ALL LIFE BANK PROJECT

2021

OBJECTIVES

- To identify different segments in the existing customer based on their spending patterns

- Build Clustering Algorithms (K-Means and Agglomerative Hierarchical) to deduce the number of clusters and apply Cluster Profiling to determine customers in each cluster and their characteristics

- To determine which Cluster of customers should be targeted more to improve the perceived poor image of the bank

- To proffer councel to the bank on how best to target and market these customers along with the best customer service offering peculiar to them.

- Generate a set of insights and recommendations to be deployed for better service and customer retention

BUSINESS PROBLEM OVERVIEW

BACKGROUND

- AllLife Bank's Executive Management seeks to establish factors responsible for its perceived poor image as well as improve customer service in a bid to endear new customers along with facilitating Customer Retention while expanding their customer base and ultimately maximizing revenue and profit through additional processing fees and credit card charges.

- As a senior data scientist at AllLife Bank, my aim is to deploy clustering algorithms to identify customers who are disenfranchised in a bid to serve them better and endear new customers to the bank as well as upgrade existing ones to drive retention and improve the bank's bottom-line, revenue and profit ultimately.

SOLUTION APPROACH (UNSUPERVISED MACHINE LEARNING CLUSTERING ALGORITHM)

- Define the problem and perform an Exploratory Data Analysis

- Illustrate the insights based on EDA

- Data pre-processing/Feature Engineering

- Building Clustering Algorithms and identifying Clusters

- Comparing Clustering Algorithms and Clusters

- Actionable Insights & Recommendations

BUSINESS IMPLICATIONS

- Identify major influencers of the perceived poor reputation of AllLife Bank

- Optimize processes for ease of card migration and subscription while improving contacts and relationship with customers in general

- Identifying loopholes in customer service through proper anlaysis of shortfalls in parameters under consideration

- To effectively allocate resources and introduce perks, loyalty points or coupons to influence retention while engaging a more robust and targeted marketing

- Increased Customer patronage, referral and traffic which ultimately impacts on AllLife Bank's bottom-line

- Tracking and Trapping potential prospects with new and customised product offerings along with additional alternatives based on isightful recommendations from the Data Scientist

- Optimize administrative and overhead costs by applying the preferred Clustering models smartly on perculiarity basis

- Optimize Sales and Profit Margin per new card subscriptions

- Effective Allocation and redistribution of resources especially for product re-design, pricing and Ad-campaigns

- Developing potential markets for more revenue

DATA MANIPULATION

- Overview of the Dataset

- Identification of Missing Values

- Fixing and Dropping Columns: (Sl_No and Customer_Key) were dropped ultimately prior to Clustering

- Identification and Removal of Duplicate Values

Variable	Description
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 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 calls made: Total number of calls made by the customer to the bank or its customer service department (yearly)

Observations	Variables
660	7

Data Type
All the variables are of Int64 Data Type

EXPLORATORY DATA ANALYSIS

Avg_Credit_Limit
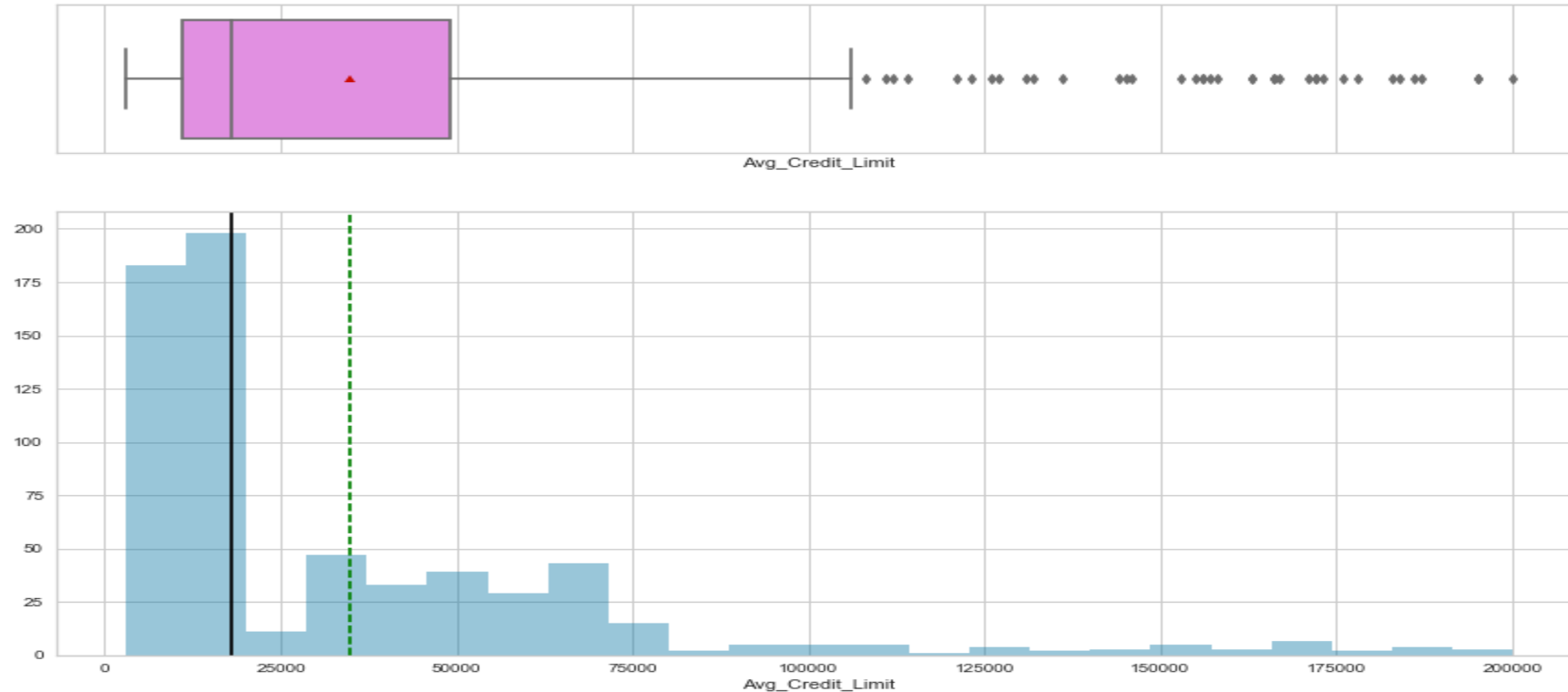
- As can be inferred, there is an obvious presence of outliers as expected, hence the right skew nature
- The mean Avg_Credit_limit (34878) is greater than the median (18000)
- The outliers are quite close and may ultimately impact or form clusters of their own. The upper outliers start at 105000 with the maximum at 200,000
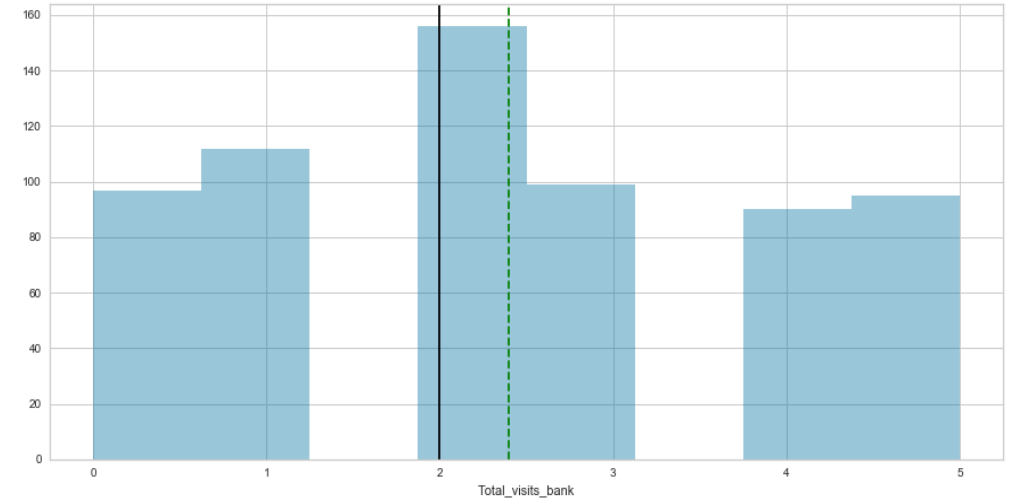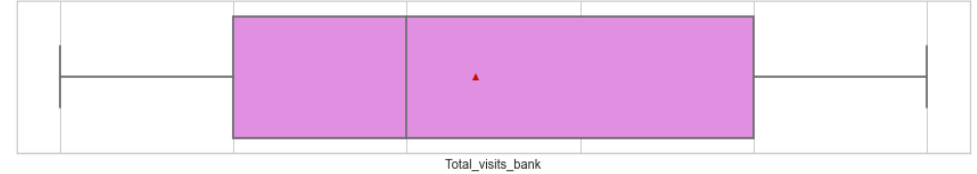
EXPLORATORY DATA ANALYSIS

TOTAL_CREDIT_CARDS

TOTAL_VISITS_BANK
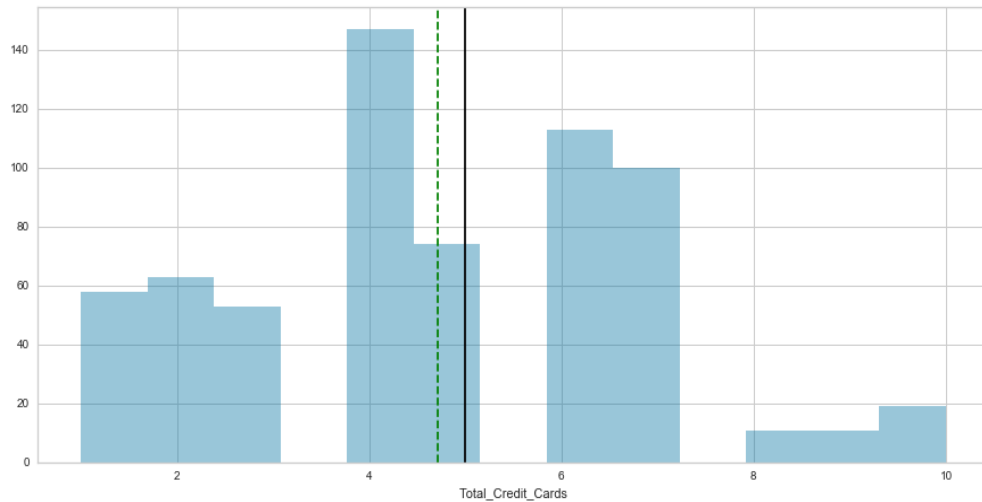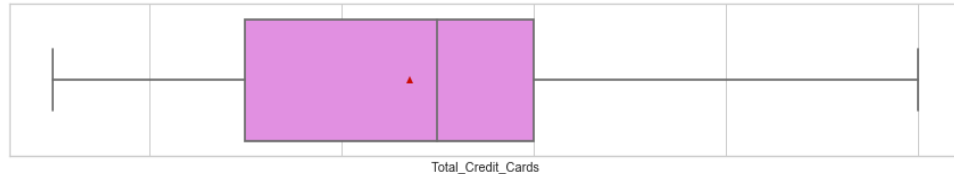
- The average Total_Credit_Cards is lower than the Mean position. Hence we have a Left Skew distribution
- The are no presence of outliers and 50% 0f the distribution has at least 5 credit cards with the maximum at 10 cards
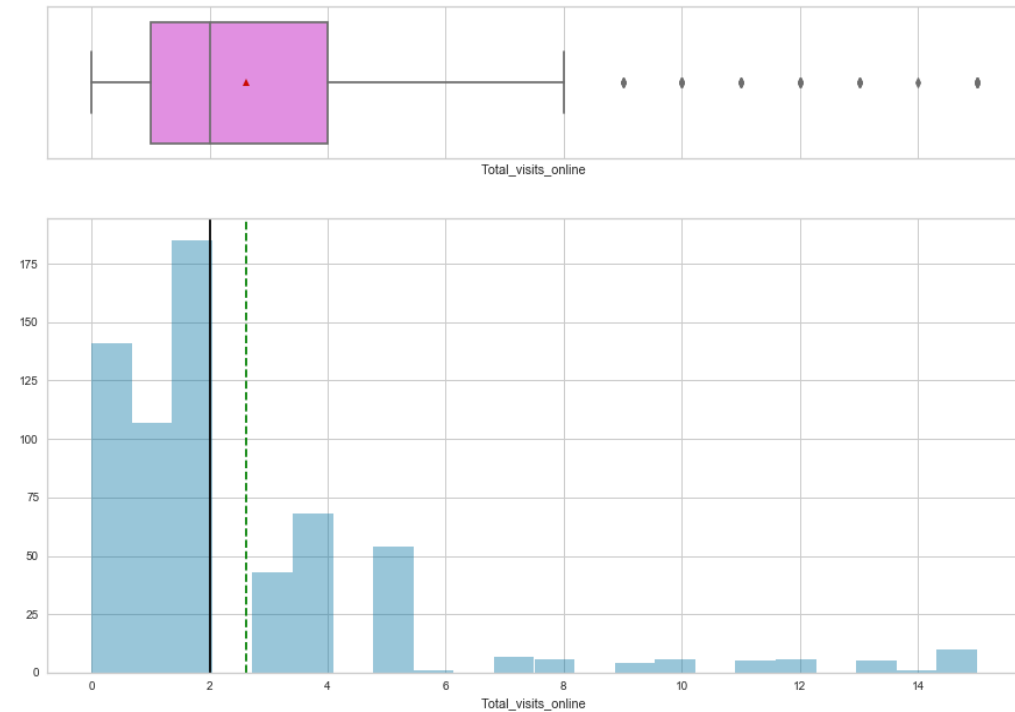- This speaks to a highly geared customer distribution

- There are no presence of outliers here
- The mean of 2.4 is greater than the median of 2.
- 75% 0f the distribution made at least 4 visits or less to the bank with the most being at 2
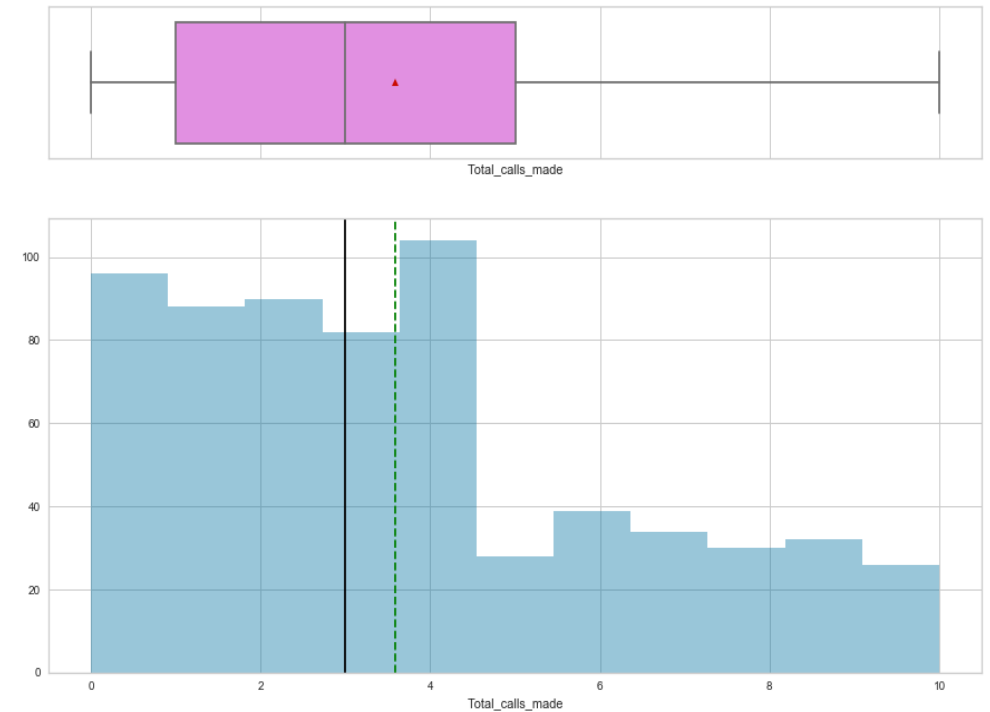
EXPLORATORY DATA ANALYSIS

TOTAL_VISITS_ONLINE

TOTAL_CALLS_MADE

- The frequency of visits by most cusomers online is at 2 with 75% of them with at least 4 online banking visits
- There are clear presence of outliers with the maximum at 15
- The distribution of online visits by customers is right-skewed.
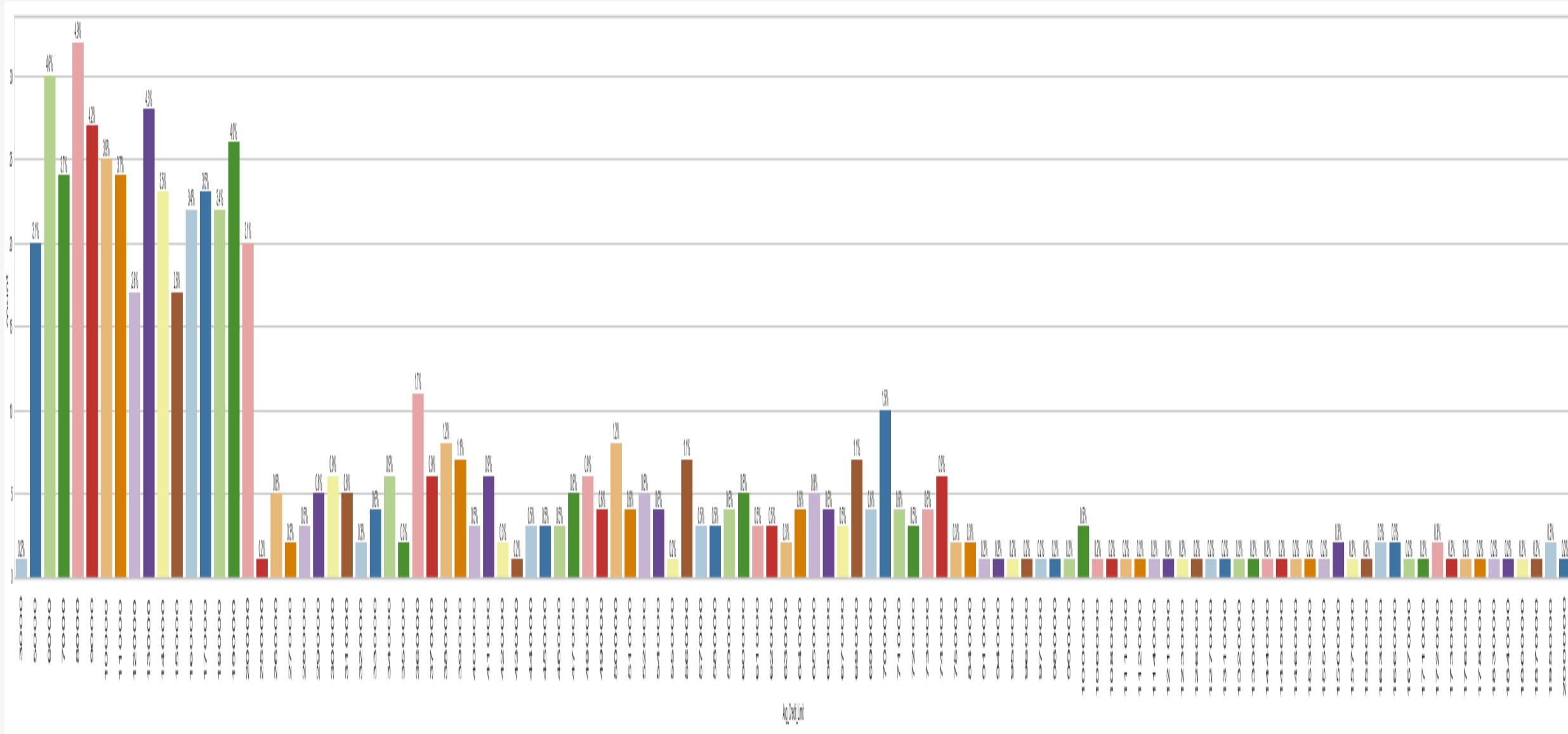
- There are no outliers
- The distribution is right-skewed
- The mean Total_calls_made is 3.6 against the median of 3.
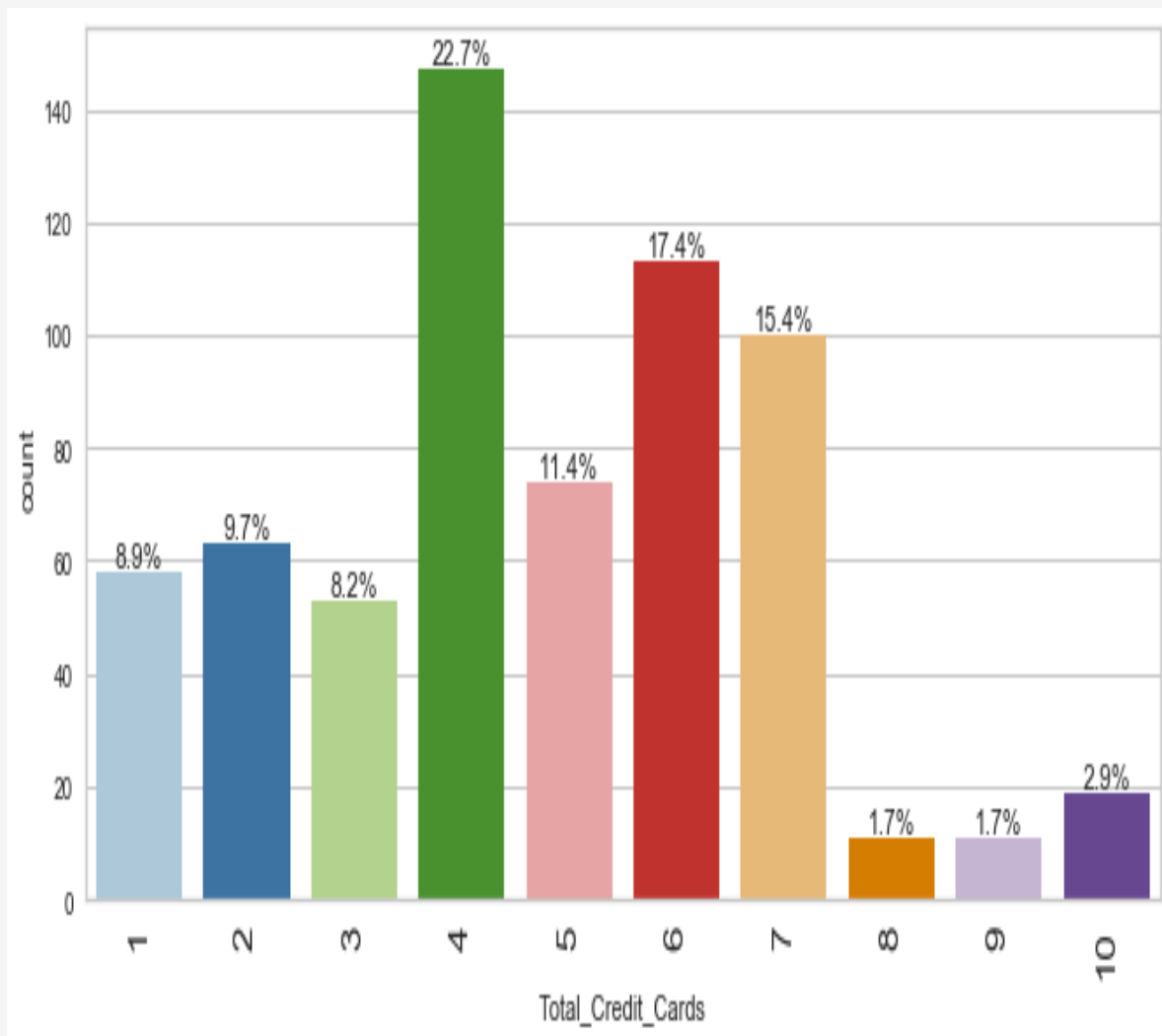- 75% of the distibution called at least 5 times

AVG_CREDIT_LIMIT

90% of the distribution have an Avg_Credit_Limit of $75,000

25% of the customers have 3 cards or less and about 51% with 4 to 6 card and about 24% between 7 to 10 cards
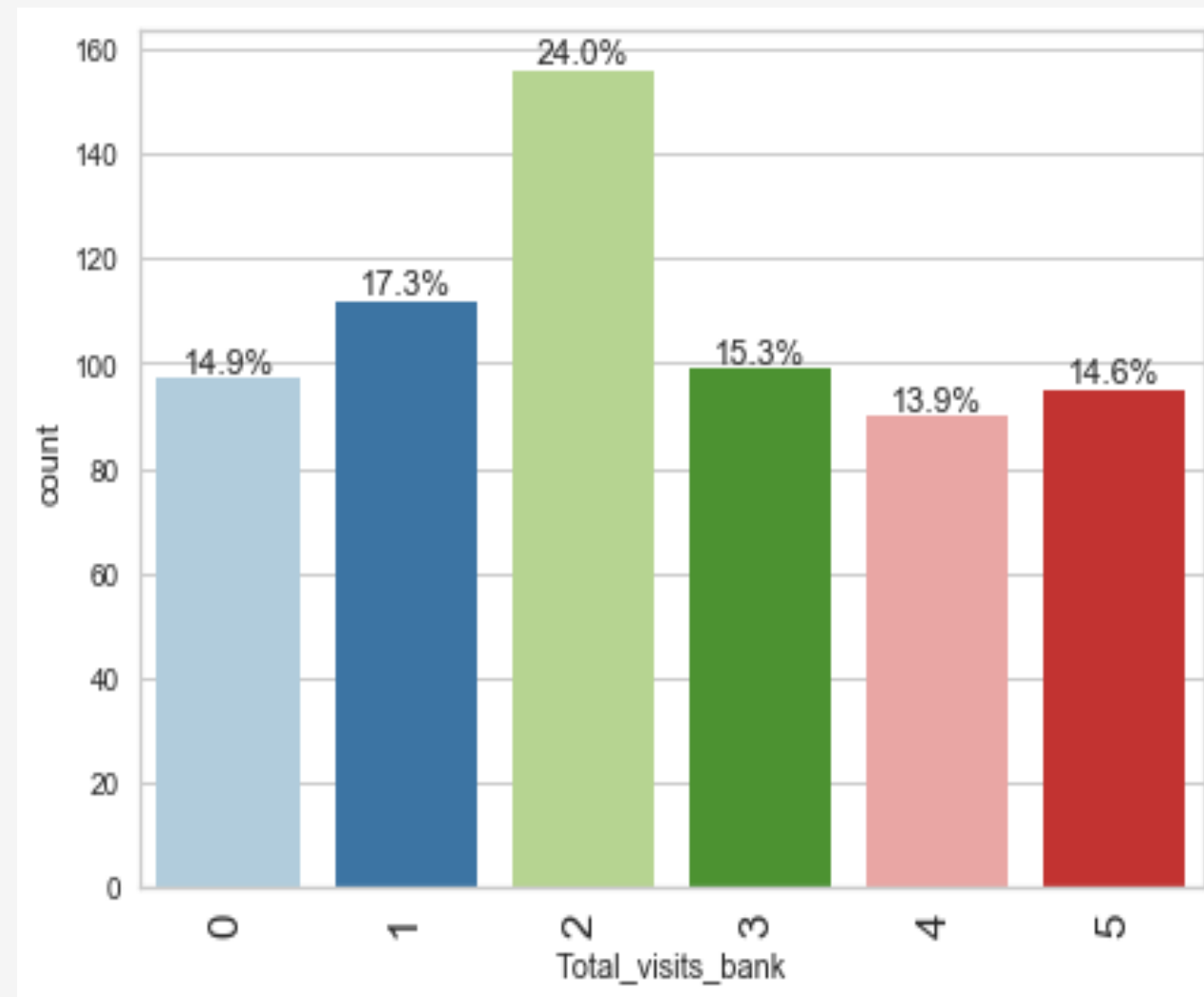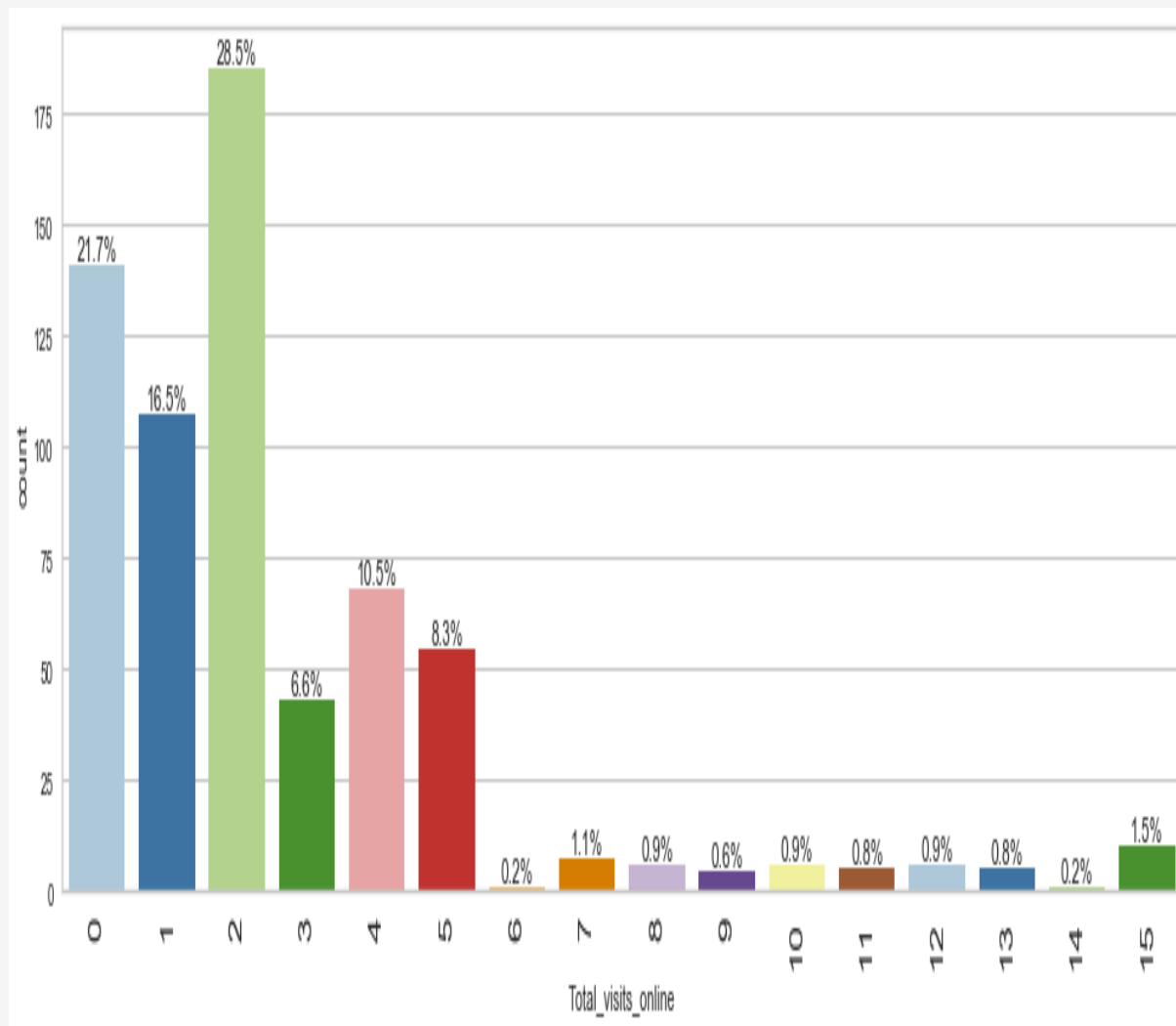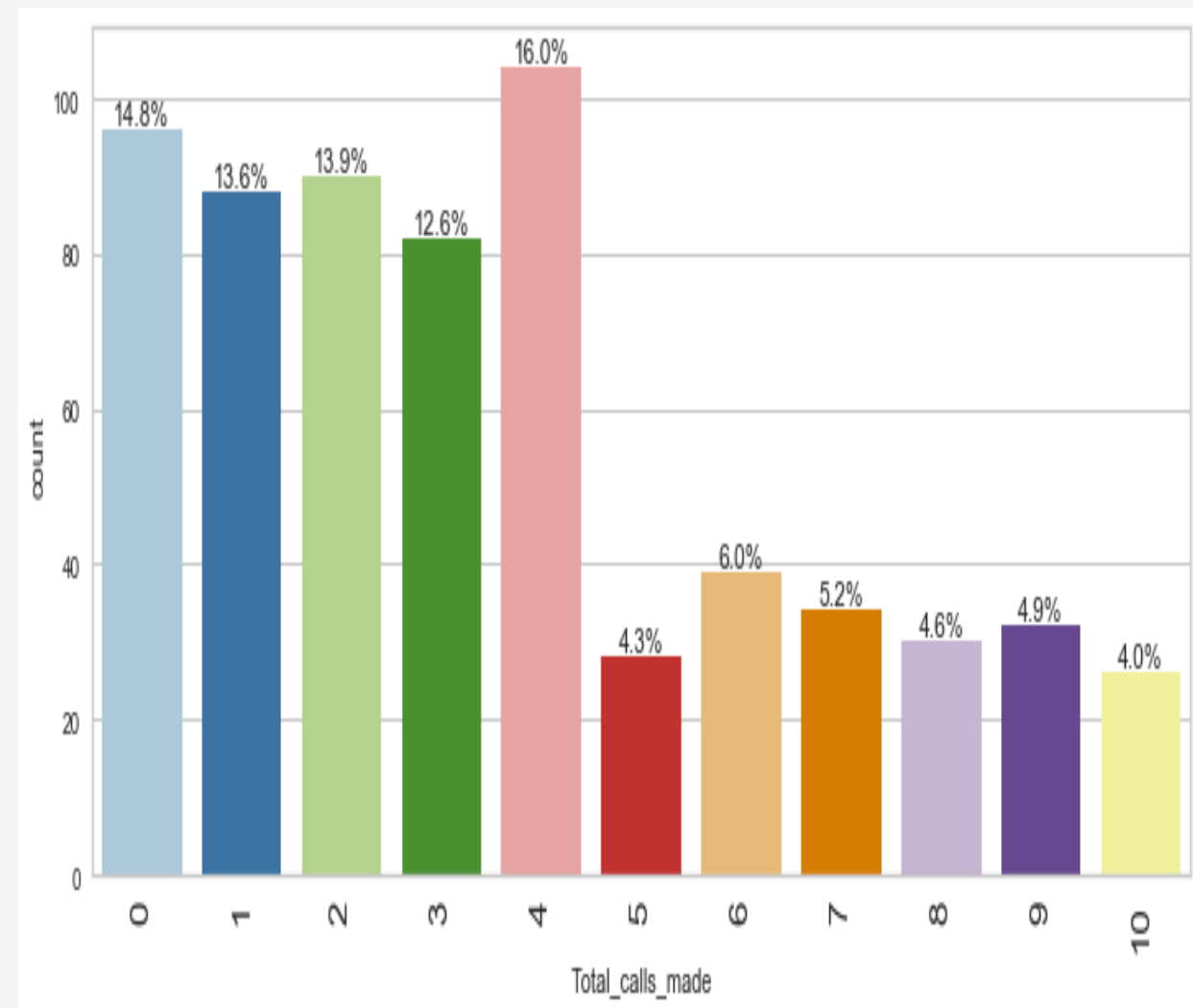
About 80% of the customers visited the banks 4 times or less in the distribution in one year

EXPLORATORY DATA ANALYSIS-PERC_BAR PLOTS OF CLUSTER COLUMNS

- 90% of the distribution made online visits 6 times or less in a year

- 80% of the distribution made 6 calls or less to the bank in one year

CDF plot of numerical variables
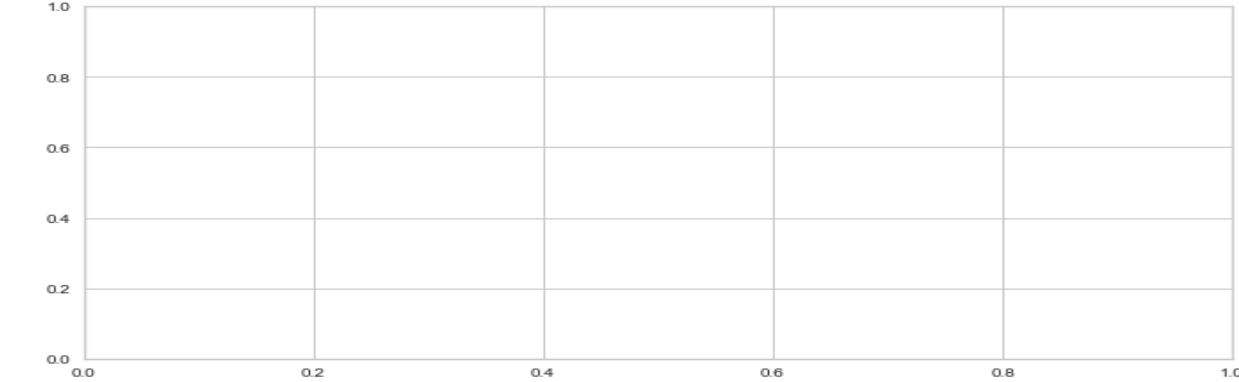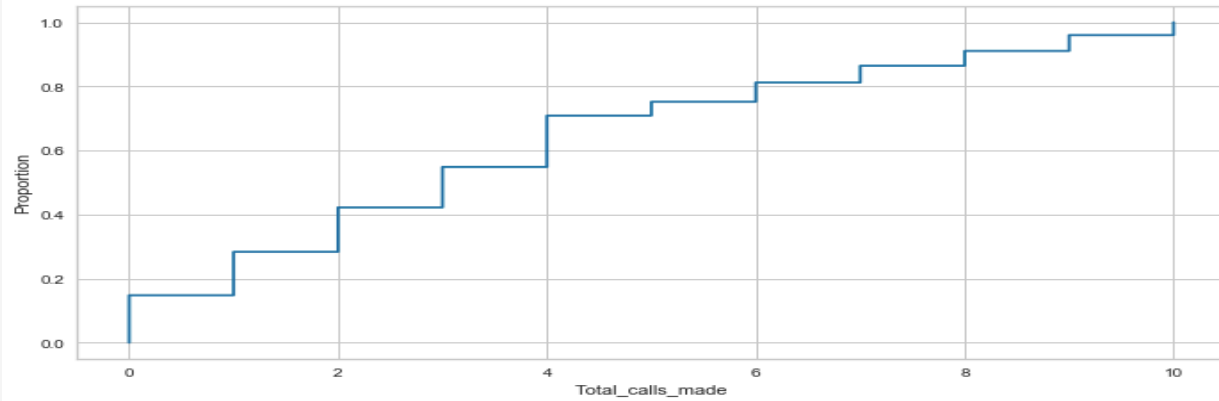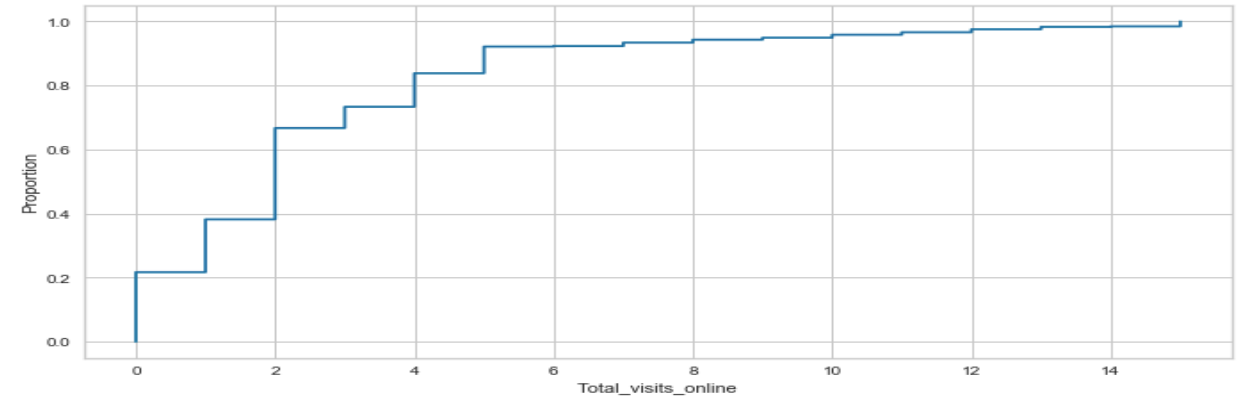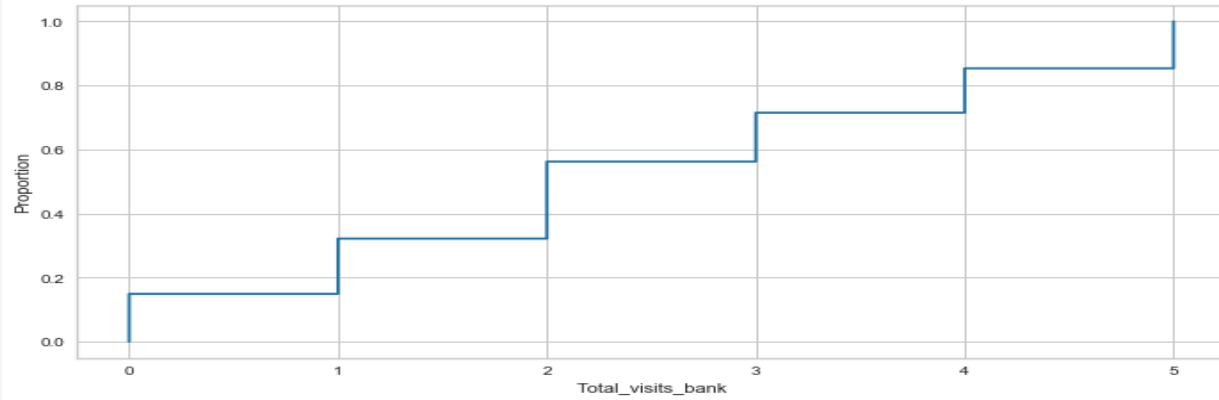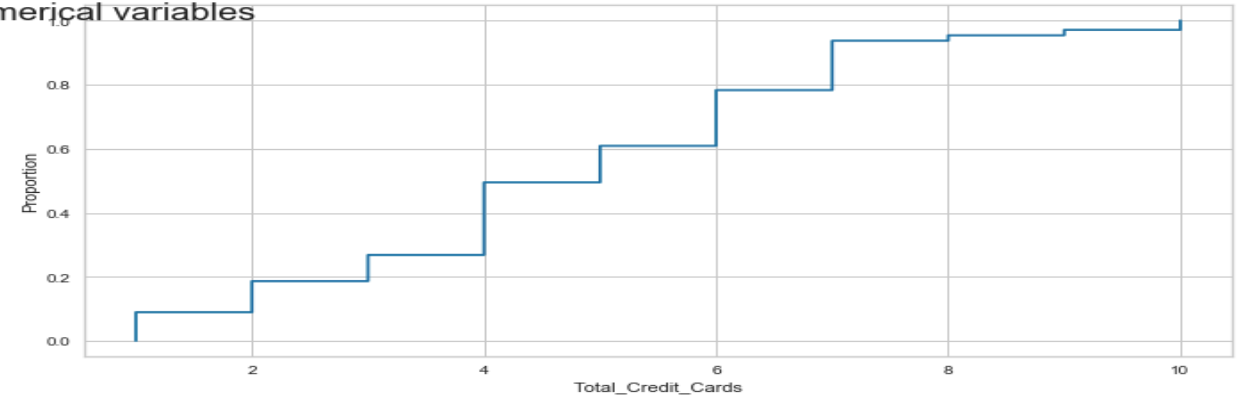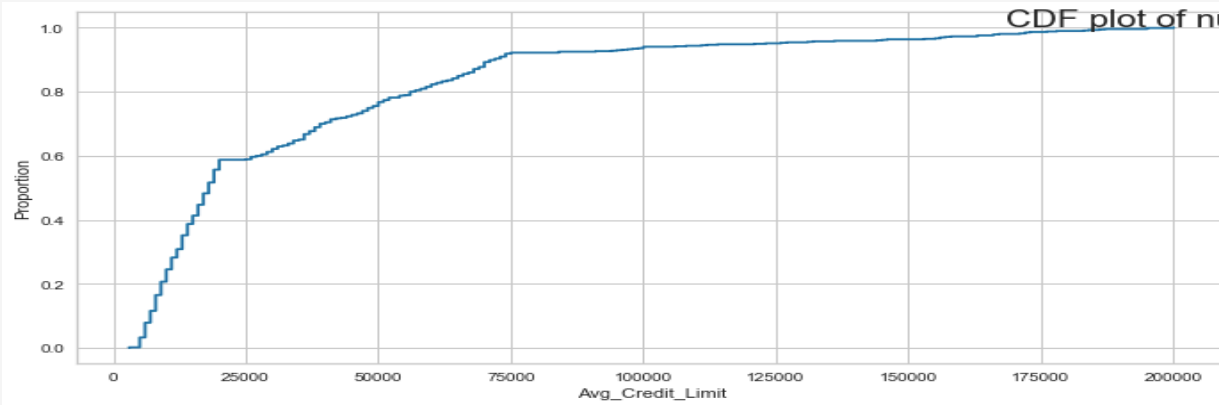
This buttresses the position of the Perc_Bar Plot

EXPLORATORY DATA ANALYSIS- HEAT MAP

- There is a fairly strong correlation between Total_visits_online and Avg_Credit_Limit
- There is quite a strong correlation between Avg_Credit_Limit and Total_Credit_Cards

EXPLORATORY DATA ANALYSIS-PAIR PLOT
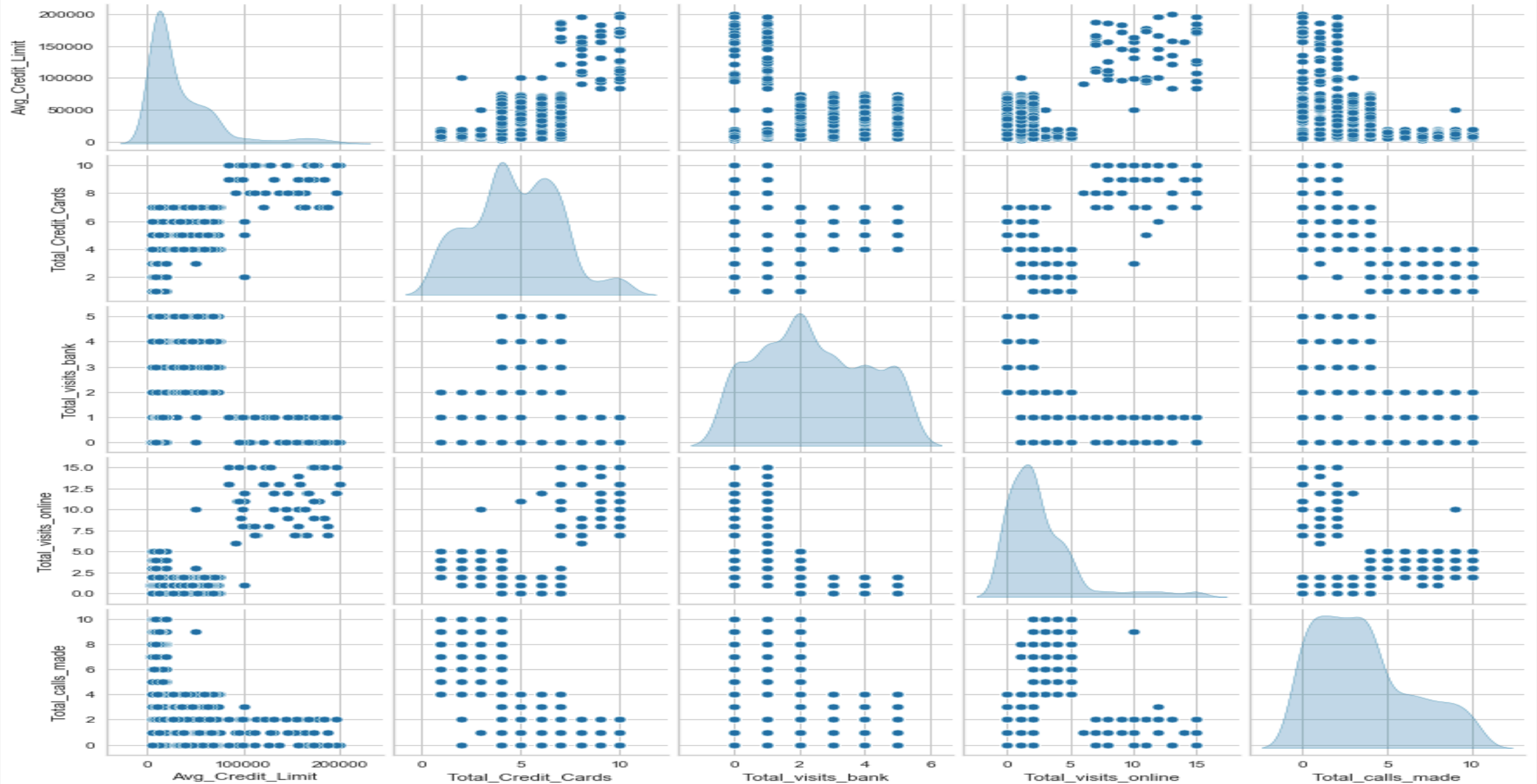
• The Scatter plots as well as the crests on the KDE Plots are clearly suggestive of obvious clusters.

K-MEANS CLUSTERING

Selecting k with the Elbow Method
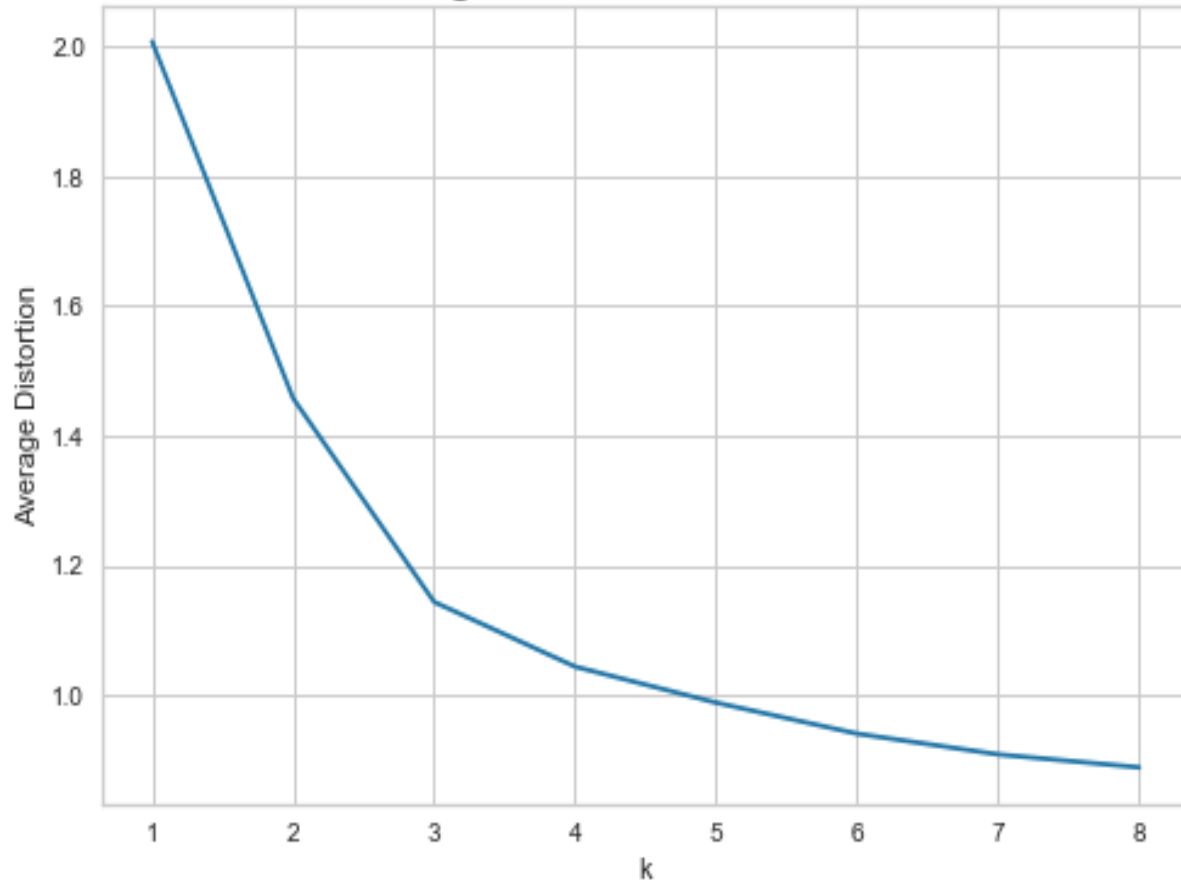
- From the Elbow curve we could notice knicks at 2,3,4,5,6
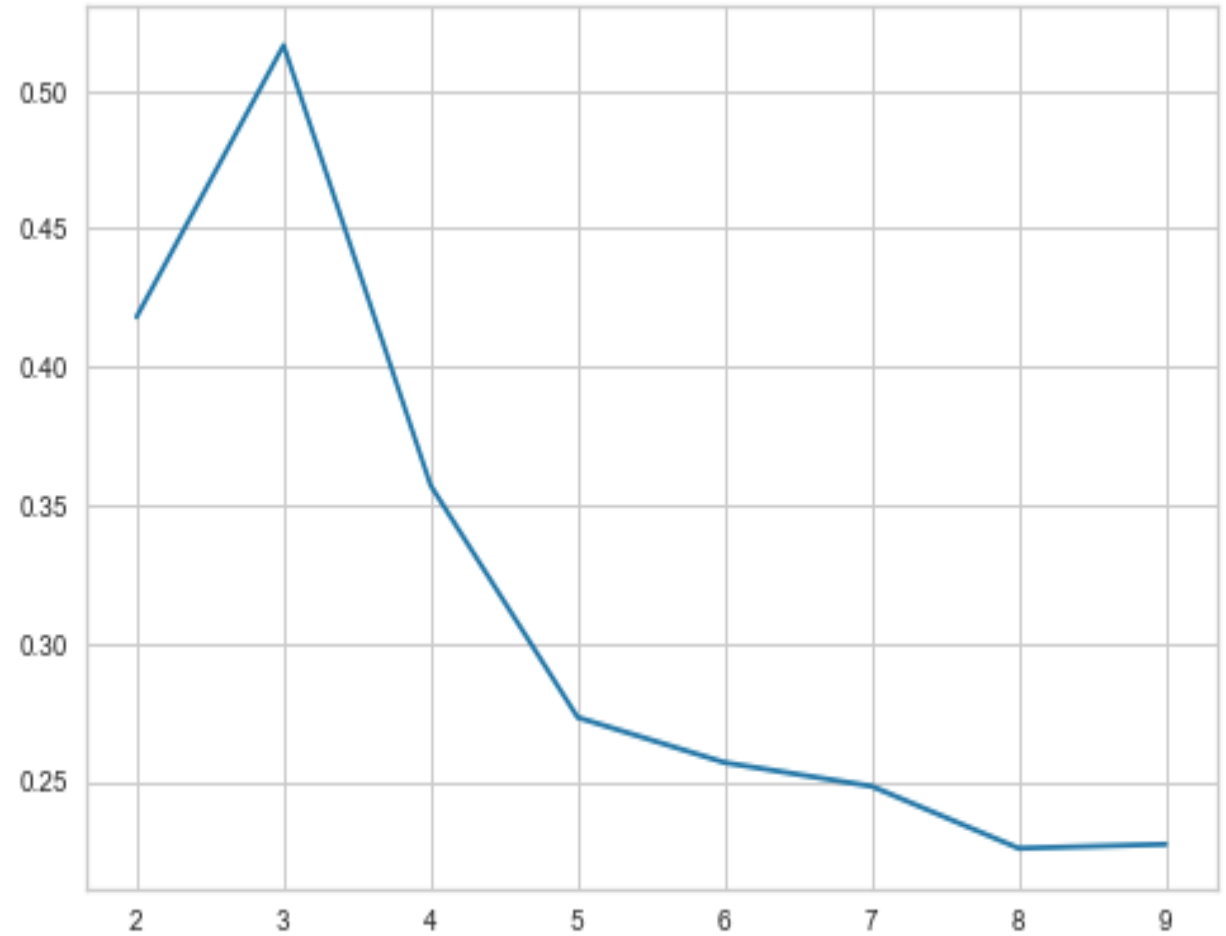
- From the Silhouette visualizer, the 7 was used as a good start for K iterations and eventually chose k= 4 as the value with the most variability in Clusters

Boxplot of Numerical variables for each cluster

We can see the consistency in the plot

CLUSTER PROFILING (K-MEANS)

Clusters 0

There are 196 customers in this cluster
The Avg_Credit_limit for this cluster is roughly 36,000 which is way above the median limit
Customers in this cluster have approximately 6 credit cards
2 visits were made by these customers to the bank in a year
Roughly 1 visit was made online by customers in this cluster in a year
Customers in this cluster made 2 calls to the customer service of the bank in a year

Cluster 1

There are 185 customers in this cluster
The Avg_Credit_limit for this cluster is roughly 32,000 which is way above the median limit
Customers in this cluster have approximately 5 credit cards
5 visits were made by these customers to the bank in a year
1 visit was made online by customers in this cluster in a year
Customers in this cluster made 2 calls to the customer service of the bank in a year

Cluster 2

There are 218 customers in this cluster
The Avg_Credit_limit for this cluster is roughly 12,500 which is way below the median limit
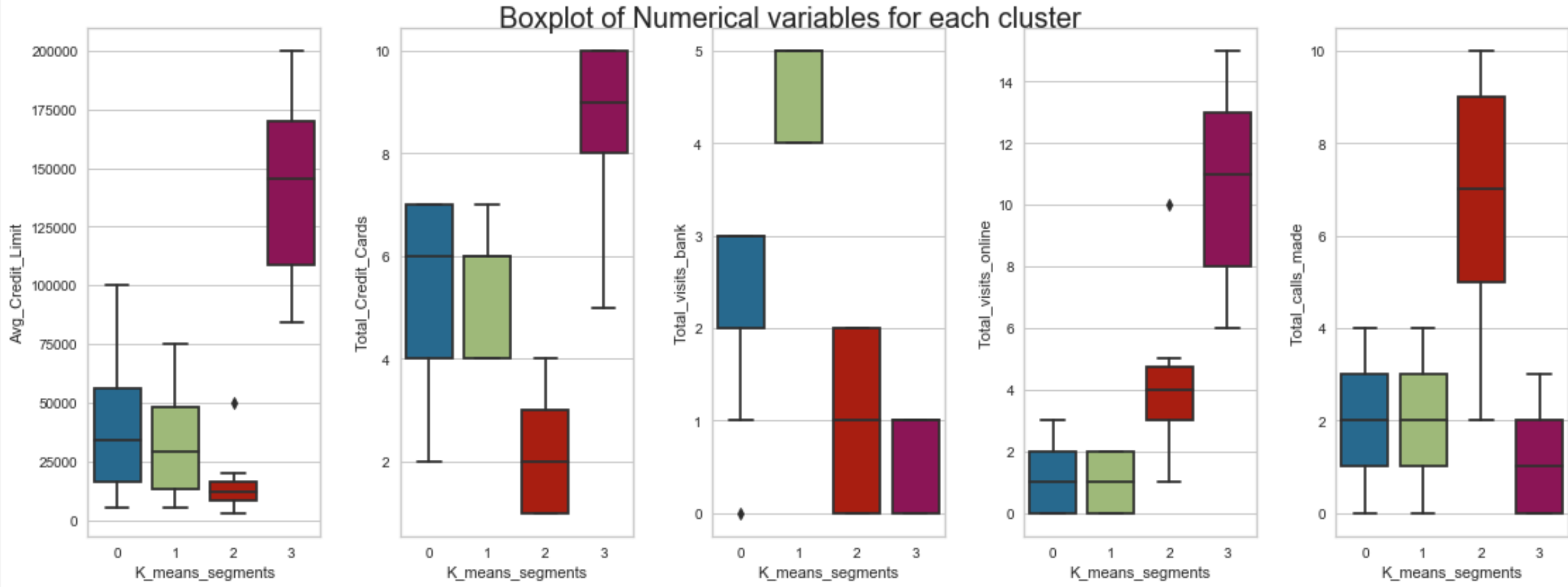Customers in this cluster have approximately 2 credit cards
1 visit was made by these customers to the bank in a year
Roughly 4 visits were made online by customers in this cluster in a year
Customers in this cluster made approx. 7 calls to the customer service of the bank in a year

Cluster 3

There are 50 customers in this cluster
The Avg_Credit_limit for this cluster is roughly 141,000 which is way above the 75th percentile limit
Customers in this cluster have approximately 9 credit cards
1 visit was made by these customers to the bank in a year
Roughly 11 visit were made online by customers in this cluster in a year
Customers in this cluster made 1 call to the customer service of the bank in a year

- Upon analysis, the cophenetic correlation for different Linkages and distances were determined and the Eucledian Distance and Average Linkage Returned the highest score

- Next , different Linkages were analysed using the Eucledian distance only and the Average linkage returned with the highest correlation

- Dendrograms for several linkages were plotted and the Ward linkage gave the most variability in Clusters(very distinct and separate) as can be seen on the Plot .No.5 from the top

- Finally, specific clusters were derived and cluster profiles established

Boxplot of scaled numerical variables for each cluster
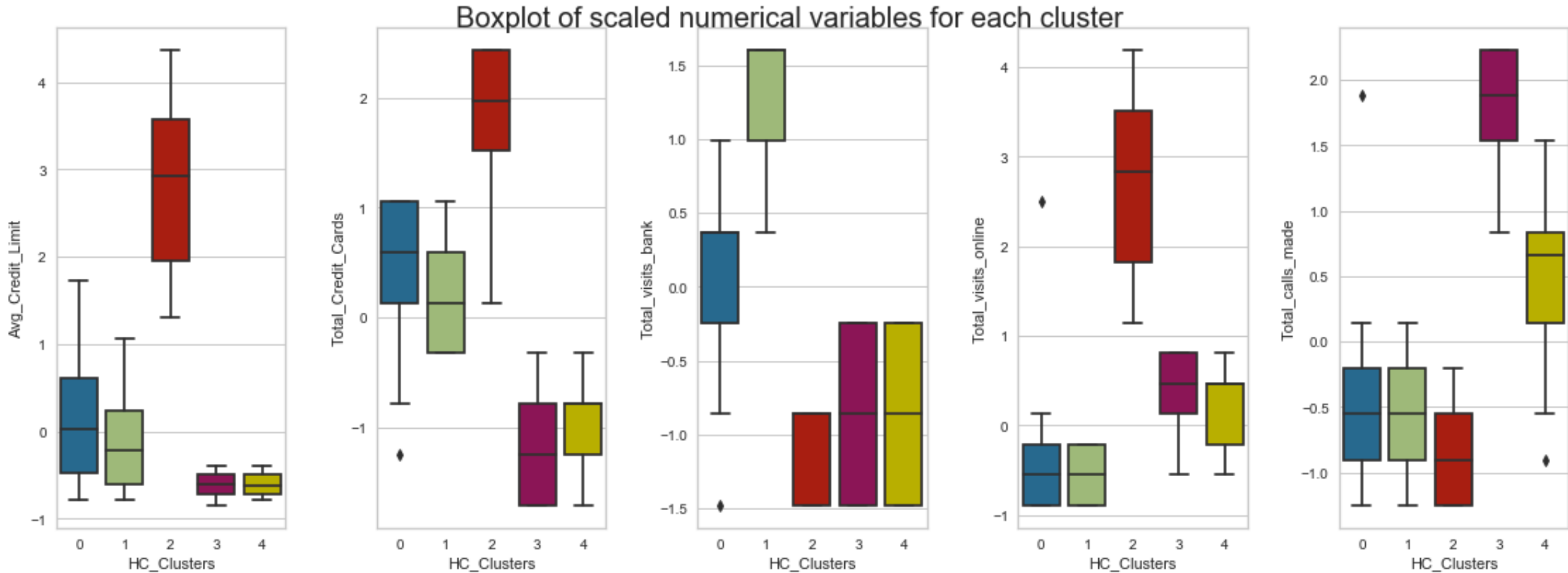
Boxplot of original numerical variables for each cluster
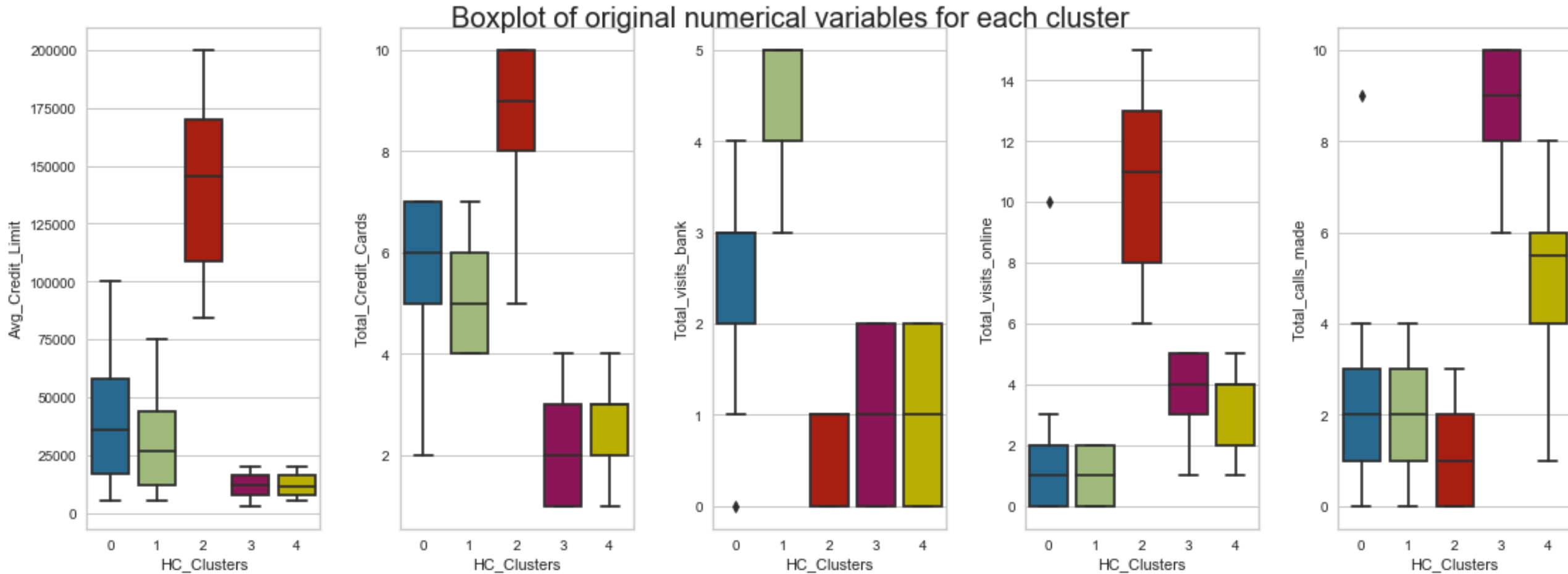
- As seen both plots , Scaled and Original are consistent

COMPARISON OF K-MEANS AND HIERARCHICAL CLUSTERING

K_MEANS

- This algorithm gave the best Clusters at k=4 i.e 4 Clusters

- Cluster 3 returned the population of customers with the highest Avg_Credit_Limit with 50 customers and the highest Total_Credit_Cards of roughly 9 as well as 11 Total_Online_visits

- Cluster 1 captures the number of customers with the highest number of Total_visits_bank at approximately 5 with 185 customers

- Cluster 2 returned the highest Total_calls_made at approximately 7 and a total of 218 customers being the highest in across the distribution

- Cluster 0 seems quite unique as it returns the next highest population in the distribution as well as the in value of the Avg_Credit_Limits and probably speaks to the mid-executives, Senior managers or Startup/SME Entreprenuers

- Finally, computational time was faster

AGGLOMERATIVE(HIERARCHICHAL)

- This algorithm resulted in k=5 i.e 5 Clusters

- Cluster 2 in this case returned same number of customers as K-Means at 50 and the highest Avg_Credit_Limit but with a lesser number of Total_Credit_Cards at approx.8 along with 11 Total_Online_visits

- Cluster 1 , just as in K-Means registered the highest Total_visits_bank across all clusters at 4 with 186 customers

- Cluster 0 records the highest population of customers in the distribution at 194

- Cluster 4 captures the customers in the distribution with the lowest Avg_Credit_Limit at roughly 12,000

- While Cluster 3 surprisingly has the highest number of Total_calls_made at 9 but still returns a low Avg_Credit_Limit of about 12,200. it is indeed similar to Cluster 4 in many regards.

- Computational time was far slower

CLUSTER PROFILING

- There are 7,917 Existing customers

- The Blue Card customers average 46 years in age with 5101 females and 4335 Males

- They have an average of 2 dependents per customer with 4433 who are Married.

- There are 2899 Graduate Blue card holders with 3403 of these class of card holders earning less than $40k pa

- There are 9436 Blue card holders with an average of 36 months in the Bank's books with an avg. of approx. 4 other products

- This category has an average of 2 contacts and 2 inactive months in a year

- They have an average credit limit of 7364 with a Total revolving Balance of 1157.8 per month

- With an Average Open to Buy of 6205 and a Total mean Transaction amount of 4225 amounting to 64 transaction counts in total, the Blue Card category indeed is quite a dominant feature

- There are 473 Existing customers

- The Silver Card customers average 45.6 years in age with 210 females and 345 Males

- They have an average of 2 dependants per customer with 251 who are Singles.

- There are 185 Graduate Silver card holders with 3403 of these class of card holders earning less than $40k pa

- There are 555 Silver card holders with an average of 35.5 months in the Bank's books with an avg. of approx. 3 other products

- This category has an average of 2 contacts and 2 inactive months in a year

- They have an average credit limit of 25278 with a Total revolving Balance of 1206 per month

- With an Average Open to Buy of 24072 and a Total mean Transaction amount of 6590 amounting to 75 transaction counts in total, the Silver Card category indeed is quite a promising pool of young prospects.

BUSINESS RECOMMENDATIONS

Cluster 0 and **Cluster 1(Similar to Cluster 0 in K-Means Clustering)

These customers Average Credit Limit is way above the median limit across the whole distribution. They maintain the most presence across the distribution at 194 and cannot be ignored. They are waht businesses would call critical mass components and account for a sizeable income and turnover of the bank. In reality, clusters of this nature speaks to middle and senior managers, start-up executives and the trendy urban geeks.**AllLife must ensure an increase in these number of customers as well as ultimate retention capabilities through improved customer service offerings, robust banking alternatives, discounted fees tied to card activities along with promotional product offerings in a bid to drive retention, revenue, profitability and the track-record of the bank. This cluster retains a huge migration potential to Cluster 2 if given the much anticipated value-driven service.**

Cluster 2(Similar to Cluster 3 in K-Means Clustering)

This cluster clear accounts for the outliers we have across the distribution. They are 50 in number across both Clustering Algorithms. They equally register a great penchant for online banking activities with the highest Total_visits_online along with the highest Total_Credit_Cards.cluster also accounts for the highest Average Credit Limit across the distribution. This clearly speaks to very High-Networths, Executive Management or Directors in realit who rarely find the time to make physical visits but spend most of their time traveling, meeting and on gadgets/devices.**They are the toast of major competitors of AllLife Bank and maintain an extensive network of contemporaries and are equally at liberty to grant referrals if offered premium service. These customers purely seek the most desired attention when transacting. As such, it is quite critical that since most time is spent online and in transit ,resources should be galvanized to develop IT solutions and Automated Serveice packages that cater to their very busy schedules and timelines to influence their relationship with the bank. In addition, perks and occassional value-oriented initiatives like periodic calls by the bank along with summarised monthly/quarterly economic or industry briefs sent via e-mails or personalised prompts on the AllLife Banking applications to keep them up to date will go a long way to adding the much need value that will inturn crystalize into referrals and corporate accounts by way of additional credit facilities where the bank will witness more traffic, transaction volume and income.Personalized ad-campaigns and customized product offering will create more opportunities. Additional Credit limit upgrades will also go a long way.**

Cluster 3 and **Cluster 4(Similar to Cluster 2 in K-Means Clustering)

The customers in these clusters have the lowest credit limits and number of credit cards across all clusters as well as the highest number of calls. These are the majority of the customers accross the distribution at 219 (101 and 118). Their attempt to communicate their reservations to the bank may have been ignored. As such, the disenchantment would clearly have taken toll on their transactions with the bank. **Efforts and feed-back mechanisms ought to be put in place to imediately despatch solutions, incentives or alternatives that would ameliorate any issues registered.Furthermore, targeted media and ad-campaigns on these clusters will track and trap quite a chunk of these.In addition,with damage-control measures like fee reviews and loyalty points per card utilization with an option of assured upgrade in Credit limits, these customers will be endeared to increasing their transactional activity/turnover,which in turn influences retention in AllLife Bank as against attrition. Resources to develop Online retention should be deployed using the requisite technology to actively increase Online Engagement. These customers are prospects for Cluster 0 if well handled.**

THANK YOU