2025

AllLife Bank Report

Prepared by
Karthick R

Table of Contents

List of Tables

List of Tables

List of figures

List of figures

List of figures

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.

AllLife Bank aims to enhance its penetration in the credit card market by:

- Running personalized marketing campaigns for existing and potential customers.
- Improving customer service experience based on customer interaction patterns.

To support these strategic goals, the Data Science team is tasked with identifying distinct customer segments using unsupervised learning techniques like clustering. These segments will be based on:

- Spending behavior
- Service interaction history

Data overview

- The dataset has 660 rows and 7 columns
- All 7 columns are numeric colmns

1. Average Credit Limit

- **What it tells us:** The customer's financial power and spending potential.
- **Insight**:
  - High credit limits → High-value customers, likely with better creditworthiness.
  - Low credit limits → New, riskier, or less-engaged customers.

2. Total Credit Cards

- **What it tells us:** Level of engagement with the bank’s credit offerings.
- **Insight**:
  - Customers with multiple cards are likely more financially active or loyal.
  - One-card holders may be new or low-risk trial users.

3. Total Visits to Bank (yearly)

- **What it tells us:** Preference for in-person banking or complexity of needs.
- **Insight**:
  - High visits may indicate issues, preference for face-to-face service, or older demographics.
  - Low visits could reflect satisfaction or a preference for digital channels.

4. Total Visits Online

- **What it tells us:** Digital engagement level.
- **Insight**:
  - Frequent online users may be younger, tech-friendly, and self-service oriented.
  - Low digital engagement could indicate potential for digital onboarding.

5. Total Calls Made

- **What it tells us:** Service needs and possible customer frustration.
- **Insight**:
  - High number of calls may indicate unresolved issues, dissatisfaction, or complex problems.
  - Low number may indicate high satisfaction or low engagement.

Displaying the first 5 rows of the dataset

	Sl_No	Customer Key	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made
0	1	87073	100000	2	1	1	0
1	2	38414	50000	3	0	10	9
2	3	17341	50000	7	1	3	4
3	4	40496	30000	5	1	1	4
4	5	47437	100000	6	0	12	3

Table 1: Top five rows of dataset

Checking the shape of the dataset

- The dataset contains 660 rows and 7 columns.

Checking the data types of the columns for the dataset

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 660 entries, 0 to 659
Data columns (total 7 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Sl_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)
memory usage: 36.2 KB
```

Table 2: data types of the columns

- All seven columns are numerical columns

Statistical summary of the dataset

	count	mean	std	min	25%	50%	75%	max
Sl_No	660.0	330.500000	190.669872	1.0	165.75	330.5	495.25	660.0
Customer Key	660.0	55141.443939	25627.772200	11265.0	33825.25	53874.5	77202.50	99843.0
Avg_Credit_Limit	660.0	34574.242424	37625.487804	3000.0	10000.00	18000.0	48000.00	200000.0
Total_Credit_Cards	660.0	4.706061	2.167835	1.0	3.00	5.0	6.00	10.0
Total_visits_bank	660.0	2.403030	1.631813	0.0	1.00	2.0	4.00	5.0
Total_visits_online	660.0	2.606061	2.935724	0.0	1.00	2.0	4.00	15.0
Total_calls_made	660.0	3.583333	2.865317	0.0	1.00	3.0	5.00	10.0

Table 3: Statistical summary of the dataset

Average Credit Limit

- Huge range: from ₹3,000 to ₹2,00,000 — a very wide spread.
- Median (₹18,000) is much lower than mean (₹34,574) → right-skewed, a few high-value customers inflate the average.
- There are a small number of premium clients with very high limits.

Total Credit Cards

- Average: ~4.7 cards per customer
- Spread: Ranges from 1 to 10 cards.
- Customers vary significantly in product usage.
- Some customers are highly engaged with multiple cards (loyal users), while others have just one (new or low-need users).

Total Visits to Bank

- Median: 2 visits/year
- 75% of users visit the bank 4 times or less
- Most customers prefer minimal in-person interaction.
- May reflect shift toward digital or satisfaction with self-service.
- Frequent branch visitors could be older or have complex service needs.

Total Visits Online

- Median: 2, but Max = 15
- Some customers are very digitally active — likely younger or tech-savvy.
- A notable portion has 0 online visits, suggesting digital onboarding opportunities.

Total Calls Made

- Median: 3 calls/year
- Max: 10 calls
- Wide variability → some customers contact support frequently, others not at all.
- High-calling customers could be dissatisfied, confused, or have complex issues.

Value counts of dataset

```
Value counts for 'Total_Credit_Cards':
Total_Credit_Cards
1      59
2      64
3      53
4     151
5      74
6     117
7     101
8      11
9      11
10     19
Name: count, dtype: int64

Value counts for 'Total_visits_bank':
Total_visits_bank
0    100
1    112
2    158
3    100
4     92
5     98
Name: count, dtype: int64
```

```
Value counts for 'Total_visits_online':
Total_visits_online
0     144
1     109
2     189
3      44
4      69
5      54
6       1
7       7
8       6
9       4
10      6
11      5
12      6
13      5
14      1
15     10
Name: count, dtype: int64

Value counts for 'Total_calls_made':
Total_calls_made
0      97
1      90
2      91
3      83
4     108
5      29
6      39
7      35
8      30
9      32
10     26
Name: count, dtype: int64
```

Table 4: Value counts of dataset

Checking for duplicate values

```
332     47437
398     96929
432     37252
541     50706
632     97935
Name: Customer Key, dtype: int64
```

Table 5: duplicate values

- There are duplicate values in the “Customer Key” column but we can ignore it because we are not going the use the column for clustering.

Checking for 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

dtype: int64

Table 6: Data on null values

- There are no null values in the dataset.

Exploratory Data Analysis

Univariate analysis

Analysis on Avg Credit Limit

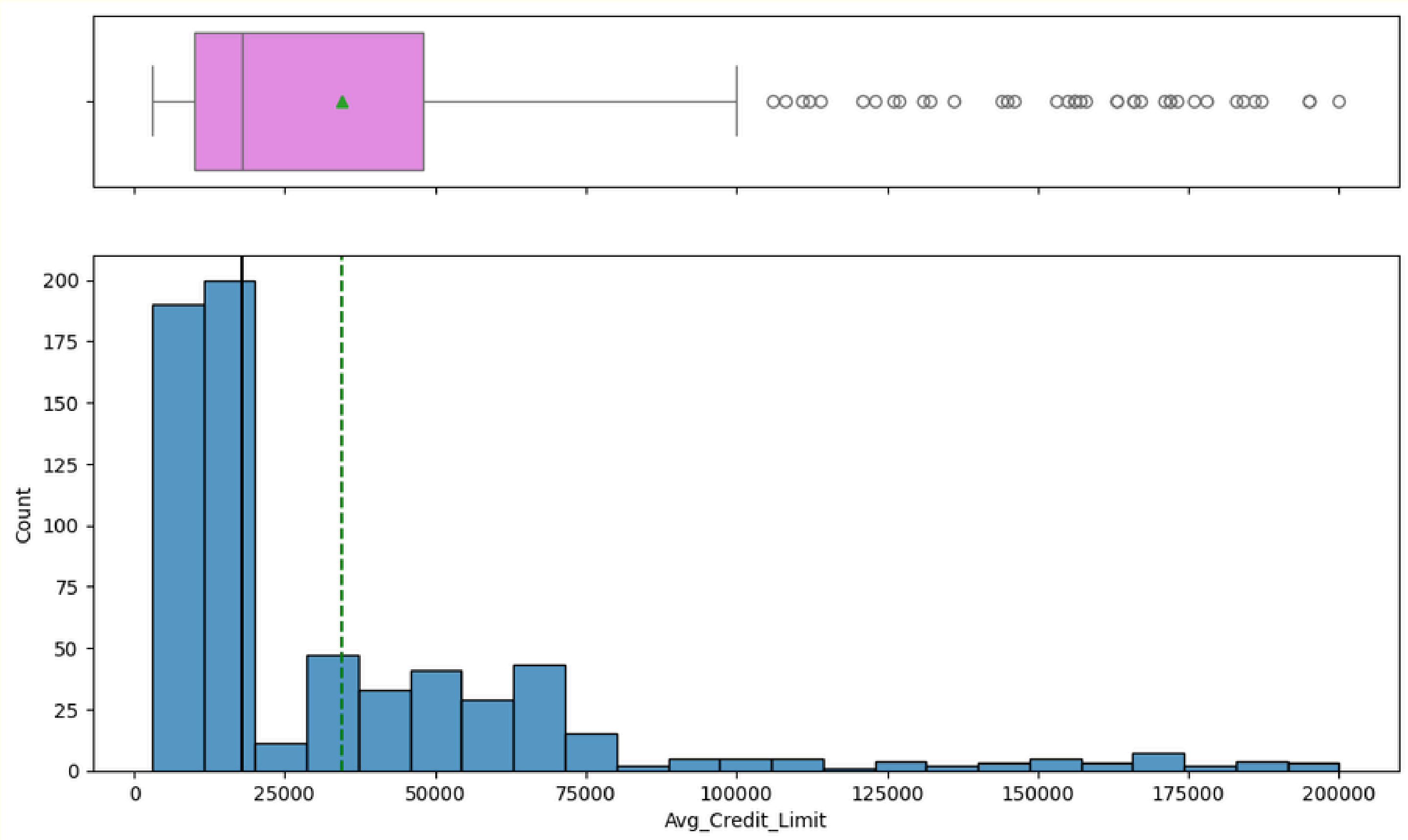


Figure 1: Avg Credit Limit

Observations on Avg Credit Limit

- The distribution of Avg_Credit_Limit is right-skewed, with most customers having limits below ₹25,000.
- There are numerous high-value outliers, as seen in the boxplot's extended right tail.

Analysis on Total Credit Cards

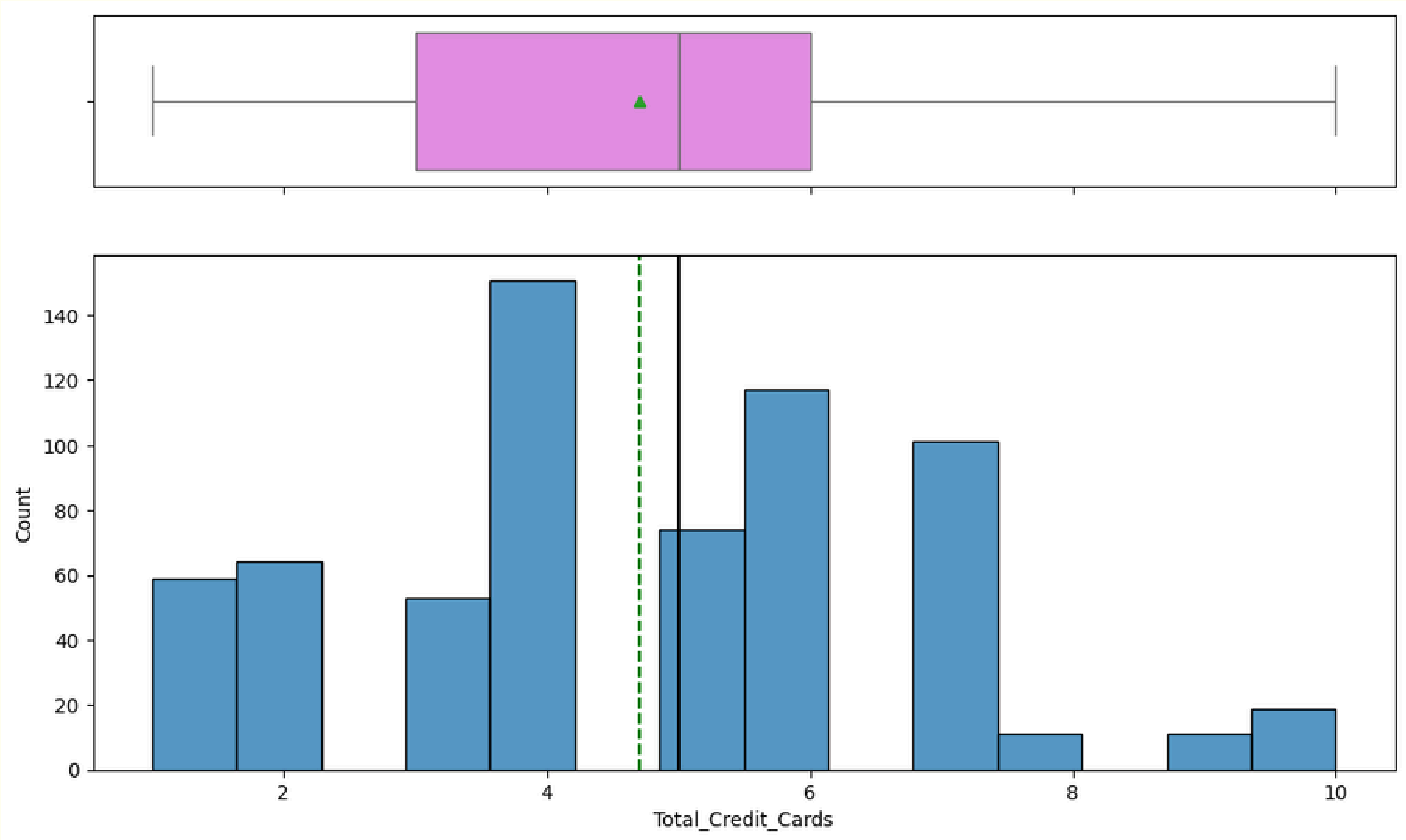


Figure 2: Total Credit Cards

Observations on Total Credit Cards

- The distribution of Total_Credit_Cards is slightly right-skewed, with most customers holding between 4 to 7 cards.
- There are no significant outliers, and the spread is relatively compact compared to the previous variable.
- The mean is slightly higher than the median, indicating a few customers have a larger number of credit cards.

Analysis on Total visits bank

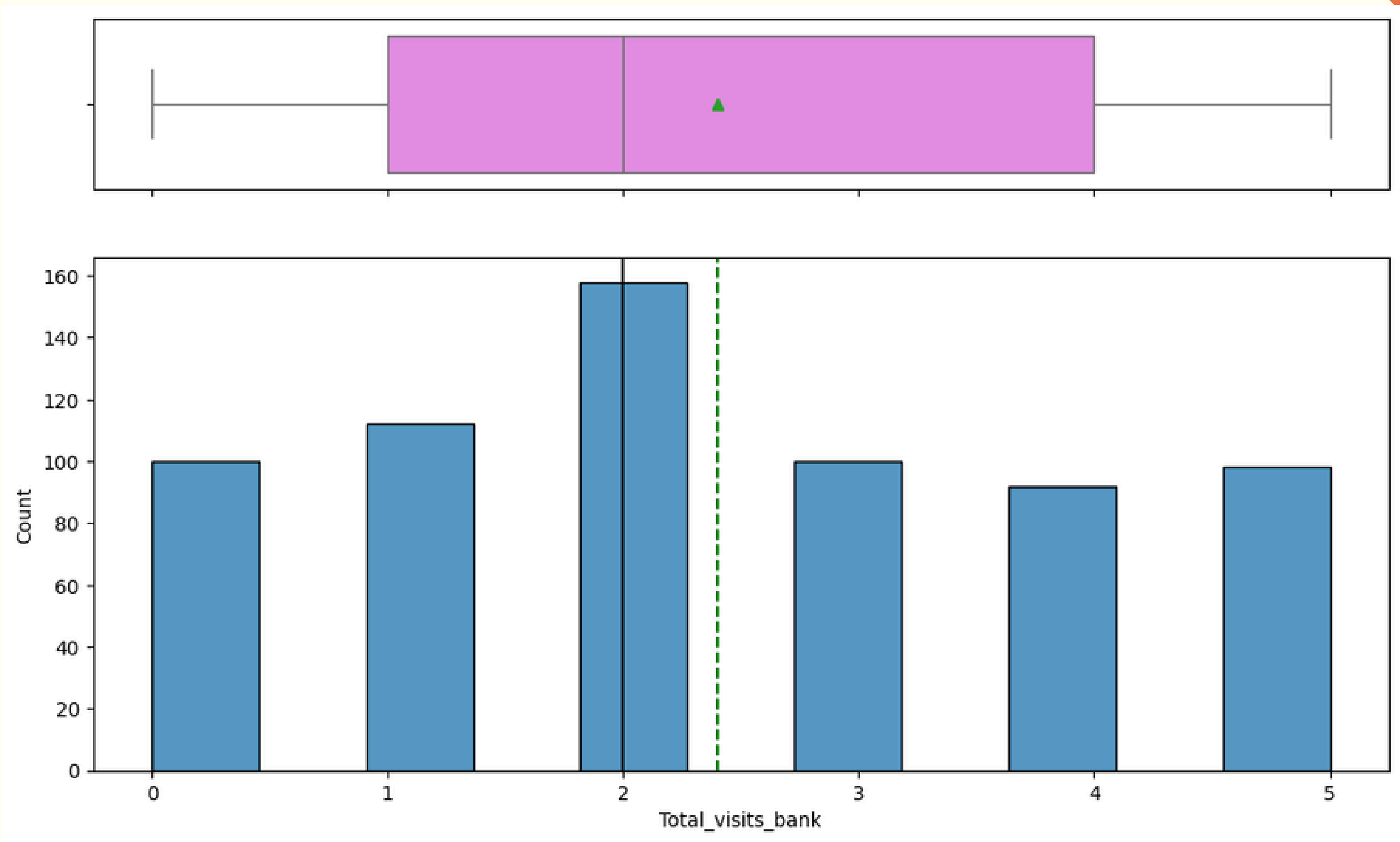


Figure 3: Total visits bank

Observations on Total visits bank

- The distribution of Total_visits_bank is fairly uniform, with visits ranging from 0 to 5 and no visible outliers.
- The median is around 2, and the mean is slightly higher, suggesting a mild right skew.
- Most customers visit the bank between 1 to 3 times, showing moderate engagement with in-person services.

Analysis on Total visits online

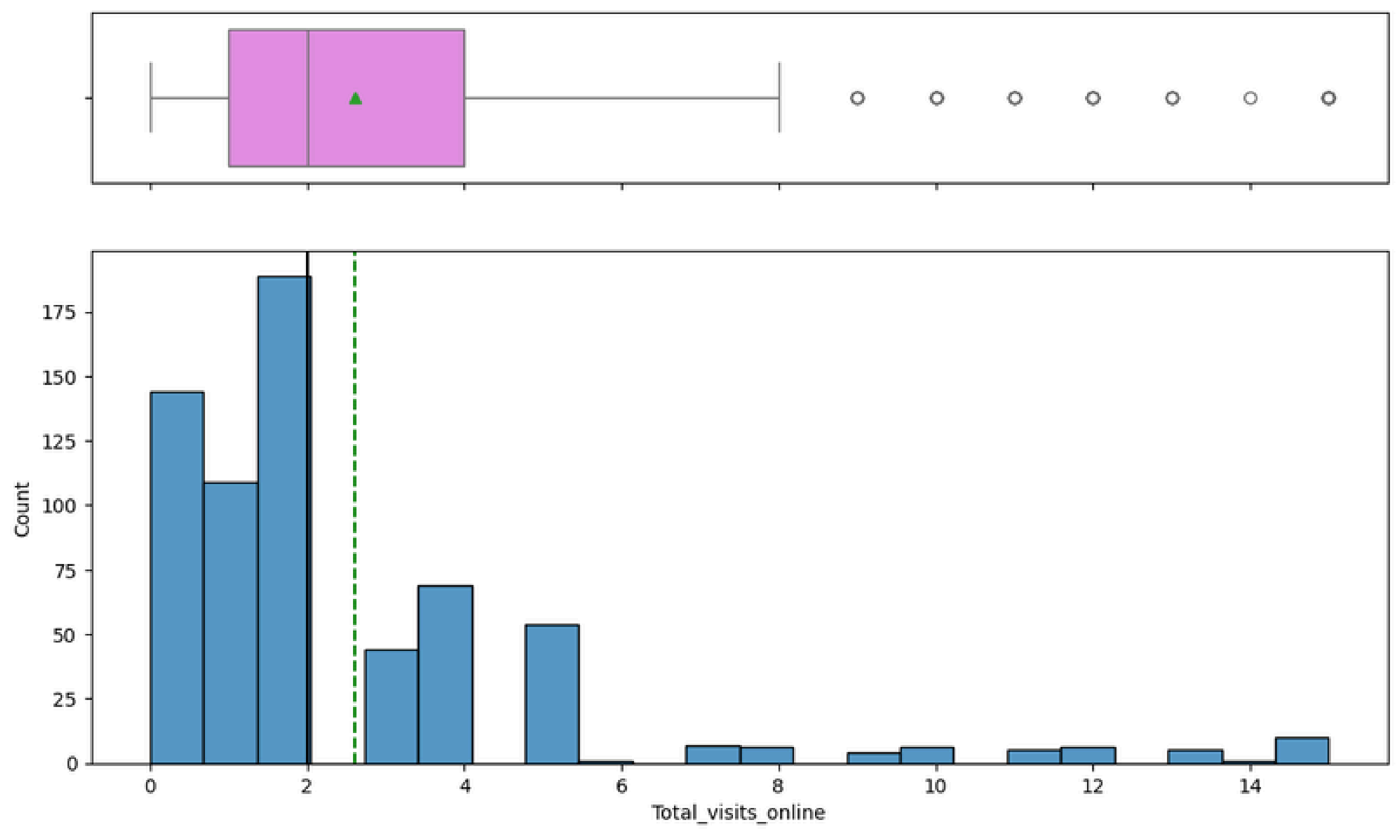


Figure 4: Total visits online

Observations on Total visits online

- The distribution of Total_visits_online is right-skewed, with most customers having 0 to 2 online visits.
- There are several high-value outliers beyond 7 visits, indicating a small group of highly active online users.
- The mean is greater than the median, again showing that a few high-frequency users are increasing the average.

Analysis on Total calls made

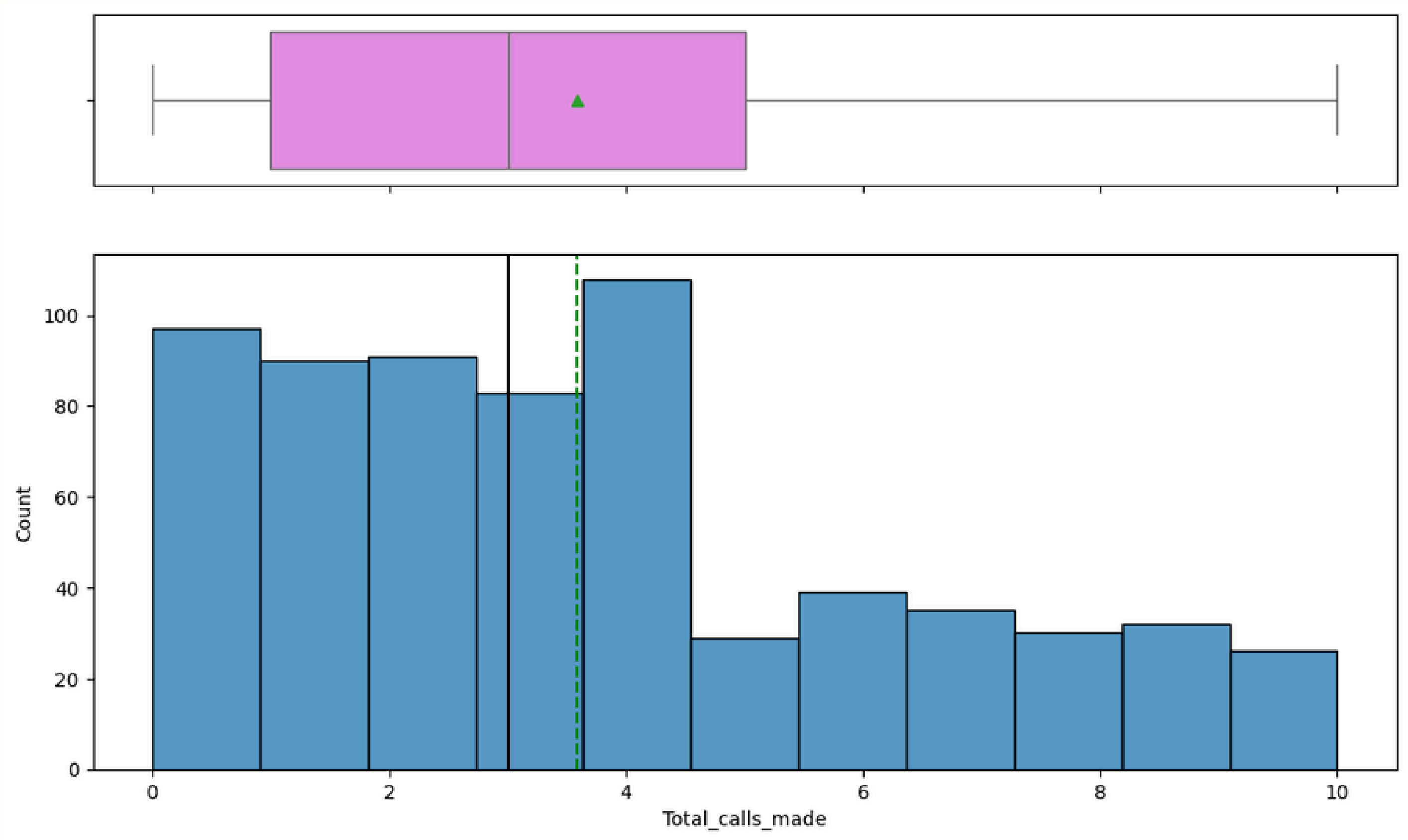


Figure 5: Total calls made

Observations on Total calls made

- The distribution of Total_calls_made appears roughly uniform with a mild right skew.
- Most customers make between 0 to 5 calls, with fewer making frequent calls beyond 6.
- There are no significant outliers, and the mean is slightly higher than the median.

Bivariate analysis

Analysis between Avg Credit Limit and Total Credit Cards

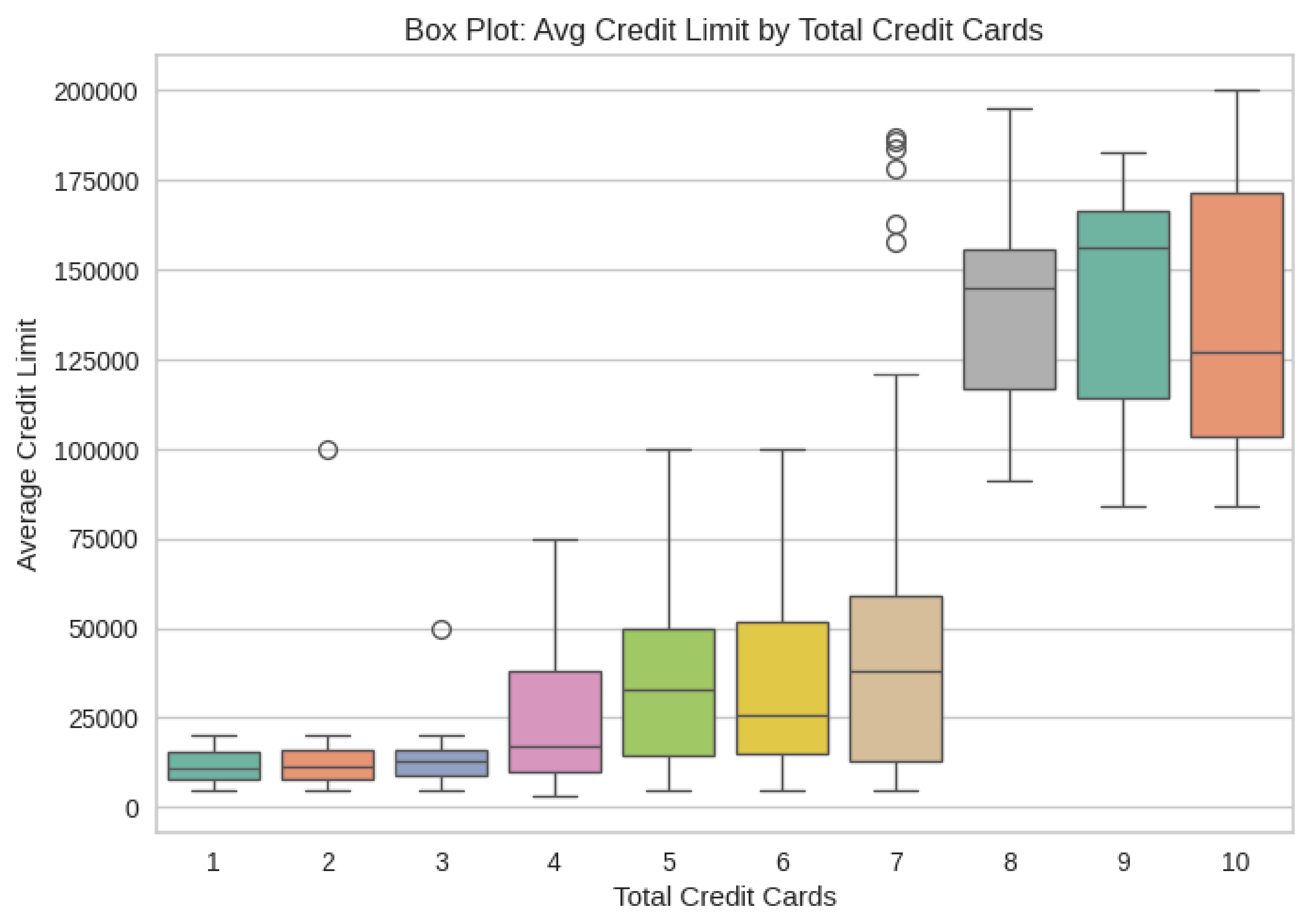


Figure 6: Avg Credit Limit and Total Credit Cards

Observations on Avg Credit Limit and Total Credit Cards

- Positive Relationship: There is a clear upward trend—customers with more credit cards tend to have higher average credit limits.
- Outliers Present: Outliers are more noticeable for customers with fewer cards (e.g., 2–4 cards), where a few individuals have unusually high credit limits compared to their peers.

Analysis between Avg_Credit_Limit vs Total_visits_online

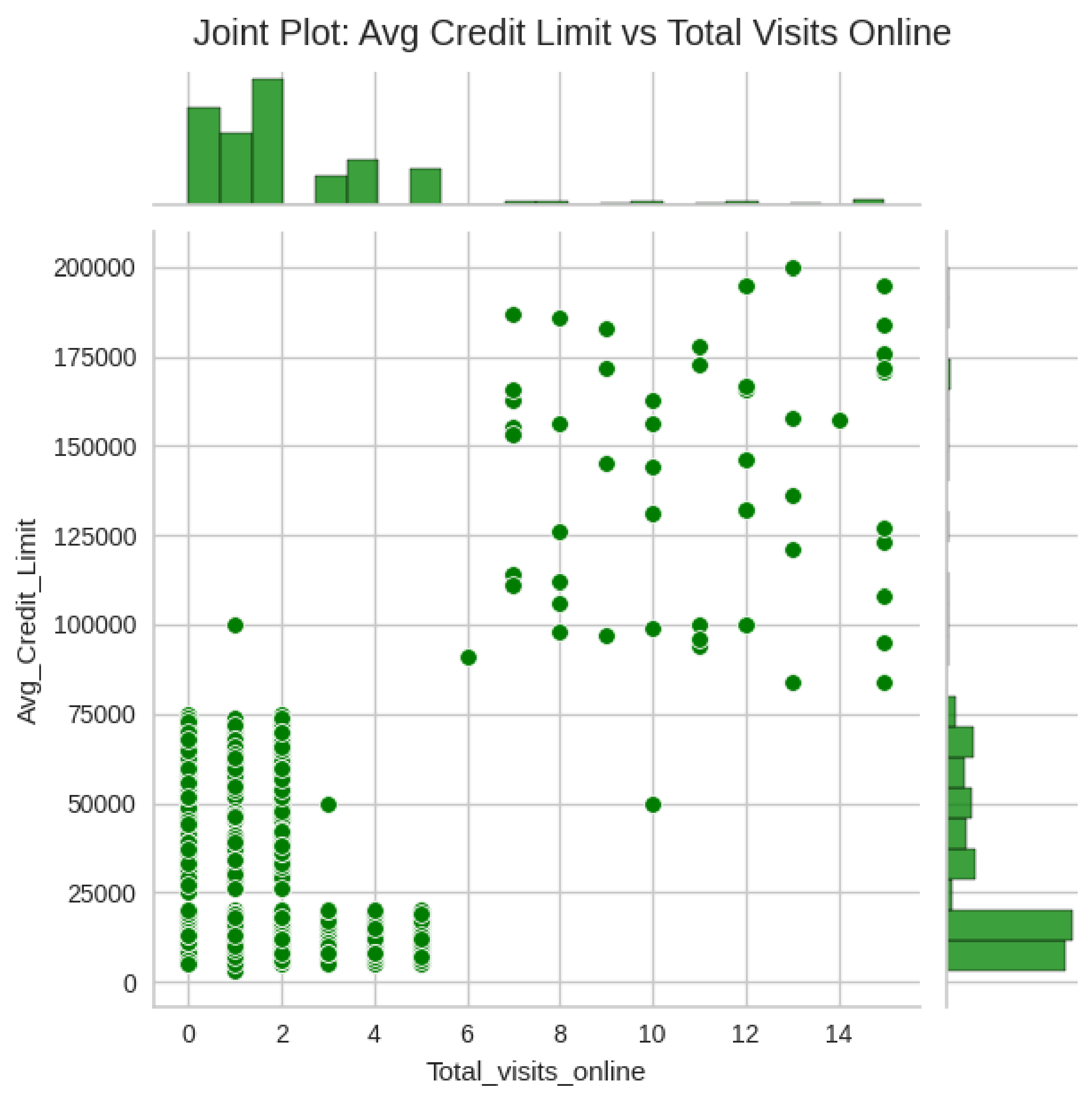


Figure 7: Avg_Credit_Limit vs Total_visits_online

Observations

- Customers who visit online more frequently tend to have higher average credit limits, suggesting a potential link between digital engagement and creditworthiness.
- Skewed Distribution: Both Total_visits_online and Avg_Credit_Limit are right-skewed, with a concentration of customers in the lower ranges and a few high-value outliers.

Pair plot

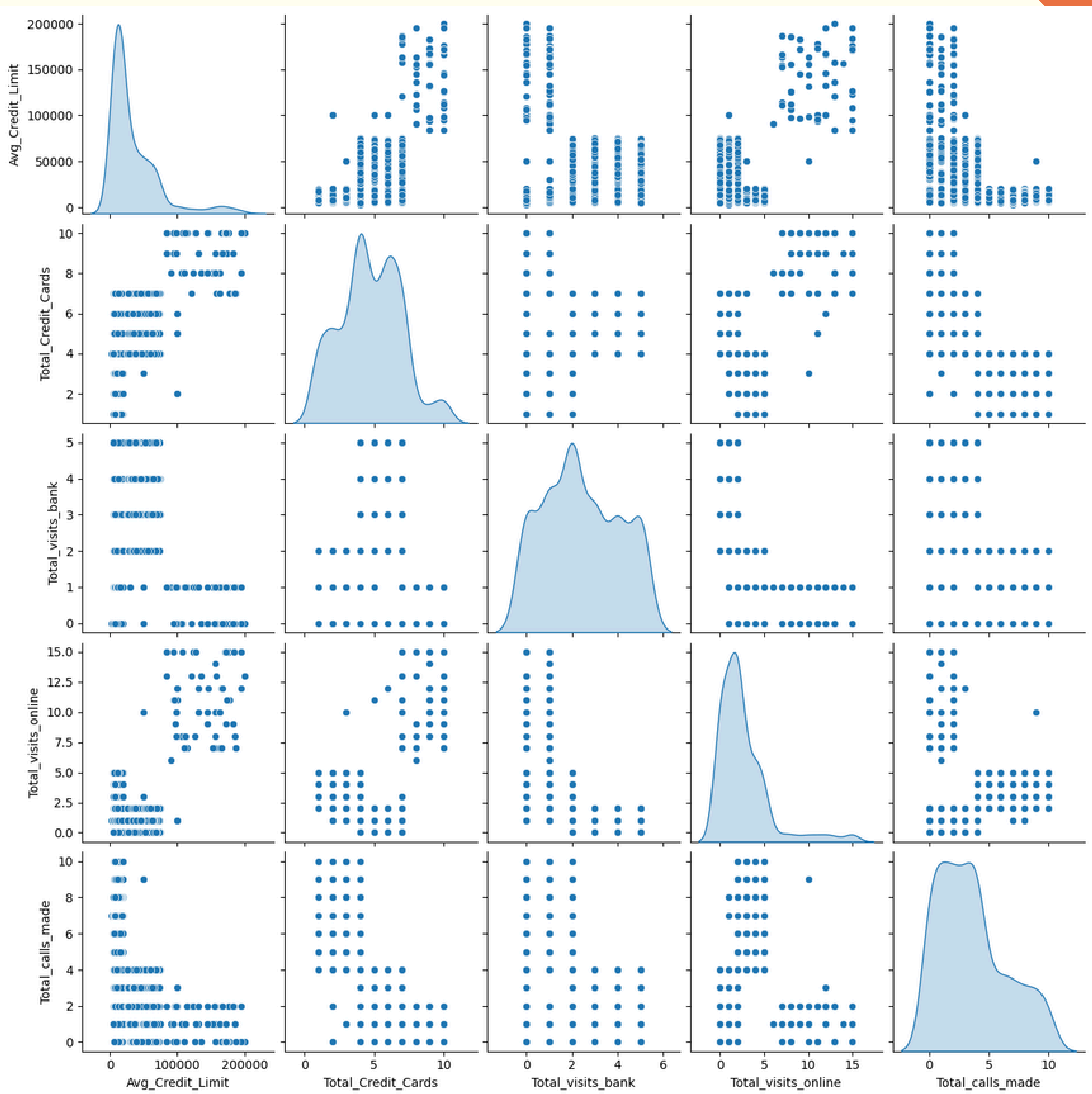


Figure 8: Pair plot

Observations

- Avg_Credit_Limit shows a visible positive correlation with Total_Credit_Cards and Total_visits_online, indicating that higher credit limits are generally associated with more credit cards and greater online engagement.
- There are mild negative trends between Total_calls_made and both Total_Credit_Cards and Avg_Credit_Limit,

Correlation matrix

Figure 9: Correlation matrix

Observations on Correlation matrix

- Avg_Credit_Limit is moderately positively correlated with Total_Credit_Cards and Total_visits_online (both at 0.6).
- Total_Credit_Cards has a strong negative correlation with Total_calls_made (-0.7), indicating that customers with more cards tend to call less.
- Total_visits_bank and Total_visits_online show a strong negative correlation (-0.6), suggesting customers prefer either online or in-person visits, but not both.

Data preprocessing

Outlier Detection

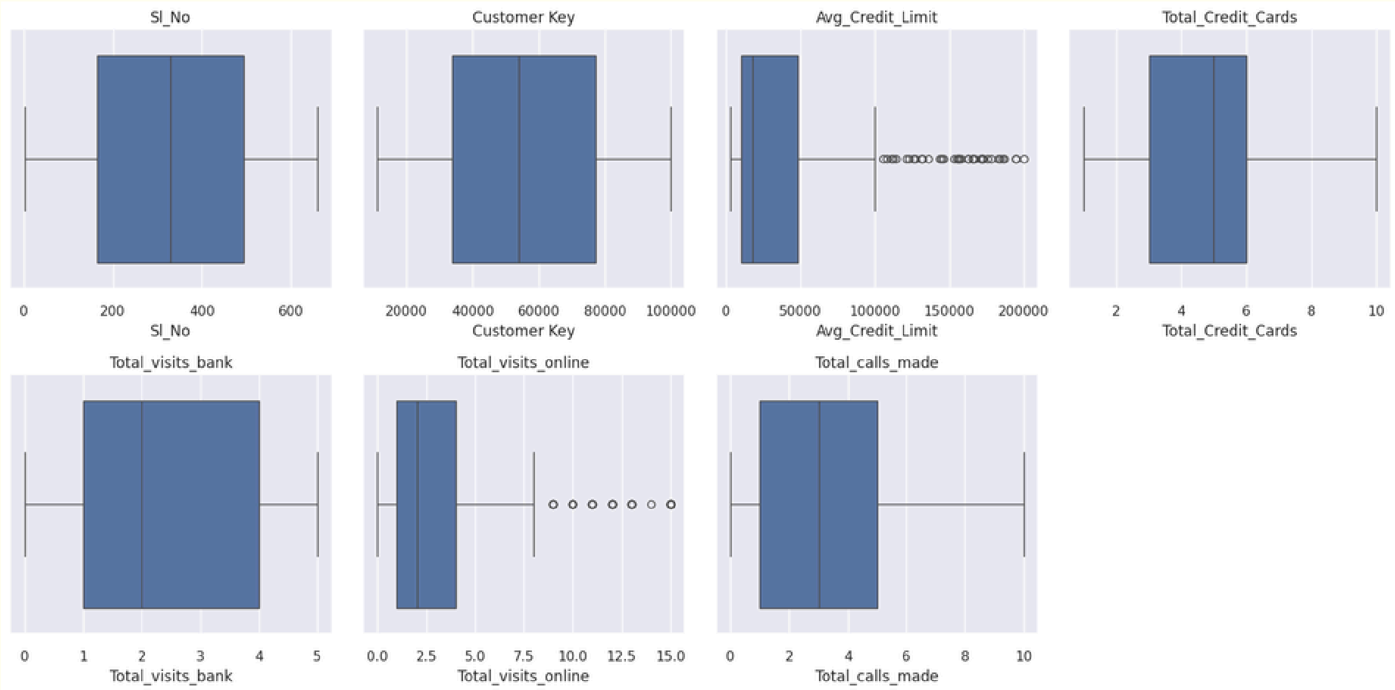


Figure 10: Outlier Detection

- There are some outlier but we are not treating them for business needs.

Checking for duplicate values

- There are duplicate values in the “Customer Key” column but we can ignore it because we are not going the use the column for clustering.

Checking for 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

dtype: int64

Table 7: Checking Missing values

- There are no missing values in the data.

Feature engineering

- Removing the columns “Sl_No” and “Customer Key” from the dataset because they dont contribute anything for the clustering.

Data Scaling

- Using Standard Scaler to scale the dataset.

	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

Table 8: Scaling

K-means Clustering

- Giving the range of the clusters between 2 to 10 to find the optimal K value.

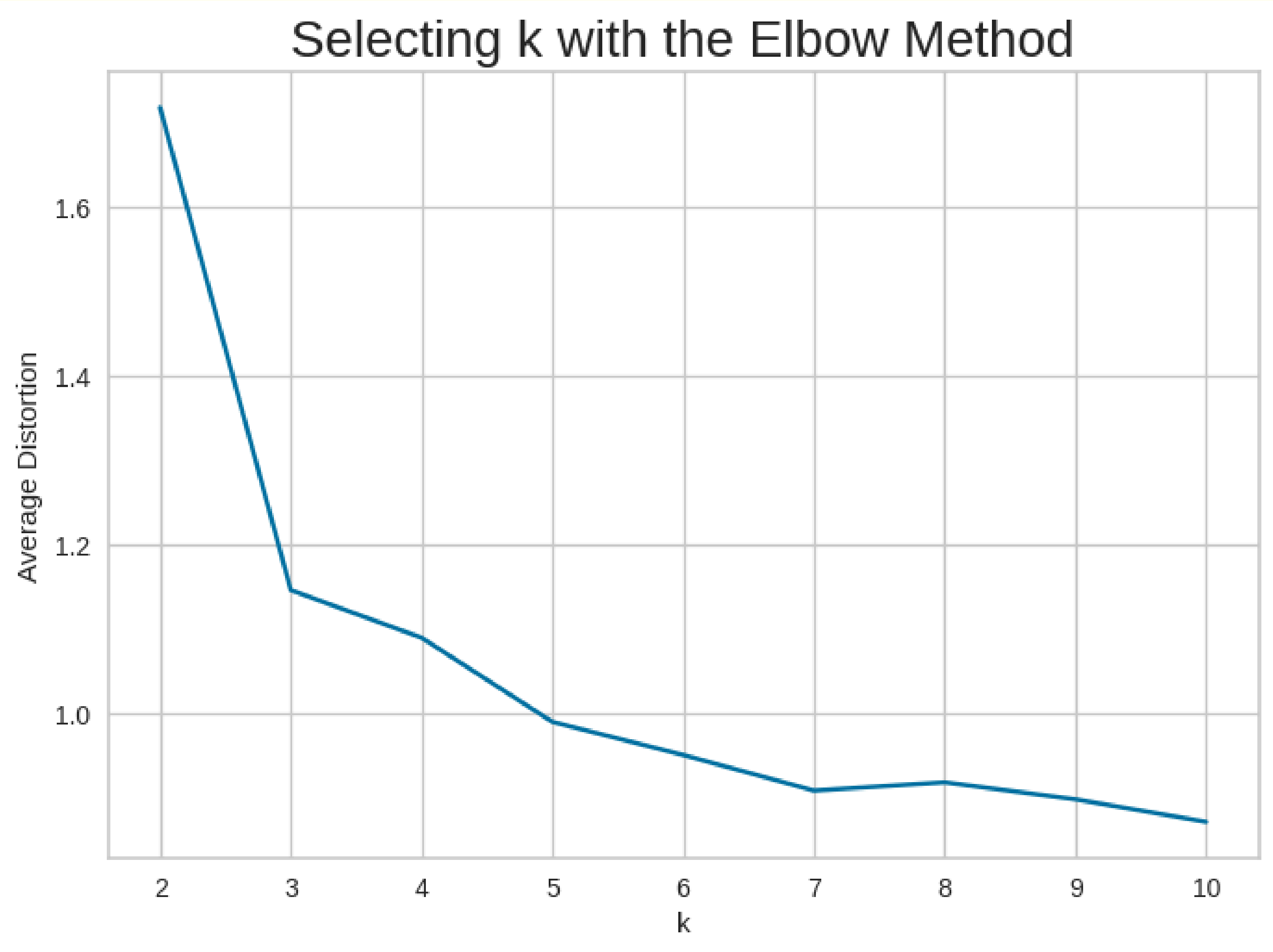


Figure 11: Elbow plot

Average Distortion value:

- 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

Insights

1. Optimal Number of Clusters ≈ 3
   - The elbow point — where the distortion sharply decreases and then levels off — appears at k = 3.
   - This suggests that 3 clusters likely provide the best balance between model complexity and performance.

2. Diminishing Returns Beyond k = 3
   - After k = 3, the reduction in Average Distortion becomes more gradual.
   - This indicates that adding more clusters yields only marginal improvements, making them less efficient for practical purposes.

Distortion Score Elbow with fit time

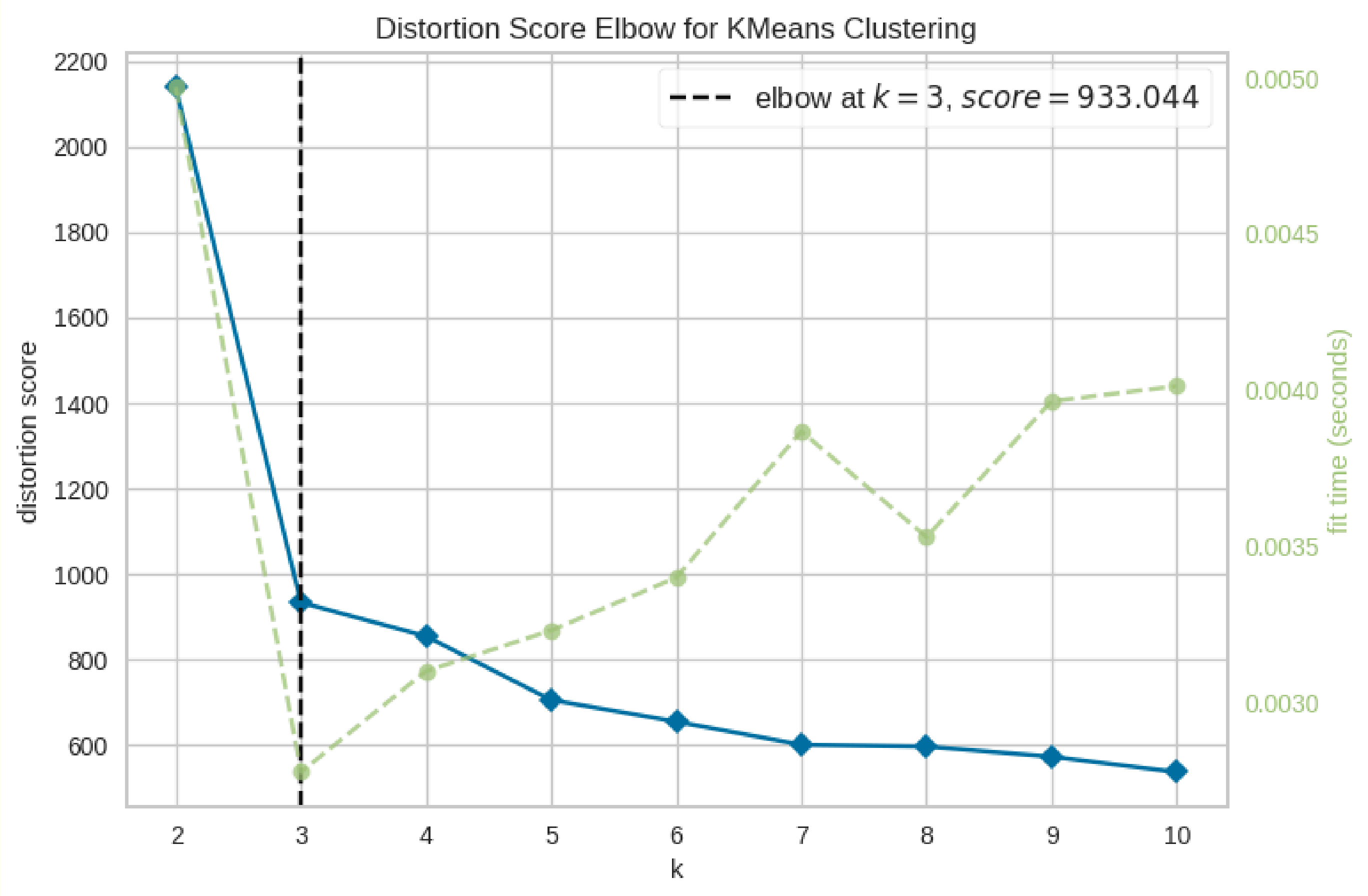


Figure 12: Distortion score with fit time

Optimal Number of Clusters ≈ 3

- 3 cluster give us the lowest fit time with the Distortion Score of 933.

The time taken for execution of the K means for the three clusters is 0.0032

Checking Silhouette Scores

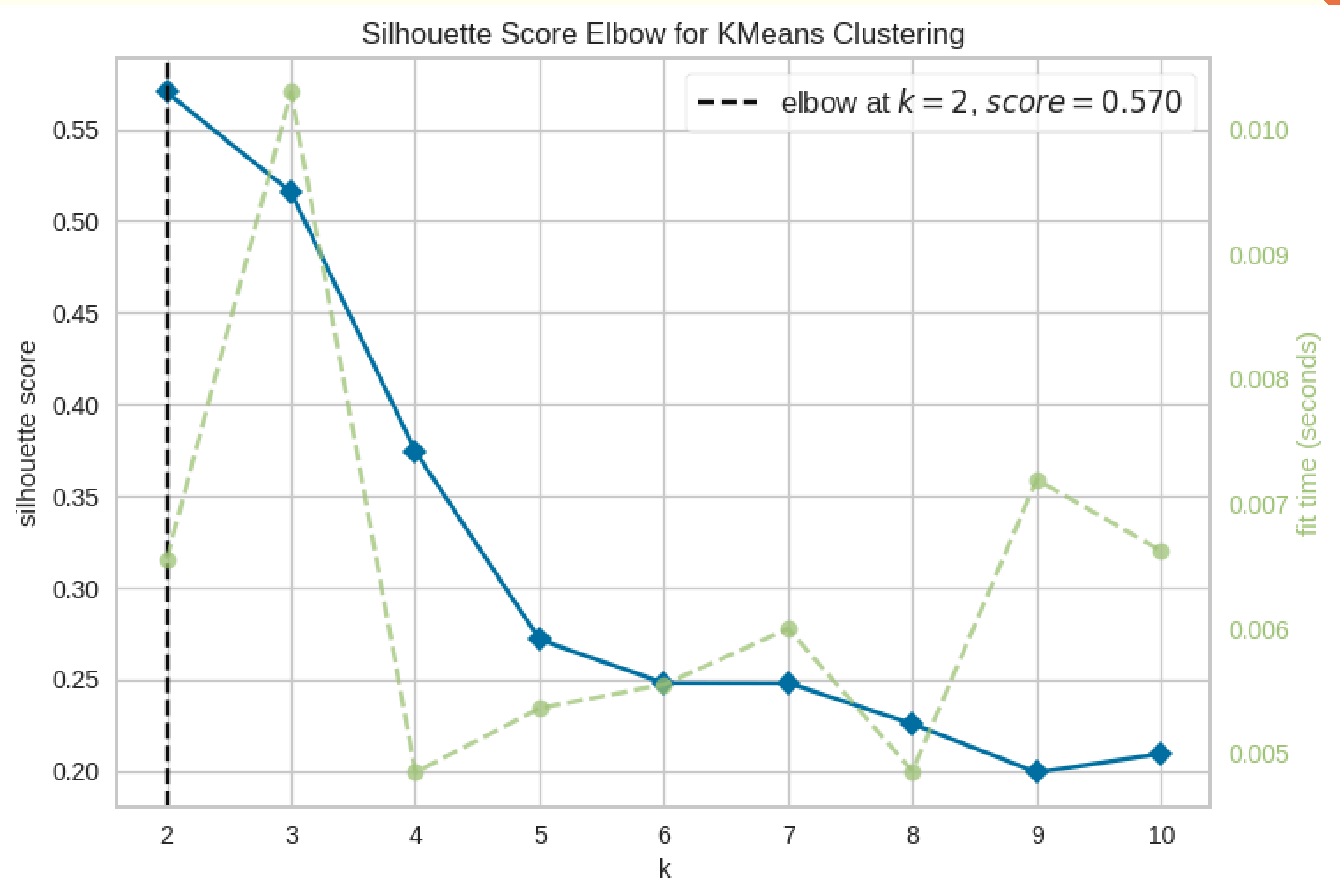


Figure 13: Silhouette Scores

- While k = 2 gives the highest silhouette score (0.57), k = 3 still offers a fairly high score (~0.52).
- This is only a slight drop in clustering quality but gains in more granular grouping, which may provide better actionable insights.
- The Elbow Method (distortion plot) clearly shows an elbow at k = 3, indicating a point of diminishing returns.
- This means that moving from k = 2 to k = 3 leads to a significant reduction in distortion, improving how well the clusters fit the data.

- In many real-world applications (e.g., customer segmentation), two clusters might oversimplify.
- Three clusters can offer better segmentation like:
  - High-value vs mid-value vs low-value customers
  - Frequent vs moderate vs infrequent users
  - or other such stratified behavior groups.

Silhouette Plot

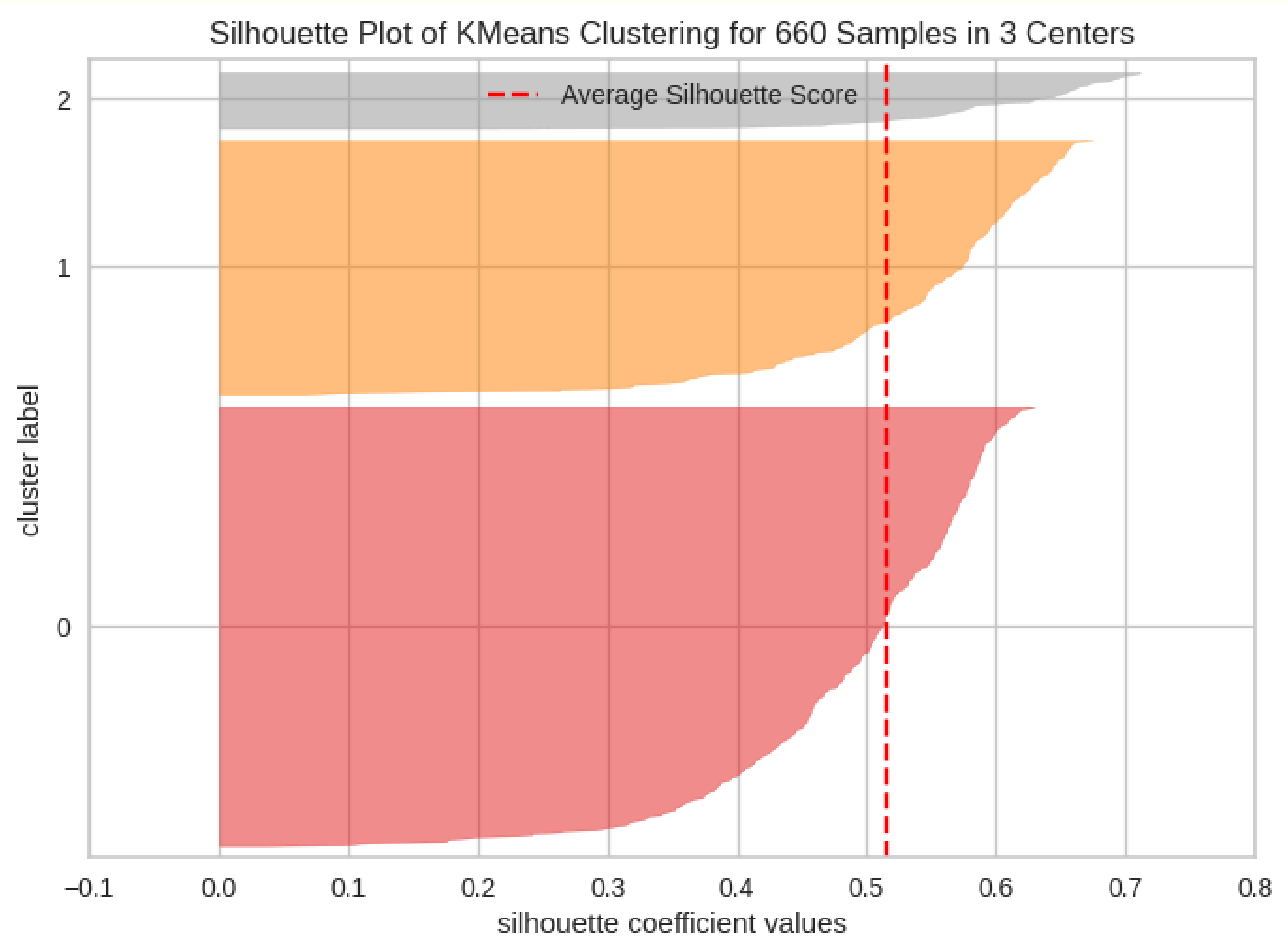


Figure 14: Silhouette Plot

1. Clear Separation of Clusters
    - Each color band (red, orange, gray) represents a distinct cluster.
    - All clusters show positive silhouette coefficients, mostly > 0.25, indicating good separation and cohesiveness.

2. Average Silhouette Score ≈ 0.5 (Red Dashed Line)
    - This score indicates moderately strong structure in the data.
    - A score above 0.5 suggests reasonably well-defined clusters.
    - None of the clusters show major overlap, reinforcing the quality of segmentation.

3. Balanced Cluster Sizes
    - The width of each colored region represents the number of samples in that cluster.
    - Cluster 0 is the largest, but Clusters 1 and 2 are also sufficiently populated—indicating that clustering is not skewed toward a single group.

- **The silhouette plot confirms that k = 3 is a statistically sound and interpretable choice.**
- **It provides well-separated, meaningful clusters with relatively high silhouette values for most data points.**

Cluster Profiling

KM_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank
0	33782.38	5.52	3.49
1	12174.11	2.41	0.93
2	141040.00	8.74	0.60

KM_segments	Total_visits_online	Total_calls_made	Cluster_Size
0	0.98	2.00	386
1	3.55	6.87	224
2	10.90	1.08	50

Table 9: Cluster Profiling

Cluster 0 – Mid-Tier Customers (Majority Segment – 384 customers)

- Moderate credit limit and card count.
- Prefer in-person banking over online services.
- Moderate engagement level overall.

Cluster 1 – Low-Tier Customers (Value Seekers – 224 customers)

- Lowest credit limits and card ownership.
- Heavily depend on phone calls and online support.
- Possibly cost-sensitive or newer to banking.

Cluster 2 – High-Value Digital Customers (Small but Elite – 50 customers)

- Extremely high-value customers with large credit limits and multiple cards.
- Prefer online channels almost exclusively.
- Minimal phone/bank visits, highly self-reliant.

Customers in cluster 0 : 384 (Customer Key)

87073 17341 40496 54838 35254 46635 97825 83125 35483 15129 83290 56486 31903 45909 14263 46813 81878 35549 85799 39122 81531 69965 18595 44398 32352 40898 27101 33457 45088 23302 27408 65372 21531 56843 17165 89328 20072 71402 47496 24808 17036 67193 34423 97109 55382 51811 53936 66504 53207 18514 51319 36340 36934 95925 49771 22919 21233 74544 52025 45652 73952 49418 77026 49331 75775 54906 94666 11698 34677 95610 41380 38033 85337 38994 67911 92956 77641 57565 53814 30712 19785 31384 16374 50878 78002 83459 91987 51552 24998 45673 11596 87485 28414 81863 33240 11466 23881 44645 49844 92782 22824 26767 26678 50412 17933 34495 47437 22610 41159 64672 62483 85614 96548 19137 69028 70779 38244 67046 64897 46223

36628 17565 77381 11799 81940 66706 87838 94437
33790 44402 29886 66804 47866 61996 15318 89635
71681 71862 96186 22348 36243 88807 82376 98126
80347 17649 62807 92522 57459 44579 45476 61994
11398 24702 27824 45878 72431 19215 23409 16418
85122 55060 55478 65574 31113 96929 78912 68439
62864 31515 77954 88207 78618 31551 75792 29864
45440 97954 90189 55090 17703 33991 88884 45808
50706 92140 88123 53932 65908 25321 87456 48602
97530 48657 76209 49913 53002 61122 82807 93496
64519 31950 23110 96297 28408 37252 41287 52460
26604 58019 87219 36839 12663 48667 42887 14439
60851 41266 37438 65747 81166 20570 14816 11265
24980 37934 70707 84351 89446 17325 64774 53166
45341 94595 55170 92489 92933 36504 40508 15798
70101 77613 84360 48402 46776 67258 44804 29919
65781 12456 62649 74446 36632 76024 75065 51682
18397 29102 56367 95147 44379 76957 42921 23102
61324 49690 20043 44144 53552 62530 41741 22842
65825 77826 61216 83192 82023 73000 64550 90131
17382 27117 94529 21717 81910 76492 43000 48692
27476 15086 43034 99131 13140 99437 91242 39285

63710 90860 35585 58708 57451 69868 43679 30256 26334 47848 17377 39644 29176 55706 51771 83585 51867 68040 75417 34775 85645 83545 44157 38125 75398 90999 70376 33295 80942 26493 97850 43841 79885 59316 83466 81510 35268 11734 88411 96269 87683 26063 42479 58116 67282 84888 75366 14377 59074 96534 31870 24748 68920 67637 60839 59170 90586 56270 87670 47703 35421 58511 76398 93310 36836 46373 94700 67860 99473 68862 93381 46548 74083 48660 13720 72339 99284 47198 67415 44403 58276 85234 31948 90191 49341 11562 16253 80623

Customers in cluster 1 : 224

38414 58634 37376 82490 44770 52741 52326 92503 25084 68517 55196 62617 96463 39137 14309 29794 87241 30507 61061 24001 68067 65034 14854 81130 29112 13999 32550 82164 61517 28254 30888 46388 74126 52142 37659 83132 20119 52363 50769 68502 99026 83326 62040 37252 74625 51182 60301 96386 43886 78503 68419 32828 17937 71632 81566 29759 36929 70248 91673 61355 60403 85868 76205 66524 69214 21976 35149 27120 18821 33187 93482 90168 71881 59656 12026 99589 38970 57990 39447 79694 79403 47296 37559 38165 49198 18007 59619 37016

91099 74704 25742 11937 52736 88338 18916 92501
96213 26599 73007 97935 26089 14946 74795 73435
41634 84069 83244 87291 18086 33369 15310 98499
35256 89007 93997 16577 25440 81116 63663 69811
36111 39454 70199 11602 49697 28701 61627 34103
14248 31256 45583 52750 95507 23743 53410 53898
66200 58389 61347 59151 37802 60475 95489 77758
23768 87471 85707 97951 54785 97011 35103 18564
61009 24054 63751 52758 78473 80457 59783 64241
32374 97536 33110 36978 54281 98602 97687 28842
38410 38261 20524 37671 25330 41787 11412 55892
95495 41946 86410 76718 98969 77143 38205 53851
52783 63405 48510 97463 18145 14398 98288 69704
29058 15546 16715 87350 13215 20593 56624 33317
99596 72430 16676 40486 90958 67212 44226 94251
61776 55275 18609 54477 12122 28208 68003 79632
73811 72892 51773 96163 61234 55849 56156 72156

Customers in cluster 2 : 50

47437 48370 94391 50598 40019 77910 89832 98216 54495 47650 32107 84192 53916 32584 97285 20337 15585 20620 75009 76203 33837 14916 97935 16180 49493 70974 40217 88442 17538 90839 99843 27212 91575 60190 18519 48762 58392 79953 13315 30570 78996 78404 28525 51826 65750 51108 60732 53834 80655 80150

Visual representation of clusters

KM_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	33782.383420	5.515544	3.489637	0.981865	2.000000	386
1	12174.107143	2.410714	0.933036	3.553571	6.870536	224
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50

Table 10: Visual representation of clusters

Insights

Cluster 0 – Traditional Mid-Tier Customers

Count: 386 (Largest segment)

- Avg_Credit_Limit: 33,782 (Mid)
- Total_Credit_Cards: 5.52 (Mid)
- Highest in Bank Visits: 3.49
- Online Visits: Low (0.98)
- Calls Made: Moderate (2.00)

- **They prefer branch visits, have a mid-level credit limit, and moderate engagement.**

Cluster 1 – Support-Heavy Low-Value Customers

Count: 224

- Avg_Credit_Limit: 12,174 (Lowest)
- Total_Credit_Cards: 2.41 (Fewest)
- Bank Visits: Low (0.93)
- Online Visits: Moderate (3.55)
- Highest in Calls Made: 6.87

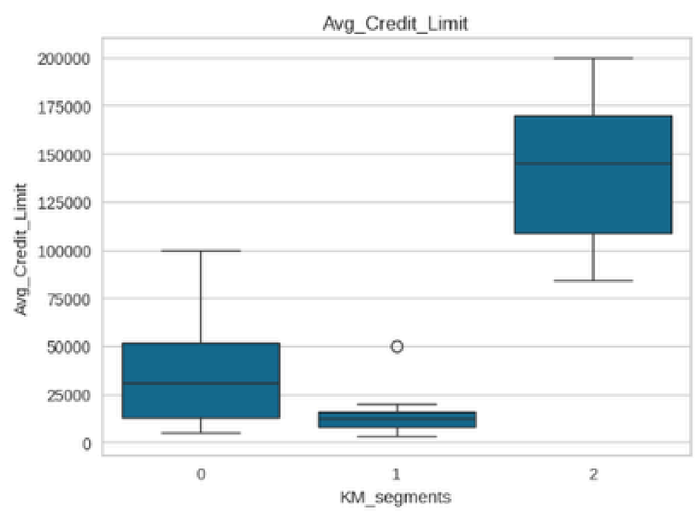
- **They rely on phone support, are cost-sensitive, and may lack digital proficiency.**

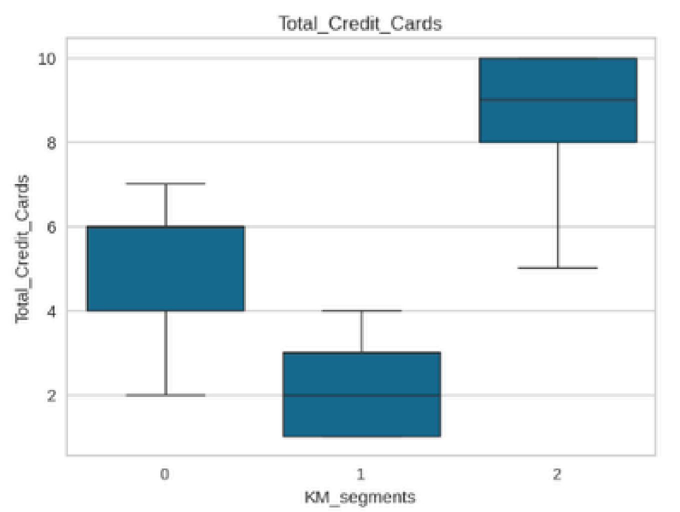
Cluster 2 – High-Value Digital Natives

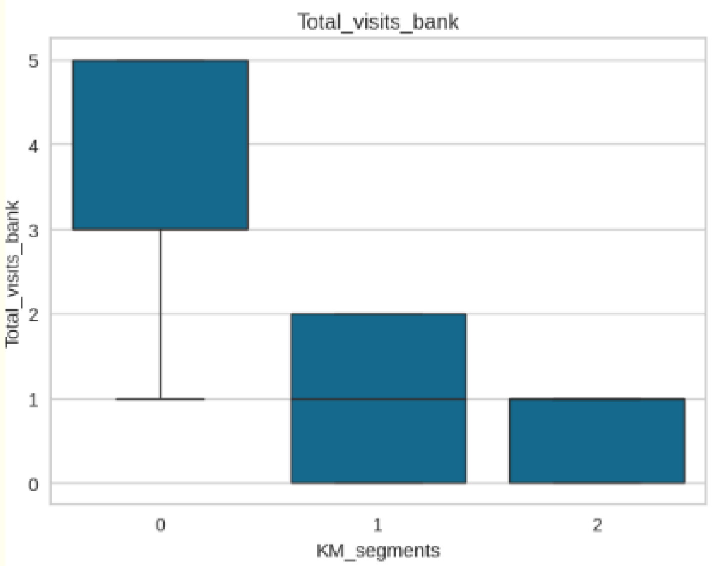
Count: 50 (Smallest segment)

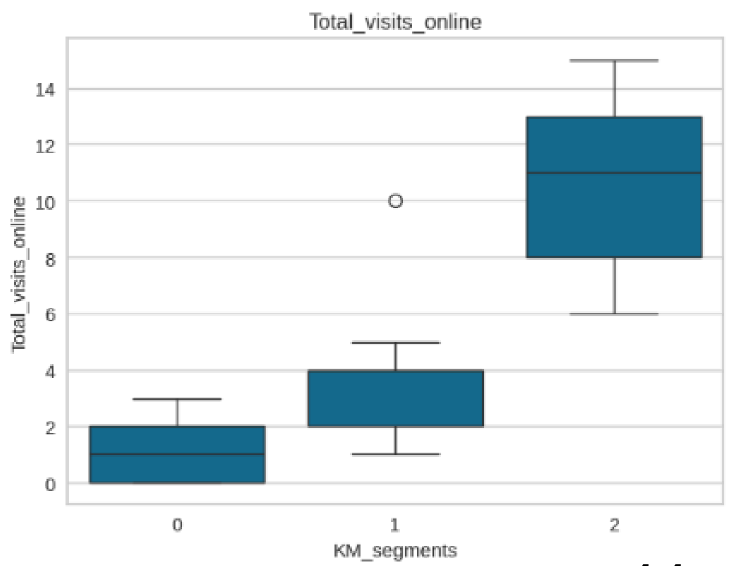
- Highest Credit Limit: ₹141,040
- Most Cards: 8.74
- Highest in Online Visits: 10.9
- Lowest in Bank Visits & Calls: 0.6 & 1.08

- **High-value, digitally fluent, self-service customers.**









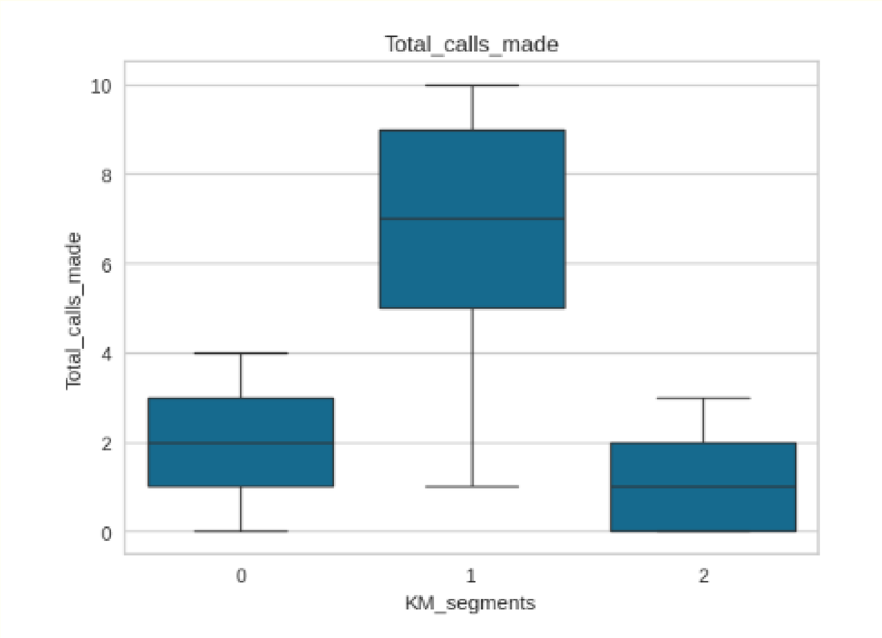


Figure 15: Plot for clusters

Insights

Cluster 0

Spending Behavior

- These customers have moderate credit limits — not low enough to be risk-averse, but not premium either.
- They carry a moderate-to-high number of credit cards, indicating some level of trust or engagement with credit.

Engagement Style

- Heavily branch-reliant: Their frequent bank visits suggest they prefer in-person interactions or are not tech-savvy.
- Low digital engagement: Rarely use online services, indicating limited adoption of mobile/internet banking.
- Moderate support interaction: Reasonable number of calls, likely to resolve doubts that could've been solved online.

Cluster 1

Spending Behavior

- Very limited credit access: These customers operate with small credit limits and minimal card ownership.
- Likely newer or lower-value customers from a revenue perspective.

Engagement Style

- Low branch usage, indicating either preference for remote interaction or lack of interest in services.
- High dependency on support calls, which suggests:
  - Frustration with services
  - Lack of understanding of products
  - Possible dissatisfaction

Cluster 2

Spending Behavior

- These are the top-tier customers: very high credit limits and large number of cards.
- Indicates strong trust, long-term engagement, and high revenue potential.

Engagement Style

- Strongly digital-first: almost all interactions are online, very few branch visits or phone calls.
- Low support dependency suggests comfort with self-service tools and high digital maturity.

Hierarchical Clustering

- Checking in each distance metrics "euclidean", "chebyshev", "mahalanobis", "cityblock"

- Appling Hierarchical clustering with different linkage methods("single", "complete", "average", "weighted")

Best result

- Highest cophenetic correlation is 0.89770, which is obtained with **Euclidean distance and average linkage.**

Checking Dendrograms

Dendrograms (Single linkage)

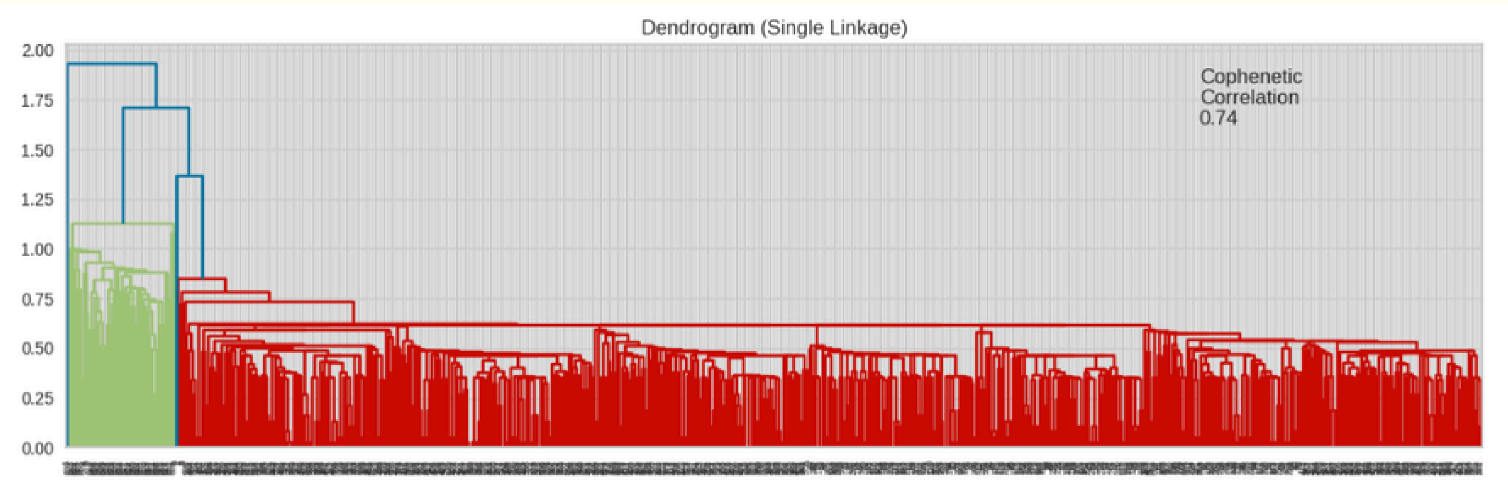


Figure 16: Dendrograms (Single linkage)

Dendrograms (Complete linkage)

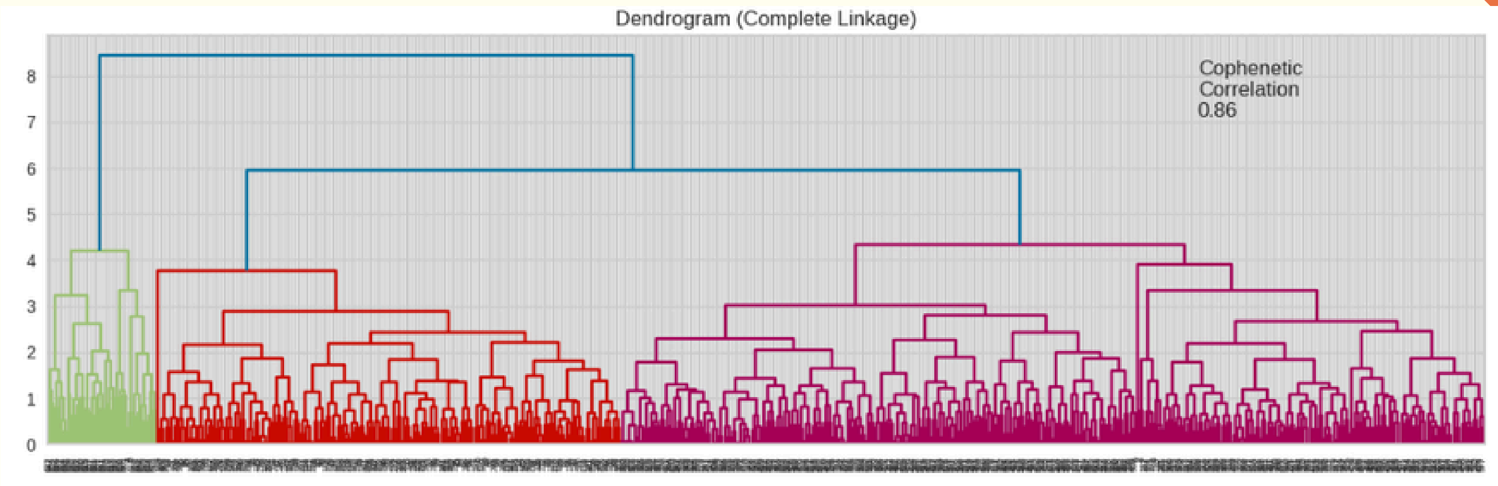


Figure 17: Dendrograms (Complete linkage)

Dendrograms (Average linkage)

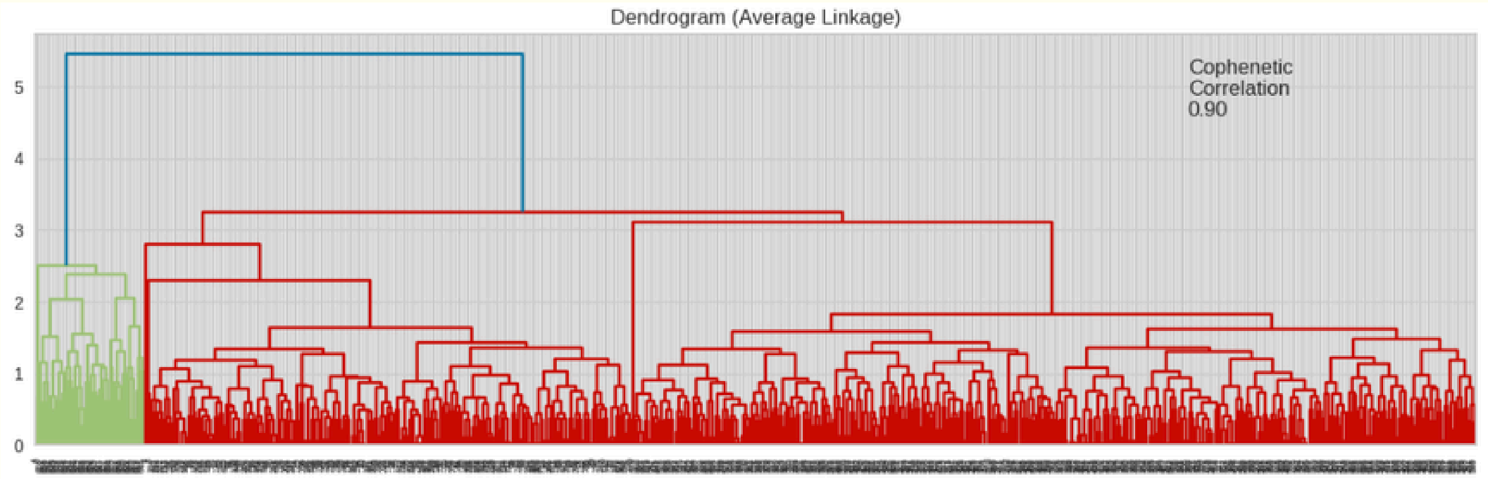


Figure 18: Dendrograms (Average linkage)

Dendrograms (Centroid linkage)

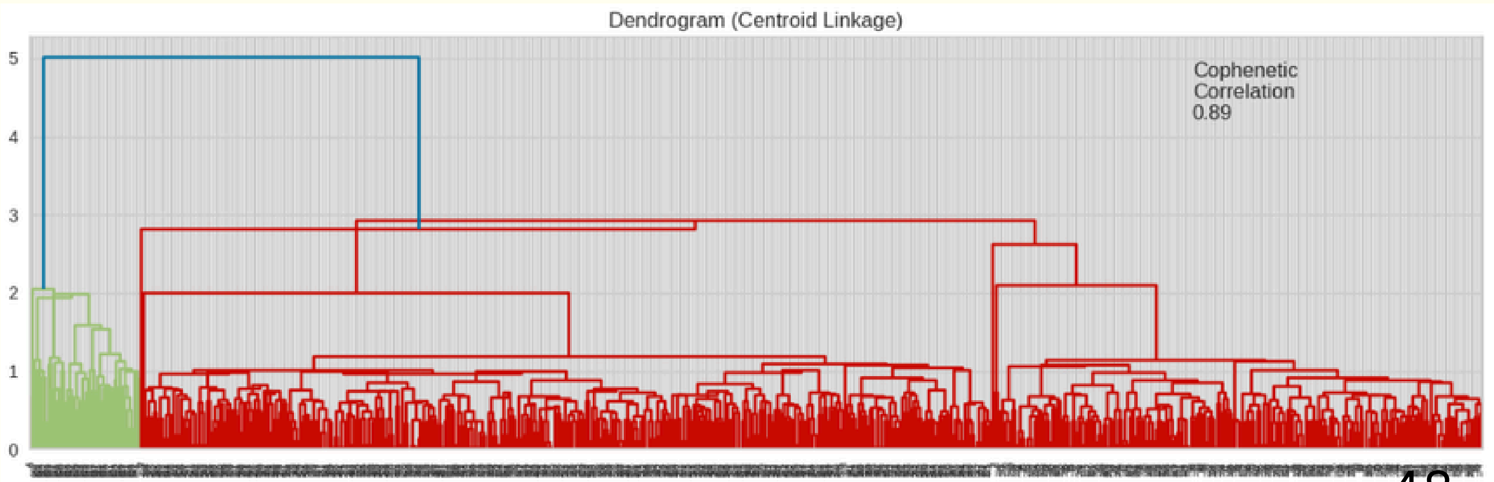


Figure 19: Dendrograms (Centroid linkage)

Dendrograms (Weighted linkage)

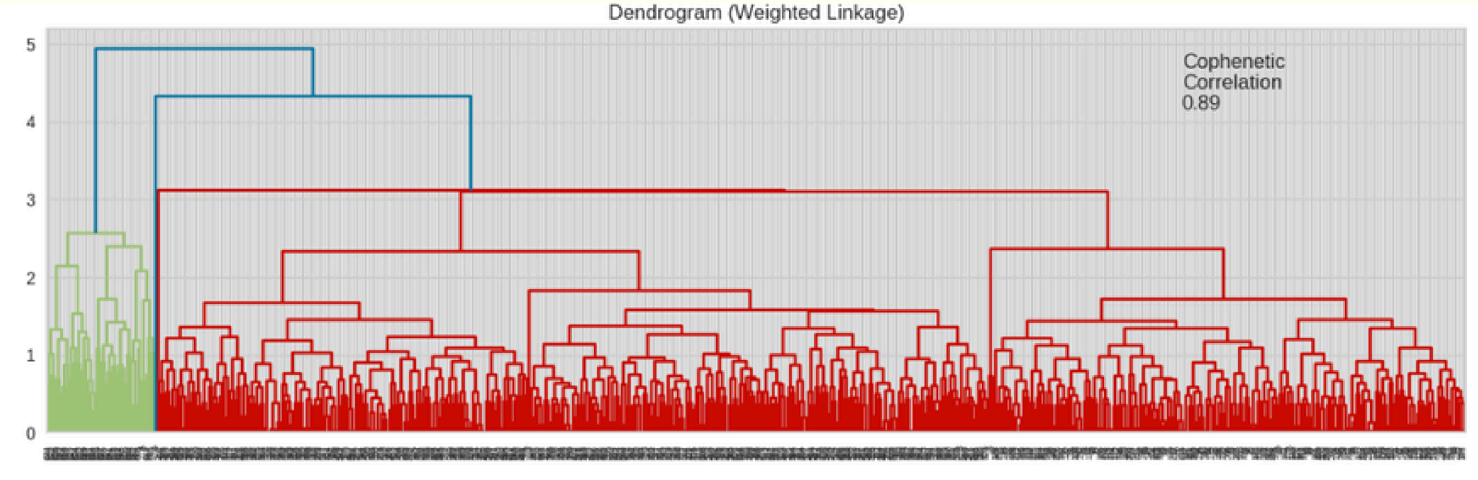


Figure 20: Dendrograms (Weighted linkage)

Dendrograms (Ward linkage)

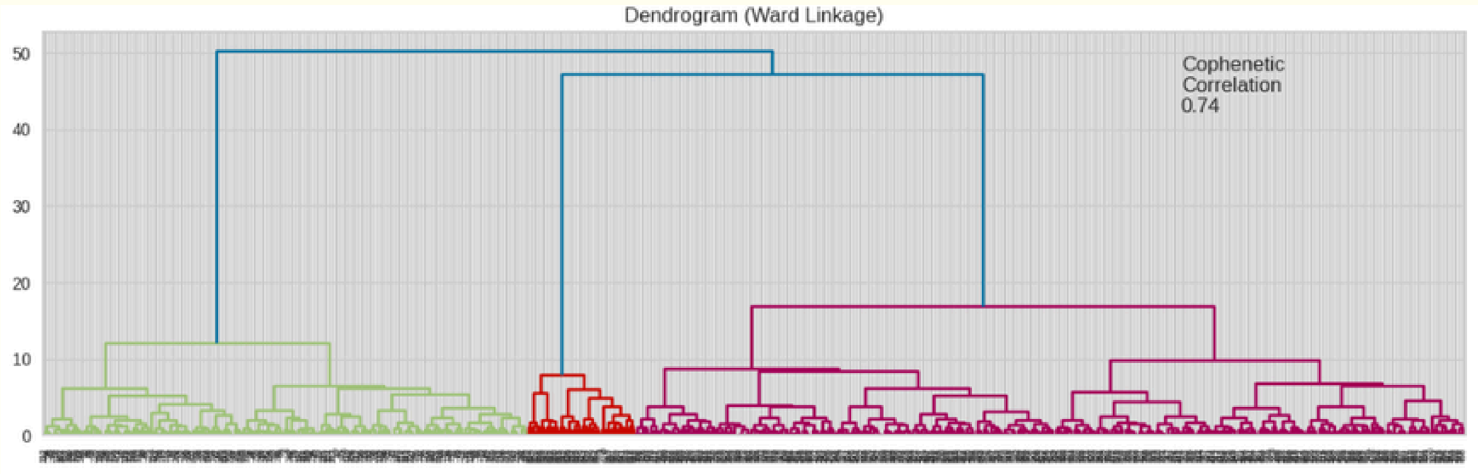


Figure 21: Dendrograms (Ward linkage)

Observations

- The cophenetic correlation is highest for average and centroid linkage methods.
- We will move ahead with average linkage.
- **3 appears to be the appropriate number of clusters from the dendrogram for average linkage.**

- **Not Choosing 4 and 5 clusters from average linkage because it splits 4th and the 5th which has only one customer for each clusters.**
- **So sticking to 3 clusters.**

Silhouette score

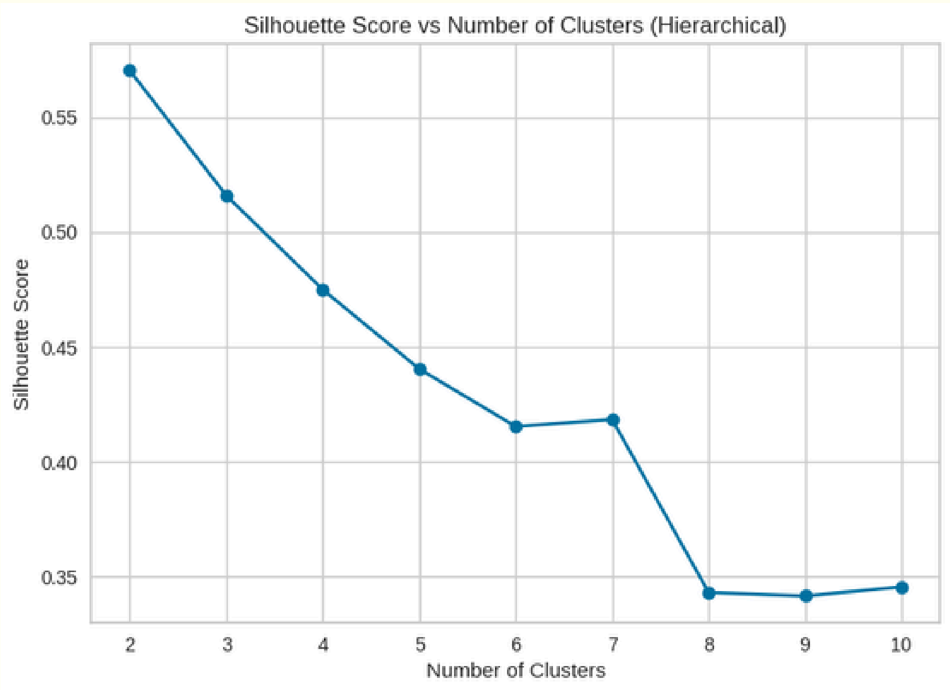


Figure 22: Silhouette score

- The silhouette score for 2 clusters is 0.5703 and the silhouette score for 3 clusters is 0.51592.
- Although 2 clusters scores are better segmenting into two cannot serve purpose so using 3 clusters to show detailed patterns in the dataset.

Cluster Profiling

HC_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	Cluster_Size
0	33713.18	5.51	3.49	0.98	2.01	387
1	141040.00	8.74	0.60	10.90	1.08	50
2	12197.31	2.40	0.93	3.56	6.88	223

Table 11: Cluster Profiling

Cluster 0 – Mid-Tier, Traditional Users (Majority Segment – 387 customers)

- Avg_Credit_Limit: 33,713
- Credit Cards: 5.51
- Bank Visits: High (3.49)
- Online Visits: Low (0.98)
- Calls Made: Moderate (2.01)

These customers are branch-reliant and moderately active financially.

Cluster 1 – High-Value Digital Users (Smallest Segment – 50 customers)

- Avg_Credit_Limit: ₹141,040 (Highest)
- Credit Cards: 8.74 (Highest)
- Bank Visits: Lowest (0.60)
- Online Visits: Highest (10.9)
- Calls Made: Lowest (1.08)

Highly digital, elite customers with strong self-service habits.

Cluster 2 – Low-Value, Support-Dependent Users (moderate Segment – 223 customers)

- Avg_Credit_Limit: ₹12,197 (Lowest)
- Credit Cards: 2.40 (Fewest)
- Bank Visits: Low (0.93)
- Online Visits: Moderate (3.56)
- Calls Made: Highest (6.88)

Low credit holders who heavily rely on support and may be new or struggling users.

The time taken for execution of the Hierarchical Clustering for the three clusters is 0.0025

Customers in cluster 0 : 385 (Customer Key)

87073 17341 40496 54838 35254 46635 97825 83125 35483 15129 83290 56486 31903 45909 14263 46813 81878 35549 85799 39122 81531 69965 18595 44398 32352 40898 27101 33457 45088 23302 27408 65372 21531 56843 17165 89328 20072 71402 47496 24808 17036 67193 34423 97109 55382 51811 53936 66504 53207 18514 51319 36340 36934 95925 49771 22919 21233 74544 52025 45652 73952 49418 77026 49331 75775 54906 94666 11698 34677 95610 41380 38033 85337 38994 67911 92956 77641 57565 53814 30712 19785 31384 16374 50878 78002 83459 91987 51552 72156 24998 45673 11596 87485 28414 81863 33240 11466 23881 44645 49844 92782 22824 26767 26678 50412 17933 34495 47437 22610 41159 64672 62483 85614 96548 19137 69028 70779 38244 67046 64897 46223 36628 17565 77381 11799 81940 66706 87838 94437 33790 44402 29886 66804 47866 61996 15318 89635 71681 71862 96186 22348 36243 88807 82376 98126 80347 17649 62807 92522 57459 44579 45476 61994 11398 24702 27824

45878 72431 19215 23409 16418 85122 55060 55478
65574 31113 96929 78912 68439 62864 31515 77954
88207 78618 31551 75792 29864 45440 97954 90189
55090 17703 33991 88884 45808 50706 92140 88123
53932 65908 25321 87456 48602 97530 48657 76209
49913 53002 61122 82807 93496 64519 31950 23110
96297 28408 37252 41287 52460 26604 58019 87219
36839 12663 48667 42887 14439 60851 41266 37438
65747 81166 20570 14816 11265 24980 37934 70707
84351 89446 17325 64774 53166 45341 94595 55170
92489 92933 36504 40508 15798 70101 77613 84360
48402 46776 67258 44804 29919 65781 12456 62649
74446 36632 76024 75065 51682 18397 29102 56367
95147 44379 76957 42921 23102 61324 49690 20043
44144 53552 62530 41741 22842 65825 77826 61216
83192 82023 73000 64550 90131 17382 27117 94529
21717 81910 76492 43000 48692 27476 15086 43034
99131 13140 99437 91242 39285 63710 90860 35585
58708 57451 69868 43679 30256 26334 47848 17377
39644 29176 55706 51771 83585 51867 68040 75417
34775 85645 83545 44157 38125 75398 90999 70376
33295 80942 26493 97850 43841 79885 59316 83466
81510 35268 11734 88411 96269 87683 26063 42479
58116 67282 84888 75366 14377 59074 96534 31870
24748 68920 67637 60839 59170 90586 56270 87670
47703 35421 58511 76398 93310 36836 46373 94700
67860 99473 68862 93381 46548 74083 48660 13720
72339 99284 47198 67415 44403 58276 85234 31948
90191 49341 11562 16253 80623

Customers in cluster 2 : 223

38414 58634 37376 82490 44770 52741 52326 92503
25084 68517 55196 62617 96463 39137 14309 29794
87241 30507 61061 24001 68067 65034 14854 81130
29112 13999 32550 82164 61517 28254 30888 46388
74126 52142 37659 83132 20119 52363 50769 68502
99026 83326 62040 37252 74625 51182 60301 96386
43886 78503 68419 32828 17937 71632 81566 29759
36929 70248 91673 61355 60403 85868 76205 66524
69214 21976 35149 27120 18821 33187 93482 90168 71881
59656 12026 99589 38970 57990 39447 79694 79403
47296 37559 38165 49198 18007 59619 37016 91099
74704 25742 11937 52736 88338 18916 92501 96213
26599 73007 97935 26089 14946 74795 73435 41634
84069 83244 87291 18086 33369 15310 98499 35256
89007 93997 16577 25440 81116 63663 69811 36111
39454 70199 11602 49697 28701 61627 34103 14248
31256 45583 52750 95507 23743 53410 53898 66200
58389 61347 59151 37802 60475 95489 77758 23768
87471 85707 97951 54785 97011 35103 18564 61009
24054 63751 52758 78473 80457 59783 64241 32374
97536 33110 36978 54281 98602 97687 28842 38410
38261 20524 37671 25330 41787 11412 55892 95495
41946 86410 76718 98969 77143 38205 53851 52783
63405 48510 97463 18145 14398 98288 69704 29058
15546 16715 87350 13215 20593 56624 33317 99596
72430 16676 40486 90958 67212 44226 94251 61776
55275 18609 54477 12122 28208 68003 79632 73811
72892 51773 96163 61234 55849 56156

Customers in cluster 1 : 50

47437 48370 94391 50598 40019 77910 89832 98216 54495 47650 32107 84192 53916 32584 97285 20337 15585 20620 75009 76203 33837 14916 97935 16180 49493 70974 40217 88442 17538 90839 99843 27212 91575 60190 18519 48762 58392 79953 13315 30570 78996 78404 28525 51826 65750 51108 60732 53834 80655 80150

Visual representation of clusters

HC_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	33713.178295	5.511628	3.485788	0.984496	2.005168	387
1	141040.000000	8.740000	0.600000	10.900000	1.080000	50
2	12197.309417	2.403587	0.928251	3.560538	6.883408	223

Table 12: clusters values

Insights

Cluster 0 – Traditional Mid-Tier Customers

Count: 386 (Largest segment)

- Avg Credit Limit: ₹33,713
- Total Credit Cards: 5.51
- Most visits to bank (3.49)
- Lowest online usage (0.98)
- Moderate call activity (2.00)

- **Users prefer branch interaction, have average credit capacity, and are moderately engaged.**

Cluster 1 – Premium Digital Natives

- Count: 50 (Smallest segment)
- Highest Avg Credit Limit: ₹141,040
- Most Credit Cards: 8.74
- Highest online engagement: 10.90
- Least bank visits (0.6)
- Lowest calls (1.08)

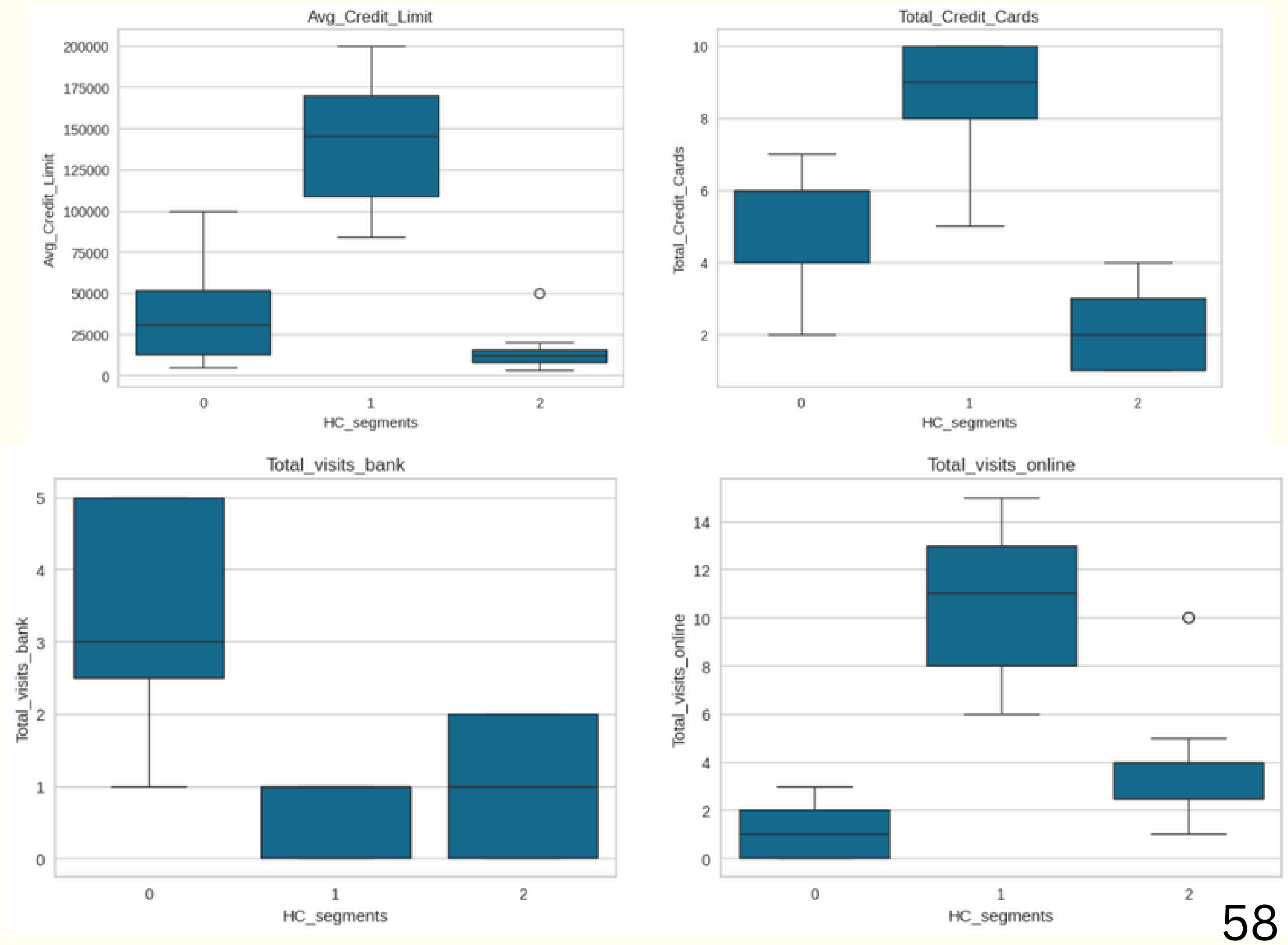
- **Affluent, tech-savvy customers who prefer self-service.**

Cluster 2 – Low-Value, Support-Heavy Users

Count: 223(Mid segment)

- Avg Credit Limit: ₹12,197 (Lowest)
- Credit Cards: 2.40 (Fewest)
- Moderate online (3.56)
- Low bank visits (0.93)
- Most calls made (6.88)

- **These users rely heavily on support, possibly due to limited experience or comfort with banking tools.**



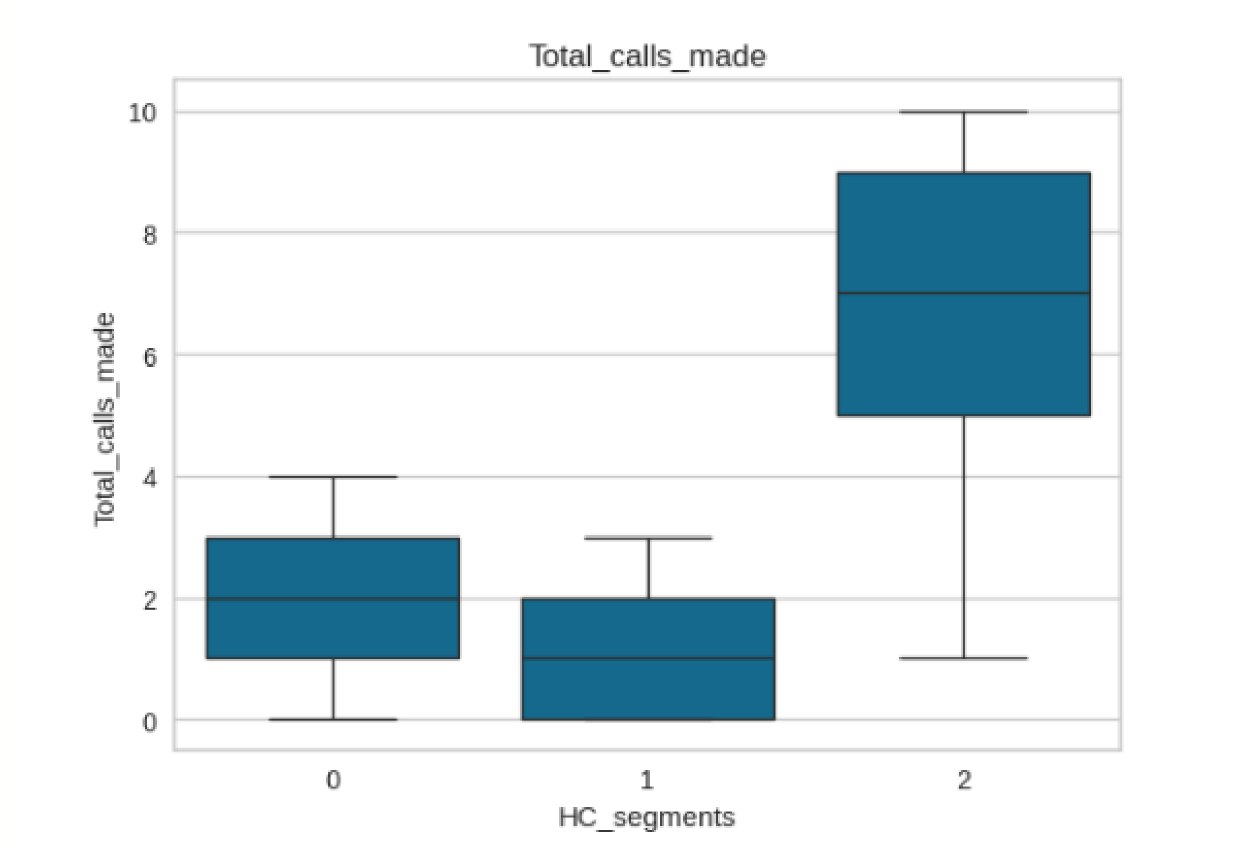


Figure 23: Plot for clusters

Insights

Cluster 0

Spending Behavior

- These customers are middle-of-the-road spenders: not premium, but not low-value either.
- They possess multiple credit cards, showing moderate engagement with financial products.

Engagement Style

- Strong reliance on in-person banking — the highest number of branch visits.
- Extremely low digital interaction, indicating either:
  - Preference for face-to-face service
  - Low comfort with digital channels
- Moderate number of support calls — may seek assistance when self-service isn't viable.

Cluster 1

Spending Behavior

- These are your top-tier customers with very high credit limits and the highest number of credit cards.
- Likely experienced, trusted, and financially mature users who handle multiple financial products with ease.

Engagement Style

- Digital-first: They strongly prefer online interactions with the bank and are confident in managing their accounts digitally.
- Minimal support dependency: Few calls made, indicating self-reliance and high comfort with systems.
- Almost no in-branch visits, reinforcing their preference for seamless remote banking.

Cluster 2

Spending Behavior

- These are low-credit customers with minimal card usage, likely new or risk-averse.
- Their limited credit limit suggests they’re not targeted with premium financial products yet.

Engagement Style

- Heavy reliance on support calls: They reach out frequently, possibly due to:
- Difficulty understanding services

- Lack of digital literacy
- Frustration or confusion
- Moderate online usage shows they may have started using digital tools but aren't fully comfortable.
- Occasional branch visits round out a mixed-mode engagement style, neither fully digital nor traditional.

K-means vs Hierarchical Clustering

KM_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	33782.383420	5.515544	3.489637	0.981865	2.000000	386
1	12174.107143	2.410714	0.933036	3.553571	6.870536	224
2	141040.000000	8.740000	0.600000	10.900000	1.080000	50

HC_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	33713.178295	5.511628	3.485788	0.984496	2.005168	387
1	141040.000000	8.740000	0.600000	10.900000	1.080000	50
2	12197.309417	2.403587	0.928251	3.560538	6.883408	223

Table 13: K-means vs Hierarchical Clustering

- The customers in cluster 2 of K-means is similar to the cluster 1 of hierarchical clustering. So changing the clusters name of the K means.

The execution time for the k-means is 00032 and the execution time for hierarchical clustering is 0.0025. So choosing hierarchical clustering for better execution.

Comparison

K-means clustering	Hierarchical clustering
The execution time for the k-means is 00032	The execution time for hierarchical clustering is 0.0025.
K-means clusters gave same clusters like Hierarchical clustering	Hierarchical clustering clusters gave same clusters like K-means
For n_clusters = 2, the silhouette score is 0.5703 and For n_clusters = 3, the silhouette score is 0.5157	For n_clusters = 2, the silhouette score is 0.5703 and For n_clusters = 3, the silhouette score is 0.5159
There are 384 customer is cluster 0, 224 customers in cluster 1, 50 in cluster 2.	There are 385 customer is cluster 0, 50 customers in cluster 1, 223 in cluster 2.
Three number of clusters are obtained for k-means	Three number of cluster are obtained for Hierarchical clustering

Table 14: Cluster Profiling

K-means vs Hierarchical Clustering after changing the cluster row

KM_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	33782.383420	5.515544	3.489637	0.981865	2.000000	386
1	141040.000000	8.740000	0.600000	10.900000	1.080000	50
2	12174.107143	2.410714	0.933036	3.553571	6.870536	224

HC_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segment
0	33713.178295	5.511628	3.485788	0.984496	2.005168	387
1	141040.000000	8.740000	0.600000	10.900000	1.080000	50
2	12197.309417	2.403587	0.928251	3.560538	6.883408	223

Table 15: K-means vs Hierarchical Clustering after changing the cluster row

- The customers in the clusters are same after swapping the clusters.

- There are 384 customers in cluster 0
- There are 234 customers in cluster 2.
- There are 50 customers in cluster 1.

Choosing Hierarchical Clustering because it has better execution time, better segmentation and better silhouette score (0.5159).

PCA for Visualization

Not applying PCA because the dataset is small. Using only for visualization.

- Assigning the columns in two principal components to to know the variance in the dataset.
- The first two principal components explain 83.16% of the variance in the data.

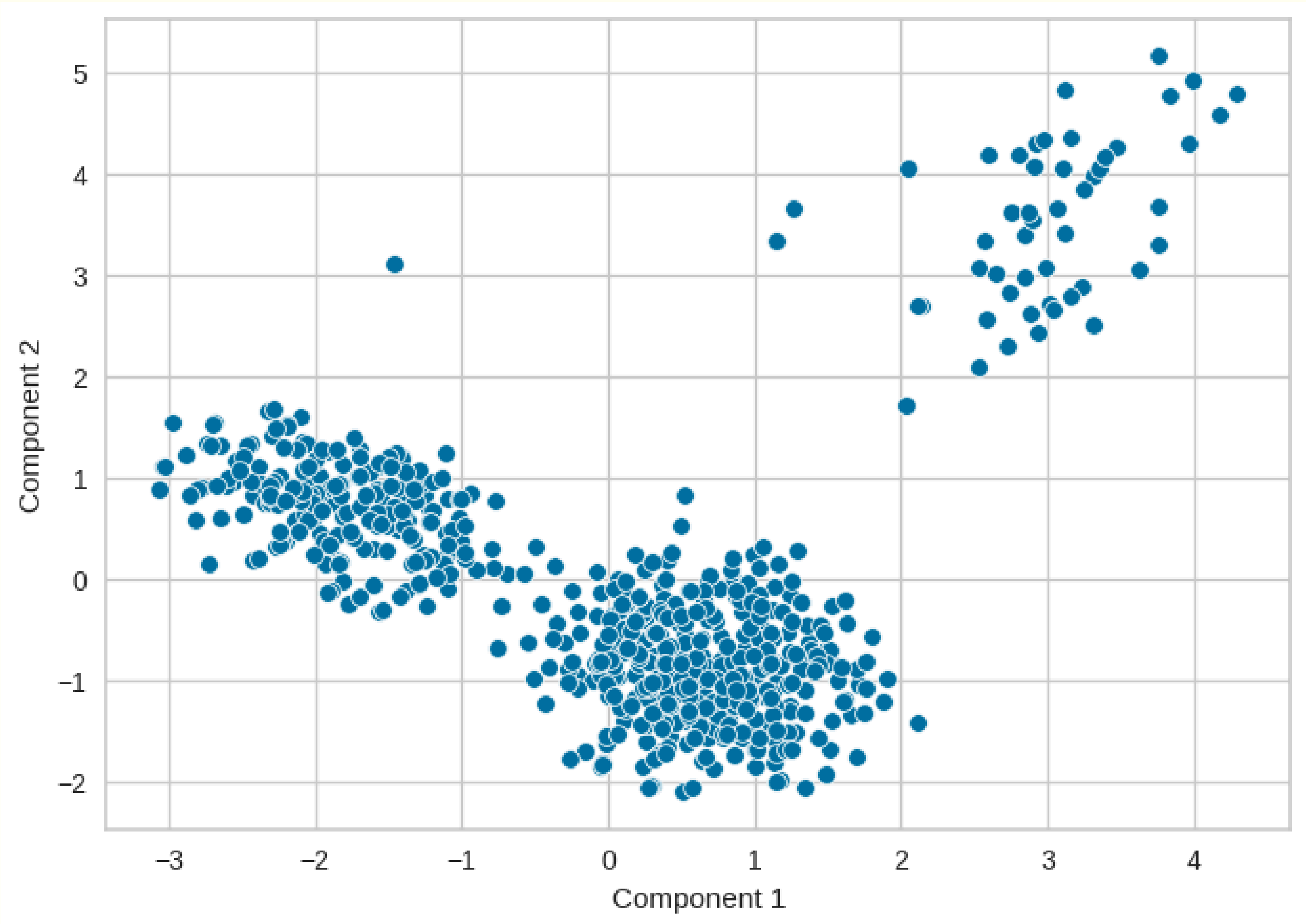


Figure 24: PCA for Visualization

Clear Cluster Separation:

- The data points form three visibly distinct clusters, suggesting a natural grouping in the data.
- This visually supports the three-cluster segmentation shown earlier in the KM and HC segment tables.

Cluster Density & Spread:

- One cluster (leftmost) is compact and dense, indicating similar behavior or feature values among its members.
- Another (center) is also dense but more spread out, showing more variability within.
- The third (top-right) is more scattered, possibly representing outliers or a small premium group (matches with Segment 2 in profiling).

Implication on Segmentation Validity:

- The clear visual separation reinforces that the clustering algorithm (especially Hierarchical Clustering) is effective in segmenting this dataset.
- This validates the 3-segment customer segmentation strategy.

So choosing Hierarchical Clustering to segment the dataset into 3 clusters.

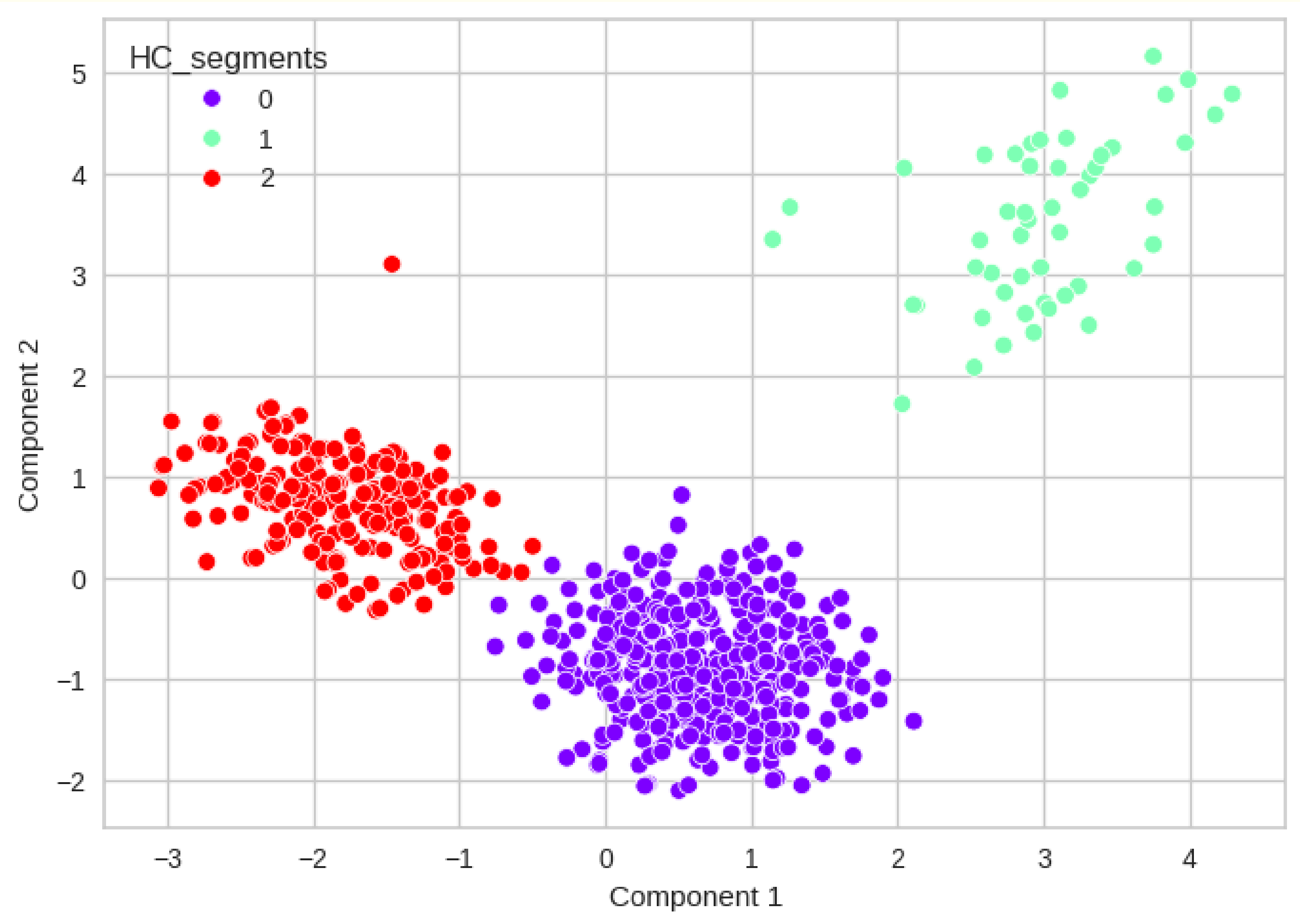


Figure 25: Segmentation of the customers

Insights

Segment 0 (Purple) — Mass Market Users

- Largest cluster in terms of population.
- Compact and dense, indicating highly similar users.
- From the profiling table:
- Moderate in credit limit (~33,700) and credit cards (~5.5).
- Moderate bank visits (~3.5), low online (~0.98), and low calls (~2.0).
- This cluster likely represents a stable, homogeneous group. In a customer segmentation context, these might be:
- Loyal customers with consistent behavior

- Similar income, usage pattern, or purchase frequency

🟢 **Segment 1 (Light Green) — Premium/Digital Users**

- Smallest cluster, clearly distinct and separate.
- Appears more spread (behaviorally diverse).
- From the table:
  - Highest credit limit (₹141,040) and most credit cards (~8.7).
  - Highest online usage (10.9 visits), lowest in calls and bank visits.
- Possibly a distinct, niche group
- Could represent high-value or unusual behavior (e.g., VIPs or outliers)
- Also might indicate less data density in this segment

🔴 **Segment 2 (Red) — Support-Seeking or Low Value Users**

- Moderate-sized cluster.
- Slightly dispersed, indicating moderate variability.
- From profiling:
  - Lowest credit limit (~₹12,197) and card ownership (~2.4).
  - High call frequency (~6.8), lowest in online usage.
- Also a cohesive group but with different characteristics than Cluster 0. Possible interpretations:
- A second group of customers with lower spending or engagement
- Different preferences or demographics

Actionable Insights & Recommendations

Best Cluster for Focus: Segment 0 (Purple - Mass Market Users)

- Large in size (387 customers) — represents the majority of the customer base.
- Moderate in credit card ownership (~5.5) and credit limit (~₹33.7k) — clear potential for upselling.
- Moderate usage of physical banking and low digital engagement — scope to nudge toward digital behavior.
- Likely under-leveraged in terms of product usage and engagement — low-hanging fruit for marketing.

Recommended Actions:

- Run targeted upsell campaigns for additional credit cards or card-linked services.
- Introduce educational campaigns or offers to promote digital adoption (app usage, online transactions).

Support Service Improvement Focus: Segment 2 (Red - Traditional/Support-Seeking)

- High call volume (~6.8) but low credit value — indicates frustration or support dependency.
- Small to medium in size (~223 customers) — manageable to address with targeted interventions.
- Poor digital and bank visit engagement — possibly less tech-savvy or unaware of services.

Recommended Actions:

- Improve call center responsiveness and provide multi-channel support (chatbots, WhatsApp, email).
- Offer onboarding help for self-service tools, tutorials, or agent-assisted demos.
- Monitor call patterns to identify and proactively resolve common pain points.

- **Nurture, Not Target for Scale: Segment 1 (Light Green - Premium/Digital)**
- Already high credit card usage (8.74), online visits (10.9), and low service needs.
- Likely happy customers — focus on retention and exclusive perks, not expansion.