ALLLIFE BANK – PROJECT 7

SARAH A. THOMAS

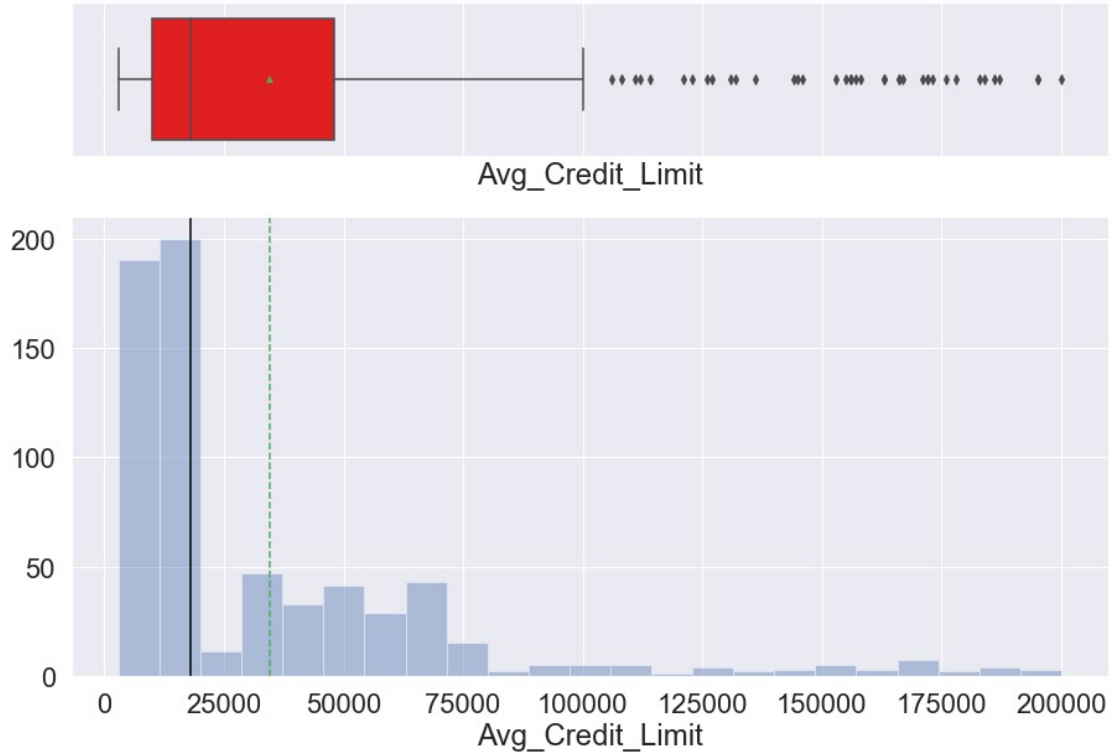
BUSINESS PROBLEM OVERVIEW AND SOLUTION APPROACH

It has come to the attention of the bank through the Marketing Research team, that penetration in the market can be improved. The Marketing team would like to create new targeted campaigns. In addition, the Marketing team found that service is perceived poorly. Thus, the Operations team would like to improve service. Towards all these efforts, the Data Science team has created segmentation models through the K-Means and Hierarchical Clustering algorithms using Python.

DATA OVERVIEW

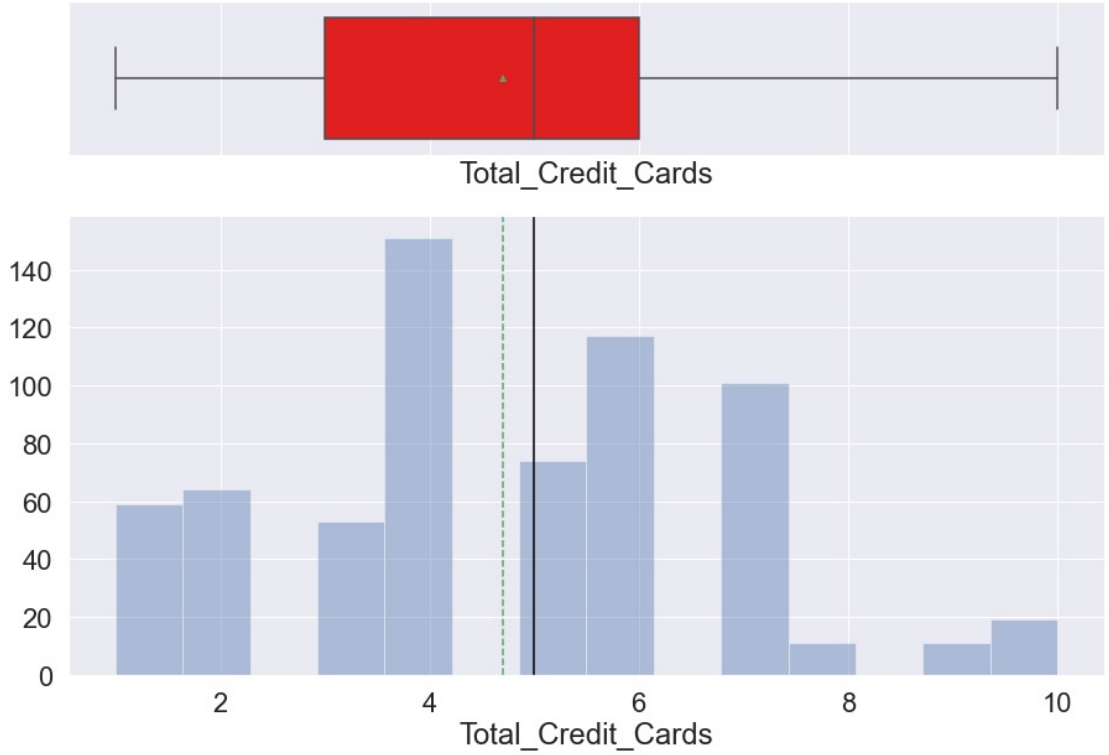
Variable	Description
Sl_No	Serial Number – Primary key of the records
Customer Key	Customer identification number
Average Credit Limit	Average credit limit for each customer for all credit cards
Total Credit Cards	Total number of credit cards for each customer
Total Visits Bank	Total number of yearly visits in person to the bank made by the customer
Total Visits Online	Total number of yearly visits made online by the customer
Total Calls Made	Total number of calls made by the customer to the bank

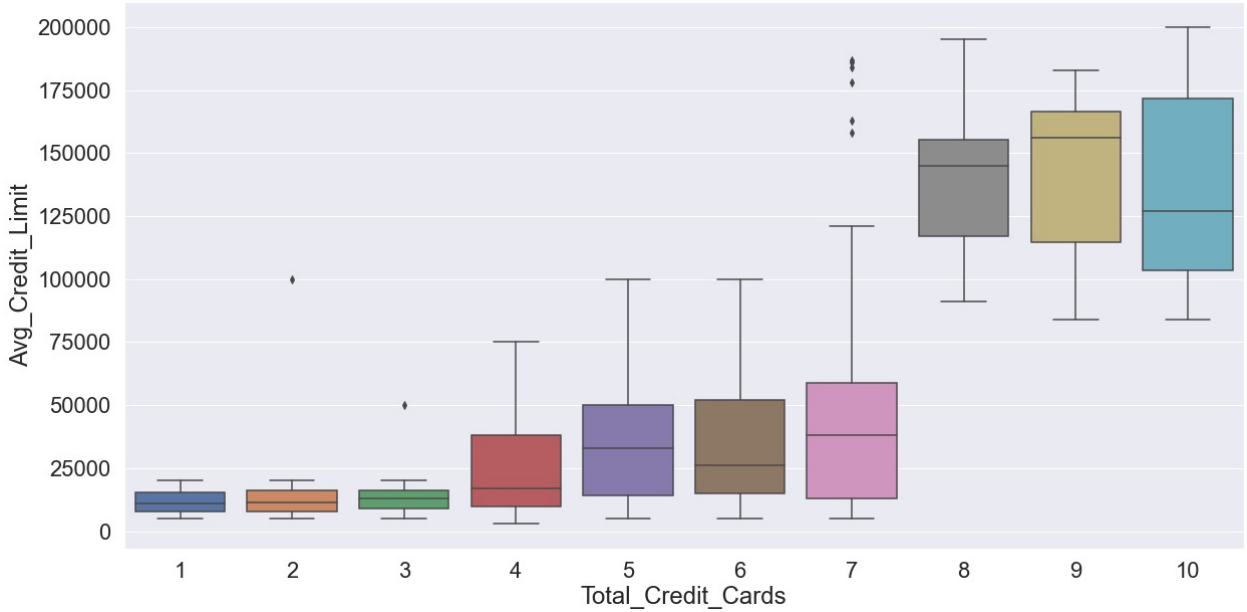
AVG_CREDIT_LIMIT



* Avg_Credit_Limit: Mean (34,574.24) is greater than median (18,000.00), indicating a right-skewed distribution.
* Outliers exist to the right and are treated.

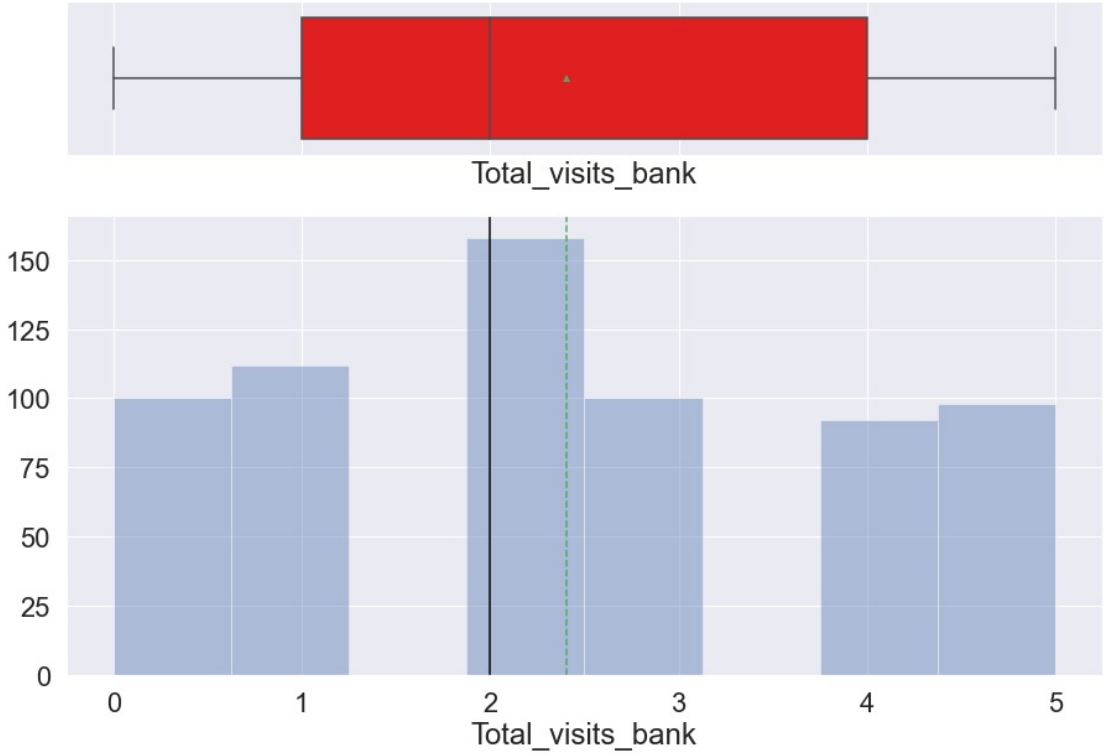
TOTAL_CREDIT_CARDS

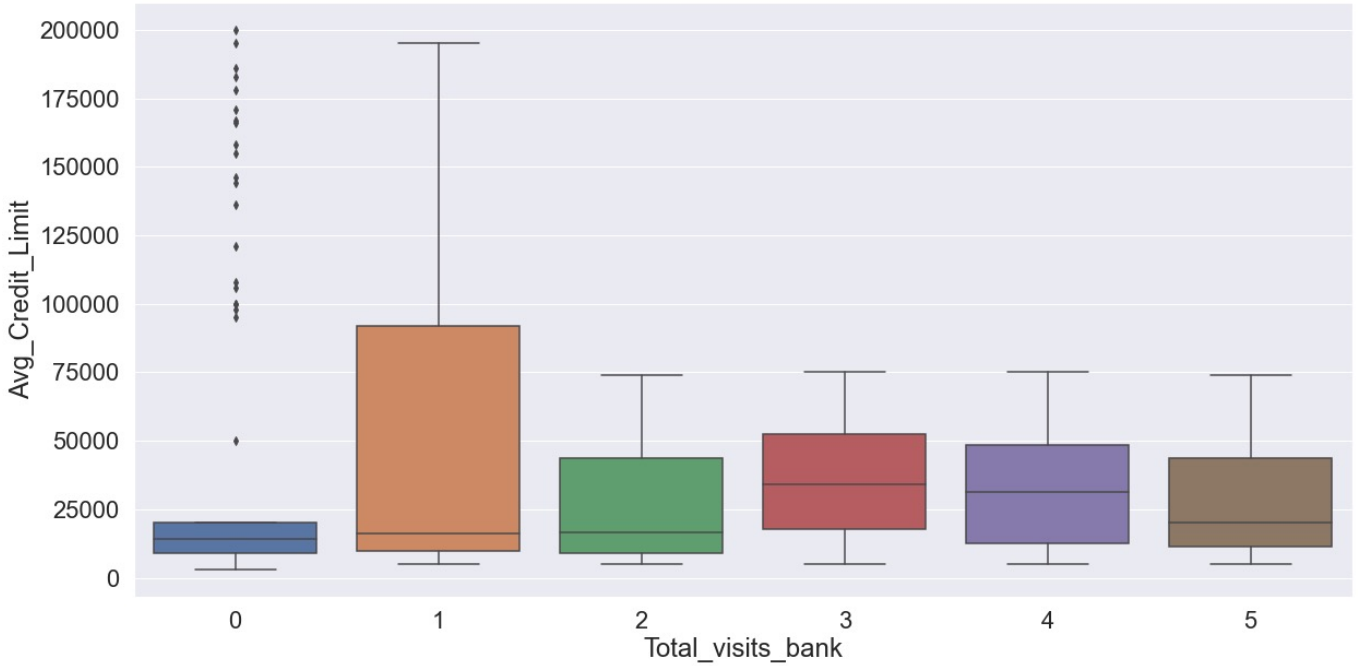




* Total_Credit_Cards: Median (5.0) is greater than mean (4.71), indicating a slightly left-skewed distribution.
* As would be expected, the average credit limit increases with the number of total credit cards.

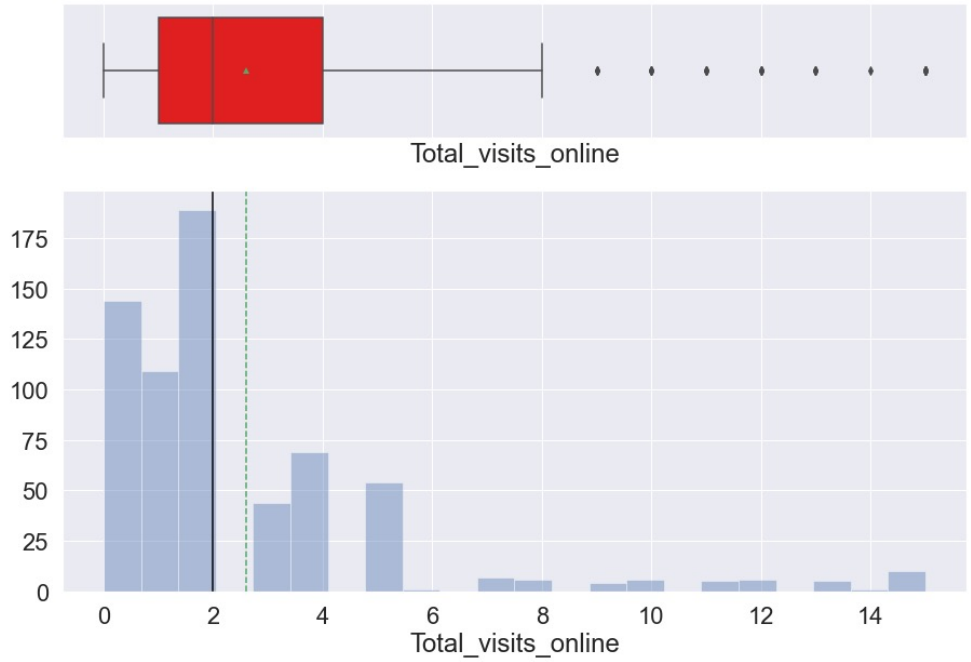
TOTAL_VISITS_BANK

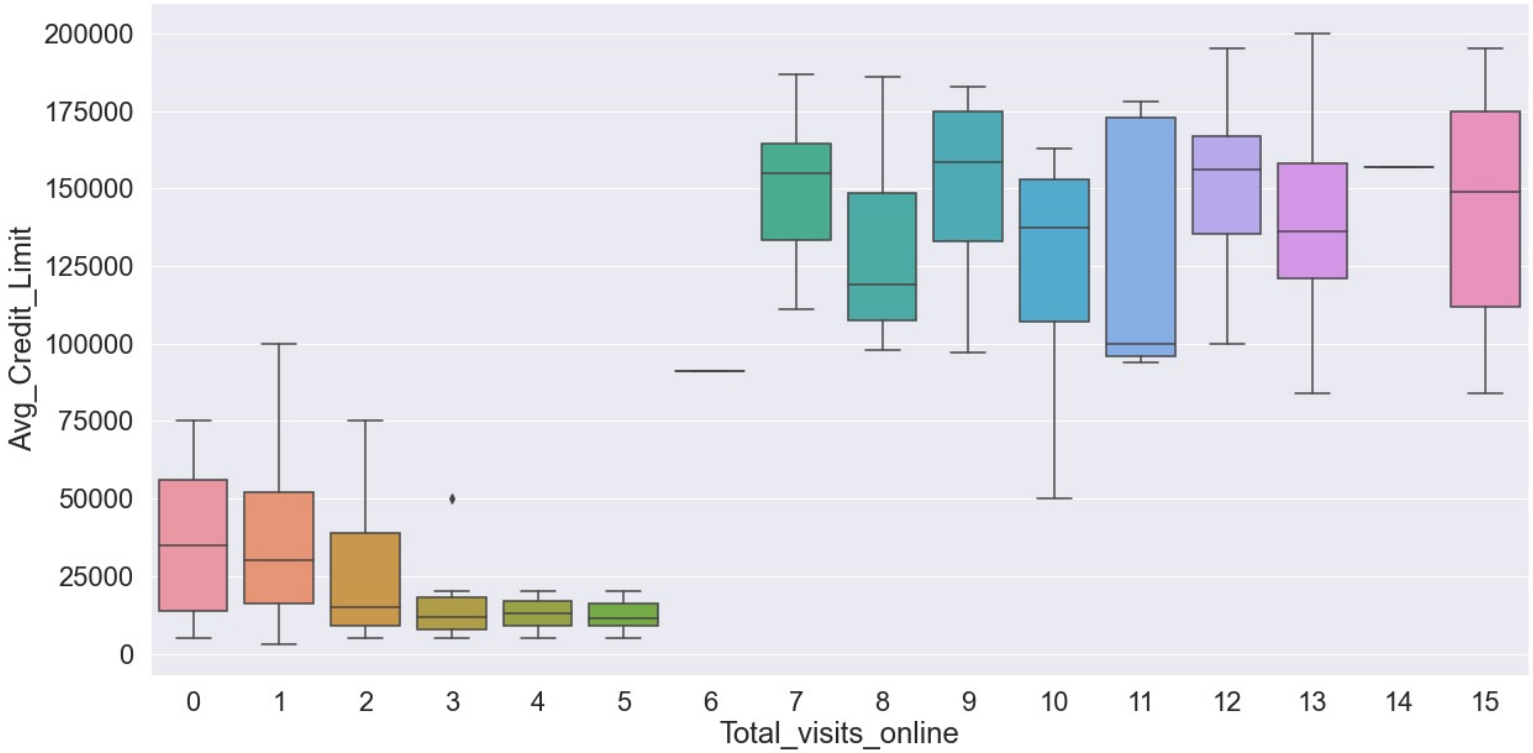




- Total_visits_bank: Mean (2.40) is greater than mean (2.0), indicating a right-skewed distribution.
- Interestingly, those with higher credit limits visit the bank less.

TOTAL_VISITS_ONLINE

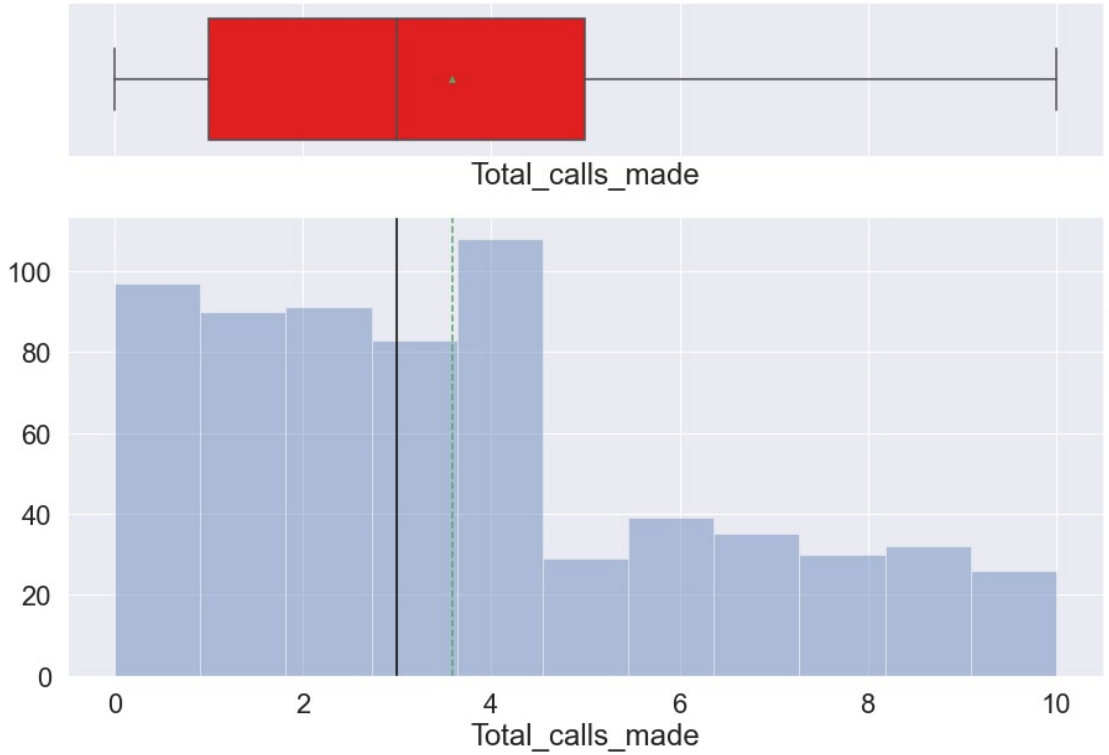


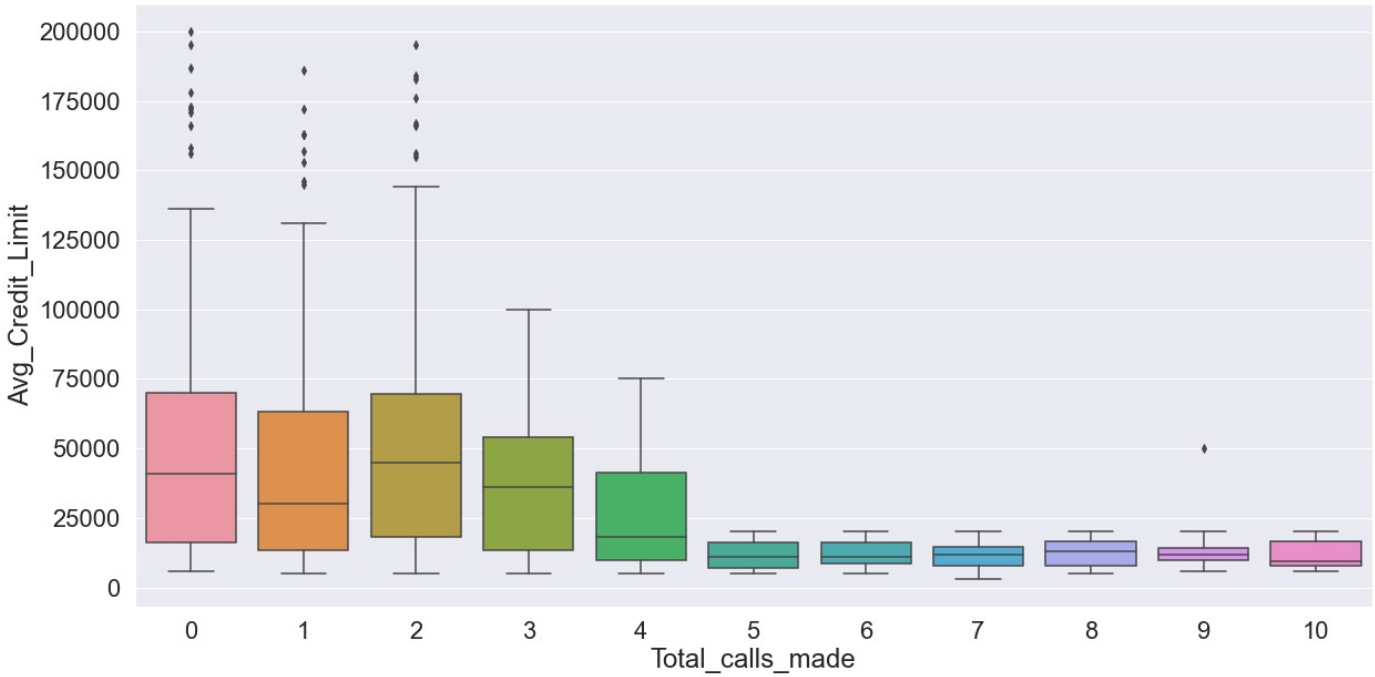


* Total_visits_online: Mean (2.61) is greater than median (2.0), indicating a right-skewed distribution.

- Outliers exist to the right but the values are close to the other values so are not treated.
- Those with higher credit limits do online banking at a higher rate.

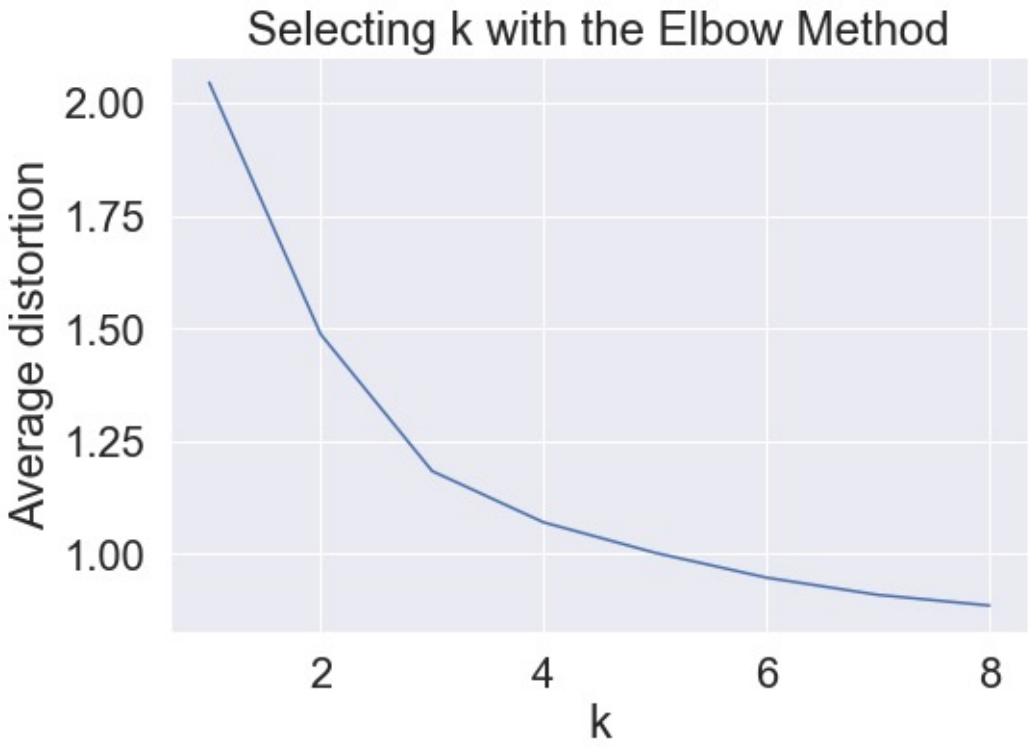
TOTAL_CALLS_MADE

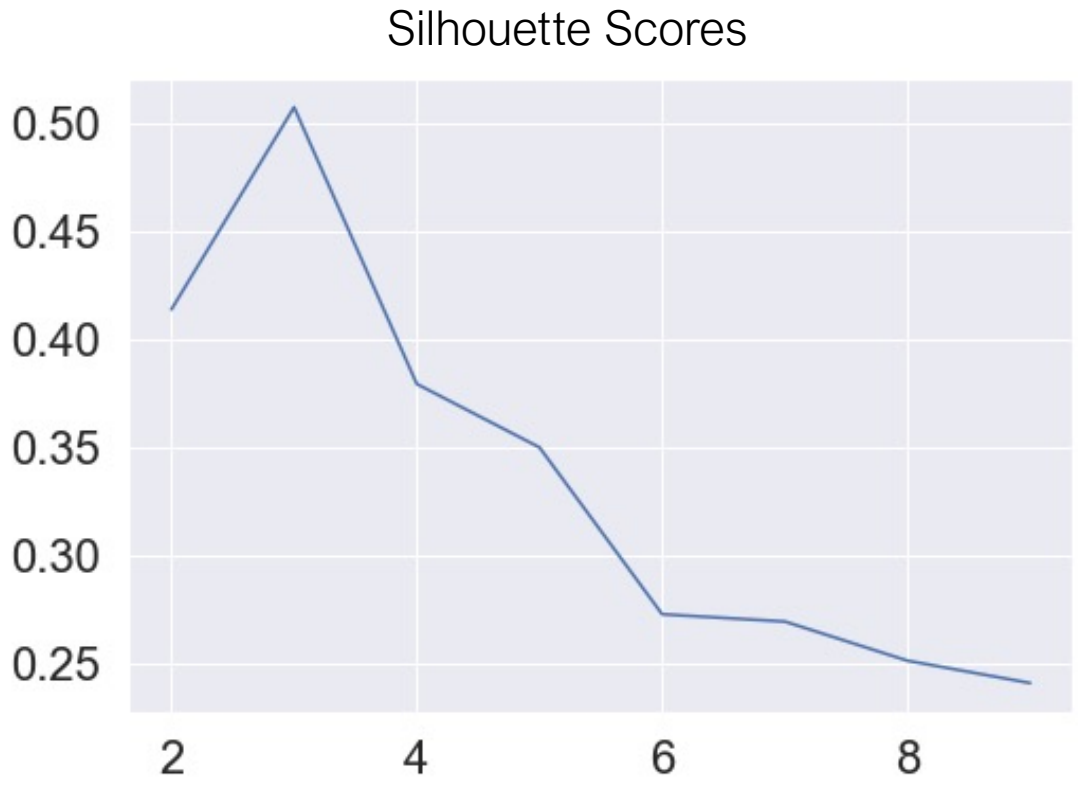




- Total_calls_made: Mean (3.58) is greater than median (3.0), indicating a right-skewed distribution.
- Those with lower credit limits made more calls to the bank.

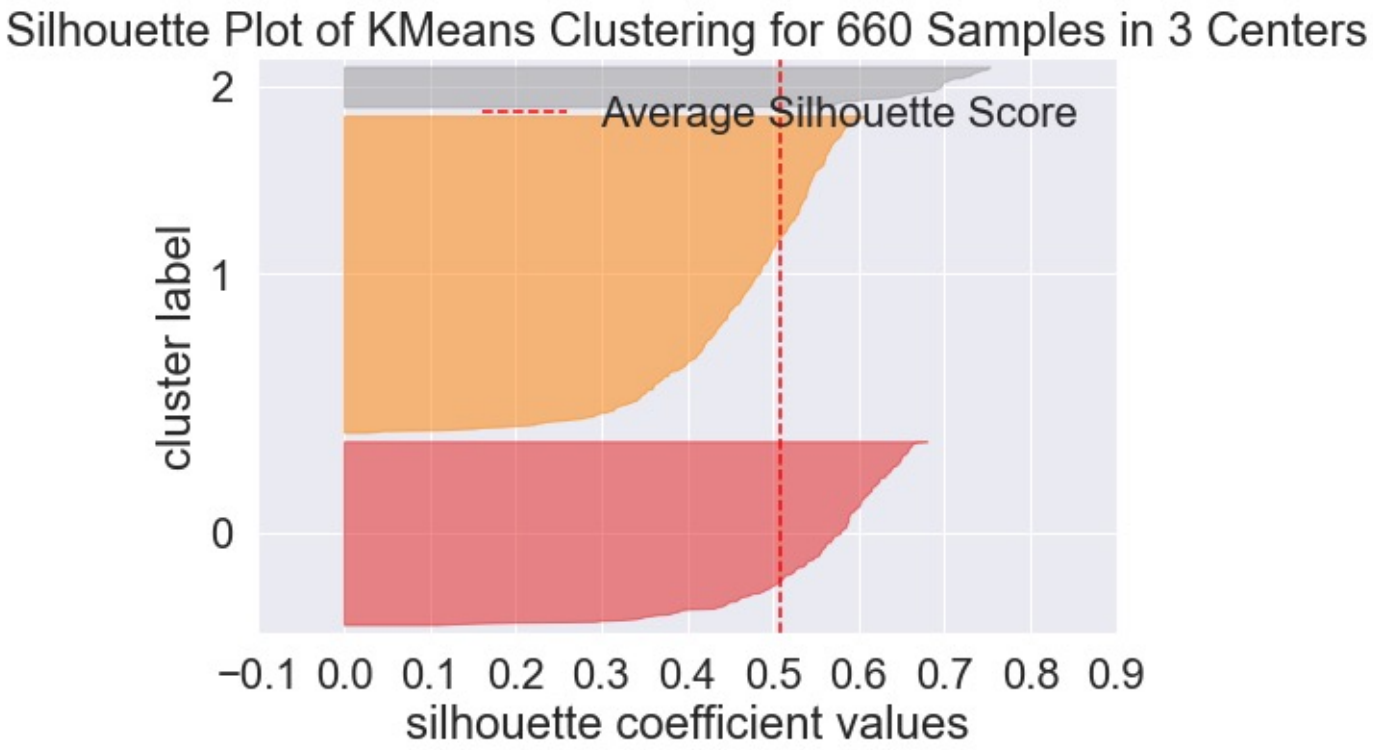
K-MEANS CLUSTERING





Based on the Elbow Method and Silhouette Scoring, we chose 3 as the number of clusters.

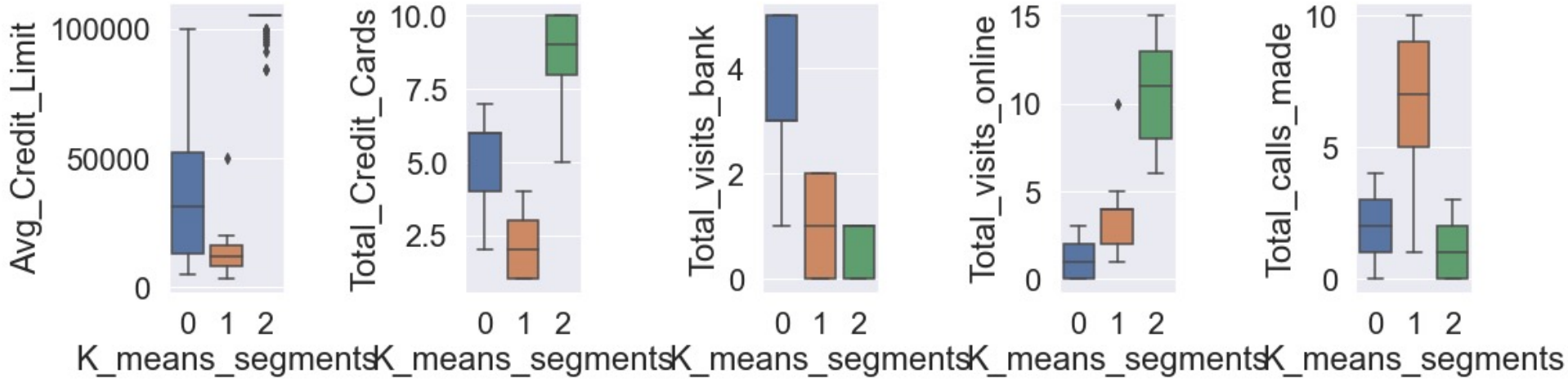
K-MEANS CLUSTERING



K-MEANS CLUSTER PROFILING

K_means_segments	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
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	102660.000000	8.740000	0.600000	10.900000	1.080000	50

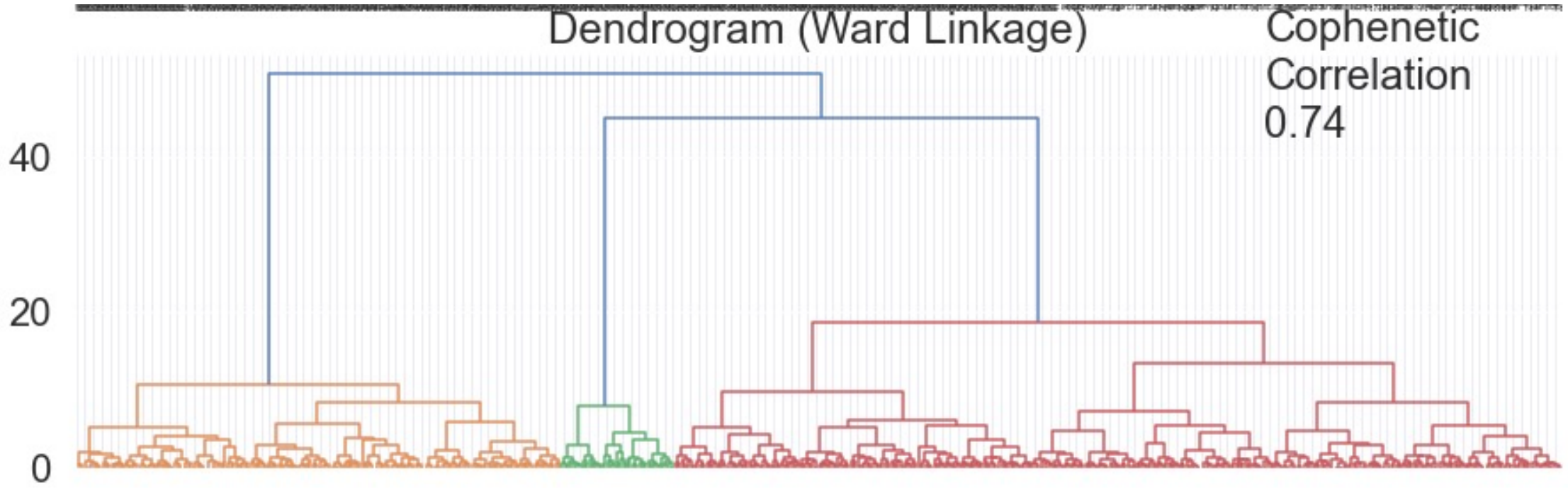
Boxplot of original numerical variables for each cluster



Insights for K-Means Clustering

- **Cluster 0:**
  - Avg_Credit_Limit: median is about $25,000 (approx. range $6,250-$100,000). This seems to be the middle range of all the clusters.
  - Total_Credit_Cards: median is about 6 (approx. range 2-7). This seems to be the middle range of all the clusters.
  - Total_visits_bank: values range from approx. 1-6 in this cluster, more than in the other clusters.
  - Total_visits_online: values range from approx. 0-3, less than in the other clusters.
  - Total_calls_made: values range from approx. 0-4. This seems to be the middle range of all the clusters.
- **Cluster 1:**
  - Avg_Credit_Limit: range is approx. $5,000-$12,500, the lowest range of all the clusters. (outlier(s) at $50,000)
  - Total_Credit_Cards: range is approx. 1-3, the lowest of all the clusters.
  - Total_visits_bank: range is approx. 0-2. This seems to be the middle range of all the clusters.
  - Total_visits_online: range is approx. 1-5 (with outlier(s) at 10). This seems to be the middle range of all the clusters.
  - Total_calls_made: range is approx. 1-10, the highest of all the clusters.
- **Cluster 2:**
  - Avg_Credit_Limit: Outliers exist ranging from approx. $87,500-$100,000. Primary values are above the $100,000 ranage, the highest of all the clusters.
  - Total_Credit_Cards: range is approx. 5-10, the highest of all the clusters.
  - Total_visits_bank: range is approx. 0-1, the lowest of all the clusters.
  - Total_visits_online: range is approx. 6-15, the highest of all the clusters.
  - Total_calls_made: range is approx. 0-3, the lowest of all the clusters.

HIERARCHICAL CLUSTERING

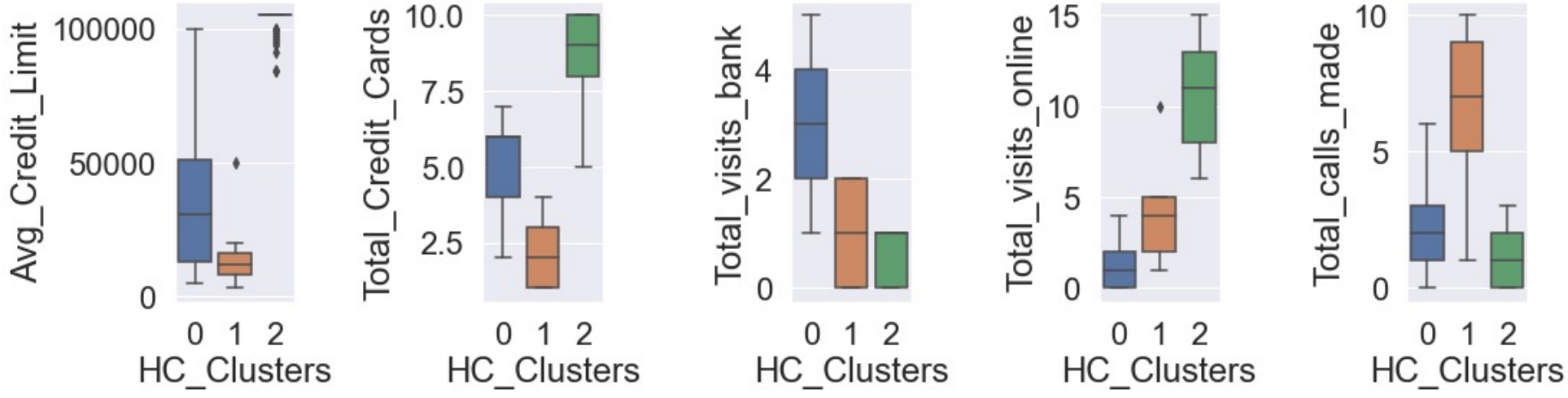


After viewing dendograms, we chose the Ward linkage method and 3 clusters.

HIERARCHICAL CLUSTER PROFILING

HC_Clusters	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	K_means_segments	count_in_each_segments
0	33355.329949	5.484772	3.459391	1.025381	2.060914	0.020305	394
1	12152.777778	2.351852	0.893519	3.569444	6.939815	1.000000	216
2	102660.000000	8.740000	0.600000	10.900000	1.080000	2.000000	50

Boxplot of original numerical variables for each cluster



Insights from Hierarchical Clustering

- **Cluster 0:**
  - Avg_Credit_Limit: median is about $25,000 (approx. range is $6,250-$100,000). This seems to be the middle range of all the clusters.
  - Total_Credit_Cards: median is about 6 (approx. range 2-7). This seems to be the middle range of all the clusters.
  - Total_visits_bank: values range from approx. 1-5 in this cluster, more than in the other clusters.
  - Total_visits_online: values range from approx. 0-4, less than in the other clusters.
  - Total_calls_made: values range from approx. 0-6. This seems to be the middle range of all the clusters.
- **Cluster 1:**
  - Avg_Credit_Limit: range is approx. $5,000-$12,500, the lowest range of all the clusters. (outlier(s) at $50,000)
  - Total_Credit_Cards: range is approx. 1-3, the lowest of all the clusters.
  - Total_visits_bank: range is approx. 0-2. This seems to be the middle range of all the clusters.
  - Total_visits_online: range is approx. 1-5 (with outlier(s) at 10). This seems to be the middle range of all the clusters.
  - Total_calls_made: range is approx. 1-10, the highest of all the clusters.
- **Cluster 2:**
  - Avg_Credit_Limit: Outliers exist ranging from approx. $75,000-$105,000. Primary values are above the $100,000 range, the highest of all the clusters.
  - Total_Credit_Cards: range is approx. 5-10, the highest of all the clusters.
  - Total_visits_bank: range is approx. 0-1, the lowest of all the clusters.
  - Total_visits_online: range is approx. 6-15, the highest of all the clusters.
  - Total_calls_made: range is approx. 0-3, the lowest of all the clusters.

BUSINESS RECOMMENDATIONS

Note: Since the insights from both clustering techniques were practically identical, the clusters below may reference either clusters from K-Means or Hierarchical clustering.

First, we can summarize our customer segmentation in terms of service calls and credit card limit/number of credit cards as follows:

- Cluster 0 represents the customers who make the most in-person visits to the bank (middle range in terms of credit limit and number of credit cards).
- Cluster 1 represents the customers who make the most calls but have the lowest credit limits (and lowest number of credit cards).
- Cluster 2 represents the customers that make the most online visits and have the highest credit limits (and highest number of credit cards).

Ideally, we would want customers to convert to Cluster 2 so that they have higher credit limits/credit cards and thus usage that can be more profitable for the bank. Some recommendations:

- Incentivize seeking support online vs. in person or on the phone by ensuring that online account registration is (1) easy and convenient, (2) and perhaps add promotions for online users such as discounts at local merchants/restaurants or bonus point programs.
- Ensure that when customers utilize phone or in-person support, they are informed of the online support option.

At this time, most customers fall into Clusters 0 (394) and 1 (216). Recommendations for improving service:

- For all modes of support (online, in-person, phone), ensure that staff are properly and extensively trained.
- For all modes of support (online, in-person, phone), ensure that staff are upselling services - possibly incentivize this for staff by offering rewards for certain number of services sold.
- In-person and phone visits: reduce wait times for service by ensuring enough staff are scheduled for busier times of day - may need to do an analysis to determine which times of day are busiest.
- Phone support: the customers who seek this option the most (Cluster 2) have the lowest number of cards - upselling/marketing is especially important for these customers.