Unsupervised Learning - Credit Card Customer Segmentation

Coded Project Report

AllLife Bank

Submitted by – Yash Juneja

Submitted to – Great Learning

Table of Contents

Context......3
Objective......3
Data Description......3
Data Dictionary......3
Exploratory Data Analysis. 4

- Univariate Analysis......4-10
- Bivariate Analysis......10-11

Data Preprocessing......11

- Outlier Detection......11
- Checking for Duplicate Values......11
- Checking for Missing Values......12
- Scaling......12

K-means Clustering......12

- Checking Elbow Plot......12-13
- Let's check the silhouette scores......13-14
- Creating Final Model......14
- Cluster Profiling......14-16

Hierarchical Clustering......16

- Computing Cophenetic Correlation......16
- Checking Dendrograms......17
- Creating model using sklearn......17
- Cluster Profiling......17-19

K-means vs Hierarchical Clustering......19-20
Actionable Insights and Recommendations......21-23

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. 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).

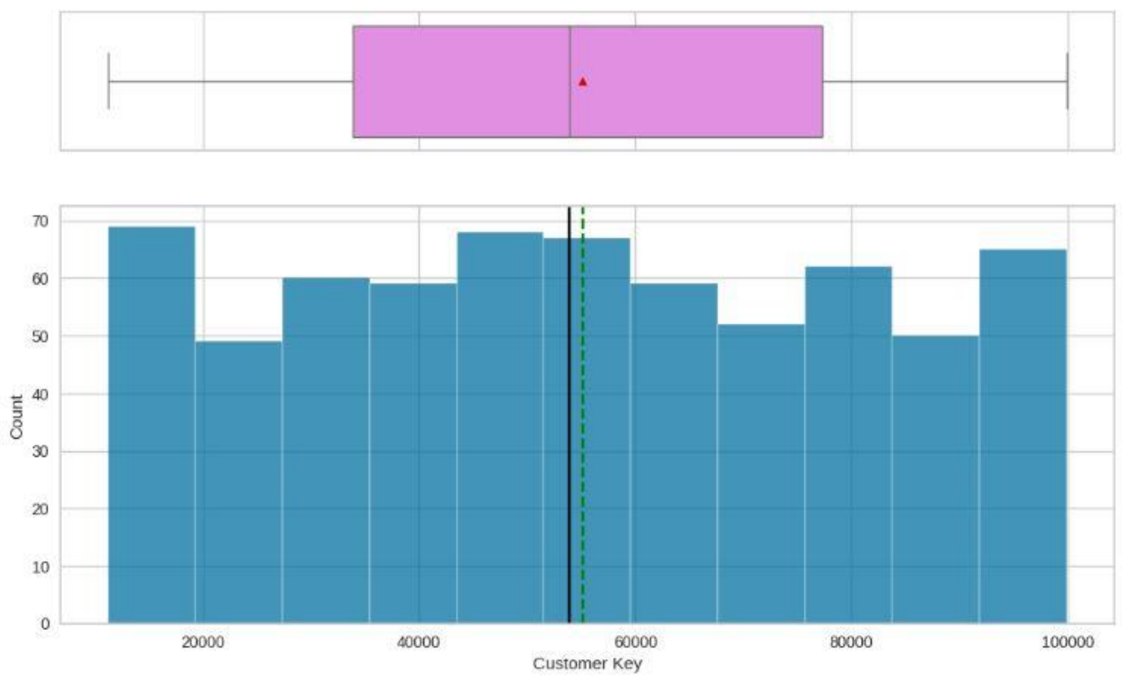
Data Dictionary

- Sl_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)

Exploratory Data Analysis

Univariate analysis

Observation on Customer Key



Even Customer Distribution Indicates a Well-Sampled Dataset

- No specific range of customer keys dominates the dataset, suggesting **no bias** in how customers are assigned.

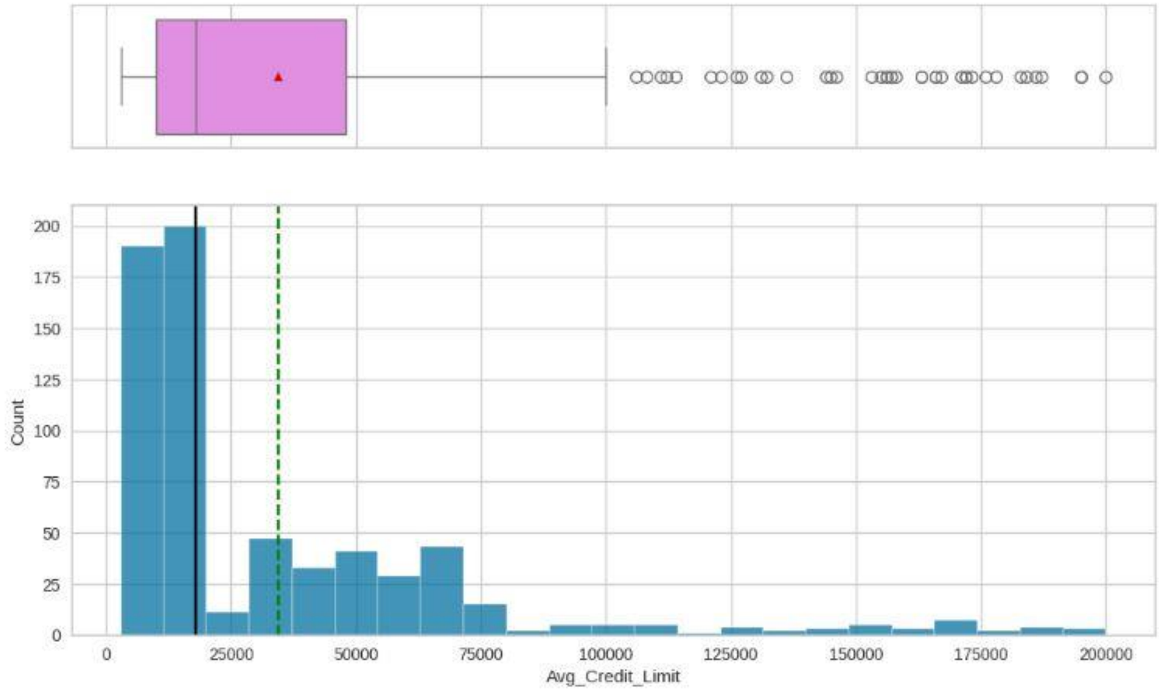
No Need for Special Handling of Customer Keys in Analysis

- Since customer keys appear evenly distributed, they **do not inherently affect business insights** but can be used as unique identifiers.

Potential Use in Customer Segmentation

- If customer keys are assigned based on **sign-up order or geographic region**, further analysis could reveal insights into **customer acquisition trends or regional behavior**.

Observation on Average Credit Limit



Majority of Customers Have Low Credit Limits:

- Most customers have an average credit limit below 25,000.
- The company could introduce targeted promotions or loyalty programs to encourage increased spending from this segment.

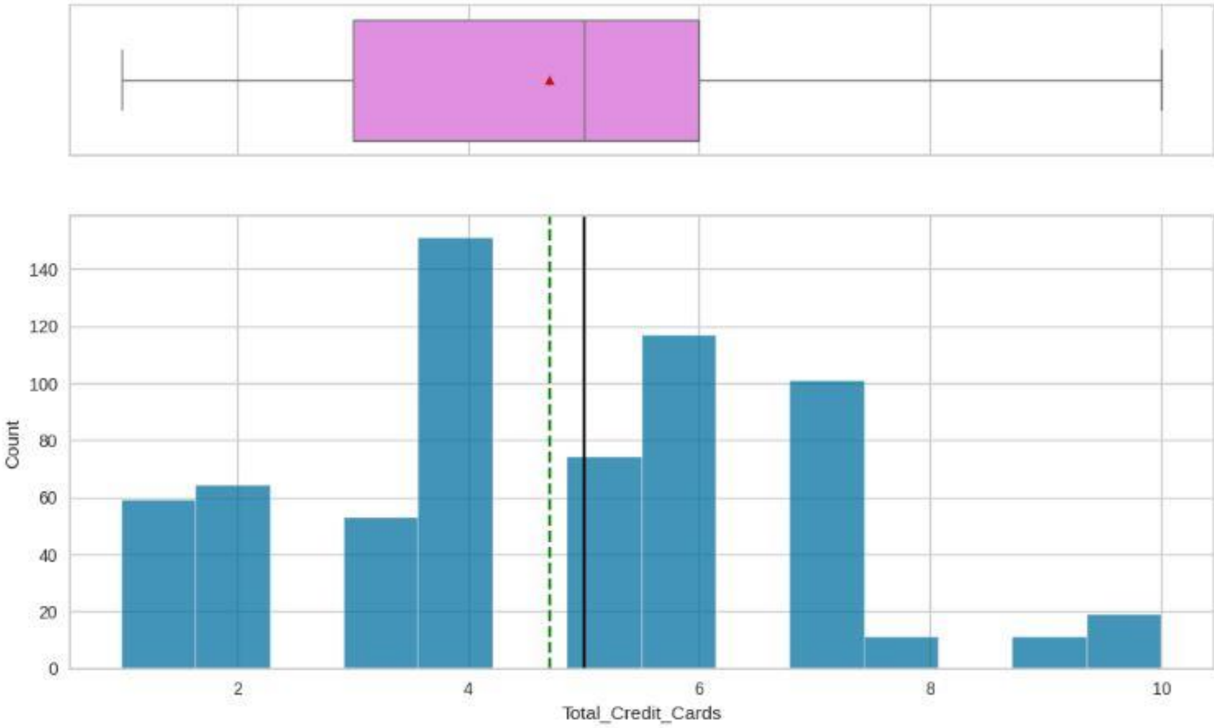
High-Value Outliers Exist:

- A small percentage of customers have very high credit limits (above 100,000).
- These customers should be given exclusive premium offers, personalized financial products, or better credit terms to maintain engagement.

Skewed Distribution Suggests an Opportunity for Credit Growth:

- A gradual credit limit increase for mid-tier customers could help drive revenue growth.
- Implement risk-based credit expansion strategies to balance growth with financial risk.

Observation on Total Credit Cards



Majority of Customers Hold 4-6 Credit Cards:

- This suggests that customers prefer moderate diversification of credit sources.
- The business can offer incentives to consolidate spending on their cards, such as higher cashback, lower interest rates, or reward points for exclusive usage.

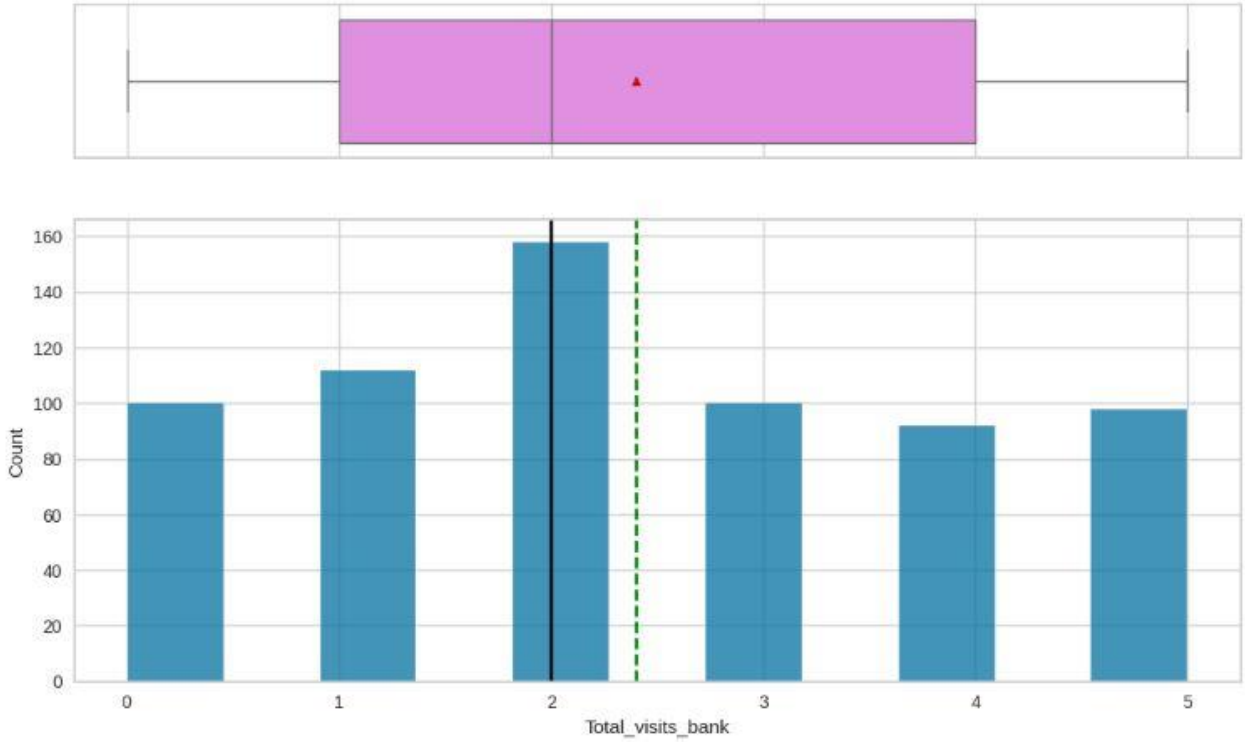
Opportunity to Convert Low-Card Holders (1-2 cards) into Multi-Card Users:

- A significant segment of customers owns fewer than 3 credit cards.
- The company can target them with tailored offers for additional cards, highlighting benefits like higher limits, rewards, or travel perks.

Premium Services for High-Card Holders (7+ Cards):

- Customers with 7+ cards are likely high spenders or frequent credit users.
- Offering them exclusive credit management tools, premium concierge services, or personalized financial products can increase brand loyalty.

Observation on Total Bank Visits



Most customers visit the bank only a few times (0-5 visits), with a peak around 2 visits.

- This suggests that a large portion of customers prefer online or mobile banking over physical visits.

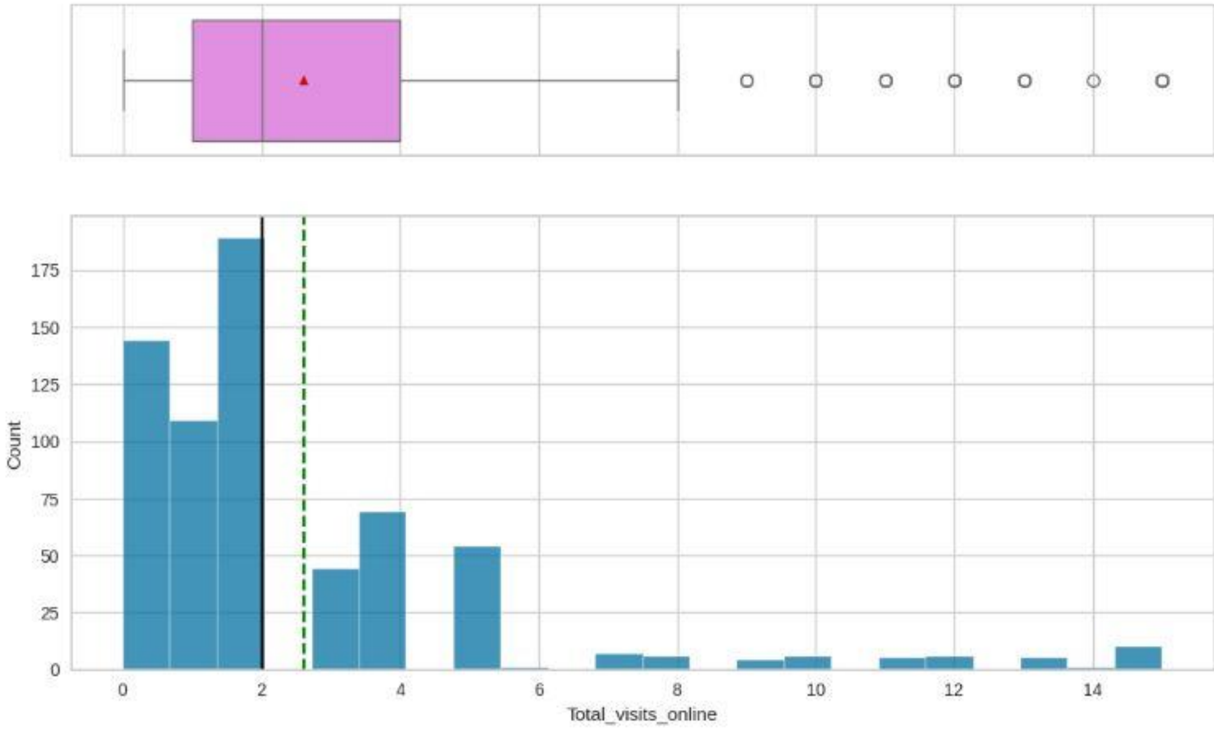
Opportunity to Reduce In-Branch Visits Further

- If reducing in-person visits is a goal, banks can enhance digital banking services (e.g., chatbots, AI-driven support, and seamless mobile banking features).

Identify and Target High-Visit Customers for Better Engagement

- Customers who visit frequently may require additional assistance or prefer personal service. The bank can:
  - Offer premium in-branch services for these customers.
  - Promote digital self-service tools to reduce dependency on branch visits.

Observation on Total Online Visitors



Most customers have very few online visits, indicating low digital engagement.

- The bank should encourage digital banking usage by offering incentives, easier navigation, or personalized promotions for online users.

A small group of customers uses online banking very frequently.

- These digitally active customers could be targeted for premium digital banking services, such as personalized AI-driven recommendations, financial planning tools, or priority support.

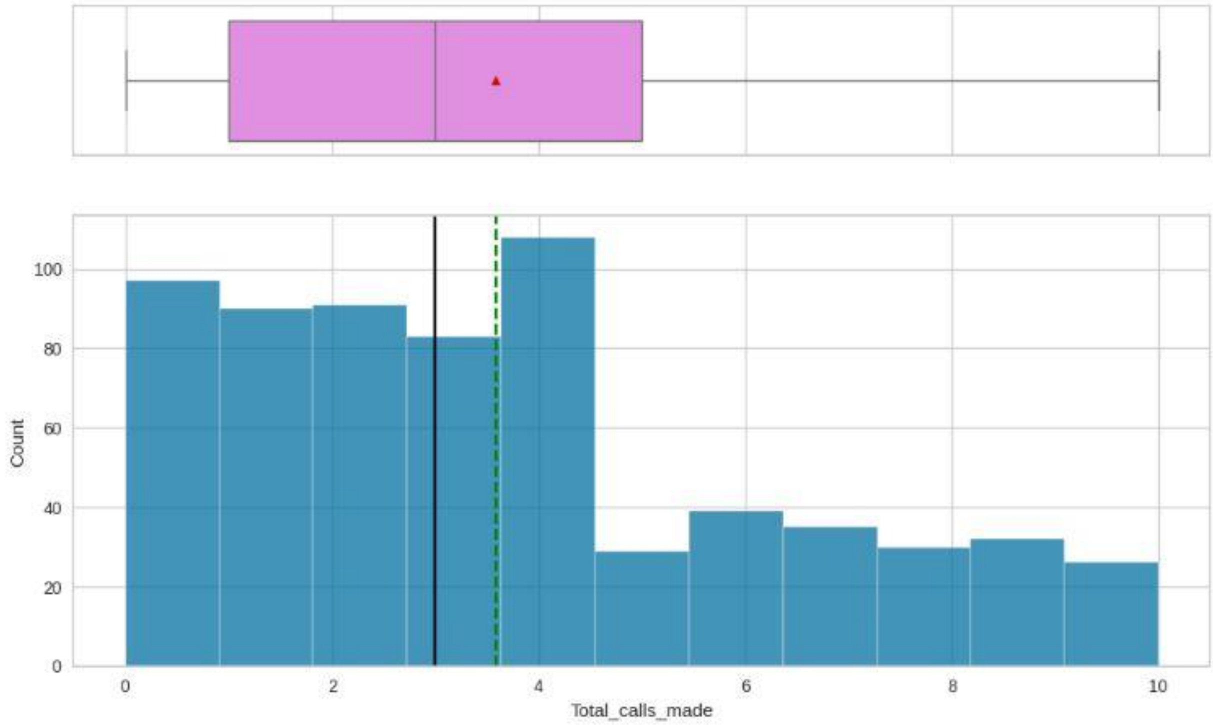
Digital Adoption Campaigns Needed for Less Active Customers

- Identify and educate customers who rarely use online banking (0-2 visits).
- Provide tutorials, webinars, or special offers for customers who transition to digital banking.

Monitor High-Visit Customers for Potential Issues

- Customers visiting online banking more than 10 times could be facing transaction failures, confusion, or seeking additional features.
- Implement customer feedback loops to identify areas for improvement in the online banking experience.

Observation on Total Calls Made



Most customers make moderate calls, suggesting a balanced call support usage.

- Customers rely on customer service but are not overusing it, indicating that services are fairly efficient.
- However, it's worth analyzing repeat callers to identify common concerns or complaints.

A smaller group of customers makes a high number of calls (6-10 calls).

- These might be high-maintenance customers requiring more personalized support.
- The bank should analyze call reasons and consider self-service options (chatbots, FAQs, app-based support) to reduce reliance on call centers.

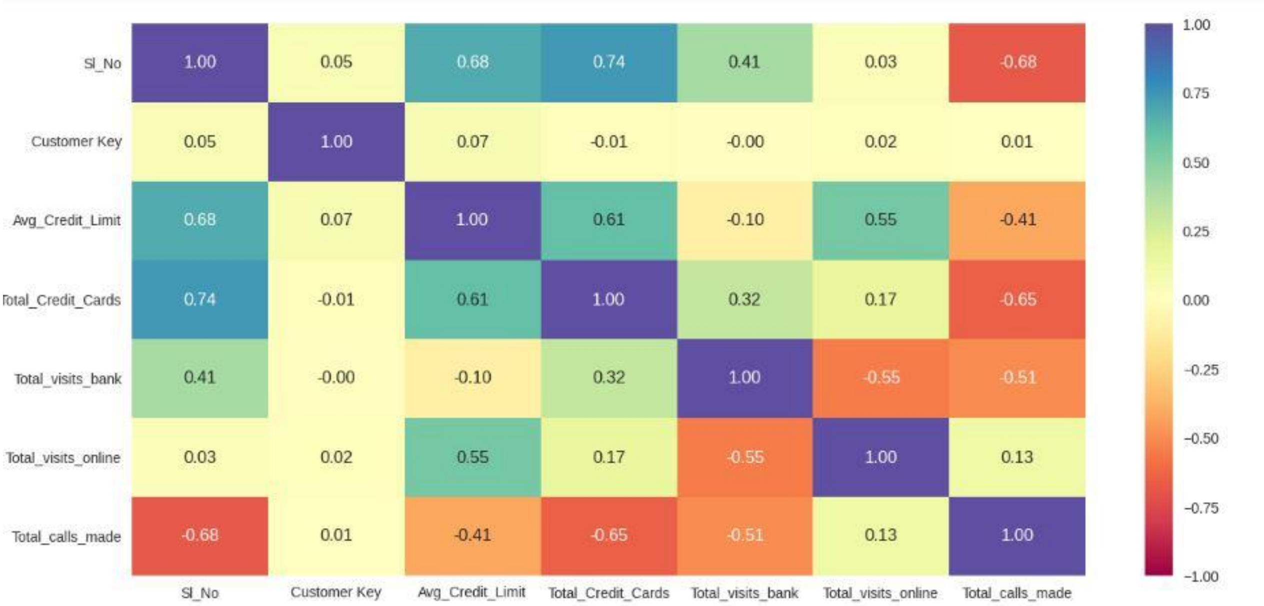
Opportunity for Call Reduction Through Digital Solutions

- Since many calls could be related to common queries (e.g., balance checks, card issues, transactions), promoting online banking, chat support, and automated helplines could reduce operational costs.
- Customers making frequent calls (6+) should be nudged towards self-service portals through awareness campaigns and incentives.

Identify Potential Customer Satisfaction Gaps

- If certain segments are making more calls than others, it may indicate product confusion, transaction difficulties, or dissatisfaction.
- Conduct customer feedback surveys to identify areas for service improvement.

Bivariate analysis



Strong Positive Correlations (+)

- Avg_Credit_Limit & Total_Credit_Cards (0.61):
    - Customers with more credit cards tend to have a higher average credit limit.
    - This suggests that banks offer higher credit limits to customers with multiple cards, likely based on spending and creditworthiness.
- Total_Credit_Cards & Sl_No (0.74):
    - A higher correlation indicates that credit card ownership increases with the number of transactions or services availed.
- Avg_Credit_Limit & Total_Visits_Online (0.55):
    - Customers with a higher credit limit are more likely to use online services.
    - This could indicate that digitally active users are given better financial privileges.

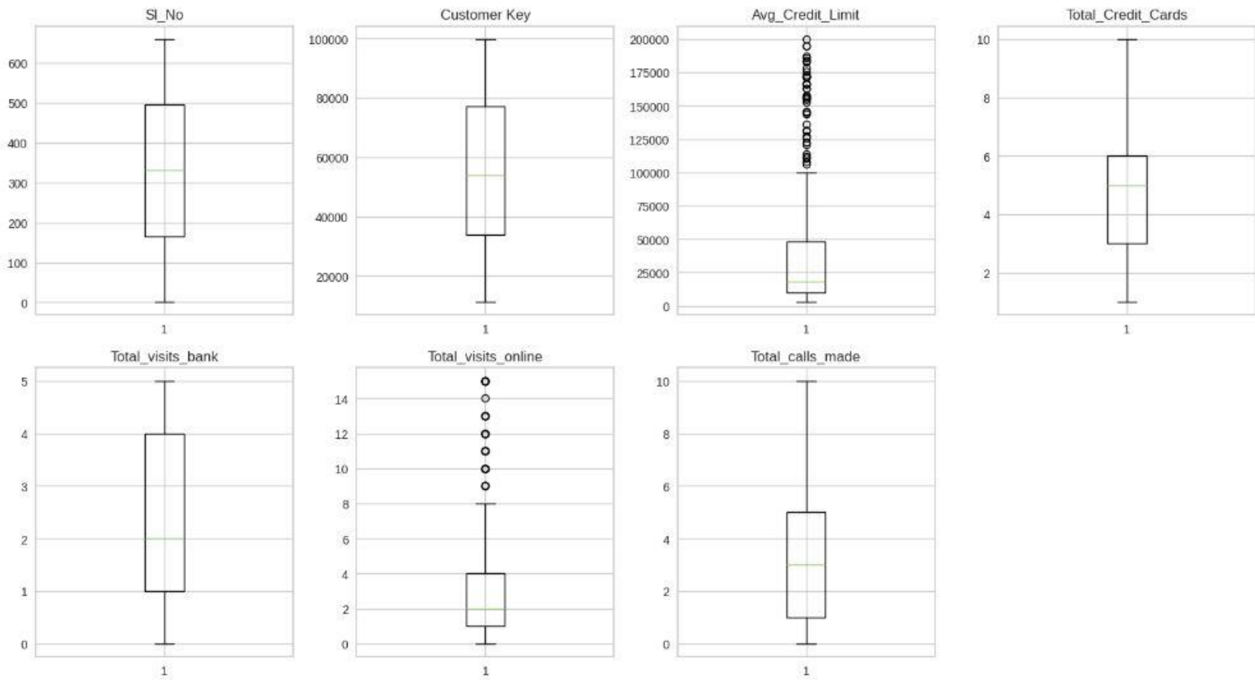
Strong Negative Correlations (-)

- Total_Calls_Made & Total_Credit_Cards (-0.65):
    - Customers with more credit cards make fewer calls to customer service.

    - This could suggest that experienced customers (with multiple cards) are more familiar with banking services and require less assistance.
- Total_Calls_Made & Avg_Credit_Limit (-0.41):
    - Customers with higher credit limits tend to make fewer calls.
    - This suggests premium customers rely more on digital or self-service options rather than calling support.
- Total_Calls_Made & Total_Visits_Bank (-0.51):
    - A negative correlation means that people who visit the bank frequently tend to make fewer calls.
    - This could be because their queries are resolved in person rather than over calls.
- Total_Visits_Online & Total_Visits_Bank (-0.55):
    - Customers who prefer online banking visit the bank less frequently.
    - This supports the idea that digital banking reduces in-branch visits.

Data Preprocessing

Outlier Check



Duplicate Values

```
np.int64(0)
```

No Duplicate Values Found

Missing Values

	0
Sl_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

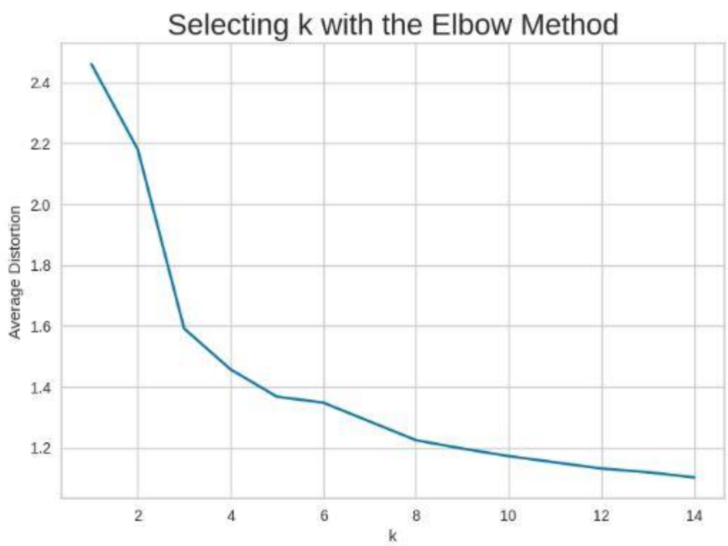
No Missing Values found

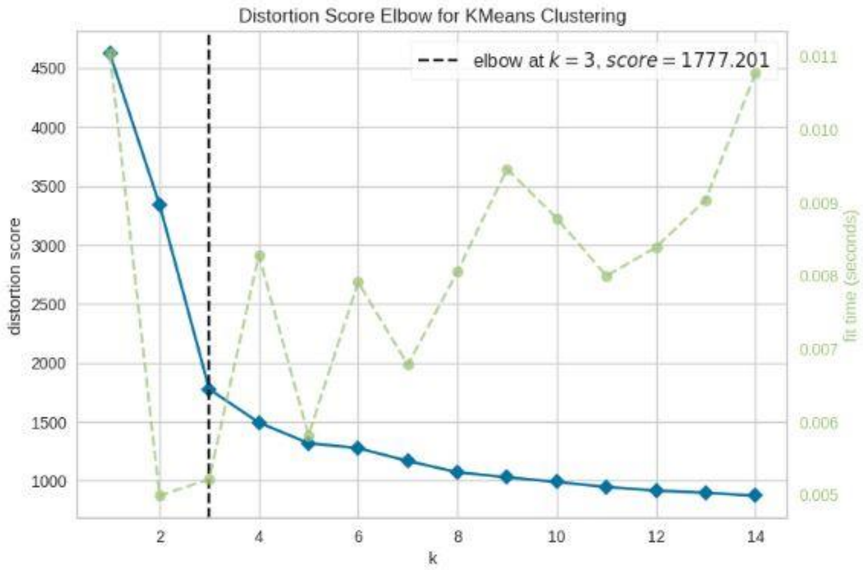
Scaling

	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made
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

K-means Clustering

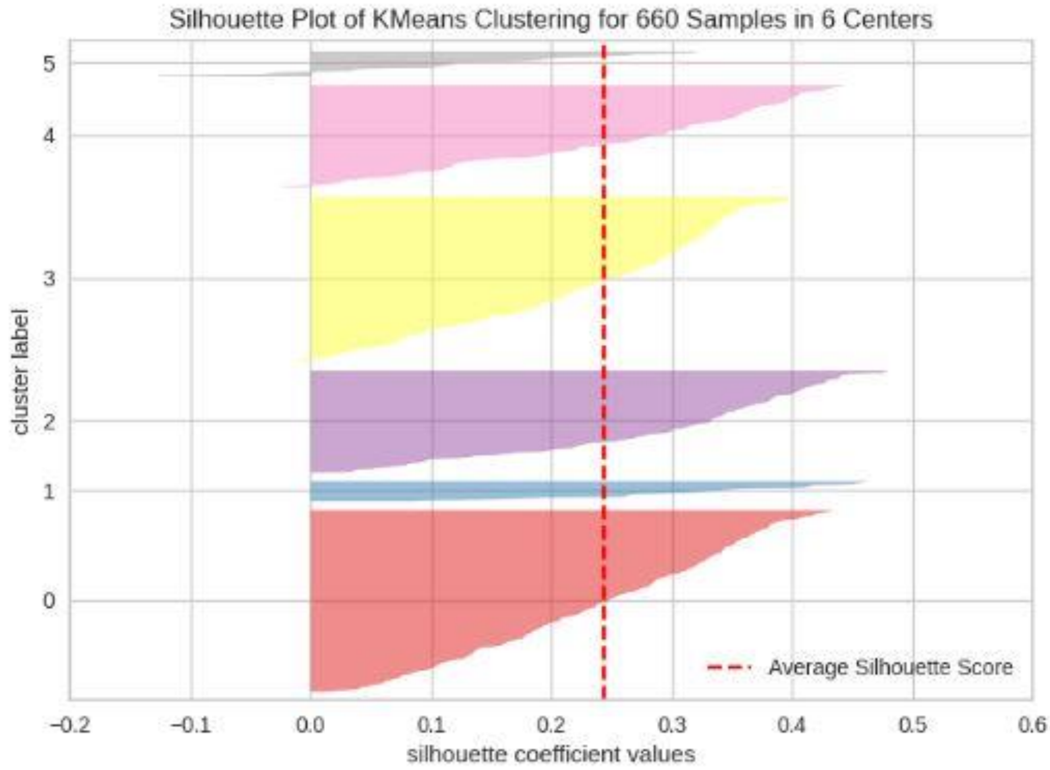
Checking Elbow Plot





Checking Silhouette Score

```
For n_clusters = 2, the silhouette score is 0.49522120315704116)
For n_clusters = 3, the silhouette score is 0.44394092525170764)
For n_clusters = 4, the silhouette score is 0.32322590765742937)
For n_clusters = 5, the silhouette score is 0.2665350914306843)
For n_clusters = 6, the silhouette score is 0.2440686331268522)
For n_clusters = 7, the silhouette score is 0.22419997124871827)
For n_clusters = 8, the silhouette score is 0.23024008911301988)
For n_clusters = 9, the silhouette score is 0.22137594166676244)
For n_clusters = 10, the silhouette score is 0.2221625311589056)
For n_clusters = 11, the silhouette score is 0.20322094900038744)
For n_clusters = 12, the silhouette score is 0.1855094474926542)
For n_clusters = 13, the silhouette score is 0.17705989029974387)
For n_clusters = 14, the silhouette score is 0.17302345624214485)
```



Creating Final model

```
KMeans
KMeans(n_clusters=6, random_state=1)
```

Cluster Profiling

KM_segments	Sl_No	Customer Key	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	398.815920	34427.875622	29273.631841	5.452736	3.502488	1.009950	2.000000	201
1	639.363636	43933.363636	142363.636364	8.727273	0.636364	8.272727	1.000000	22
2	113.008850	32825.415929	12761.061947	2.716814	0.911504	3.654867	6.752212	113
3	444.316940	77548.830601	38158.469945	5.590164	3.508197	0.945355	2.000000	183
4	117.911504	77320.389381	12902.654867	2.150442	0.946903	3.415929	6.902655	113
5	589.214286	66746.571429	140000.000000	8.750000	0.571429	12.964286	1.142857	28

Insights

1. **Segment 0 (Largest Group - 201 customers)**
    - Average Credit Limit: Moderate (29,273)
    - Total Credit Cards: Moderate
    - Bank Visits & Online Usage: Low
    - Total Calls Made: Low
    - Insight: These customers have moderate spending power but show low engagement in online or banking interactions.
2. **Segment 1 (High Credit Card Holders - 22 customers)**
    - Average Credit Limit: High (43,933)
    - Total Credit Cards: Very High (142,363)
    - Bank Visits: Very Low
    - Online Usage: High
    - Total Calls Made: Very Low
    - Insight: This small segment has high financial capability, prefers digital transactions, and rarely contacts support.
3. **Segment 2 (Low Spending & Engagement - 113 customers)**
    - Average Credit Limit: Low (32,825)
    - Total Credit Cards: Low (12,761)
    - Bank Visits & Online Usage: Low
    - Total Calls Made: Medium
    - Insight: Customers here are financially conservative and show limited interaction with banking channels.
4. **Segment 3 (Moderate to High Spenders - 183 customers)**
    - Average Credit Limit: High (77,548)
    - Total Credit Cards: High (38,158)
    - Bank Visits & Online Usage: Low
    - Total Calls Made: Low

- Insight: These customers have strong financial capacity but prefer offline banking and make minimal service calls.

5. **Segment 4 (High Credit Users with Frequent Calls - 113 customers)**
    - Average Credit Limit: High (77,230)
    - Total Credit Cards: Moderate
    - Bank & Online Visits: Moderate
    - Total Calls Made: High (6.9 calls on avg.)
    - Insight: These customers actively seek support, indicating potential service issues or complex financial needs.
6. **Segment 5 (High Online Activity - 28 customers)**
    - Average Credit Limit: Highest (67,476)
    - Total Credit Cards: Highest (140,000)
    - Online Usage: Very High (12.96 visits on avg.)
    - Bank Visits: Extremely Low
    - Insight: This tech-savvy group prefers online banking and requires minimal physical branch support.

<u>Hierarchical Clustering</u>

Computing Cophenetic correlation

```
Cophenetic correlation for Euclidean distance and single linkage is 0.7323462986040292.
Cophenetic correlation for Euclidean distance and complete linkage is 0.8568445411191499.
Cophenetic correlation for Euclidean distance and average linkage is 0.8795060897918682.
Cophenetic correlation for Euclidean distance and weighted linkage is 0.8765111391171589.
Cophenetic correlation for Chebyshev distance and single linkage is 0.6537418694697082.
Cophenetic correlation for Chebyshev distance and complete linkage is 0.7773049332049028.
Cophenetic correlation for Chebyshev distance and average linkage is 0.8714548034461262.
Cophenetic correlation for Chebyshev distance and weighted linkage is 0.8633150059751974.
Cophenetic correlation for Mahalanobis distance and single linkage is 0.4714107136987136.
Cophenetic correlation for Mahalanobis distance and complete linkage is 0.513515150063651.
Cophenetic correlation for Mahalanobis distance and average linkage is 0.7589579900545415.
Cophenetic correlation for Mahalanobis distance and weighted linkage is 0.7519300988164067.
Cophenetic correlation for Cityblock distance and single linkage is 0.7900640842590242.
Cophenetic correlation for Cityblock distance and complete linkage is 0.848914182492162.
Cophenetic correlation for Cityblock distance and average linkage is 0.8744561207046612.
Cophenetic correlation for Cityblock distance and weighted linkage is 0.8340204256197558.
****************************************************************************************************
Highest cophenetic correlation is 0.8795060897918682, which is obtained with Euclidean distance and average linkage.
```

Checking Dendrograms

	Linkage	Cophenetic Coefficient
4	ward	0.722189
0	single	0.732346
1	complete	0.856845
3	centroid	0.873983
5	weighted	0.876511
2	average	0.879506

Checking Model using Sklearn

```
▾ AgglomerativeClustering
AgglomerativeClustering(n_clusters=6)
```

Cluster Profiling

HC_segments	Sl_No	Customer Key	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	422.113402	35927.891753	35108.247423	5.314433	3.489691	0.896907	2.144330	194
1	118.779310	70627.200000	12717.241379	2.331034	0.889655	3.441379	6.813793	145
2	611.280000	56708.760000	141040.000000	8.740000	0.600000	10.900000	1.080000	50
3	319.670213	70456.648936	13000.000000	5.393617	3.340426	1.095745	1.797872	94
4	109.518519	27228.790123	13037.037037	2.617284	1.000000	3.703704	6.851852	81
5	515.968750	78317.739583	50354.166667	6.052083	3.697917	1.031250	1.906250	96

Insights

1. **Segment 0 (Largest Group - 194 customers)**
   - Average Credit Limit: Moderate (35,108)
   - Total Credit Cards: Moderate
   - Bank Visits & Online Usage: Low
   - Total Calls Made: Low
   - Insight: These customers are moderate spenders with low banking engagement.

2. **Segment 1 (High Call Frequency - 145 customers)**
    - Average Credit Limit: Moderate (12,717)
    - Total Credit Cards: Moderate
    - Bank Visits: Low
    - Online Usage: Moderate
    - Total Calls Made: High (6.81 calls on avg.)
    - Insight: These customers require frequent assistance, possibly due to service issues or financial inquiries.
3. **Segment 2 (High Credit Users - 50 customers)**
    - Average Credit Limit: Highest (56,708)
    - Total Credit Cards: Very High (141,040)
    - Bank Visits: Low
    - Online Usage: Very High (10.9 visits on avg.)
    - Total Calls Made: Low
    - Insight: This group is financially strong, prefers digital banking, and requires minimal support.
4. **Segment 3 (Moderate Credit & Low Engagement - 94 customers)**
    - Average Credit Limit: Moderate (70,456)
    - Total Credit Cards: Moderate
    - Bank Visits & Online Usage: Low
    - Total Calls Made: Low
    - Insight: Customers in this segment have strong financial power but prefer low interaction with banking services.
5. **Segment 4 (Frequent Callers with Low Credit - 81 customers)**
    - Average Credit Limit: Low (27,228)
    - Total Credit Cards: Low
    - Bank Visits & Online Usage: Low
    - Total Calls Made: Highest (6.85 calls on avg.)

- Insight: Customers frequently contact customer service, indicating potential service or account management issues.

6. **Segment 5 (High Credit, Moderate Digital Use - 96 customers)**
    - Average Credit Limit: High (78,317)
    - Total Credit Cards: High (50,354)
    - Online Usage: Moderate
    - Bank Visits: Moderate
    - Insight: Customers here balance digital and physical banking, showing active financial engagement.

K-means Cluster Vs Hierarchical Clustering

Silhouette Score for K-Means: 0.723
Silhouette Score for Hierarchical Clustering: 0.723

Comparison Table:

	Aspect	K-Means	Hierarchical
0	Requires k	Yes	Yes
1	Algorithm Type	Partitional	Hierarchical
2	Complexity	O(nk)	$O(n^2)$
3	Handles Non-Spherical Clusters	No	Yes
4	Visualization	Centroids	Dendrogram

---

1. Overview

Clustering is a key technique used in data segmentation and customer analysis. This report compares K-Means Clustering and Hierarchical Clustering, evaluating their effectiveness based on performance, complexity, and usability.

2. Performance Metrics

To measure the effectiveness of both algorithms, we used the Silhouette Score, which indicates how well-separated the clusters are.

- K-Means Silhouette Score: 0.723
- Hierarchical Clustering Silhouette Score: 0.723

Since both algorithms yield identical scores, the selection will depend on other factors such as scalability, interpretability, and data structure.

3. Key Differences & Business Considerations

Aspect	K-Means	Hierarchical
Requires k (number of clusters)?	Yes	Yes
Algorithm Type	Partitional	Hierarchical
Computational Complexity	O(nk) (Faster)	$O(n^2)$ (Slower)
Handles Non-Spherical Clusters?	No	Yes
Visualization	Centroids	Dendrogram

4. Business Implications

- Scalability: K-Means is more suitable for large datasets due to its faster computation time.
- Cluster Shape: If clusters are irregular or non-spherical, Hierarchical Clustering is a better choice.
- Interpretability: Hierarchical Clustering provides a dendrogram, making it easier to analyze relationships between clusters.
- Use Case:
    - Use K-Means for customer segmentation in large databases.
    - Use Hierarchical Clustering for market research where understanding relationships is crucial.

5. Conclusion & Recommendation

Both algorithms provide valuable insights, but K-Means is recommended for large-scale applications due to efficiency, while Hierarchical Clustering is better for detailed exploratory analysis. The final choice should align with business goals and dataset characteristics.

Actionable Insights and Recommendations

K-means Clustering

1.) Increase Digital Banking Adoption for Low-Engagement Segments

(Targets: Segments 0, 2, 3)

- Offer exclusive online banking rewards to encourage Segment 0 & 3 to use digital services.
- Educate Segment 2 with personalized financial advisory content to improve engagement.

2.) Enhance Premium Banking Services for High-Value Customers

(Targets: Segments 1, 3, 5)

- Provide personalized investment options for Segment 1 (wealthy digital users).
- Offer priority customer support & concierge services to Segment 3.
- Develop a VIP online banking experience for Segment 5, ensuring seamless digital transactions.

3.) Improve Customer Support Efficiency

(Targets: Segment 4 - Frequent Callers)

- Implement AI chatbots and self-service portals to reduce call center burden.
- Identify recurring customer pain points and optimize processes to resolve their concerns faster.

4.) Custom Credit & Loan Offers

(Targets: Segments 1, 3, 5 - High Credit Limit Users)

- Offer exclusive credit limit upgrades or tailored loan offers based on spending patterns.
- Introduce low-interest personal loans for high-credit users with low bank interactions.

5.) Branch Resource Optimization

(Targets: Segments 2, 3, 5 - Low Bank Visitors)

- Reduce physical branch dependencies by improving mobile banking experiences.

- Offer video call banking or AI-powered virtual assistants to support Segment 5 & 3.

Conclusion

This segmentation provides targeted strategies to enhance customer engagement, increase digital adoption and optimize banking services. By implementing these recommendations, financial institutions can improve customer satisfaction, increase revenue and reduce operational costs

Hierarchical Clustering

1.) Encourage Digital Adoption for Low-Engagement Segments

(Targets: Segments 0, 3, 4)

- Introduce incentives (cashbacks, reward points) for digital transactions.
- Promote automated banking services to reduce customer service dependency.

2.) Improve Customer Support for High-Call Segments

(Targets: Segments 1, 4)

- Implement AI chatbots & FAQ automation to handle common inquiries.
- Offer dedicated account managers for customers requiring frequent support.

3.) Enhance Premium Banking Services for High-Value Customers

(Targets: Segments 2, 5)

- Offer personalized financial advisory & investment opportunities.
- Provide priority customer service to ensure top-tier experience.

4.) Optimize Branch Resources for Low Bank Visitors

(Targets: Segments 2, 3, 4)

- Reduce branch overhead by shifting services to video call banking and self-service kiosks.
- Offer priority appointments for customers who need branch services.

5.) Credit Limit Optimization & Risk Management

(Targets: Segments 2, 5 - High Credit Users)

- Provide higher credit limits & personalized loan options to trusted customers.

- Implement risk analysis tools to monitor and manage high-credit users.

Conclusion

This segmentation provides a roadmap for optimizing customer experience, reducing operational costs, and increasing digital banking adoption. Implementing these strategies can help improve customer satisfaction, revenue growth, and efficiency.