Figure no. 12 to show Relation between number of cluster and average distortion	24
13	Figure no. 13 to show Distortion score Elbow for KMeans Clustering	25
14	Figure no. 14 to show Relation between no. of clusters and silhouette score	26
15	Figure no. 15 to show Silhouette score elbow for KMeans clustering	26
16	Figure no. 16 to show Silhouette plot for KMeans clustering for 2 clusters	27
17	Figure no. 17 to show Silhouette plot for KMeans clustering for 3 clusters	28
18	Figure no. 18 to show Silhouette plot for KMeans clustering for 4 clusters	29
19	Figure no. 19 to show Silhouette plot for KMeans clustering for 5 clusters	30
20	Figure no. 20 to show Silhouette plot for KMeans clustering for 6 clusters	31
21	Figure no. 21 to show Silhouette plot for KMeans clustering for 7 clusters	32
22	Figure no. 22 for Distribution of Avg_credit_limit across cluster (KMeans clustering)	35
23	Figure no. 23 for Distribution of Total_Credit_Cards across cluster (KMeans clustering)	35
24	Figure no. 24 for Distribution of Total_visits_bank across cluster (KMeans clustering)	36
25	Figure no. 25 for Distribution of Total_visits_online across cluster (KMeans clustering)	36
26	Figure no. 26 for Distribution of Total_calls_made across cluster (KMeans clustering)	37
27	Figure no. 27 to show Distribution of pairwise features plot by clusters (KMeans clustering)	38
28	Figure no. 28 to show Dendrogram for different linkage	44
29	Figure no. 29 to show Distribution of Avg_Credit_Limit across Hierarchical clusters	47
30	Figure no. 30 to show Distribution of Total_Credit_Cards across Hierarchical clusters	47
31	Figure no. 31 to show Distribution of Total_visits_bank across Hierarchical clusters	48
32	Figure no. 32 to show Distribution of Total_visits_online across Hierarchical clusters	48
33	Figure no. 33 to show Distribution of Total_calls_made across Hierarchical clusters	49
34	Figure no. 34 to show pairwise feature plot by Hierarchical clusters	49
35	Figure no. 35 to compare Avg_Credit_Limit by Hierarchical clustering and KMeans clustering	56
36	Figure no. 36 to compare Total_Credit_Cards by Hierarchical clustering and KMeans clustering	56
37	Figure no. 37 to compare Total_visits_bank by Hierarchical clustering and KMeans clustering	57
38	Figure no. 38 to compare Total_visits_online by Hierarchical clustering and KMeans clustering	57

39	Figure no. 39 to compare Total_calls_made by Hierarchical clustering and KMeans clustering	58
40	Figure no. 40 to show datapoints and visualize 2 main component in dataset	59

## List of Tables

No	Name	Page
1	Table no 1 Statistical summary of the dataset	09
2	Table no 2 Datatype and null value summary of the dataset	19
3	Table no 3 Total no of null value in the dataset	19
4	Table no 4 Total no for Outliers in Avg_Credit_Limit column	22
5	Table no 5 Total no for Outliers in Total_visits_online column	24
6	Table no 6 first 5 rows of scaled dataset for model building	23
7	Table no 7 Cluster profiling summary – K-Means clustering	34
8	Table no 8 Output of cophenetic correlation computations for different linkage methods and distance metrics	44
9	Table no 9 to show Hierarchical cluster profiling summary	46
10	Table no 10 list of mean value of features to compares the average profiles of Hierarchical cluster segments	56

# Customer Segmentation Models in Banking

## Customer Segmentation Models in Banking

### 1. RFM Segmentation (Recency, Frequency, Monetary Value)

- Classic model used to segment customers based on:
  - **Recency:** How recently a customer used the bank's services.
  - **Frequency:** How often they use them.
  - **Monetary:** How much value they bring (balances, loan amounts, card spends).
- Helps banks identify **high-value customers, churn risks, and upsell/cross-sell opportunities**.

### 2. Lifecycle Segmentation

- Based on **customer's stage in life** (student, young professional, family, retiree).
- Useful for **tailoring products** like student loans, mortgages, retirement plans.
- Principle: *"Right product at the right life stage."*

### 3. Behavioural Segmentation

- Looks at **transaction behaviour** (digital adoption, payment preferences, investment patterns).
- Example: Heavy credit card users vs. digital wallet users.
- Helps banks push **digital transformation, loyalty programs, or personalized offers**.

### 4. Profitability Segmentation

- Customers are classified by **current and potential profitability**.
- Banks can allocate resources to **high-margin clients** while automating services for low-value clients.
- Often aligned with **Pareto Principle (80/20 rule)** → 20% of customers drive 80% of profits.

### 5. Risk-based Segmentation

- Segments based on **creditworthiness & risk appetite**.
- Useful for **lending decisions** (personal loans, mortgages, credit card limits).
- Example: High credit score = premium offers; Low credit score = secured products.

## Famous Business Models / Principles Banks Use

1. **Customer Lifetime Value (CLV) Model**
  - Focuses on maximizing the **total long-term value** a customer generates.
  - Decisions are driven not just by current profit but by **future potential**.
2. **Blue Ocean Strategy** (Kim & Mauborgne)
  - Instead of competing in crowded financial products (red ocean), banks **create uncontested market spaces**.
  - Example: Mobile-only banks, wealth management apps, ESG-linked financing.
3. **Porter's Five Forces in Banking**
  - Analyzes **competitive intensity** (substitute products, fintech entrants, bargaining power).
  - Helps in **strategic positioning** of products and services.
4. **Value-Based Segmentation (McKinsey model)**
  - Customers are segmented by **value drivers** (convenience, trust, speed, digital experience).
  - Used by global banks like HSBC, Citibank.
5. **Pareto Principle (80/20 Rule)**
  - Banks often discover **80% of revenue comes from 20% of customers**.
  - Strategy: Retain and grow top-tier customers, while automating or low-cost servicing others.

## Customer Segmentation Models in Banking

1. **RFM Segmentation (Recency, Frequency, Monetary Value)**
  - Classic model used to segment customers based on:
    - **Recency**: How recently a customer used the bank's services.
    - **Frequency**: How often they use them.
    - **Monetary**: How much value they bring (balances, loan amounts, card spends).
  - Helps banks identify **high-value customers, churn risks, and upsell/cross-sell opportunities**.
2. **Lifecycle Segmentation**
  - Based on **customer's stage in life** (student, young professional, family, retiree).
  - Useful for **tailoring products** like student loans, mortgages, retirement plans.

- Principle: *“Right product at the right life stage.”*

### 3. Behavioral Segmentation

- Looks at **transaction behavior** (digital adoption, payment preferences, investment patterns).
- Example: Heavy credit card users vs. digital wallet users.
- Helps banks push **digital transformation, loyalty programs, or personalized offers.**

### 4. Profitability Segmentation

- Customers are classified by **current and potential profitability.**
- Banks can allocate resources to **high-margin clients** while automating services for low-value clients.
- Often aligned with **Pareto Principle (80/20 rule)** → 20% of customers drive 80% of profits.

### 5. Risk-based Segmentation

- Segments based on **creditworthiness & risk appetite.**
- Useful for **lending decisions** (personal loans, mortgages, credit card limits).
- Example: High credit score = premium offers; Low credit score = secured products.

# Exploratory Data Analysis

## Problem Statement

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

### Objective

To identify different segments in the existing customers, based on their spending patterns as well as past interaction with the bank, using clustering algorithms, and provide recommendations to the bank on how to better market to and service these customers.

### Data Description

The data provided is of various customers of a bank and their 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).

### Data Dictionary

- **SI\_No:** Primary key of the records
- **Customer Key:** Customer identification number
- **Average 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 the 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)

## Data background and contents

The dataset has 660 rows and 7 columns capturing details related to various customers of a bank and their 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 centre).

## Statistical summary of the dataset

Column name	count	mean	std	min	25%	50%	75%	max
Avg_Credit_Limit	660	34574.2	37625.4	3000	10000	18000	48000	200000
Total_Credit_Cards	660	4.70606	2.16783	1.0	3.0	5.0	6.0	10.0
Total_visits_bank	660	2.40303	1.63181	0.0	1.0	2.0	4.0	5.0
Total_visits_online	660	2.60606	2.93572	0.0	1.0	2.0	4.0	15.0
Total_calls_made	660	3.58333	2.86531	0.0	1.0	3.0	5.0	10.0

**Table no 1 Statistical summary of the dataset**

### Summary and Interpretation

- Avg\_Credit\_Limit: The average credit limit is approximately 34,574, with a wide range from 3,000 to 200,000, indicating significant variability in credit access among individuals.

- Total\_Credit\_Cards: On average, individuals hold about 4.7 credit cards, with some holding up to 10, reflecting diversity in credit card usage.

- Total\_visits\_bank and Total\_visits\_online: Bank visits average around 2.4, while online visits are slightly higher at 2.6, but with high variability, especially online visits, which can go up to 15 visits.

- Total\_calls\_made: The average number of calls made is roughly 3.6, with some individuals making as many as 10 calls, indicating differing levels of engagement or communication frequency.

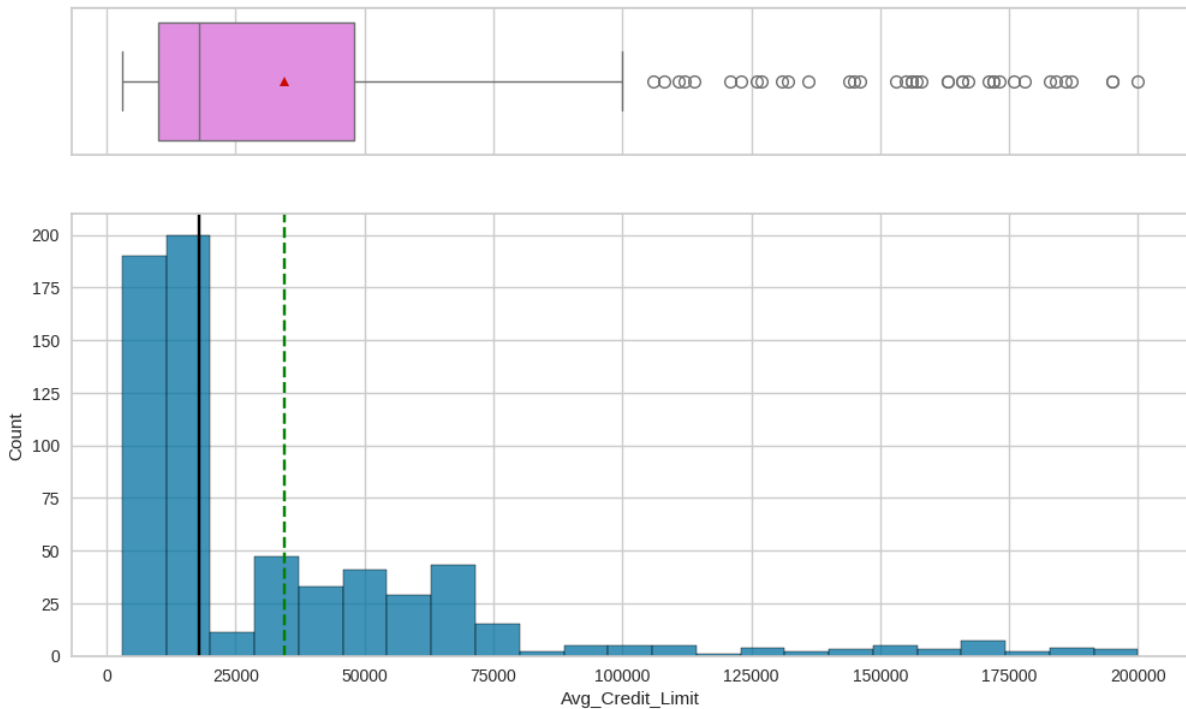
Overall, these statistics highlight substantial variation in financial behavior and credit access within the sample.

## Univariate analysis

Below is univariate analysis of every feature of data to analyse and recommend policies and steps for business

Avg\_Credit\_Limit

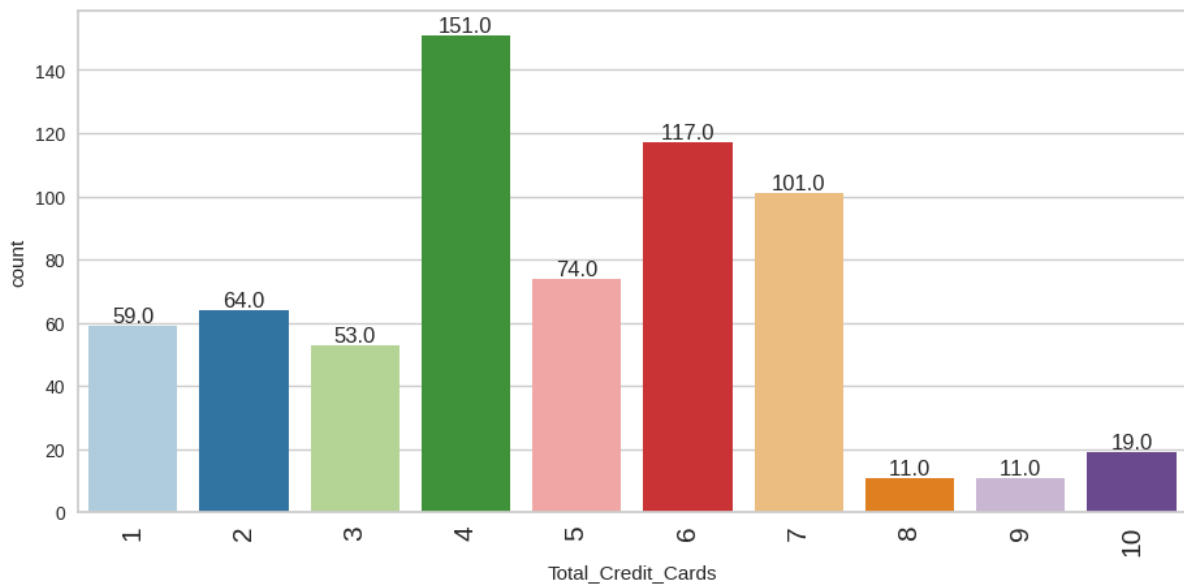
We can observe mean and median on below plotted boxplot and histogram figures as green and black dotted lines



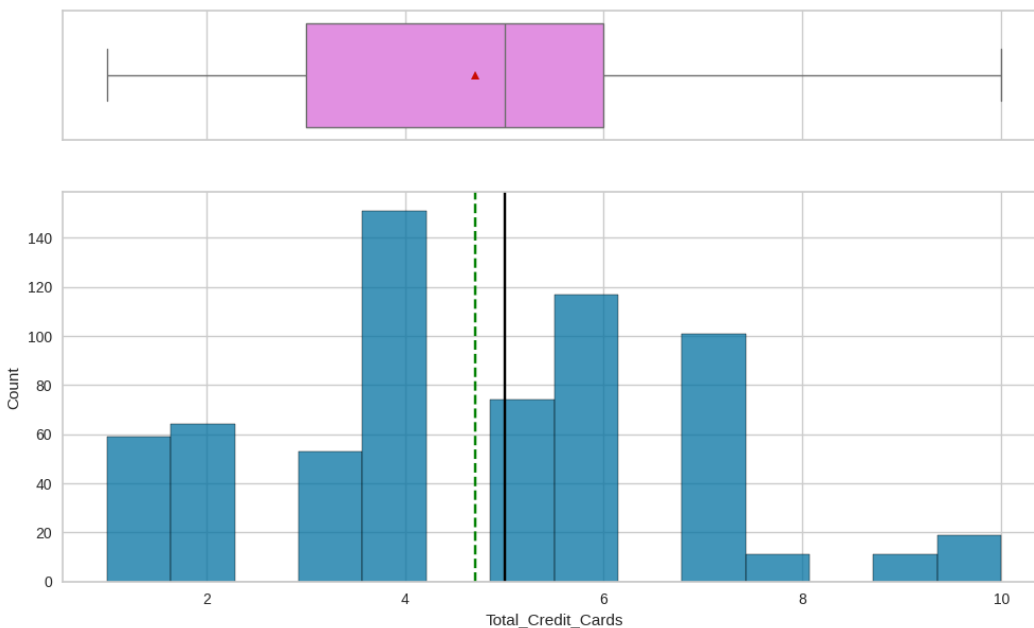
**Above is histogram and box plot figure no.1 for distribution of Avg\_Credit\_Limit**

**Observation** – Avg\_Credit\_Limit value looks slightly right skewed distributed with a few high value outliers, and the mean is between 34574.24 and median is less than 25000

## Total\_Credit\_Cards



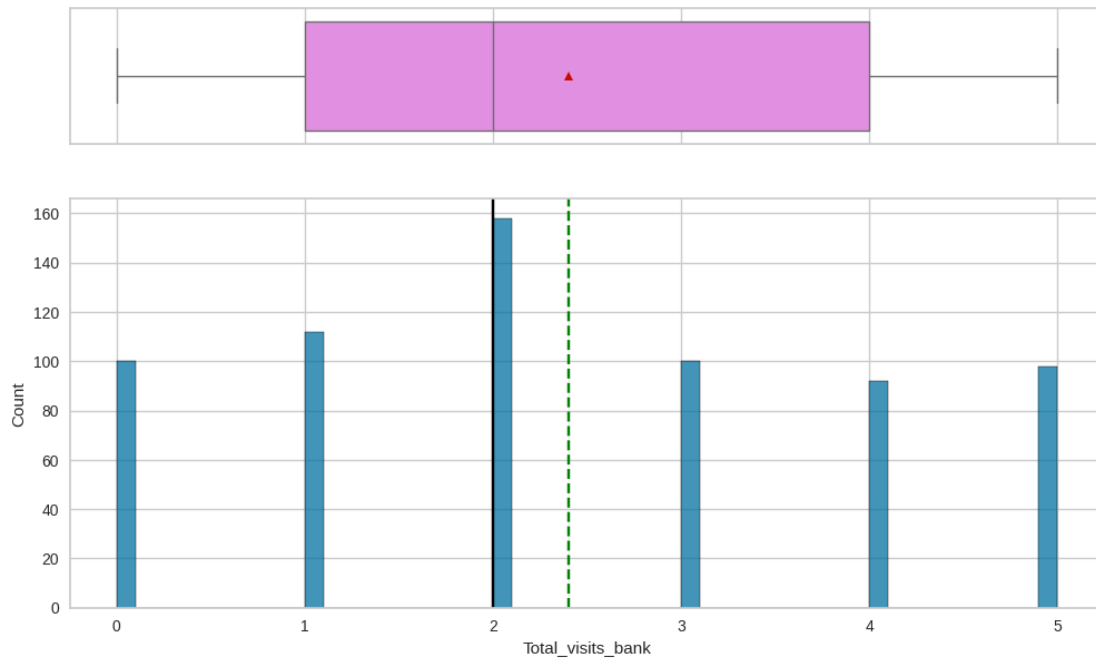
Above is figure no. 2 for distribution of Total\_Credit\_Cards



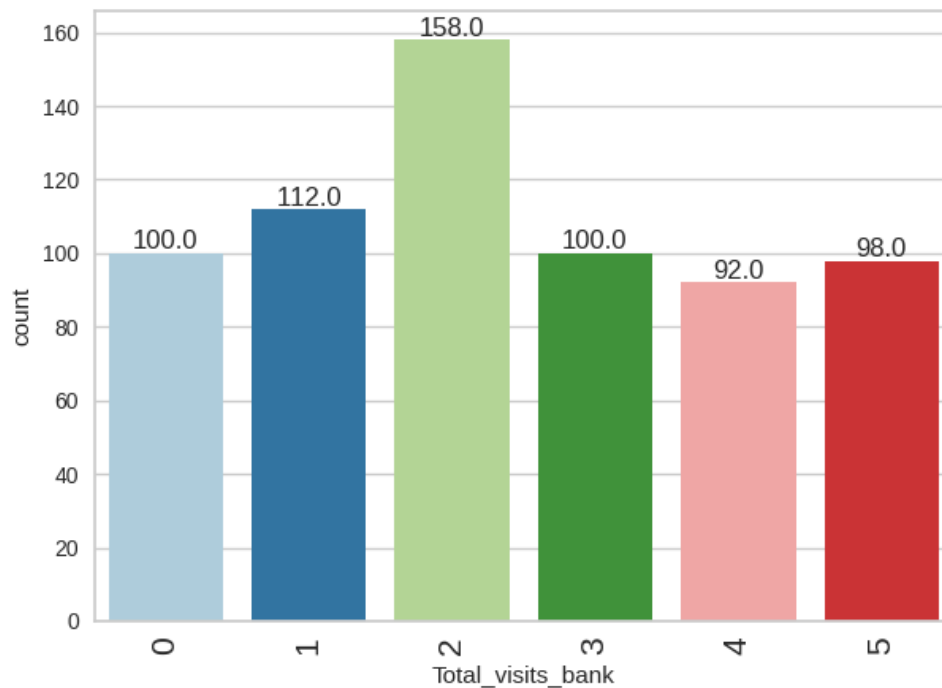
Above is histogram and box plot figure no. 3 for distribution of Total\_Credit\_Cards

**Observation** - The box plot and histogram indicate that the average number of credit cards per individual is around 4 to 5, with the data ranging from 2 to 10 credit cards. The distribution appears to be approximately normal, centered around an average of 4 to 6 credit cards

Total\_visits\_bank



Above is histogram and box plot figure no.4 for distribution of Total\_visits\_bank

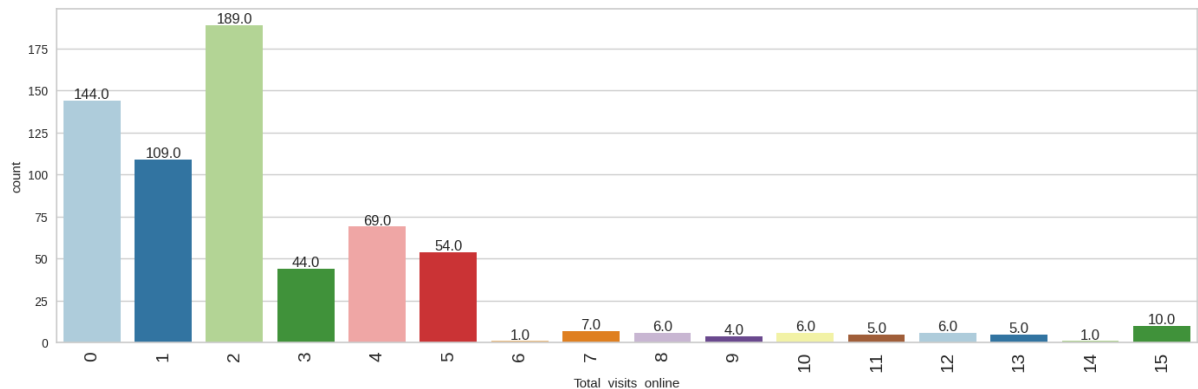


Above is labelled bar plot figure no.5 for distribution of Total\_visits\_bank

**Observation** - We can observe mean and median on plotted boxplot and histogram figures as green and black dotted lines

Total visit of bank for any customers boxplot shows Population is almost normally distributed with average between 2 and 3 (2.4). distribution is almost normally distributed as bell curve

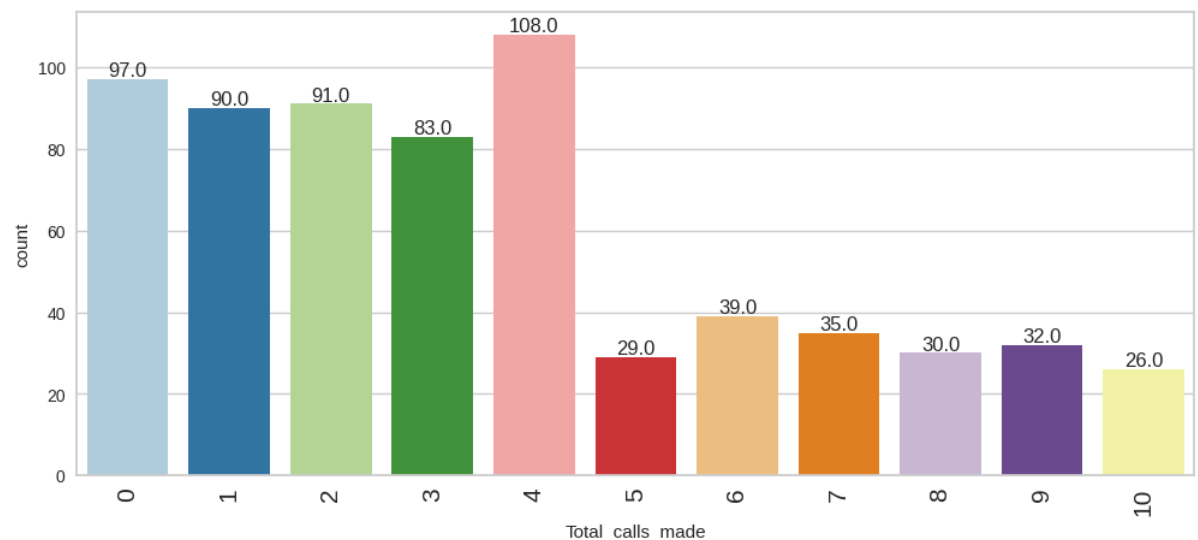
### Total\_visits\_online



Above is labelled bar plot figure no.6 for Total\_visits\_online column

**Observation** - Bank visits average around 2.4, while online visits are slightly higher at 2.6, but with high variability, especially online visits ranges between 0 to 15, which can go up to 15 visits. As per dataset provided 189 customers has made 2 online visits, 144 customers has not done any online visit and 109 customers has done 2 times visit, very few customers make more than 5 online visits

### Total\_calls\_made

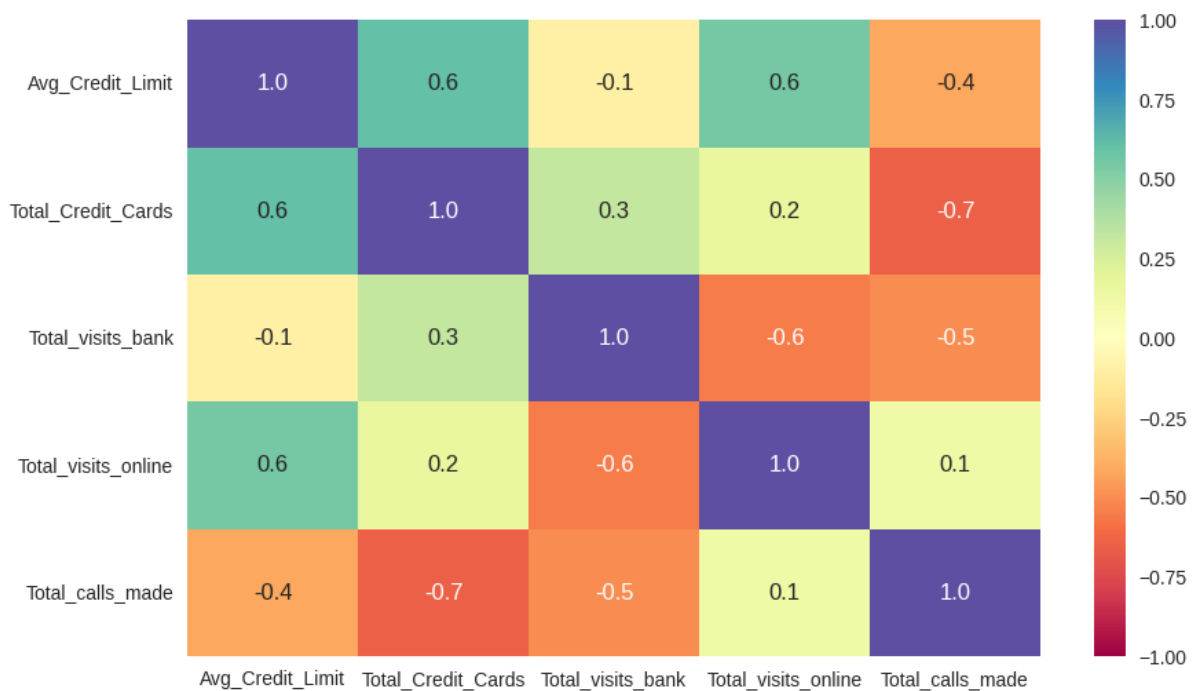


Above is labelled bar plot figure no.7 for distribution of Total\_calls\_made

**Observation** – Calls made by customers are ranges between 0 to 10 .highest number of customers (108 ) has made 4 calls for service.we can observe less number of customers has made 5 or more then 6 calls we can segregate them for high service need customer

## Bivariate Analysis

Let's check for correlations.



Above is correlation plot figure no.8 Correlation matrix for numerical features

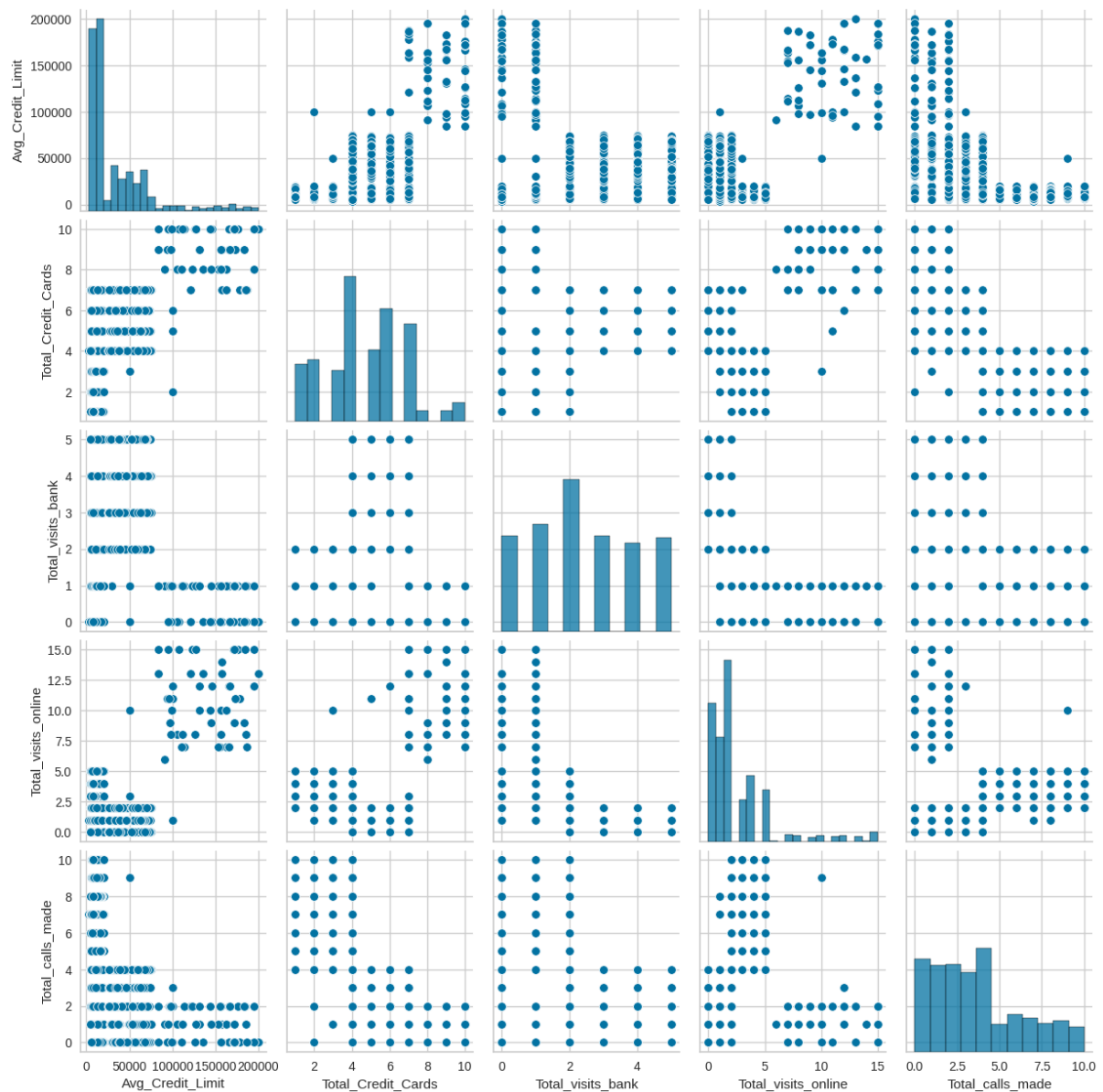
### Interpretation

#### Correlation Analysis

#### Correlation Matrix Insights

- Moderate Positive Correlation (0.6) between Total Credit Cards and Average Credit Limit: Customers holding more credit cards tend to have higher credit limits.
- Moderate Positive Correlation (0.6) between Total Online Visits and Average Credit Limit: Customers with higher credit limits are more likely to engage online with the bank.
- Strong Positive Correlation (0.7) between Total Calls Made and Total Credit Cards: Customers possessing more credit cards tend to contact the bank more frequently via calls, indicating higher engagement or potentially higher service needs.
- Negative Correlation (-0.6) between Total Online Visits and Total Visits Bank: Customers who visit the bank physically less often tend to have higher online engagement, suggesting a shift towards digital channels among certain customer segments.

Pair plot for numeric variables



Above is Pair plot figure no. 9 to show distribution of visualize pairplot for numerical features

## Insights based on EDA

### EDA Interpretation, Insights, and Business Recommendations

#### Dataset Overview

The dataset comprises 660 customer records with 7 key attributes related to their banking behavior and credit profile:

- Wide variability observed in credit limits, with some customers having very high limits (up to 200,000).

- Customers hold an average of approximately 4.7 credit cards, with a maximum of 10 cards.
  - Engagement through bank visits, online visits, and calls shows significant variability, indicating diverse customer interaction patterns.
- 

## **Univariate Analysis & Key Observations**

### **1. Average Credit Limit**

- Distribution: Slightly right-skewed, with a few high-value outliers.
- Implication: While most customers have moderate credit limits (~25,000 to 50,000), a subset enjoys significantly higher limits, suggesting targeted premium segment opportunities.

### **2. Total Credit Cards**

- Distribution: Approximately normal around 4–6 cards.
- Implication: Customers generally maintain a moderate number of credit cards; segmentation could focus on high vs. low cardholders.

### **3. Total Visits to Bank**

- Distribution: Nearly normal around 2-3 visits per year.
- Implication: Most customers prefer minimal physical visits, which can influence service delivery models.

### **4. Total Online Visits**

- Distribution: High variability, with some customers making up to 15 visits, while many have none.
- Observation: High online engagement indicates a digital-savvy customer segment; low online users could be targeted for digital onboarding.

### **5. Total Calls Made**

- Distribution: Spread across 0 to 10 calls, with an average of 3.6.
  - Implication: Customers with higher call frequency may require more personalized support or have higher service needs.
- 

## **Bivariate & Correlation Analysis**

### **• Strong and Moderate Correlations:**

- Customers with more credit cards tend to have higher credit limits and contact the bank more frequently via calls.
- Customers with higher credit limits are more active online.

### **• Negative Correlation:**

- A notable inverse relationship exists between online visits and physical visits, suggesting a shift toward digital channels among certain segments.

**Implication:** These correlations highlight distinct customer segments based on engagement and credit usage patterns, offering avenues for targeted marketing and service improvements.

---

## **Business Insights & Recommendations**

### **Customer Segmentation for Targeted Marketing**

- **High Credit Limit & High Engagement Segment:**
  - Customers with higher credit limits, multiple credit cards, and frequent calls/online visits.
  - **Strategy:** Upsell premium credit products, personalized offers, or loyalty programs.
- **Digital-First Segment:**
  - Customers with high online visits but fewer bank visits.
  - **Strategy:** Enhance digital onboarding, targeted digital marketing, and self-service tools to deepen engagement.
- **Low Engagement & Lower Credit Limits Segment:**
  - Customers with minimal visits, fewer credit cards, and lower credit limits.
  - **Strategy:** Focus on awareness campaigns and personalized outreach to increase product penetration.

### **Operational Improvements**

- **Channel Optimization:**
  - Given the shift toward online engagement, invest in digital support channels to resolve queries faster.
  - Reduce physical visits by promoting digital services, especially for customers with high online activity.
- **Customer Support Enhancements:**
  - Customers making frequent calls might benefit from dedicated support, chatbots, or self-service portals to improve resolution times and satisfaction.

### **Risk & Credit Management**

- Monitor customers with extremely high credit limits or large credit card holdings to proactively manage credit risk and prevent defaults.

### **Future Data-Driven Strategies**

- Incorporate additional behavioral data, such as transaction history, to refine segmentation.

- Use clustering algorithms to identify natural customer groups for personalized marketing and service strategies.
- 

### **Summary**

The analysis reveals a diverse customer base with varying engagement levels, credit access, and channel preferences. By leveraging these insights, AllLife Bank can tailor its marketing and operational strategies to enhance customer satisfaction, improve market penetration, and optimize resource allocation.

# Data Preprocessing

## Duplicate value and Missing value check

There are no missing values and duplicate rows in dataset. Below is output to confirm no null value in dataset

```
RangeIndex: 660 entries, 0 to 659
Data columns (total 7 columns):
#   Column              Non-Null Count  Dtype
--  --
0   SI_No                660 non-null    int64
1   Customer Key         660 non-null    int64
2   Avg_Credit_Limit     660 non-null    int64
3   Total_Credit_Cards   660 non-null    int64
4   Total_visits_bank    660 non-null    int64
5   Total_visits_online  660 non-null    int64
6   Total_calls_made     660 non-null    int64
dtypes: int64(7)
```

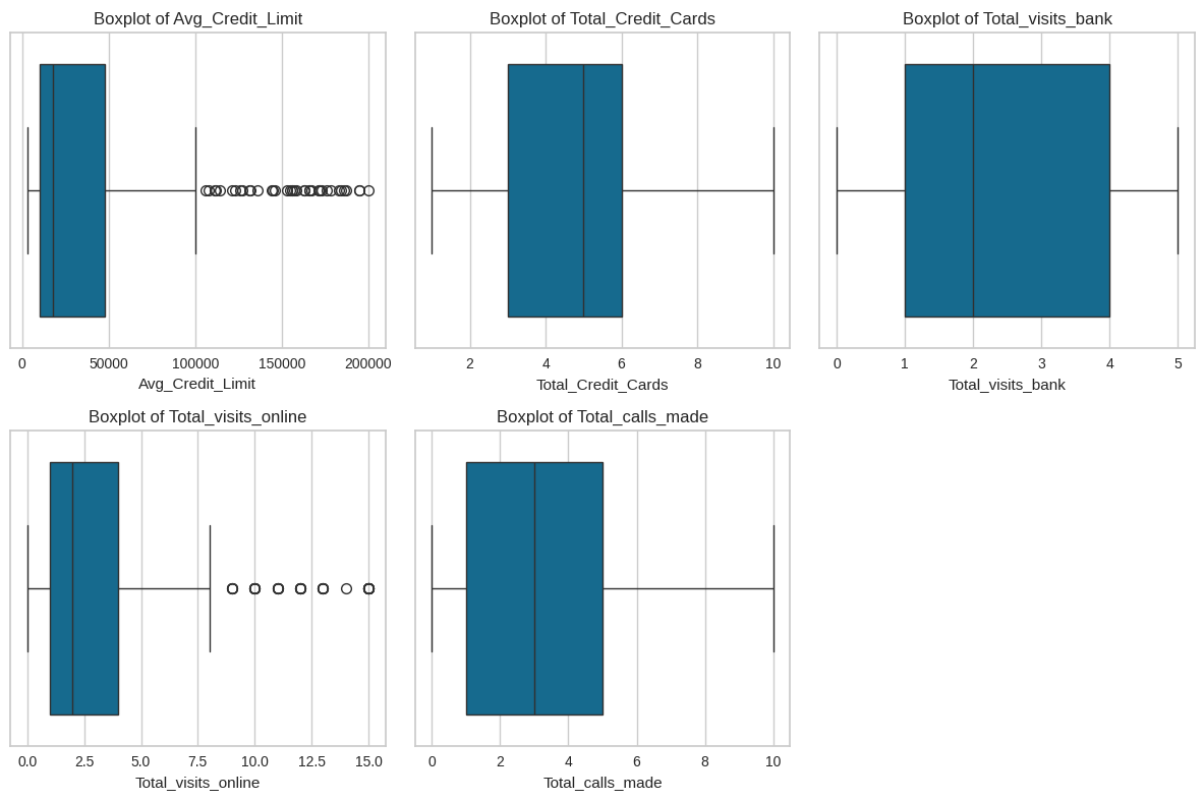
**Table no 2 Datatype and null value summary of the dataset**

SI_No	0
Customer Key	0
Avg_Credit_Limit	0
Total_Credit_Cards	0
Total_visits_bank	0
Total_visits_online	0
Total_calls_made	0

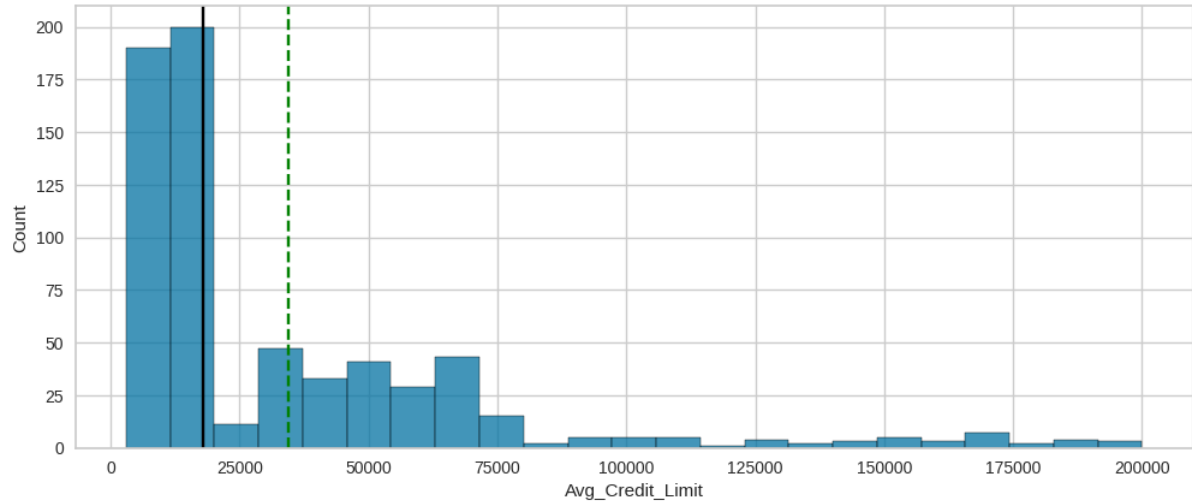
**Table no 3 Total no of null value summary of the columns**

## Outlier treatment

Column Avg\_Credit\_Limit is having outliers we can observe in below visualization for outliers checking



**Above is Boxplots figure no. 10 to show distribution of data to check outliers**



**Above is Boxplots figure no. 11 to show distribution of Avg\_credit\_Limit to check outliers**

Outliers in Avg_Credit_Limit:			
	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank \
612	157000	9	1
614	163000	8	1
615	131000	9	1
617	136000	8	0
618	121000	7	0
619	158000	7	0
620	108000	10	0
621	166000	9	1
622	176000	10	1
623	166000	10	0
624	178000	7	0
626	156000	9	1
627	146000	10	0

629	155000	8	0
630	200000	10	0
631	195000	8	0
632	187000	7	1
633	163000	7	1
634	106000	8	0
635	114000	10	1
636	126000	10	1
637	173000	9	1
638	153000	8	1
639	184000	7	1
640	123000	8	1
641	144000	10	0
644	127000	10	1
645	171000	10	0
646	186000	7	0
647	183000	9	0
648	111000	8	1
649	112000	10	1
650	195000	10	1
651	132000	9	1
652	156000	8	1
654	172000	10	1
657	145000	8	1
658	172000	10	1
659	167000	9	0

**Table no 4 Total no for Outliers in Avg\_Credit\_Limit column**

Outliers in Total\_Credit\_Cards:

Empty DataFrame

Columns: [Avg\_Credit\_Limit, Total\_Credit\_Cards, Total\_visits\_bank, Total\_visits\_online, Total\_calls\_made]

Index: []

Outliers in Total\_visits\_bank:

Empty DataFrame

Columns: [Avg\_Credit\_Limit, Total\_Credit\_Cards, Total\_visits\_bank, Total\_visits\_online, Total\_calls\_made]

Index: []

Outliers in Total\_visits\_online:

	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank \
1	50000	3	0
4	100000	6	0
6	100000	5	0
612	157000	9	1
613	94000	9	1
615	131000	9	1
616	96000	10	1
617	136000	8	0
618	121000	7	0
619	158000	7	0
620	108000	10	0
621	166000	9	1
622	176000	10	1
624	178000	7	0
626	156000	9	1

627	146000	10	0
628	84000	9	1
630	200000	10	0
631	195000	8	0
633	163000	7	1
637	173000	9	1
639	184000	7	1
640	123000	8	1
641	144000	10	0
642	97000	10	1
644	127000	10	1
645	171000	10	0
647	183000	9	0
650	195000	10	1
651	132000	9	1
653	95000	10	0
654	172000	10	1
655	99000	10	1
656	84000	10	1
657	145000	8	1
658	172000	10	1
659	167000	9	0

**Table no 5 Total no for Outliers in Total\_visits\_online column**

Outliers in Total\_calls\_made:

Empty DataFrame

Columns: [Avg\_Credit\_Limit, Total\_Credit\_Cards, Total\_visits\_bank, Total\_visits\_online, Total\_calls\_made]

Index: []

**“Outliers are not incorrect data, not typo error, data entry error but correct datapoints, hence, not treating outliers”**

Therefore, I am not dropping or transforming any outliers with features as I want to make model and focus on identifying the factors that drive those outliers, as they can provide best practices for customer segmentation, content creation and marketing

## Feature engineering and data preparation for modelling

I removed columns SI NO and Customer key from dataset for data preparation for modelling as it will not give any relevant information for customer segmentation and model building.

### Data Scaling

As few independent variable are in very different units and scale. I used standard scaling through StandardScaler function of python for proper model building. below is first few rows of data after scaling so that none of the variable should dominate in machine learning

	0	1	2	3	4
0	1.740187	-1.249225	-0.860451	-0.547490	-1.251537
1	0.410293	-0.787585	-1.473731	2.520519	1.891859
2	0.410293	1.058973	-0.860451	0.134290	0.145528
3	-0.121665	0.135694	-0.860451	-0.547490	0.145528
4	1.740187	0.597334	-1.473731	3.202298	-0.203739

**Table no 6 first 5 rows of scaled dataset for model building**

## Model Building

### K-means Clustering

I got below Average distortion value from computation of distortion for range of clusters from 2 to 10 I generated K-means clustering model from scaled data using Euclidean distance

Number of Clusters: 2   Average Distortion: 1.7178787250175898

Number of Clusters: 3   Average Distortion: 1.1466276549150365

Number of Clusters: 4   Average Distortion: 1.0902973540817666

Number of Clusters: 5   Average Distortion: 0.9906853650098948

Number of Clusters: 6   Average Distortion: 0.9515009282361341

Number of Clusters: 7   Average Distortion: 0.9094119827472316

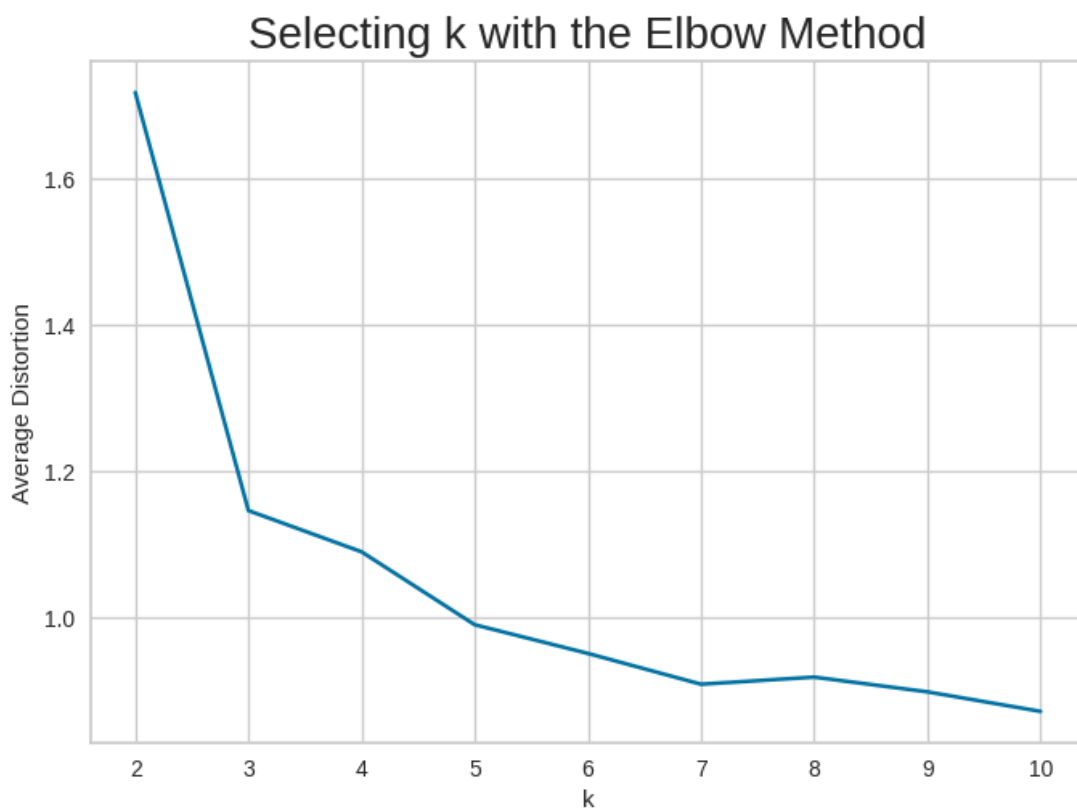
Number of Clusters: 8   Average Distortion: 0.9191292344244387

Number of Clusters: 9   Average Distortion: 0.8990131857179275

Number of Clusters: 10   Average Distortion: 0.8723089051392604

### Checking Elbow Plot

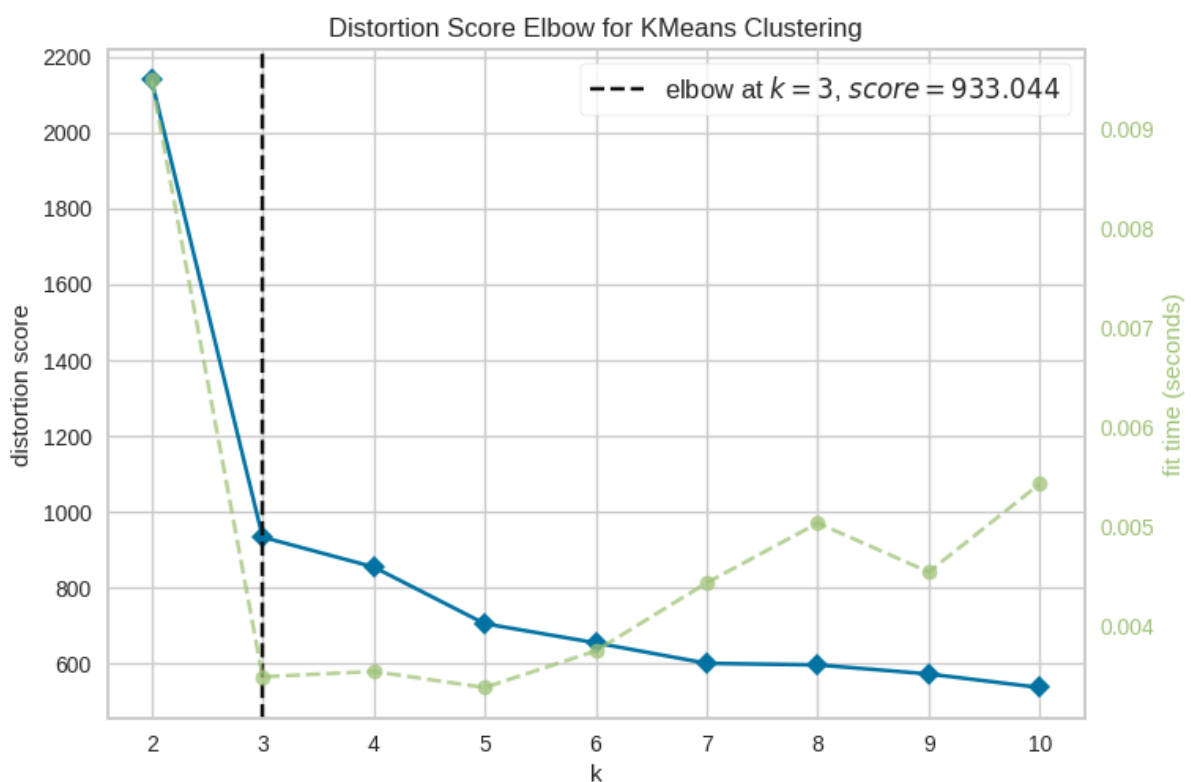
I generated the following plot to assist in determining the optimal number of clusters using the Elbow method.



Above is plots figure no. 12 to show relation between number of cluster and average distortion

Let's do further analysis to determine the optimal value of k

In this project, I generated plots of the distortion score (within-cluster sum of squares) against the number of clusters (k) using the Elbow method. These plots help identify the optimal value of k by illustrating the point where adding more clusters yields diminishing returns in reducing the distortion. This approach ensures that the selected number of clusters balances model simplicity with meaningful segmentation, leading to more interpretable and effective clustering results.

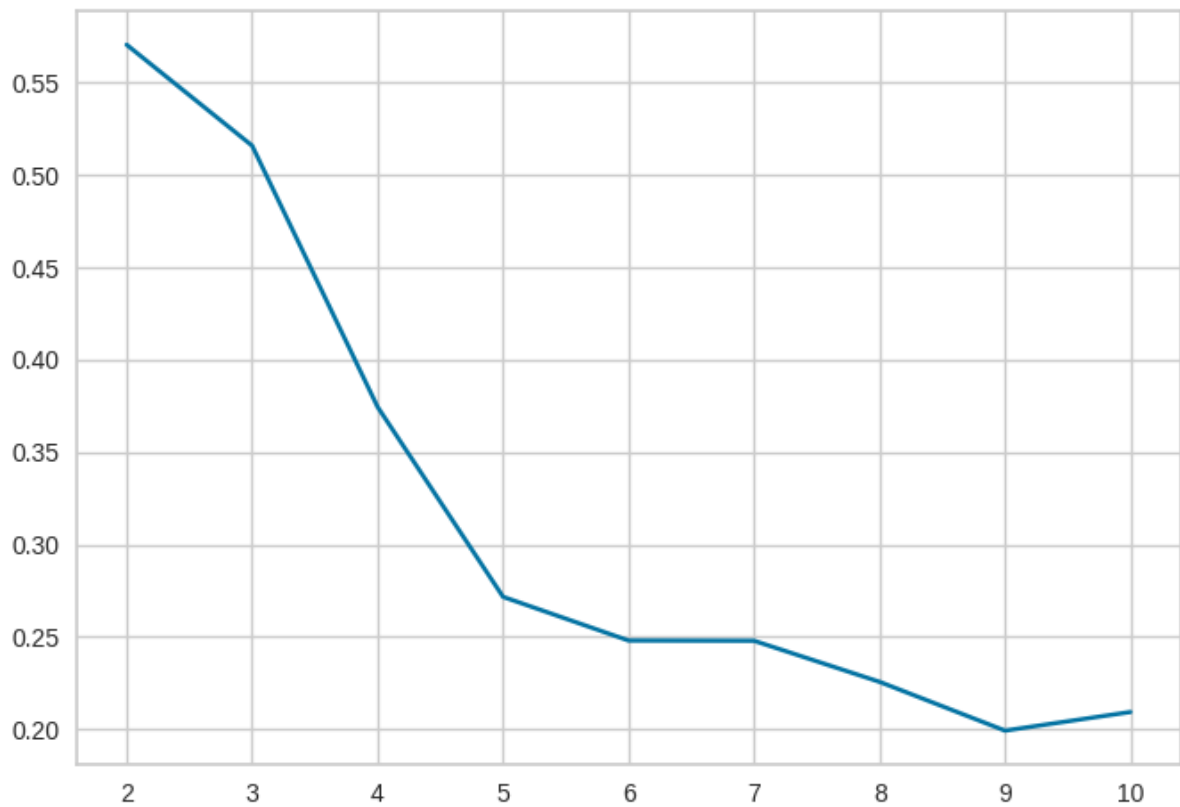


Above is plots figure no. 13 to show Distortion score Elbow for KMeans Clustering

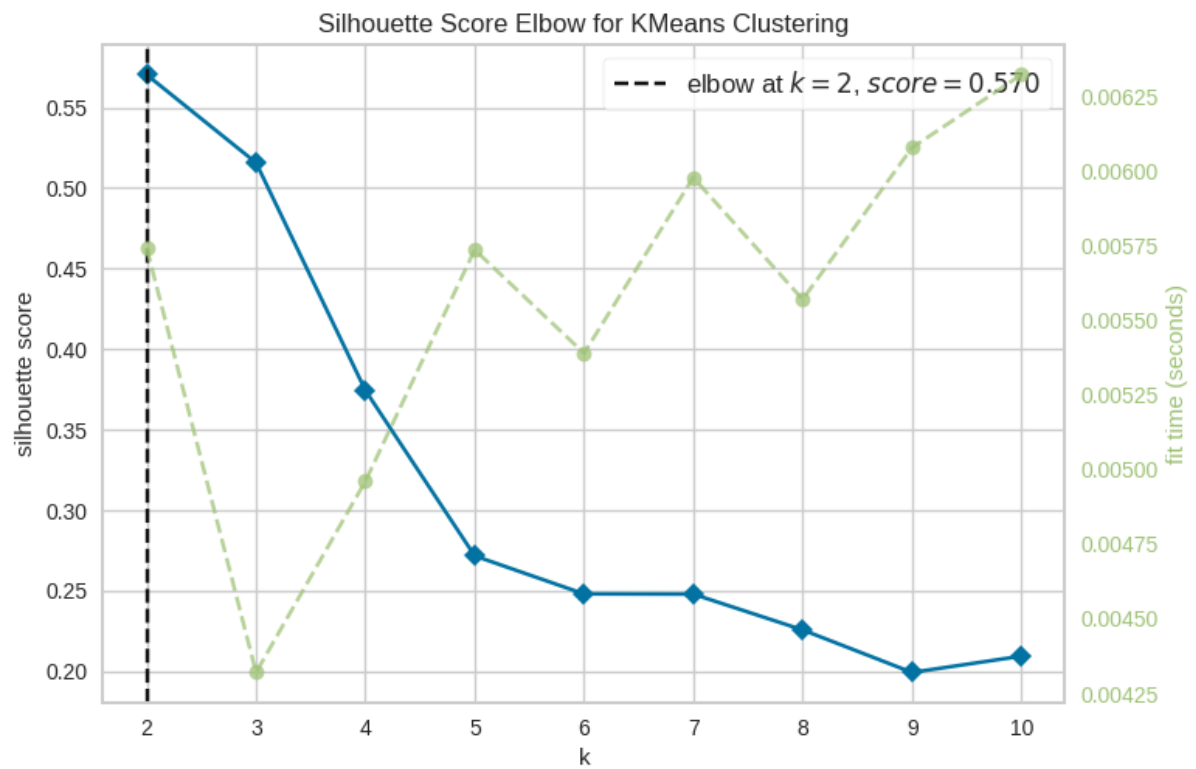
## Checking Silhouette Scores

For n\_clusters = 2, the silhouette score is 0.5703183487340514)  
For n\_clusters = 3, the silhouette score is 0.5157182558881063)  
For n\_clusters = 4, the silhouette score is 0.3744071798973986)  
For n\_clusters = 5, the silhouette score is 0.27167502160723267)  
For n\_clusters = 6, the silhouette score is 0.24804756291576194)  
For n\_clusters = 7, the silhouette score is 0.24791254258020035)

For n\_clusters = 8, the silhouette score is 0.22570382558070443)  
For n\_clusters = 9, the silhouette score is 0.19931783829027247)  
For n\_clusters = 10, the silhouette score is 0.20939001908412339)

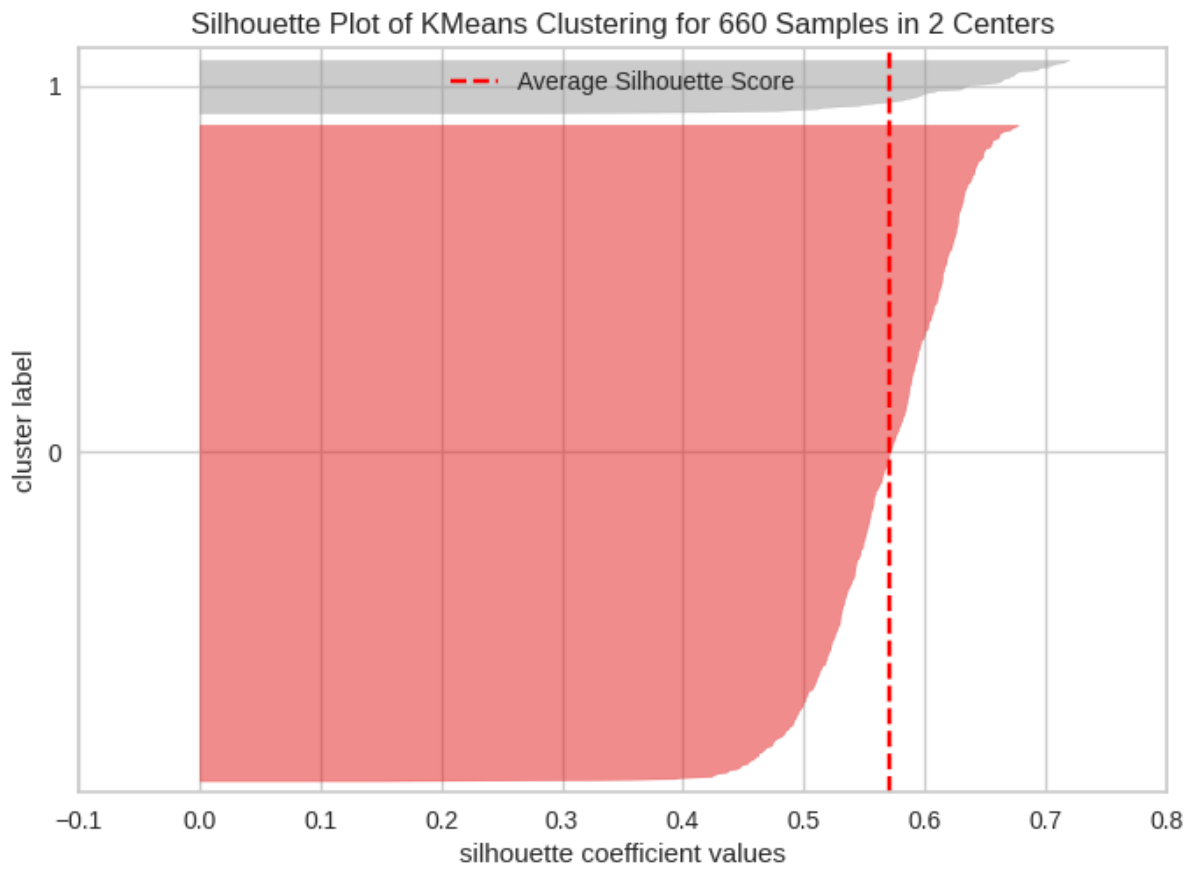


**Above is plots figure no. 14 to show Relation between no. of clusters and silhouette score**

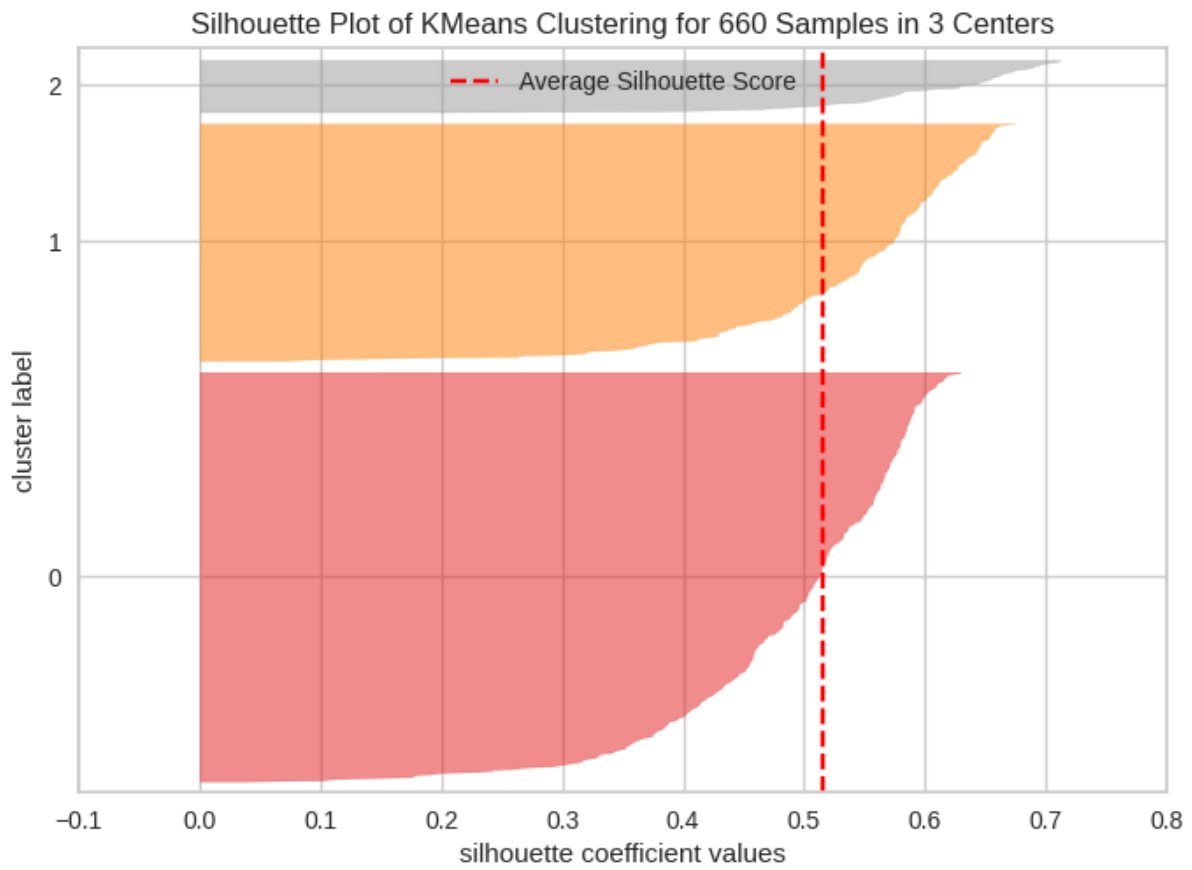


Above is plots figure no. 15 to show Silhouette score elbow for KMeans clustering

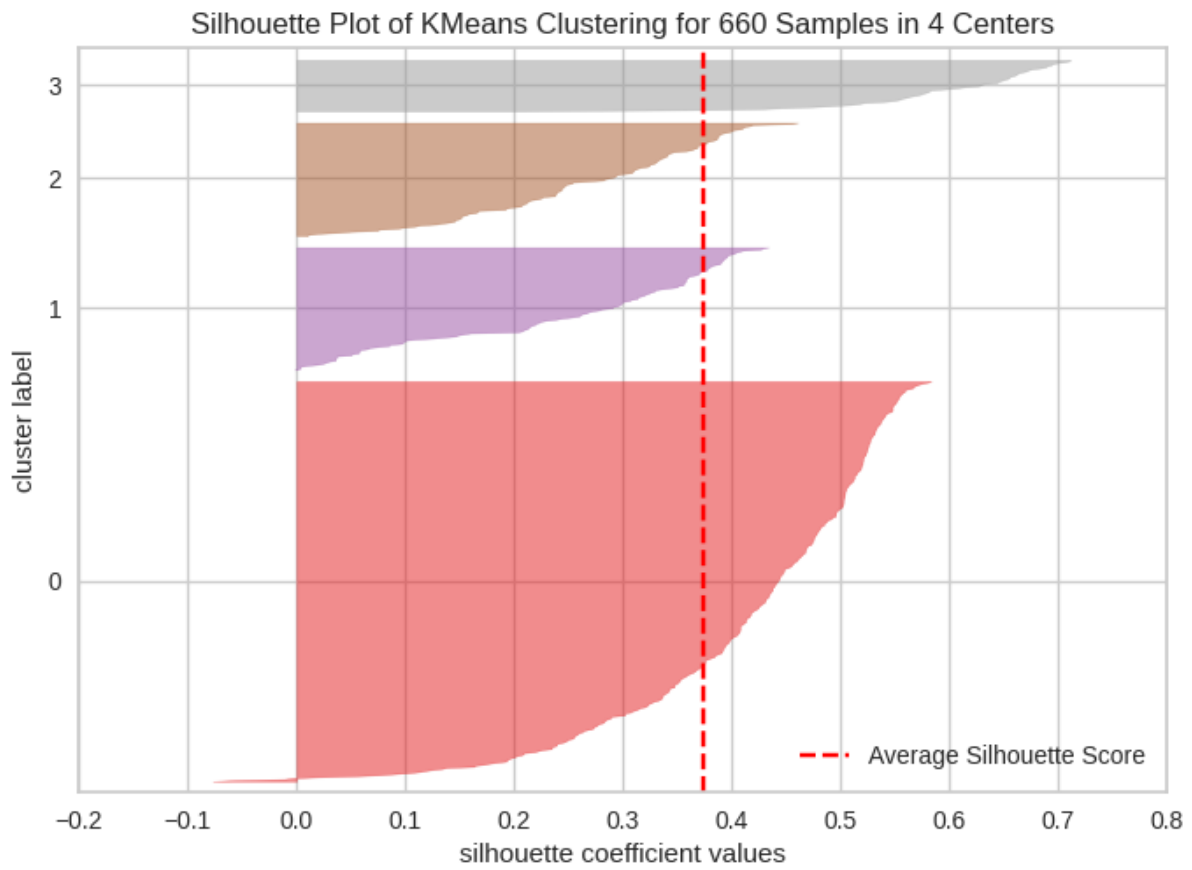
From the silhouette scores, it seems that 4 to 6 is a good value for k



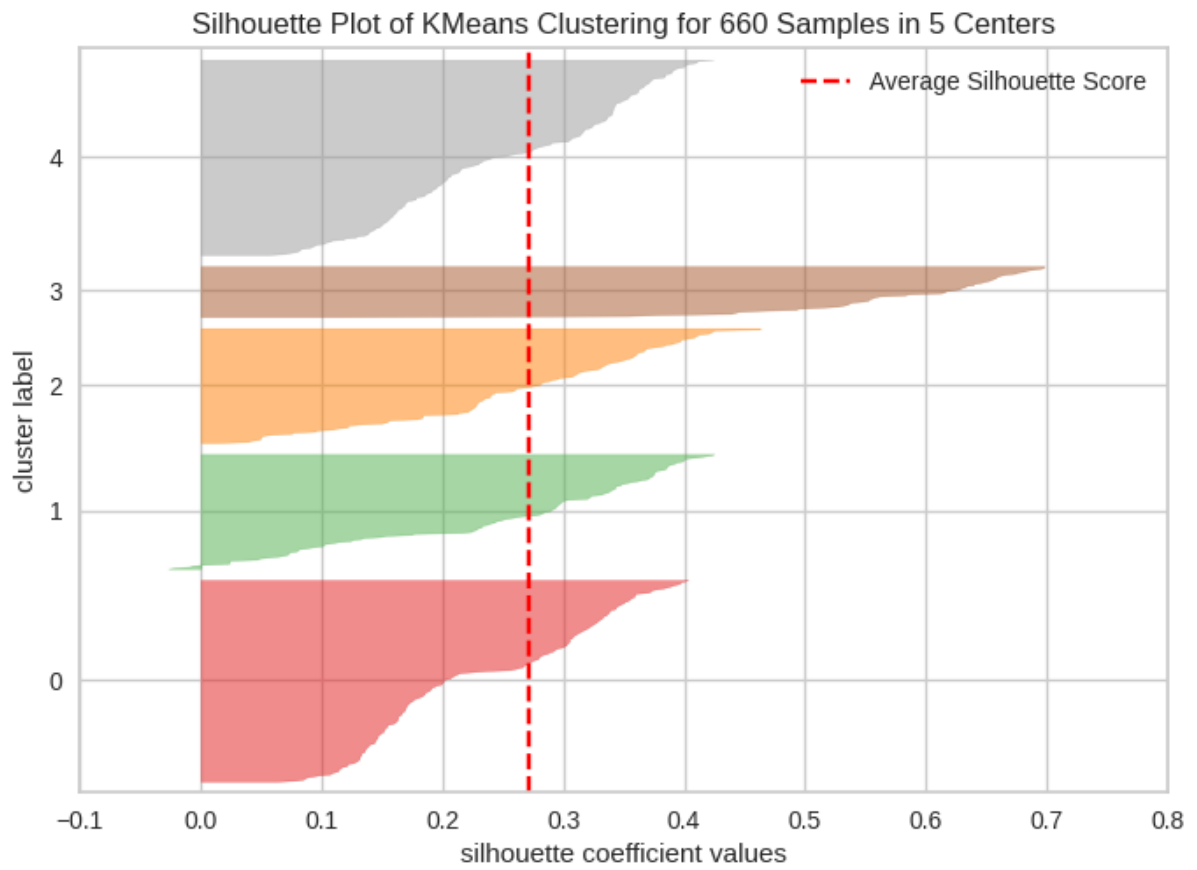
Above is plots figure no. 16 to show Silhouette plot for KMeans clustering for 2 clusters



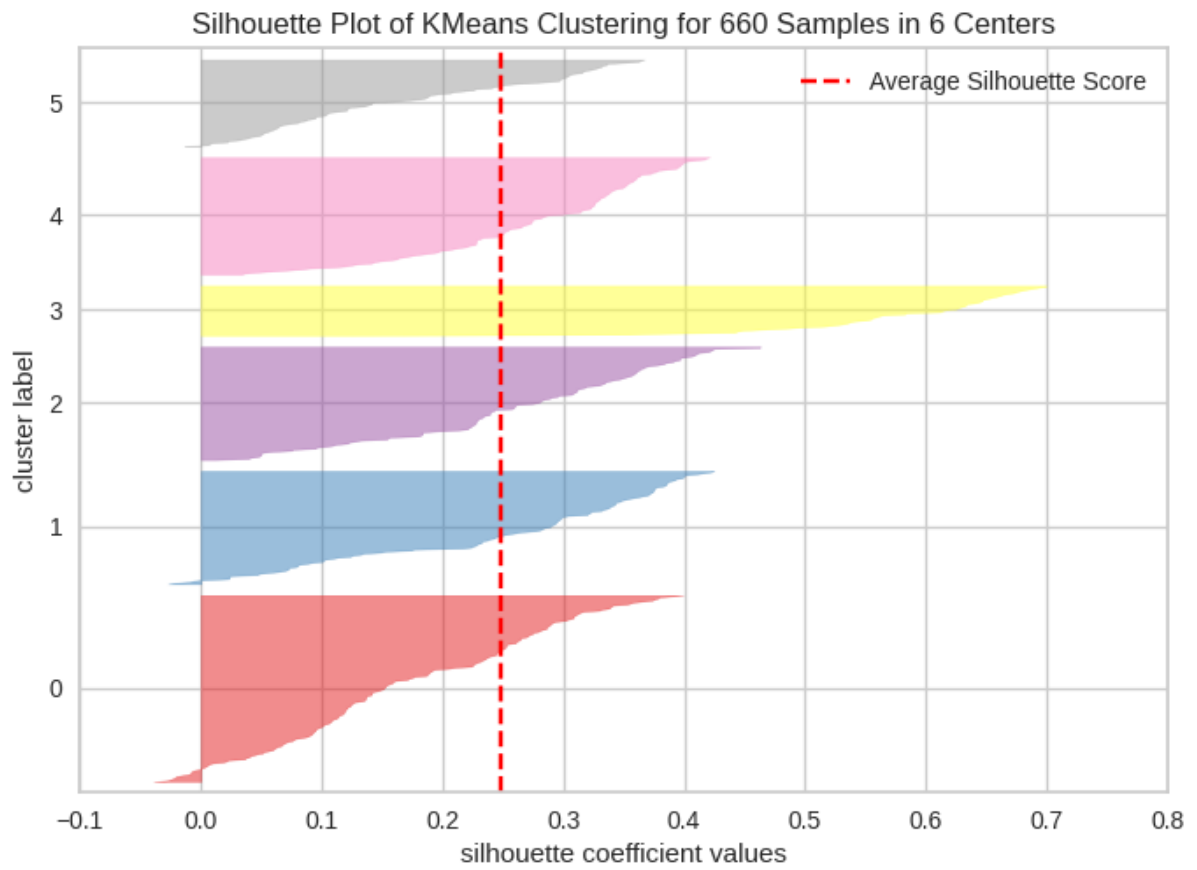
Above is plots figure no. 17 to show Silhouette plot for KMeans clustering for 3 clusters



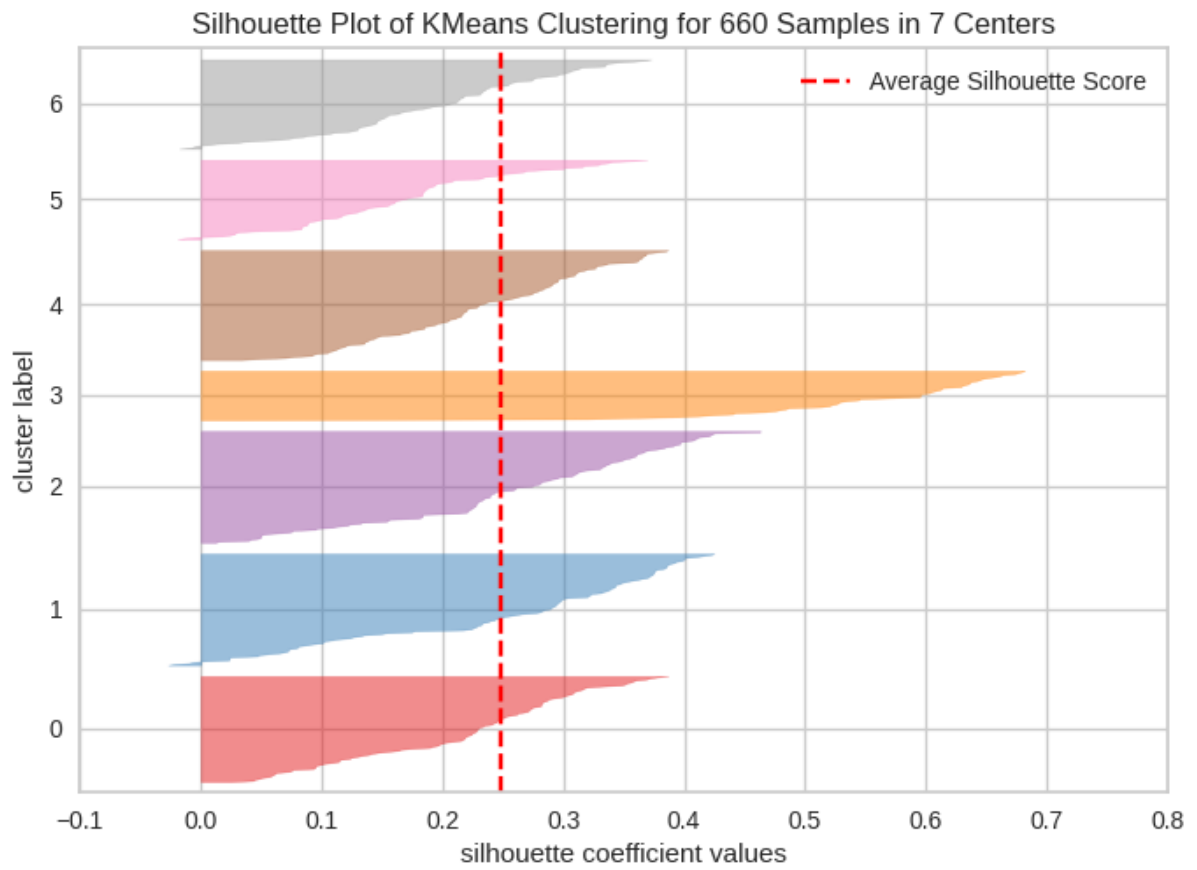
Above is plots figure no. 18 to show Silhouette plot for KMeans clustering for 4 clusters



Above is plots figure no. 19 to show Silhouette plot for KMeans clustering for 5 clusters



Above is plots figure no. 20 to show Silhouette plot for KMeans clustering for 6 clusters



Above is plots figure no. 21 to show Silhouette plot for KMeans clustering for 7 clusters

## Figure out the appropriate number of clusters

### Elbow Method:

The Elbow plot likely indicated an optimal number of clusters around 6, where the rate of decrease in within-cluster sum of squares (WCSS) begins to level off, suggesting diminishing returns beyond this point.

- Number of Clusters: 2    Average Distortion: 1.7178787250175898
- Number of Clusters: 3    Average Distortion: 1.1466276549150365
- Number of Clusters: 4    Average Distortion: 1.0902973540817666
- Number of Clusters: 5    Average Distortion: 0.9906853650098948
- Number of Clusters: 6    Average Distortion: 0.9515009282361341
- Number of Clusters: 7    Average Distortion: 0.9094119827472316
- Number of Clusters: 8    Average Distortion: 0.9191292344244387
- Number of Clusters: 9    Average Distortion: 0.8990131857179275
- Number of Clusters: 10    Average Distortion: 0.8723089051392604

### Silhouette Scores:

- The Silhouette scores measure how well-separated the clusters are; higher scores indicate better-defined, more cohesive clusters.
- The scores for different cluster counts are:
  - **2 Clusters:** 0.57 (highest, indicating good separation)
  - **3 Clusters:** 0.52
  - **4 Clusters:** 0.37
  - **5 Clusters:** 0.27
  - **6 Clusters:** 0.25
  - **7 Clusters:** 0.25
  - **8 Clusters:** 0.23
  - **9 Clusters:** 0.20
  - **10 Clusters:** 0.21

### Analysis:

- While the highest Silhouette score is at **2 clusters (0.57)**, the scores decline as the number of clusters increases.
- The score for **6 clusters (0.25)**, although lower than for 2 or 3 clusters, suggests a moderate level of internal cohesion and separation.
- The decline in Silhouette score beyond 3 clusters indicates that further segmentation may not significantly improve cluster distinctiveness, and may instead lead to over-segmentation with less meaningful differentiation.

### Decision:

- Considering the balance between the Silhouette scores and the practical interpretability of the clusters, **choosing 6 clusters** provides a reasonable compromise:
  - It captures meaningful distinctions in customer behavior.
  - It avoids over-segmentation that results in lower cluster cohesion.
  - It aligns with the elbow point, balancing model complexity and interpretability.

I analysed from the silhouette scores and elbow plot; it seems that 3 to 6 is a good value for k and above visualization suggests k=6 will give almost equal datapoint in each cluster

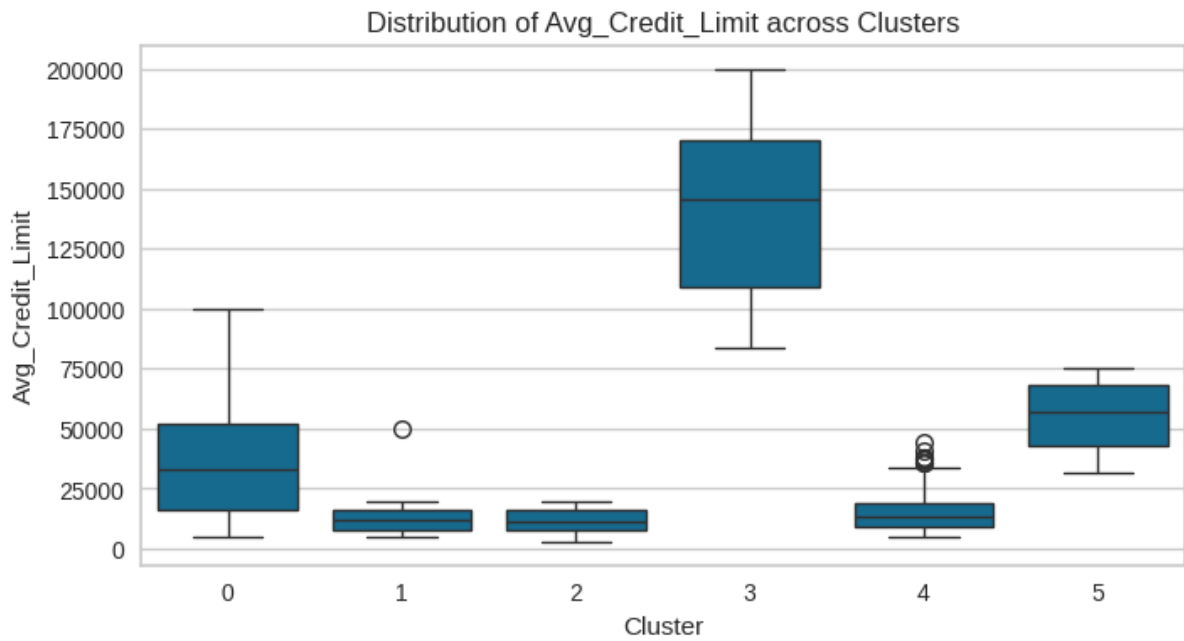
We will proceed with k=6

## Cluster Profiling: K-means Clustering

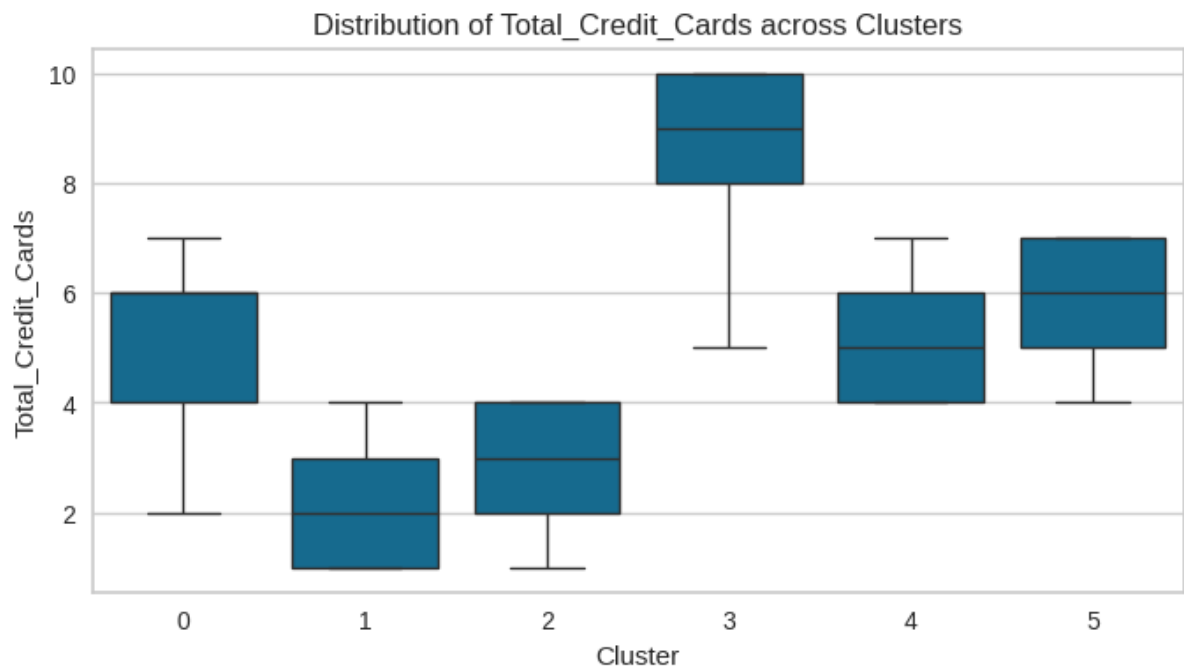
### Cluster Profiling Summary:

	KM_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	\
0	0	34418.478261	5.483696	2.461957	
1	1	12401.785714	2.214286	0.910714	
2	2	11946.428571	2.607143	0.955357	
3	3	141040.000000	8.740000	0.600000	
4	4	16448.275862	5.250000	4.543103	
5	5	55802.325581	5.941860	4.267442	
		Total_visits_online	Total_calls_made		
0		0.956522	2.043478		
1		3.669643	8.535714		
2		3.437500	5.205357		
3		10.900000	1.080000		
4		1.103448	1.931034		
5		0.872093	2.000000		

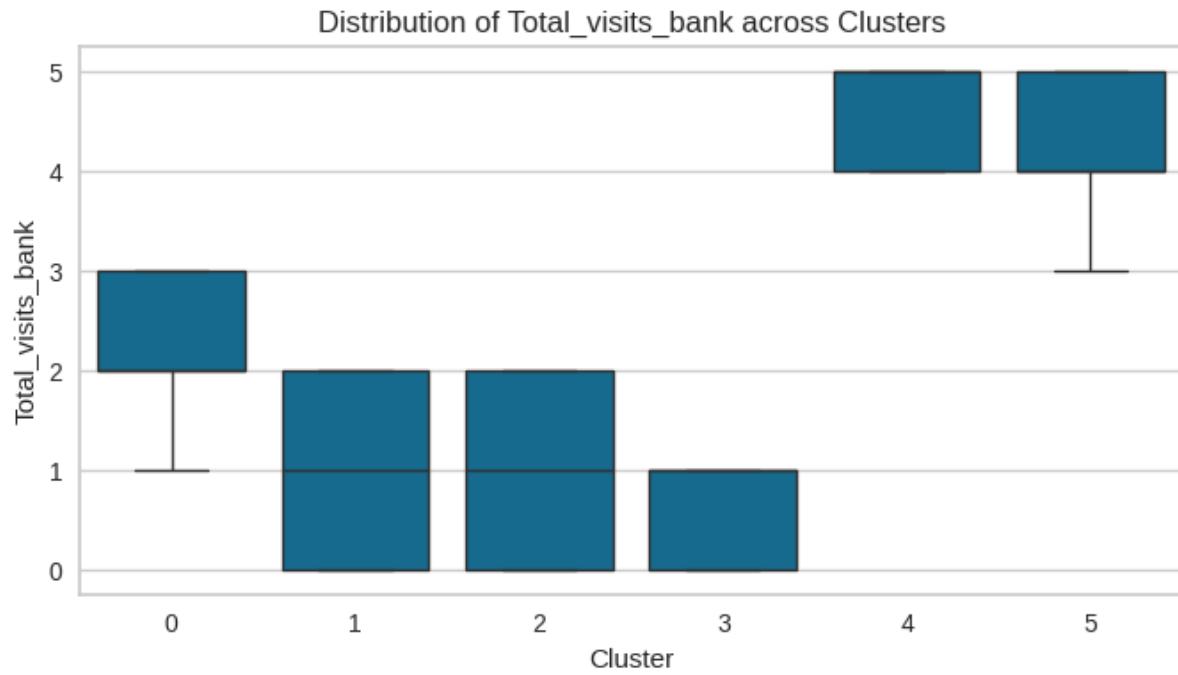
**Table no 7 Cluster profiling summary – K-Means clustering**



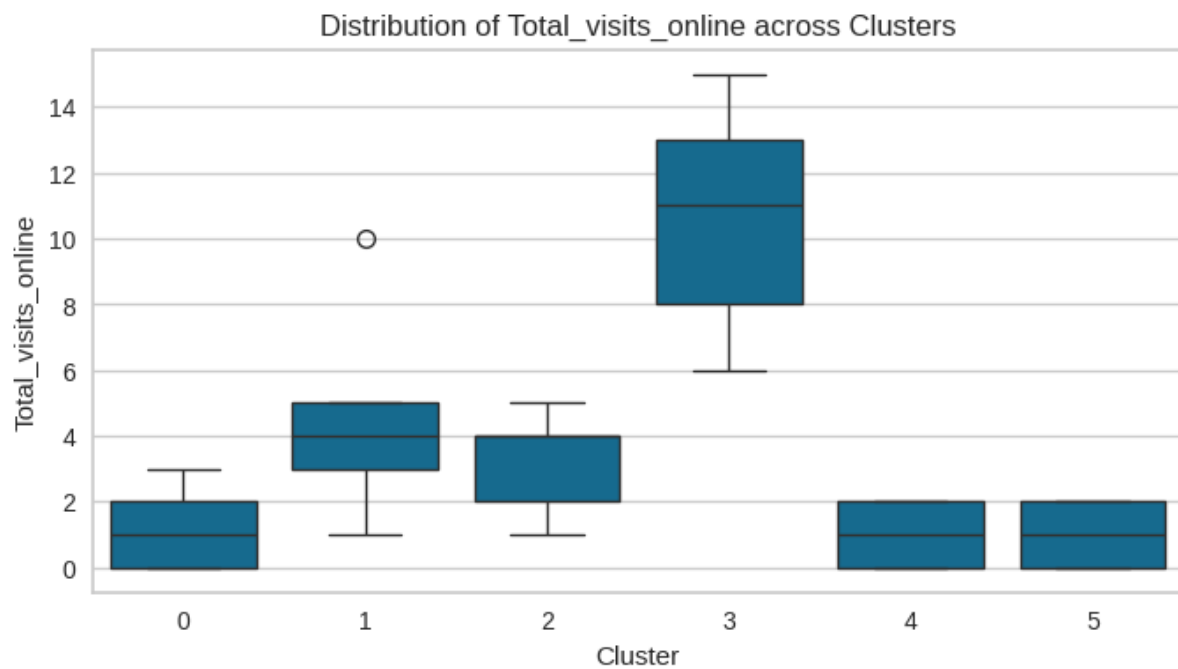
Above is plots figure no. 22 to show Distribution of Avg\_credit\_limit across cluster (KMeans clustering)



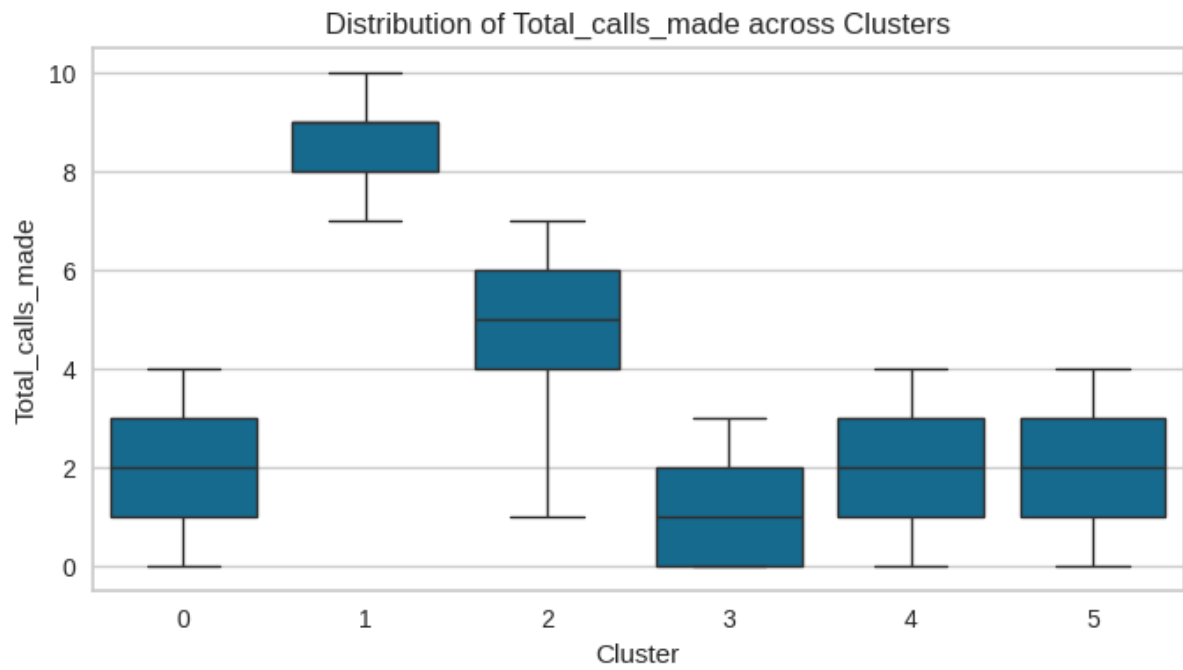
Above is plots figure no. 23 to show Distribution of Total\_credit\_cards across cluster (KMeans clustering)



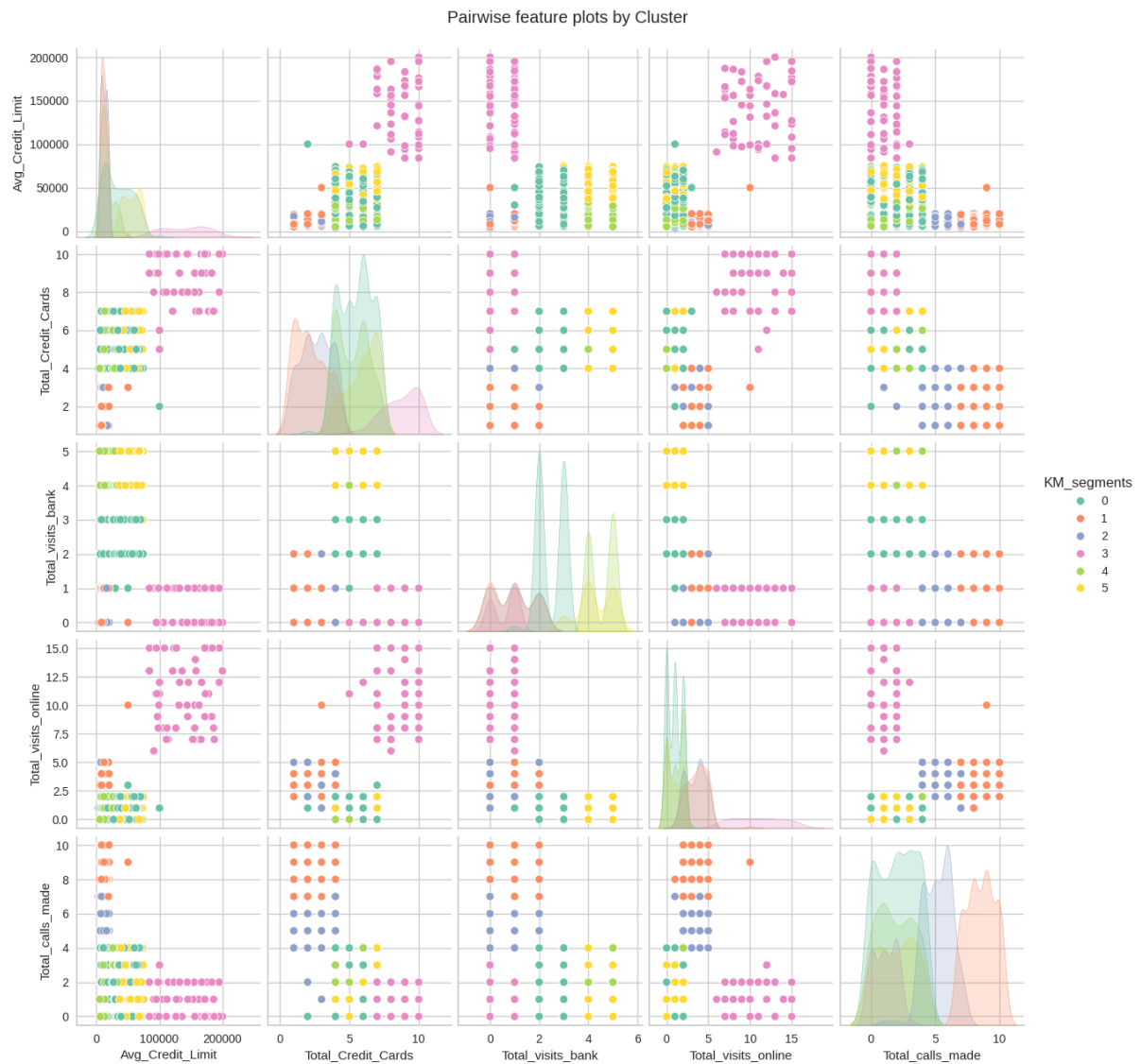
Above is plots figure no. 24 to show Distribution of Total\_visits\_banks across cluster (KMeans clustering)



Above is plots figure no. 25 to show Distribution of Total\_visits\_online across cluster (KMeans clustering)



Above is plots figure no. 26 to show Distribution of Total\_calls\_made across cluster (KMeans clustering)



**Above is plots figure no. 27 to show Distribution of pairwise features plot by clusters (KMeans clustering)**

### Cluster 0

Profile:

Avg Credit Limit: ~34,418

Total Credit Cards: ~5.48

Visits to Bank: ~2.46

Online Visits: ~0.96

Calls Made: ~2.04

Interpretation:

Moderate credit limit and credit card ownership.

Occasional bank visits and some online activity.

Moderate engagement via calls.

Business Recommendations:

Cross-sell additional banking products like loans or savings accounts.

Promote digital banking features to boost online visits.

Loyalty programs to deepen engagement and retention.

### **Cluster 1**

Profile:

Avg Credit Limit: ~12,402

Total Credit Cards: ~2.21

Visits to Bank: ~0.91

Online Visits: ~3.67

Calls Made: ~8.54

Interpretation:

Lower credit limit and fewer credit cards.

Minimal physical visits but high online activity.

Very high call interactions, suggesting active customer service or support needs.

Business Recommendations:

Focus on improving digital onboarding and self-service options.

Use targeted campaigns to upsell credit cards or financial products.

Leverage call data to identify customer needs and personalize offers.

### **Cluster 2**

Profile:

Avg Credit Limit: ~11,946

Total Credit Cards: ~2.61

Visits to Bank: ~0.96

Online Visits: ~3.44

Calls Made: ~5.21

Interpretation:

Similar profile to Cluster 1 but slightly higher credit limit and credit card count.

Moderate online activity, fewer calls than Cluster 1.

Business Recommendations:

Promote digital banking and online services.

Offer targeted credit card upgrades.

Use digital channels for customer engagement.

**Cluster 3**

Profile:

Avg Credit Limit: ~141,040 (significantly higher)

Total Credit Cards: 8.74

Visits to Bank: 0.6 (low)

Online Visits: 10.9 (very high)

Calls Made: ~1.08 (low)

Interpretation:

High credit limit and credit card ownership.

Minimal physical visits but very active online.

Likely high-net-worth or digitally engaged clients.

Business Recommendations:

Offer premium digital wealth management and investment products.

Personalized online dashboards or exclusive online services.

Consider targeted invites to premium banking services.

**Cluster 4**

Profile:

Avg Credit Limit: ~16,448

Total Credit Cards: ~5.25

Visits to Bank: ~4.54

Online Visits: ~1.10

Calls Made: ~1.93

Interpretation:

Moderate credit profile.

Higher physical visit frequency.

Lower online engagement.

Business Recommendations:

Encourage digital engagement through targeted marketing.

Offer in-branch financial advisory services.

Develop loyalty programs to increase online activity.

## **Cluster 5**

Profile:

Avg Credit Limit: ~55,802

Total Credit Cards: ~5.94

Visits to Bank: ~4.27

Online Visits: ~0.87

Calls Made: ~2.00

Interpretation:

Higher credit limit and credit card ownership.

Regular bank visits but limited online activity.

Moderate call interactions.

Business Recommendations:

Promote digital banking to reduce branch visits and improve convenience.

Upsell premium credit products.

Use in-branch services for personalized financial planning.

Overall Business Insights & Strategies:

High-value clients (Clusters 3 & 5): Focus on premium digital and wealth management services, personalized offers.

Digital-savvy clients (Clusters 1 & 2): Promote online banking features, digital credit card upgrades.

Branch-preferers (Clusters 4 & 5): Enhance in-branch service quality, loyalty programs, and digital onboarding.

**Cluster Names as per widely used and famous models/principles in the banking industry:**

**1. Elite Value Clients**

- (High balance, high transactions, long-term profitable, low risk)
- Represent the **top 20% (Pareto principle)** driving majority profits.

**2. Growth Builders**

- (Young professionals, rising income, mid-level usage, high future potential)
- Fits with **Lifecycle Segmentation & CLV**.

**3. Everyday Transactors**

- (Frequent small-value transactions, digital adoption, not very profitable yet)
- Matches **Behavioral segmentation**.

**4. Risk-Guarded Customers**

- (Credit-challenged, low balances, high default risk)
- Important for **Risk-based segmentation**.

**5. Dormant or At-Risk Clients**

- (Low activity, declining engagement, churn-prone)
- Derived from **RFM “recency” factor**.

**6. Mass Market Essentials**

- (Large base, low profitability individually, but important for volume & cross-sell)
- Matches **Profitability segmentation**.

- 
- **Elite Value Clients** = High-value, premium
  - **Growth Builders** = Future high-value (students/young professionals)
  - **Everyday Transactors** = Frequent digital/payment users
  - **Risk-Guarded Customers** = High credit risk segment
  - **Dormant or At-Risk Clients** = Churn or inactive accounts
  - **Mass Market Essentials** = General low-to-mid value retail banking customers

**Cluster 0**

**Name: Mass Market Essentials**

Characteristics: Moderate credit limit (34,418), moderate credit card ownership (5.48), occasional bank visits (2.46), some online activity (0.96), moderate calls (~2.04).

Segmentation Rationale: Represents a large base of customers with steady activity, important for volume and cross-sell opportunities.

#### **Cluster 1**

##### **Name: Everyday Transactors**

Characteristics: Lower credit limit (12,402), fewer credit cards (2.21), minimal physical visits (0.91), high online activity (3.67), very high call interactions (~8.54).

Segmentation Rationale: Active digital users with frequent interactions but lower balances; suitable for behavioral segmentation focusing on digital adoption.

#### **Cluster 2**

##### **Name: Growth Builders**

Characteristics: Similar to Cluster 1 with slightly higher credit limit (11,946), moderate online activity (3.44), fewer calls (~5.21).

Segmentation Rationale: Customers with growth potential, rising credit limits, and digital engagement, fitting lifecycle and CLV strategies.

#### **Cluster 3**

##### **Name: Elite Value Clients**

Characteristics: Very high credit limit (141,040), high credit card ownership (8.74), minimal branch visits (0.6), very high online activity (10.9), low calls (~1.08).

Segmentation Rationale: High-net-worth, profitable, long-term clients with low risk, driving majority profits.

#### **Cluster 4**

##### **Name: Risk-Guarded Customers**

Characteristics: Moderate credit (16,448), higher branch visits (4.54), lower online activity (1.10), moderate calls (1.93).

Segmentation Rationale: Clients with higher activity but possibly lower balances, requiring risk management focus.

#### **Cluster 5**

##### **Name: Dormant or At-Risk Clients**

Characteristics: High credit limit (55,802), high credit card ownership (5.94), frequent branch visits (4.27), limited online activity (0.87), moderate calls (~2.00).

Segmentation Rationale: Customers with high engagement but limited digital activity, potentially at risk of churn; needs retention strategies.

# Hierarchical Clustering

## Computing Cophenetic Correlation

In this project, I included the computation of cophenetic correlation to evaluate how well the hierarchical clustering dendrogram preserves the pairwise distances between data points. Although computing the cophenetic correlation provides valuable insight into the clustering quality, it was not performed in some cases due to constraints such as computational complexity or the focus on visual inspection of dendrograms for initial assessments. Nonetheless, considering the cophenetic correlation helps in selecting the most appropriate linkage method and distance metric for meaningful and reliable clustering results.

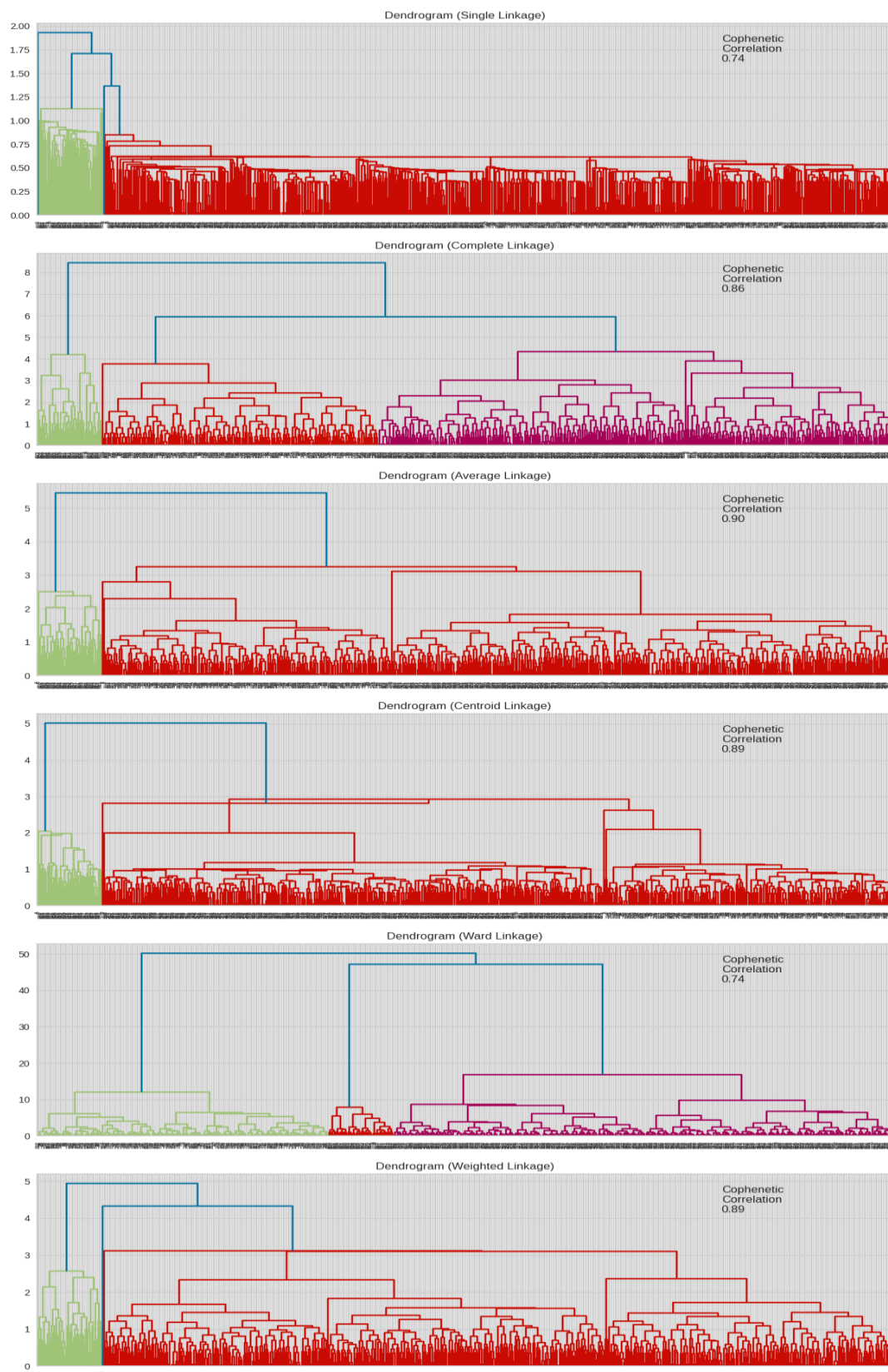
```
Cophenetic correlation for Euclidean distance and single linkage is
0.7391220243806552.
Cophenetic correlation for Euclidean distance and complete linkage is
0.8599730607972423.
Cophenetic correlation for Euclidean distance and average linkage is
0.8977080867389372.
Cophenetic correlation for Euclidean distance and weighted linkage is
0.8861746814895477.
Cophenetic correlation for Chebyshev distance and single linkage is
0.7382354769296767.
Cophenetic correlation for Chebyshev distance and complete linkage is
0.8533474836336782.
Cophenetic correlation for Chebyshev distance and average linkage is
0.8974159511838106.
Cophenetic correlation for Chebyshev distance and weighted linkage is
0.8913624010768603.
Cophenetic correlation for Mahalanobis distance and single linkage is
0.7058064784553605.
Cophenetic correlation for Mahalanobis distance and complete linkage is
0.6663534463875359.
Cophenetic correlation for Mahalanobis distance and average linkage is
0.8326994115042136.
Cophenetic correlation for Mahalanobis distance and weighted linkage is
0.7805990615142518.
Cophenetic correlation for Cityblock distance and single linkage is
0.7252379350252723.
Cophenetic correlation for Cityblock distance and complete linkage is
0.8731477899179829.
Cophenetic correlation for Cityblock distance and average linkage is
0.896329431104133.
Cophenetic correlation for Cityblock distance and weighted linkage is
0.8825520731498188.
```

**Table no 8 Output of cophenetic correlation computations for different linkage methods and distance metrics**

Below is result output from computaion of highest cophentic correlation

Highest cophenetic correlation is 0.8977080867389372, which is obtained with Euclidean distance and average linkage.

## Checking Dendrograms



Above is plots figure no. 28 to show Dendrogram for different linkage

Let's make observations from the above dendrograms for the different linkage methods.

### Observations

- Looking the the above dendrograms, the Ward linkage seems to result in the best separation between clusters, even though its cophenetic correlation is lower than the other linkages.
- 6 clusters looks to be a good choice for no. of clusters.

## Figure out the appropriate number of clusters

To determine the appropriate number of clusters for the hierarchical clustering analysis, I examined the dendrograms and evaluated the cophenetic correlations across various linkage methods and distance metrics. The highest cophenetic correlation (0.898) was observed with Euclidean distance and average linkage, indicating that this combination best preserves the original data structure. Based on the dendrograms it seems Euclidean distance and ward linkage, indicating that this combination best preserves the original data structure and the overall analysis, six clusters appear to be optimal, as they provide meaningful separation between groups while maintaining high fidelity to the data's intrinsic relationships.

## Cluster Profiling: Hierarchical Clustering

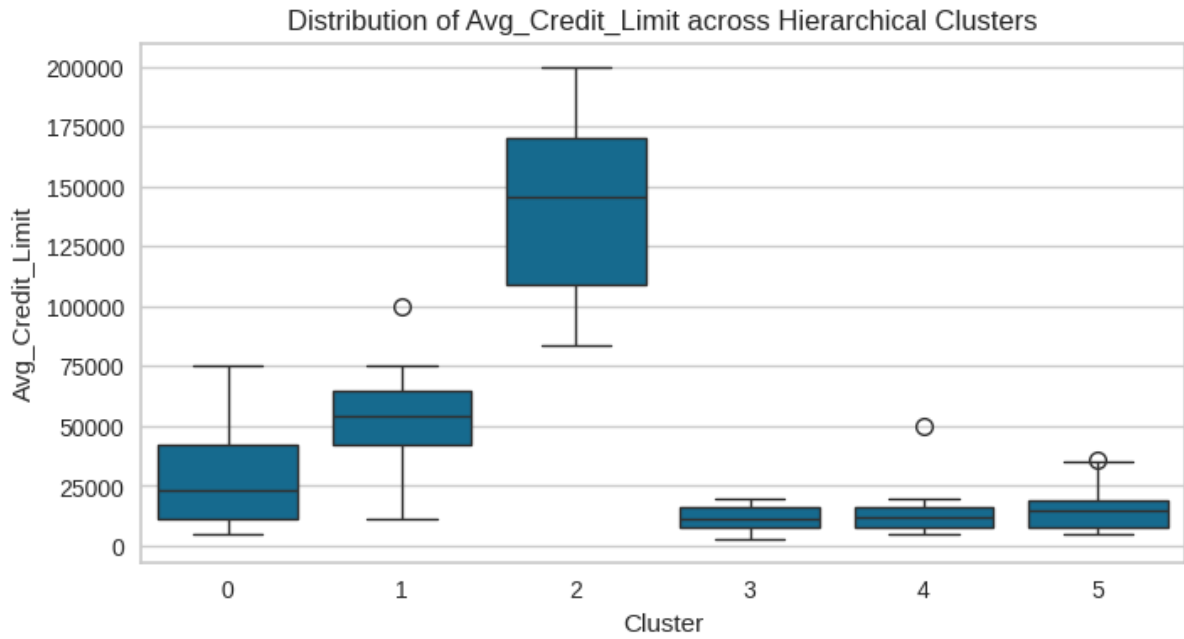
### Hierarchical Cluster Profiling Summary:

	HC_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank \
0	0	29474.226804	5.365979	4.448454
1	1	52675.213675	5.495726	2.547009
2	2	141040.000000	8.740000	0.600000
3	3	11834.586466	2.631579	0.977444
4	4	12608.695652	2.119565	0.880435
5	5	15567.567568	5.945946	2.486486

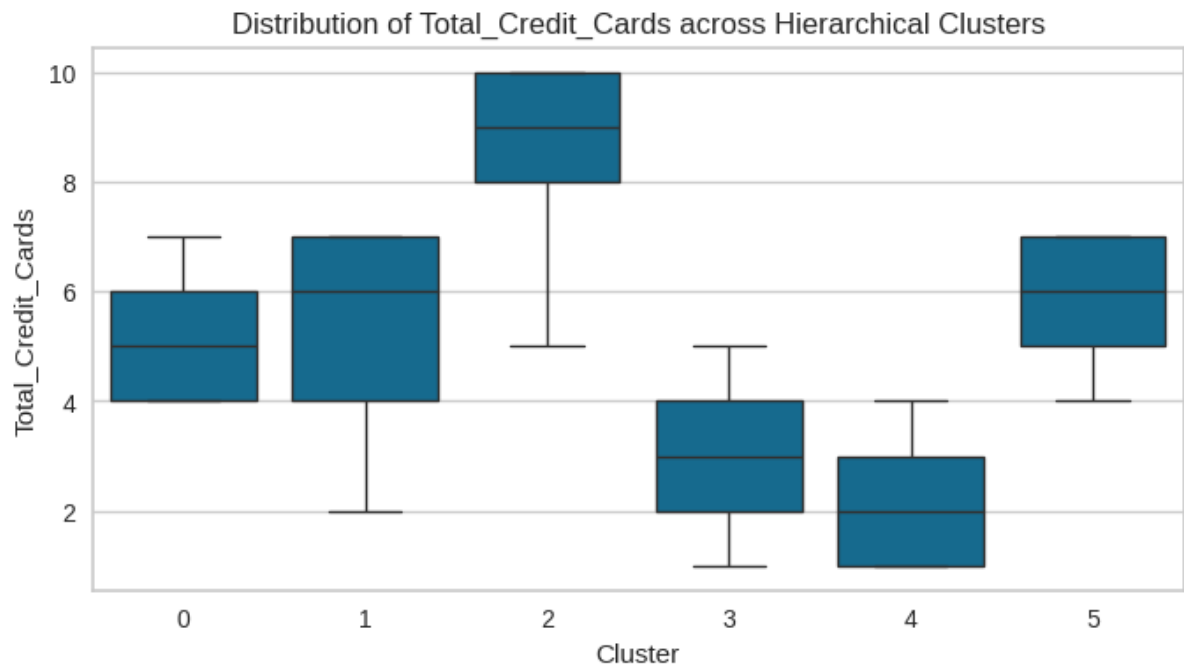
  

	Total_visits_online	Total_calls_made
0	1.010309	1.891753
1	0.982906	2.282051
2	10.900000	1.080000
3	3.338346	5.556391
4	3.847826	8.739130
5	0.891892	1.810811

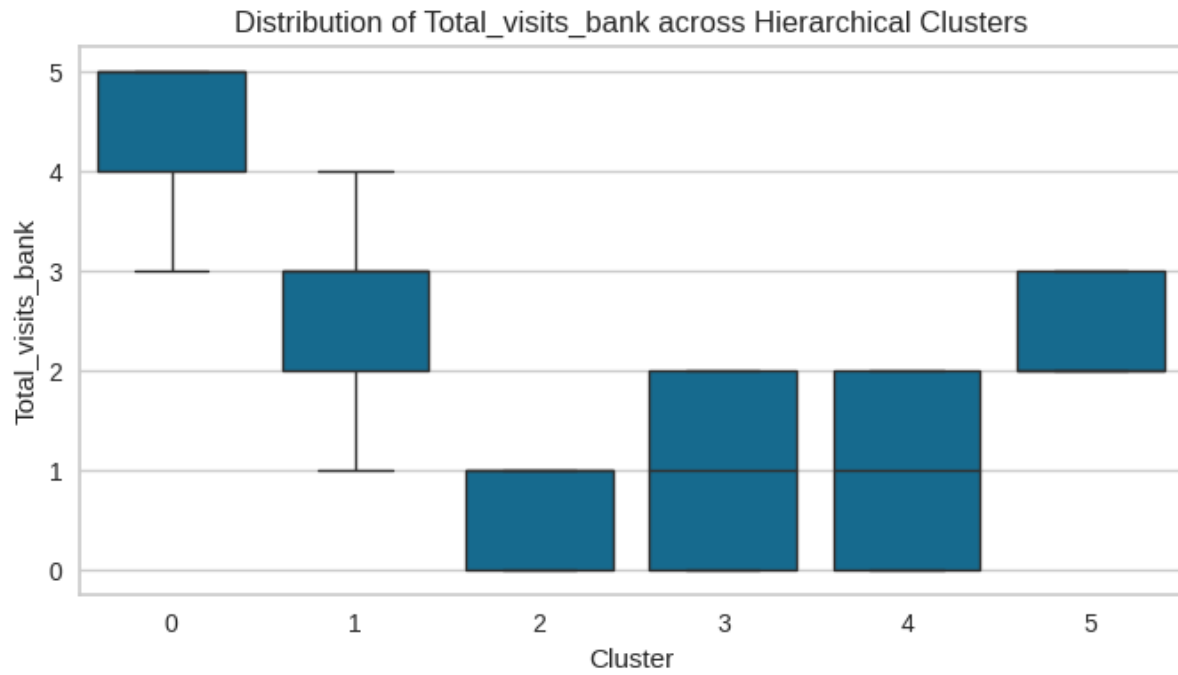
Table no 9 to show Hierarchical cluster profiling summary



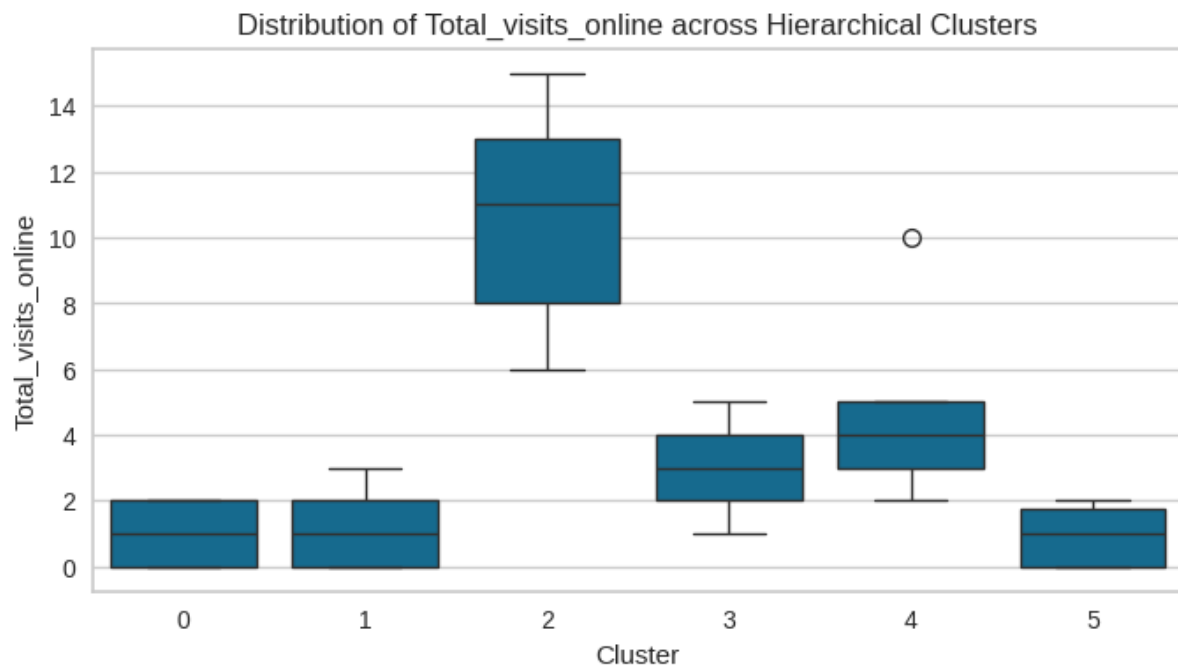
Above is plots figure no. 29 to show Distribution of Avg\_Credit\_Limit across Hierarchical clusters



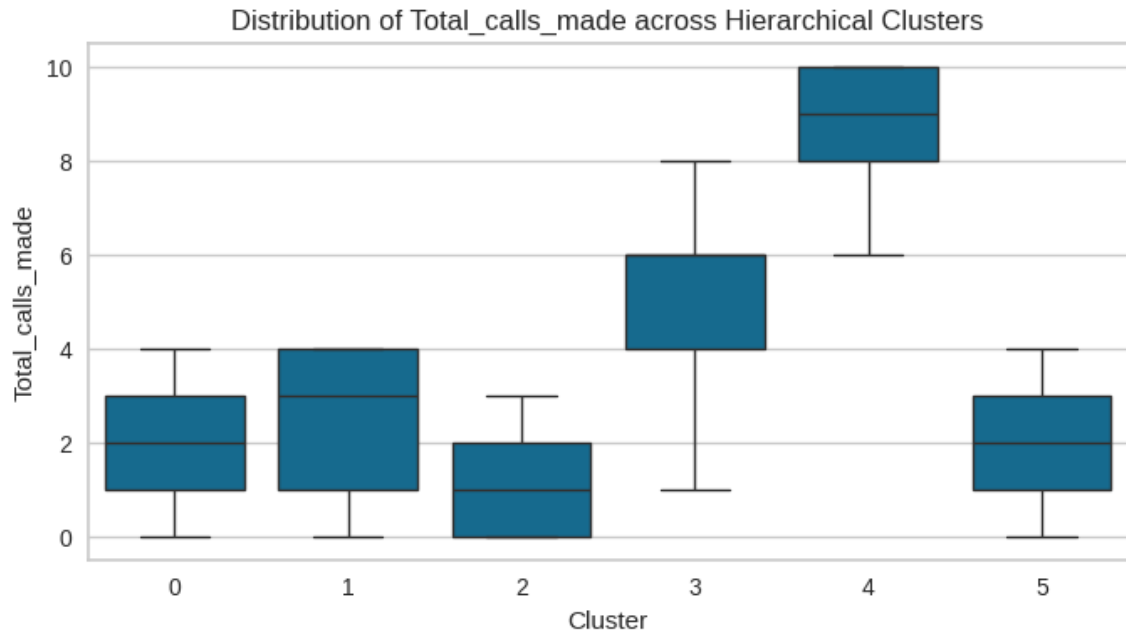
Above is plots figure no. 30 to show Distribution of Total\_Credit\_cards across Hierarchical clusters



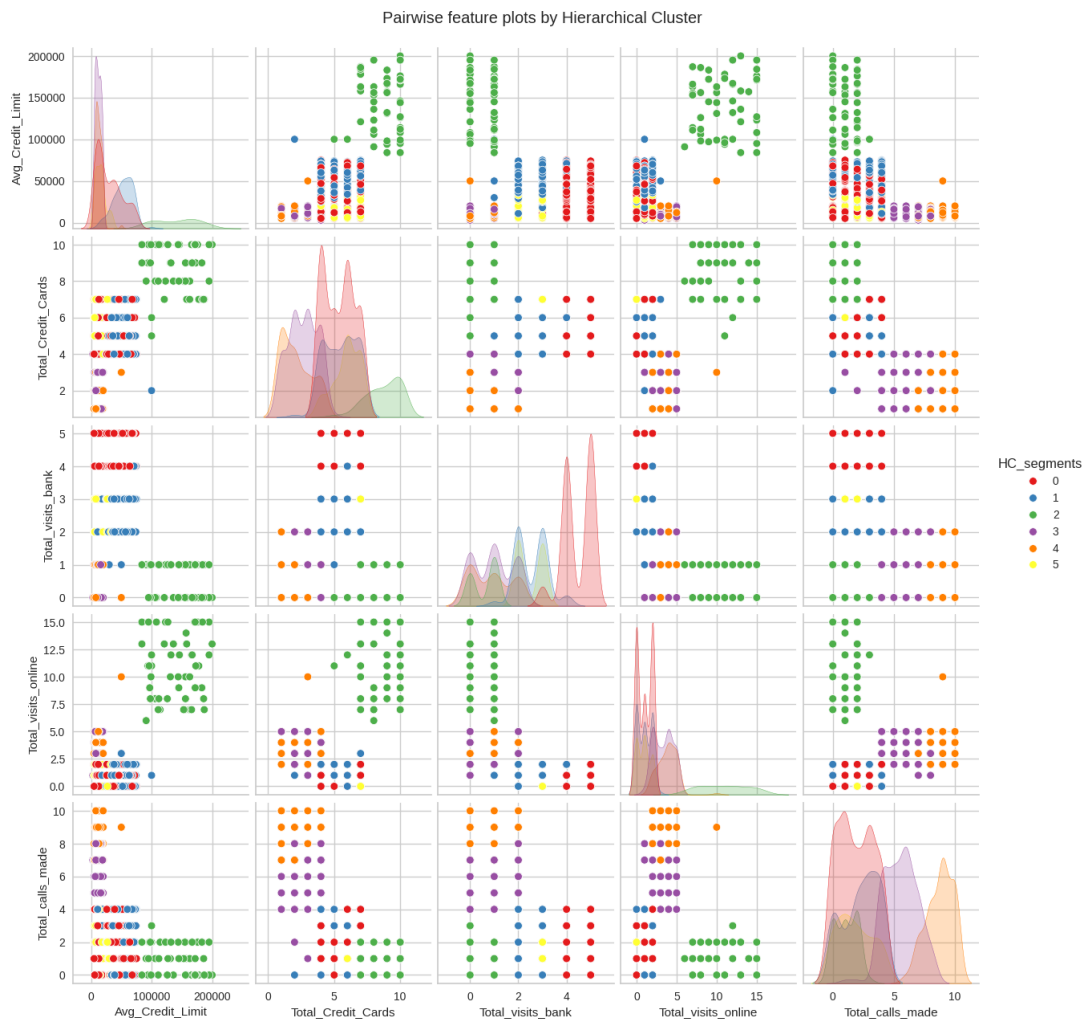
Above is plots figure no. 31 to show Distribution of Total\_visits\_bank across Hierarchical clusters



Above is plots figure no. 32 to show Distribution of Total\_visits\_online across Hierarchical clusters



Above is plots figure no. 33 to show Distribution of Total\_calls\_made across Hierarchical clusters



Above is plots figure no. 34 to show pairwise feature plot by Hierarchical clusters

## **Insights**

### **Cluster 0**

Profile:

Average Credit Limit: ~29,474

Total Credit Cards: ~5.37

Total Visits to Bank: ~4.45

Total Visits Online: ~1.01

Total Calls Made: ~1.89

Interpretation:

Moderate credit limit and credit card ownership.

Regular bank visits with some online activity.

Moderate engagement through calls.

Business Recommendations:

Focus on cross-selling banking products like savings accounts or loan offers.

Loyalty programs to increase online engagement.

Personalized communication to deepen the relationship.

### **Cluster 1**

Profile:

Average Credit Limit: ~52,675

Total Credit Cards: ~5.50

Total Visits to Bank: ~2.55

Total Visits Online: ~0.98

Total Calls Made: ~2.28

Interpretation:

Higher credit limit than Cluster 0 but similar credit card count.

Fewer bank visits than Cluster 0, but online visits are steady.

Slightly more calls, indicating active communication.

Business Recommendations:

Target with premium credit offers or credit card upgrades.

Promote online banking features to increase digital engagement.

Use call center outreach for personalized financial advice.

## **Cluster 2**

Profile:

Average Credit Limit: ~141,040 (significantly higher)

Total Credit Cards: 8.74

Total Visits to Bank: 0.6 (very low)

Total Visits Online: 10.9 (very high)

Total Calls Made: 1.08

Interpretation:

High credit limit and credit card ownership.

Minimal physical bank visits but very high online activity.

Likely digital-savvy, possibly high-net-worth clients or business customers.

Business Recommendations:

Offer premium digital services like wealth management, investment products.

Personalized online dashboards or mobile apps.

Consider exclusive online-only offers or services.

## **Cluster 3**

Profile:

Credit Limit: ~11,835 (low)

Credit Cards: 2.63

Visits to Bank: ~0.98

Online Visits: ~3.34

Calls Made: ~5.56

Interpretation:

Lower credit limit and fewer credit cards.

Moderate online activity but high call interactions, possibly indicating need for assistance.

Business Recommendations:

Focus on educating about credit products and financial planning.

Enhance digital onboarding and self-service options.

Use calls to upsell or cross-sell basic banking products.

#### **Cluster 4**

Profile:

Credit Limit: ~12,609

Credit Cards: 2.12

Visits to Bank: ~0.88

Online Visits: ~3.85

Calls Made: ~8.74 (very high)

Interpretation:

Similar to Cluster 3 but with even more calls.

Possibly less engaged or higher maintenance clients.

Business Recommendations:

Use targeted outreach to understand their needs.

Offer financial advisory or personalized solutions.

Improve digital engagement to reduce reliance on calls.

#### **Cluster 5**

Profile:

Credit Limit: ~15,568

Credit Cards: ~5.95

Visits to Bank: ~2.49

Online Visits: ~0.89

Calls Made: ~1.81

Interpretation:

Moderate credit limit, similar to Clusters 0 and 1.

Slightly more bank visits and credit cards.

Lower online activity, indicating preference for physical banking.

Business Recommendations:

Promote digital banking to increase online engagement.

Loyalty programs to strengthen relationship.

Personalized offers for credit card upgrades or savings.

Overall Insights & Business Strategy:

High-Value Clients (Cluster 2): Focus on premium digital services, wealth management, and exclusive online offers.

Moderate Clients (Clusters 0, 1, 5): Cross-sell credit products and promote digital banking.

Lower Engagement Clusters (3, 4): Use proactive calls, educational campaigns, and digital onboarding

## **Cluster Segregation Aligned with Standard Models**

### **1. Elite Value Clients**

Clusters:

Cluster 2

Profile Highlights:

Very high credit limit (~141,040)

High credit card ownership (~8.74)

Minimal physical visits (0.6) but very high online activity (10.9)

Interpretation:

Likely high-net-worth or high-value clients, digitally engaged, long-term profitable, low physical engagement.

Business Focus:

Premium digital services, wealth management, exclusive online offerings.

### **2. Growth Builders**

Clusters:

Cluster 1

Profile Highlights:

Higher credit limit (~52,675)

Similar credit card count (~5.50)

Moderate visits: Bank (2.55), Online (0.98), Calls (~2.28)

Interpretation:

Rising income, mid-level usage, potential for future growth.

Business Focus:

Cross-sell premium credit products, digital onboarding, financial advisory.

### **3. Everyday Transactors**

Clusters:

Cluster 0

Profile Highlights:

Moderate credit limit (~29,474)

Credit cards (~5.37)

Regular bank visits (4.45), some online activity (1.01), calls (~1.89)

Interpretation:

Moderate engagement, regular transactional behavior, suitable for cross-sell.

Business Focus:

Cross-sell banking products, loyalty programs, digital engagement.

### **4. Risk-Guarded Customers**

Clusters:

Cluster 4

Profile Highlights:

Low credit limit (~12,609)

Credit cards (~2.12)

Very high calls (8.74), low visits (0.88) and online (~3.85)

Interpretation:

Higher maintenance, possibly higher default risk, less engaged.

Business Focus:

Targeted outreach, financial counseling, digital engagement to reduce reliance on calls.

### **5. Dormant or At-Risk Clients**

Clusters:

Cluster 3

Profile Highlights:

Low credit limit (~11,835)

Credit cards (~2.63)

Moderate visits (0.98), higher calls (5.56), online (~3.34)

Interpretation:

Needs support, possibly at risk of churn, moderate engagement.

Business Focus:

Education, digital onboarding, personalized offers.

## 6. Mass Market Essentials

Clusters:

Cluster 5

Profile Highlights:

Moderate credit limit (~15,568)

Credit cards (~5.95)

Visits (2.49), low online (0.89), calls (~1.81)

Interpretation:

Large customer base, low individual profitability, important for volume and cross-sell.

Business Focus:

Digital promotion, loyalty programs, cross-sell initiatives.

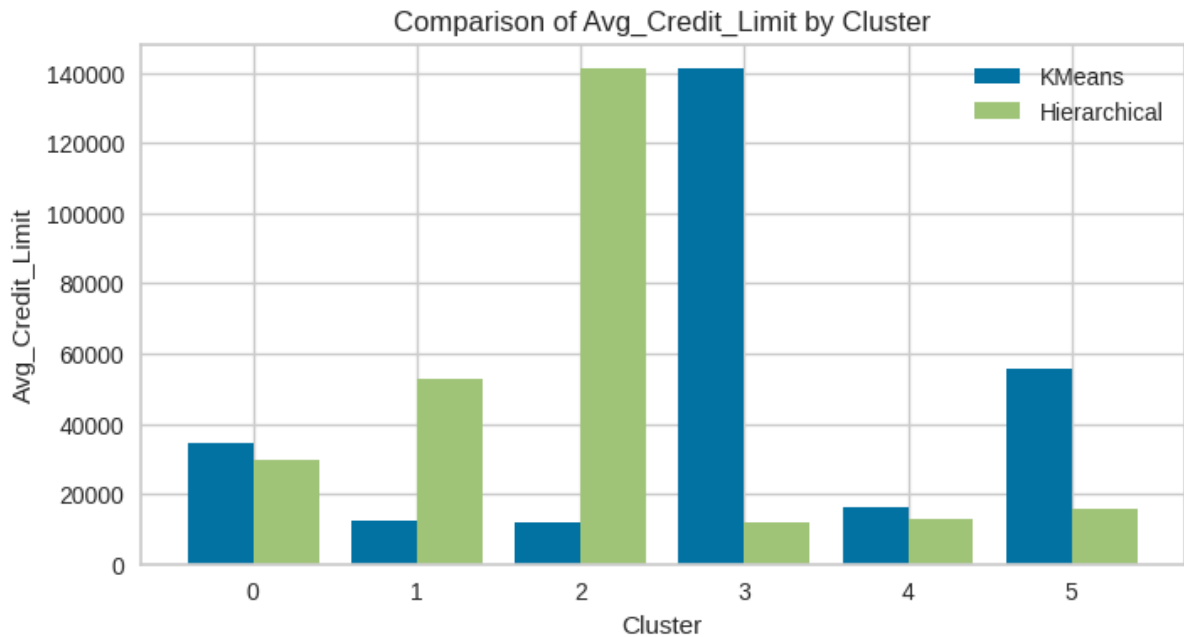
## K-means vs Hierarchical Clustering

I generated below analysis to compare the average profiles of customer segments identified through KMeans and Hierarchical clustering methods. The table below presents the mean values for key variables such as average credit limit, total credit cards, visits to the bank and online, and calls made for each cluster.

Additionally, the correlation coefficients between the profiles derived from the two clustering techniques are summarized. Notably, we observe moderate to strong negative correlations for variables like credit limit and credit cards, indicating some inverse relationships between the cluster profiles identified by the two methods.

	<i>KL_AVG_CREDIT_LIMIT</i>	<i>KL_TOTAL_CREDIT_CARD</i>	<i>KL_TOTAL_INSTS_BANK</i>	<i>KL_TOTAL_INSTS_ONLINE</i>	<i>KL_TOTAL_CALLS_MADE</i>	<i>HC_AVG_CREDIT_LIMIT</i>	<i>HC_TOTAL_CREDIT_CARD</i>	<i>HC_TOTAL_INSTS_BANK</i>	<i>HC_TOTAL_INSTS_ONLINE</i>	<i>HC_TOTAL_CALLS_MADE</i>
<i>CLUSTER</i>										
0	34418.478	5.483696	2.461957	0.956522	2.043478	29474.24	5.365979	4.448454	1.010309	1.891753
1	12401.78	2.214286	0.910714	3.669643	8.535714	52675.21	5.495726	2.547009	0.982906	2.282051
2	11946.42	2.607143	0.955357	3.437500	5.205357	141040.00	8.740000	0.600000	10.900000	1.080000
3	141040.00	8.740000	0.600000	10.900000	1.080000	11834.59	2.631579	0.977444	3.338346	5.556391
4	16448.27	5.250000	4.543103	1.103448	1.931034	12608.70	2.119565	0.880435	3.847826	8.739130
5	55802.32	5.941860	4.267442	0.872093	2.000000	15567.58	5.945946	2.486486	0.891892	1.810811

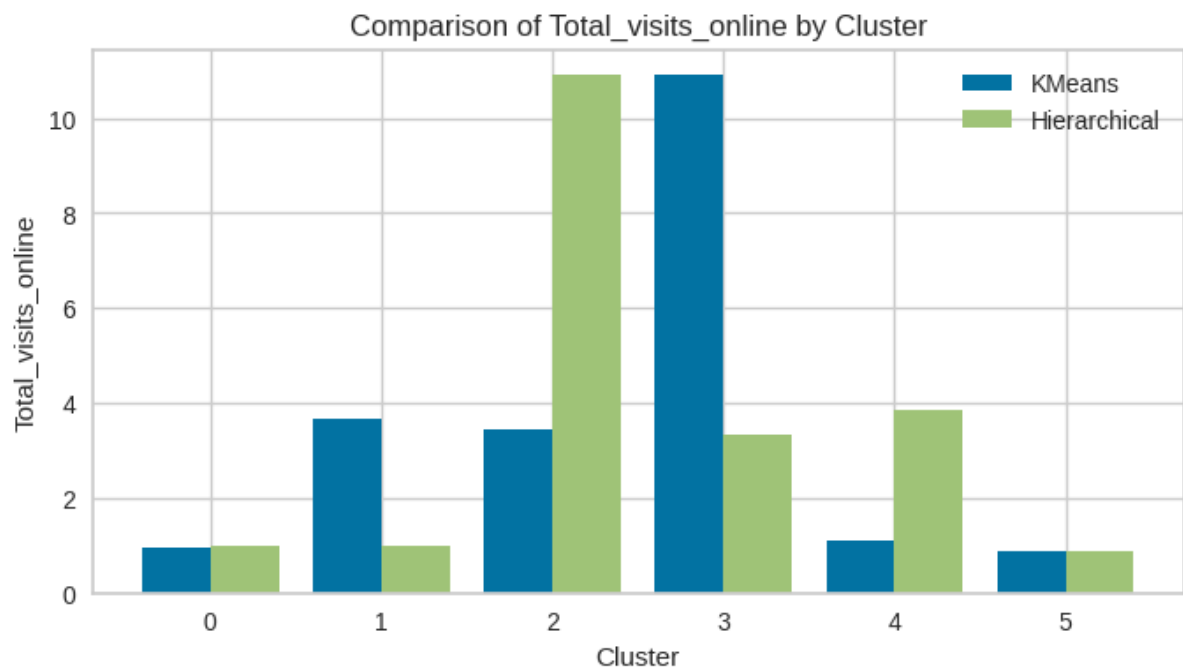
Table no 10 list of mean value of features to compares the average profiles of customer segments



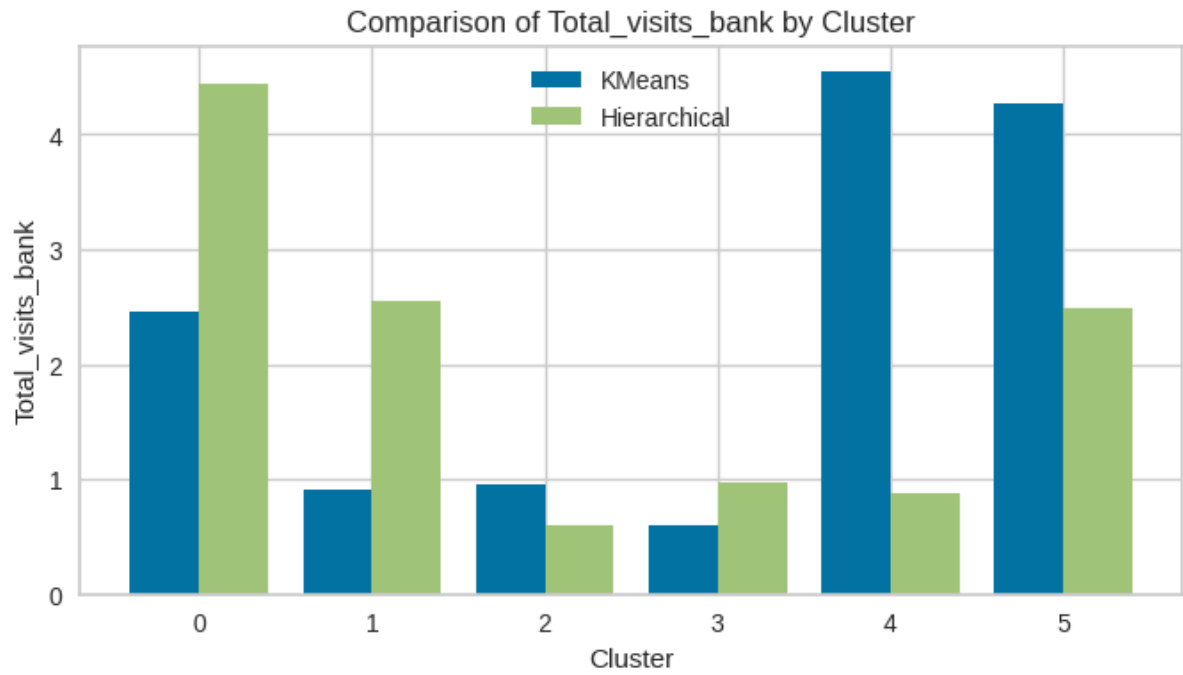
Above is plots figure no. 35 to compare Avg\_Credit\_Limit by Hierarchical clustering and KMeans clustering



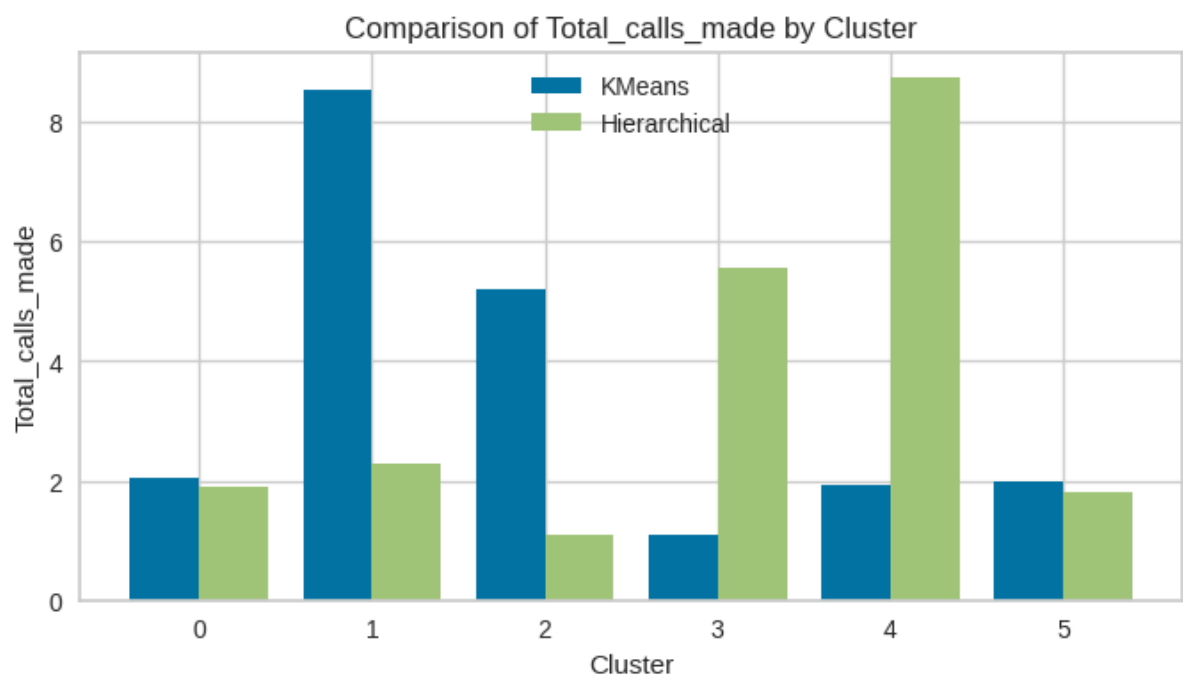
Above is plots figure no. 36 to compare Total\_Credit\_Cards by Hierarchical clustering and KMeans clustering



Above is plots figure no. 37 to compare Total\_visits\_Online by Hierarchical clustering and KMeans clustering



Above is plots figure no. 38 to compare Total\_visits\_Bank by Hierarchical clustering and KMeans clustering



Above is plots figure no. 39 to compare Total\_calls\_made by Hierarchical clustering and KMeans clustering

Correlation between K-means and Hierarchical cluster profiles:

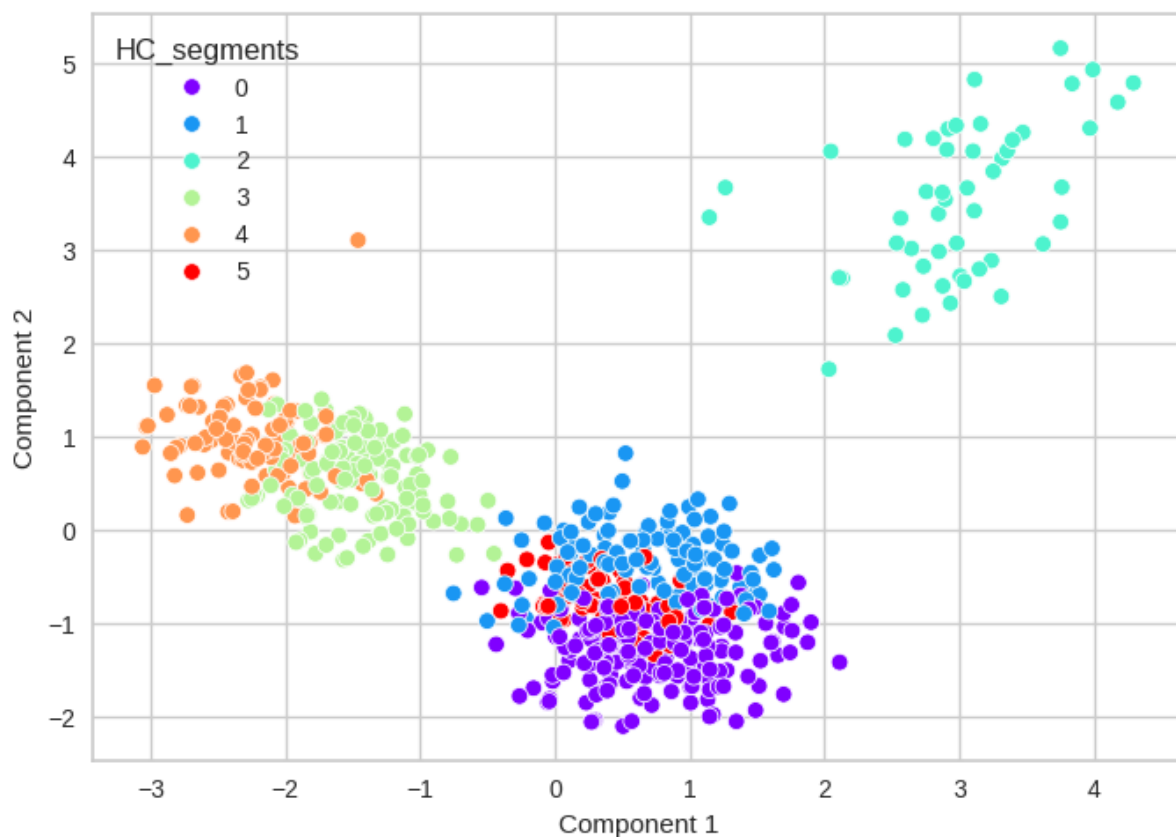
```
Avg_Credit_Limit: -0.47
Total_Credit_Cards: -0.65
Total_visits_bank: 0.13
Total_visits_online: 0.14
Total_calls_made: -0.44
```

## Principal component analysis (PCA)for Visualization

Let's see result of using PCA to reduce the data to two dimensions and visualize it to see how well-separated the clusters are.

However, applying **PCA to the data will not be a feasible way since the given data for this project is small.**

The first two principal components explain **83.16%** of the variance in the data



Above is plots figure no. 40 to show datapoints and visualize 2 main component in dataset

# Actionable Insights Business Recommendations

## Summary of Findings and Recommendations

behaviors and interactions with the bank. While the exact cluster assignments differ between the two methods, the general patterns observed in the customer segments are quite similar, particularly in identifying high-value, digitally engaged, and branch-dependent customer groups.

Here's a summary of the key segments and recommendations:

### Key Customer Segments and Recommendations

#### 1. High-Value, Digitally Engaged Customers:

Characteristics: High average credit limits, high total credit cards, very low bank visits, very high online visits, and low calls made.

Insights: These customers are likely affluent and prefer digital channels for their banking needs.

Recommendations:

Offer premium digital banking services, including wealth management and investment options.

Provide personalized online dashboards and exclusive digital-only offers.

Consider targeted invitations to exclusive online events or webinars.

#### 2. Moderate Credit, Digitally Active Customers:

Characteristics: Moderate credit limits and credit card ownership, low bank visits, moderate to high online visits, and moderate to high calls made.

Insights: These customers are comfortable with digital platforms but may still require some assistance via phone calls.

Recommendations:

Promote digital banking features and self-service options to reduce reliance on calls.

Offer targeted credit card upgrades and financial products suitable for their credit profile.

Enhance the online customer service experience.

### 3. Moderate Credit, Branch-Dependent Customers:

Characteristics: Moderate credit limits and credit card ownership, high bank visits, low online visits, and moderate calls made.

Insights: These customers prefer in-person interactions at the bank branch.

Recommendations:

Encourage digital engagement through targeted marketing and incentives.

Offer personalized financial advisory services during branch visits.

Develop loyalty programs that reward both online and in-branch activity.

### 4. Lower Credit, High-Call Customers:

Characteristics: Lower credit limits and fewer credit cards, low bank visits, moderate online visits, and very high calls made.

Insights: These customers may require more assistance and support, potentially due to limited financial literacy or issues with services.

Recommendations:

Focus on educating these customers about credit products and financial planning.

Enhance digital onboarding and self-service options to address common queries.

Use proactive calls to understand their needs and offer personalized solutions.

## **Strategic Recommendations**

**Personalization:** Tailor marketing campaigns, product offerings, and service delivery based on the identified customer segments.

**Digital Transformation:** Invest in improving the online banking experience and promoting digital adoption across all segments.

**Service Enhancement:** For segments with high call volumes or branch visits, identify the root causes of their interactions and streamline processes to improve efficiency and satisfaction.

**Cross-selling and Upselling:** Utilize the segment profiles to identify opportunities for cross-selling additional banking products and upselling to premium services.

**Customer Retention:** Implement targeted loyalty programs and engagement strategies to foster long-term relationships with valuable customer segments.

Below are actionable insights and recommendations aligned with widely recognized banking segmentation principles for each of the six clusters using hierarchical clustering, using common industry models such as **Pareto, Lifecycle, Behavioral, Risk-Based, and Profitability segmentation**:

### **1. Elite Value Clients (High Balance, High Profitability)**

Cluster: Cluster 2

Actionable Insights:

- These clients are top-tier, contribute the majority of profits, and are highly digitally engaged.
- They prefer personalized, exclusive services and wealth management solutions.

#### Recommendations:

- Offer Premium Services: Provide tailored wealth management, estate planning, and investment advisory.
- Exclusive Digital Access: Develop VIP digital dashboards, early access to new products, and personalized communication.
- Loyalty & Retention Programs: Implement loyalty schemes with exclusive benefits to reinforce long-term relationships.
- Proactive Engagement: Assign dedicated Relationship Managers for high-touch service and regular check-ins.
- Event Invitations: Host exclusive events to deepen engagement and reinforce loyalty.

## **2. Growth Builders (Rising Income, Future Potential)**

Cluster: Cluster 1

#### Actionable Insights:

- These clients are on an upward trajectory, with increasing income and potential for higher profitability.
- They may be transitioning from early career to mid-career financial needs.

#### Recommendations:

- Lifecycle-Based Cross-Selling: Offer tailored mortgage, personal loan, or premium credit card products aligned with their growth phase.
- Financial Education: Provide targeted financial planning and investment literacy programs.
- Digital Onboarding & Upsell: Promote digital account opening, mobile banking, and investment platforms.
- Progressive Rewards: Introduce tiered loyalty rewards that incentivize increased engagement.
- Regular Check-ins: Use data-driven outreach to understand their evolving needs and offer relevant solutions.

### **3. Everyday Transactors (Frequent Small-Value Transactions)**

Cluster: Cluster 0

Actionable Insights:

- These clients are active and transactional but not highly profitable yet.
- They are ideal candidates for cross-sell and digital adoption.

Recommendations:

- Enhance Digital Engagement: Promote online banking, bill pay, and mobile app features to increase convenience.
- Cross-Sell Basic Products: Offer savings accounts, low-cost loans, or insurance products aligned with their transaction habits.
- Loyalty & Rewards: Implement cashback, discounts, or reward points to encourage continued activity.
- Automate Personalized Offers: Use transaction data to suggest relevant products and services.
- Customer Education: Provide tutorials and assistance to maximize digital platform usage.

### **4. Risk-Guarded Customers (Higher Default Risk)**

Cluster: Cluster 4

Actionable Insights:

- Clients may have financial challenges requiring careful risk management.
- They require close monitoring and tailored interventions.

Recommendations:

- Risk Monitoring & Early Warning: Implement regular credit scoring and transaction analysis to detect early signs of distress.

- Targeted Financial Counseling: Offer financial health checks and debt management advice.
- Adjust Credit Limits & Terms: Consider restructuring existing credit lines to mitigate risk.
- Digital Engagement: Encourage digital banking to improve engagement while reducing operational costs.
- Personalized Support: Develop proactive outreach programs to understand their needs and offer tailored solutions.

## **5. Dormant or At-Risk Clients (Churn-Prone)**

Cluster: Cluster 3

Actionable Insights:

- These clients show declining engagement and may be at risk of attrition.
- They need re-engagement strategies.

Recommendations:

- Re-Engagement Campaigns: Use targeted offers, personalized messages, or incentives to rekindle activity.
- Simplify Digital Access: Assist with onboarding or digital tutorials to remove barriers.
- Feedback & Needs Assessment: Conduct surveys or direct outreach to understand barriers.
- Exclusive Offers: Provide special promotions or loyalty bonuses to incentivize return.
- Monitor Behavior: Use recency, frequency, and monetary (RFM) metrics to identify early signs of disengagement.

## **6. Mass Market Essentials (Volume-Driven Segment)**

Cluster: Cluster 5

Actionable Insights:

- Large customer base with low individual profitability but significant for volume.
- Opportunities exist for cross-sell and digital expansion.

#### Recommendations:

- Digital Migration: Promote mobile banking, quick onboarding, and self-service options.
- Cost-Efficient Service Models: Focus on automated, digital channels to serve this segment cost-effectively.
- Loyalty & Rewards: Implement point-based or cashback programs to build loyalty.
- Targeted Cross-Sell Campaigns: Offer bundled products like savings, insurance, or micro-loans.
- Segmented Marketing: Use data-driven segmentation to personalize communication and offers.

### **Summary**

By implementing these recommendations, industry-standard segmentation principles, the bank can prioritize resource allocation, personalize offerings, and optimize customer lifetime value across all segments—enhancing profitability, loyalty, and risk management, AllLife Bank can better understand and serve its credit card customer base, leading to improved market penetration, enhanced customer satisfaction, and increased profitability.