

PROJECT 7- UNSUPERVISED LEARNING

**DATA SCIENCE AND BUSINESS ANALYTICS
KRITHIKA SRINIVASAN**

ANALYTICS

BUSINESS OVERVIEW

- AllLife Bank wants to focus on its credit card customer base in the next financial year
- The Marketing team proposes to run personalized campaigns to target new as well as to upsell to existing customers
- Objective:

To identify different segments in the existing customer base, based on spending patterns and past interactions with the bank using clustering algorithms and providing recommendations to the bank on how to better market to these customers

SOLUTION APPROACH

- The objective can be achieved by performing different clustering algorithms on this dataset and grouping customers into different target bases
- The clustering methods applied:

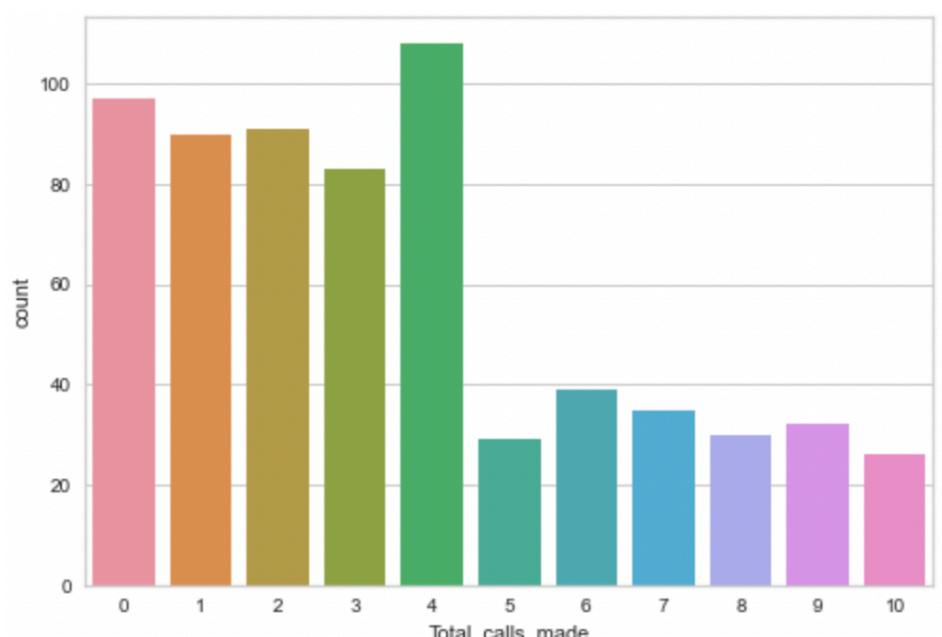
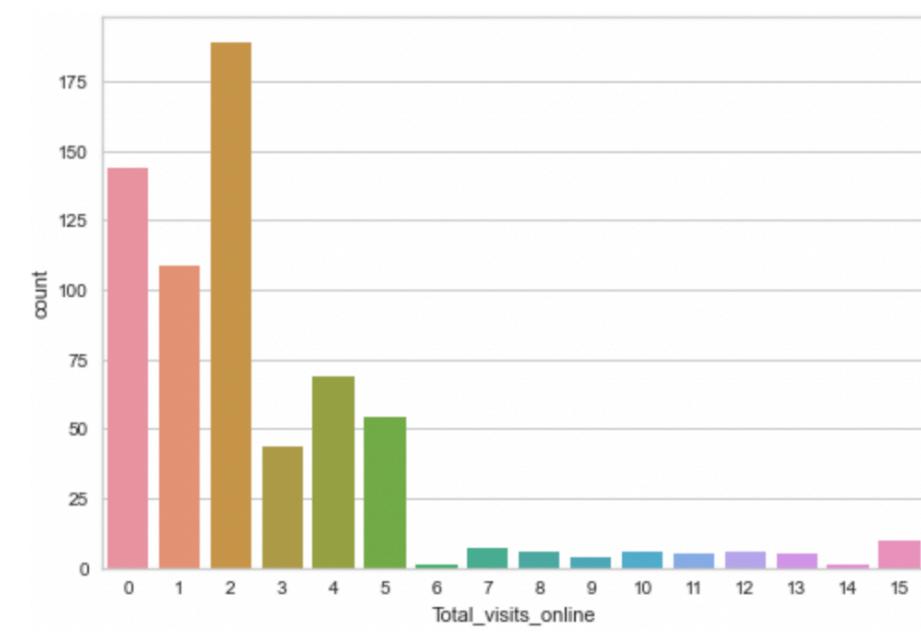
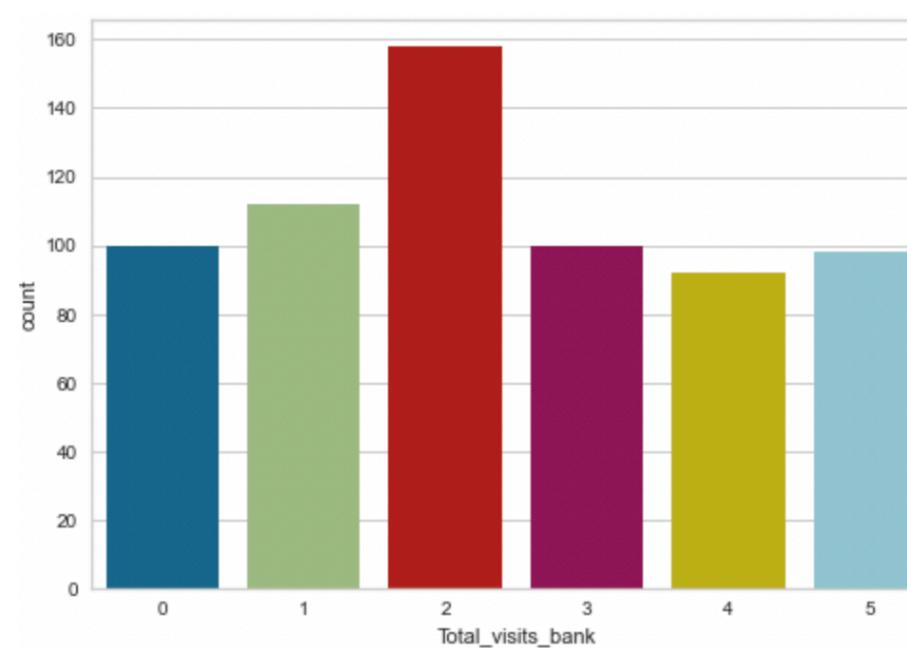
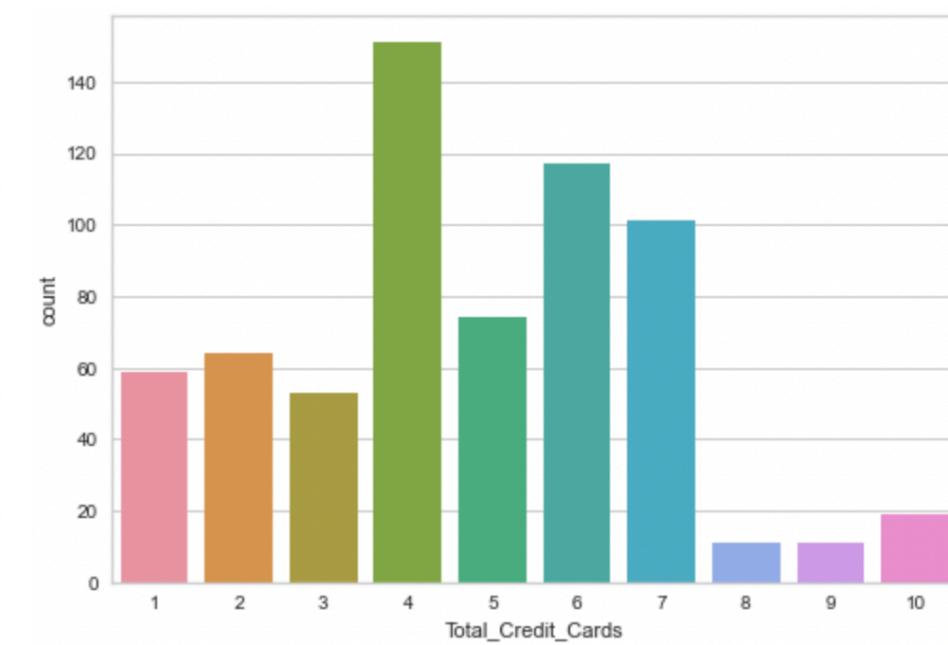
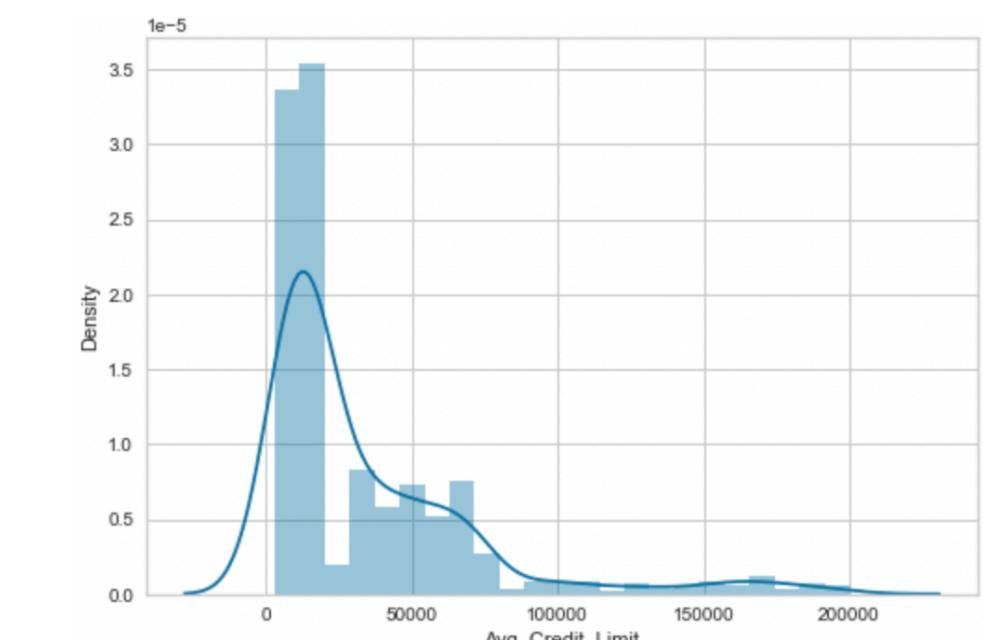
K-means clustering

Hierarchical clustering

- The results from these two algorithms are then compared and recommendations for each customer group are determined

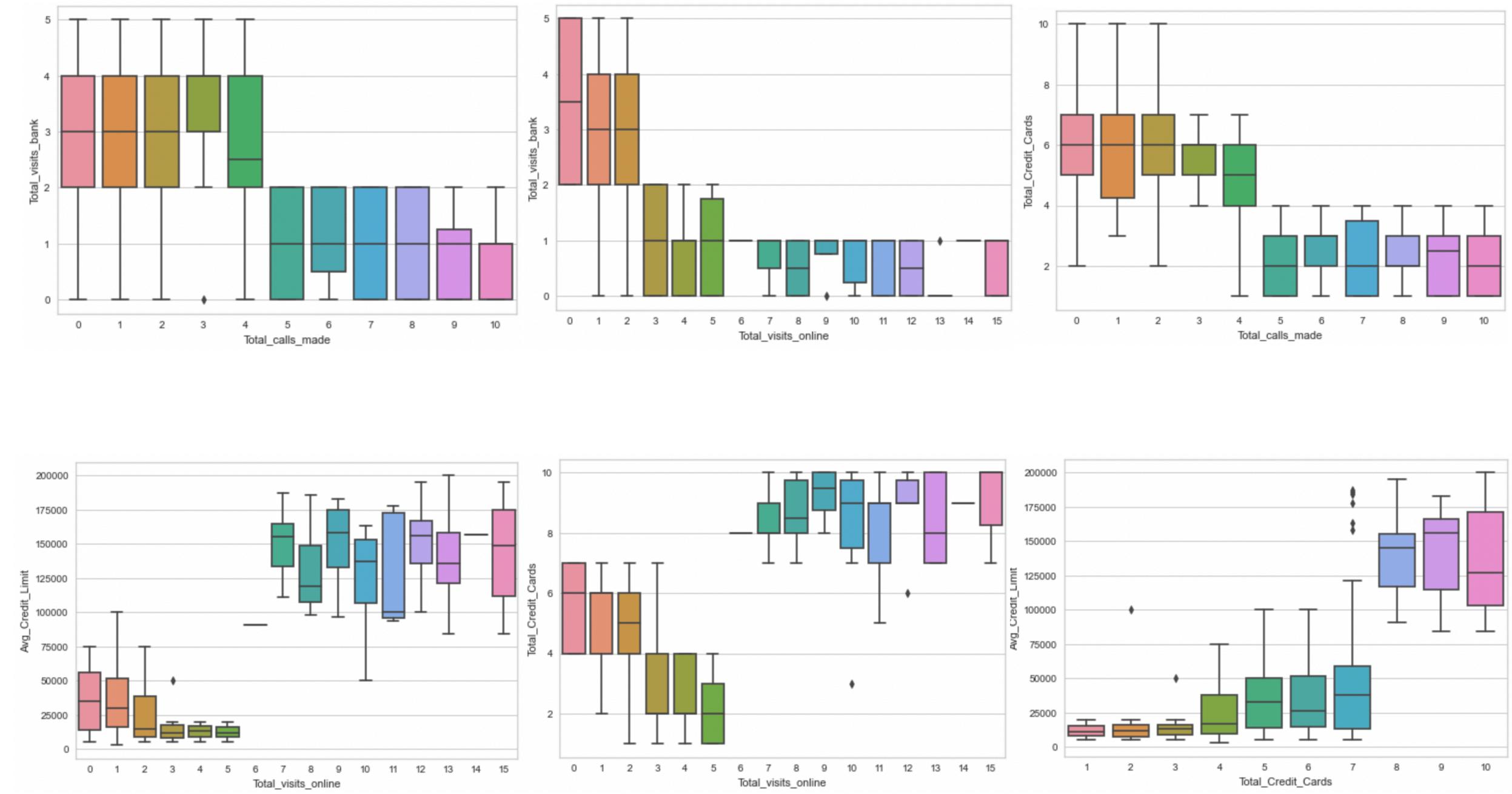
KEY FINDINGS AND INSIGHTS (EDA)

- Avg_Credit_Limit is highly right skewed, which means most of the customers have a lower Avg_Credit_Limit with a few customers having high Avg_Credit_Limit
- Most customers have 4 credit cards followed by those who have 6 credit cards
- Most customers visit the bank twice yearly followed by those customers who visit once
- Most customers visit the bank website only twice a year followed by those who never visit the website
- Most customers have called the bank 4 times in a year followed by those who never call the bank



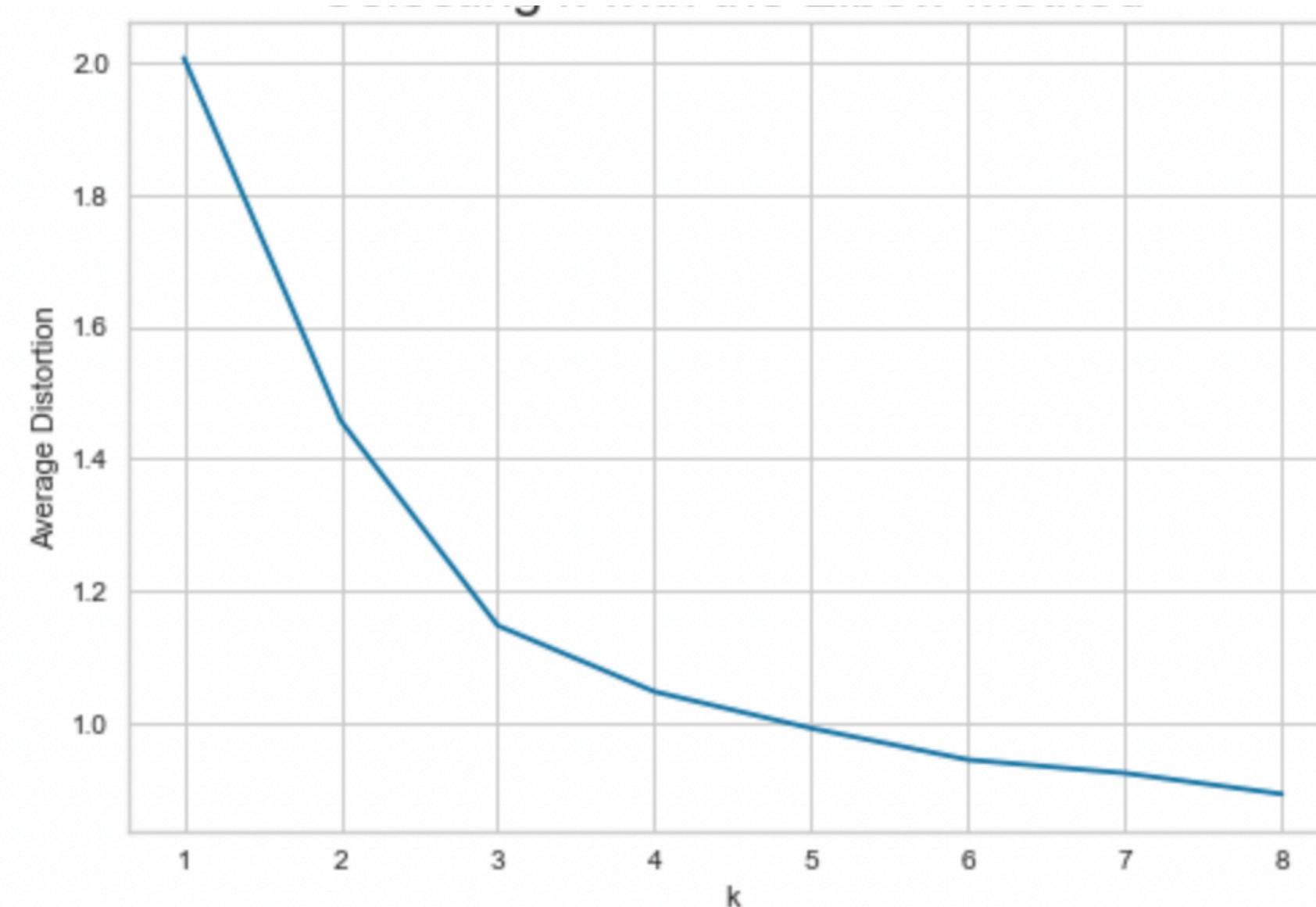
KEY FINDINGS AND INSIGHTS (EDA)

- We can see that higher the Avg_Credit_Limit, higher is the number of credit cards possessed by the customer
- We can see that higher the Avg_Credit_Limit, higher is the number of visits made to the banks website
- We can see that greater the number of credit cards, higher is the number of visits made to the banks website
- We can see that greater the number of credit cards, less calls are made to the bank
- We can see that greater the number of online visits, lesser is the number of bank visits
- We can see that smaller the number of total calls made, greater is the total number of bank visits

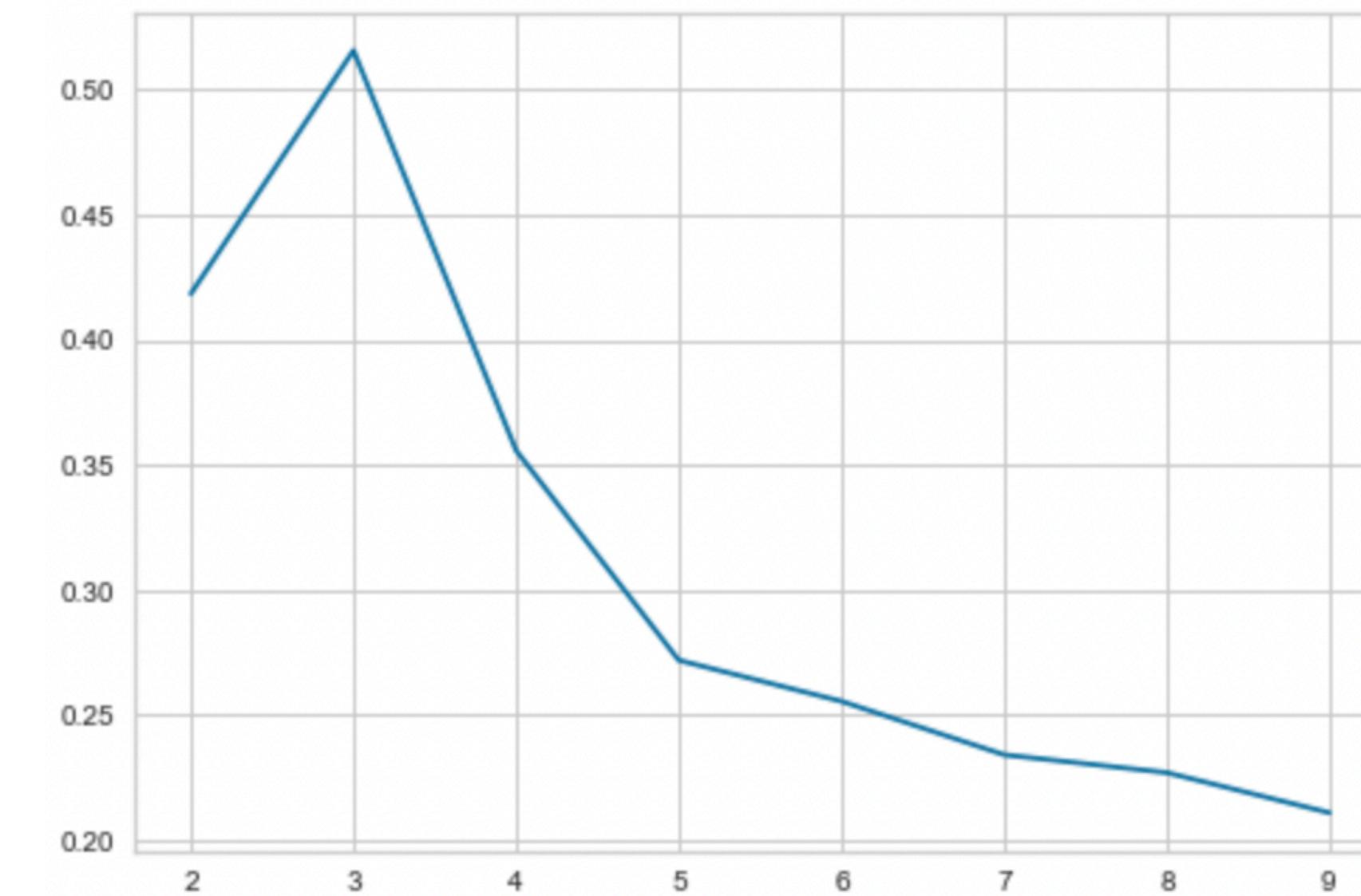


K MEANS CLUSTERING ALGORITHM

ELBOW CURVE



SILHOUETTE SCORES



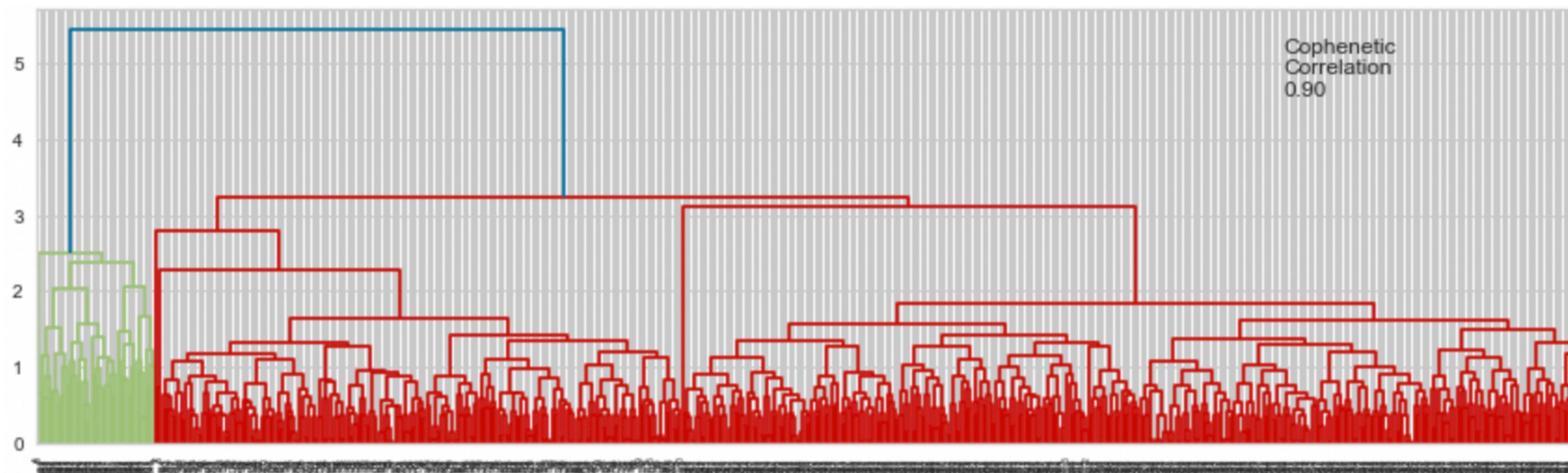
Number of clusters = 3 as there is an elbow at 3 and 3 has the highest silhouette scores

HIERARCHICAL CLUSTERING ALGORITHM

DIFFERENT LINKAGE METHODS WITH EUCLIDEAN DISTANCE

Cophenetic correlation for single linkage is 0.7391220243806552.
Cophenetic correlation for complete linkage is 0.8599730607972423.
Cophenetic correlation for average linkage is 0.8977080867389372.
Cophenetic correlation for centroid linkage is 0.8939385846326323.
Cophenetic correlation for ward linkage is 0.7415156284827493.
Cophenetic correlation for weighted linkage is 0.8861746814895477.

DENDROGRAM (AVERAGE LINKAGE)

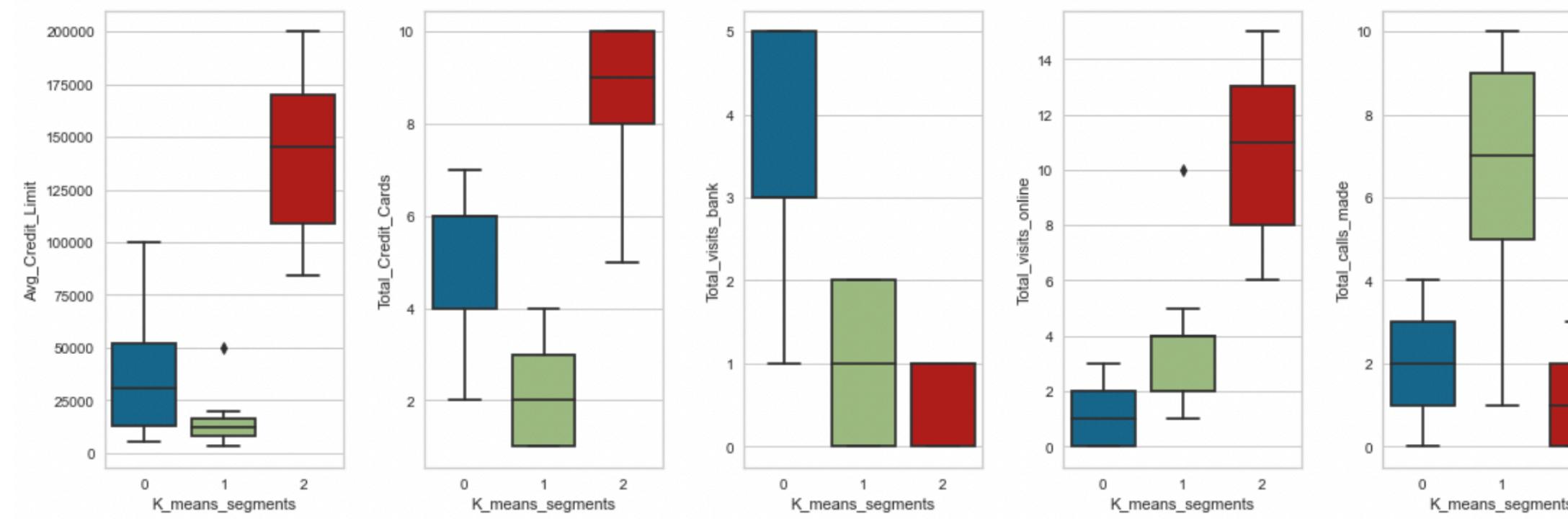


Choosing the dendrogram with average linkage since the correlation is maximum with Euclidean distance and average linkage

MODEL COMPARISON

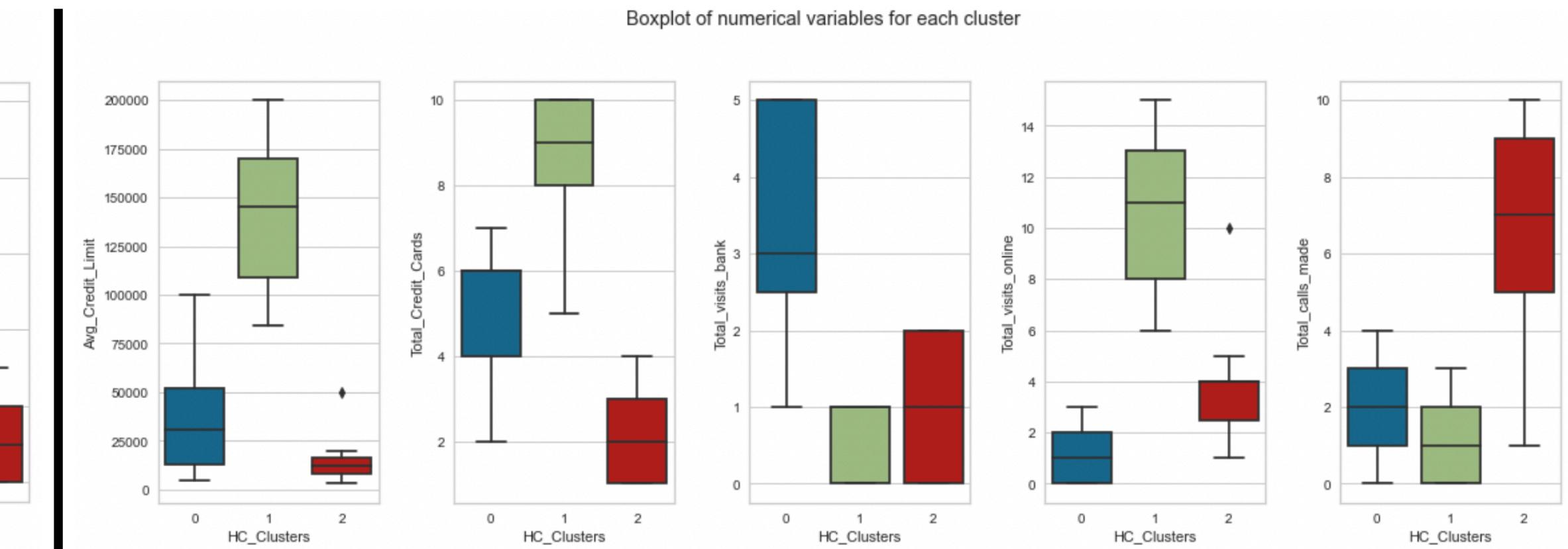
K-MEANS CLUSTERING

Boxplot of numerical variables for each cluster



HIERARCHICAL CLUSTERING

Boxplot of numerical variables for each cluster



- The clusters resulting from the two algorithms are almost identical
- The K-means clustering technique takes lesser time than Hierarchical clustering
- 3 distinct clusters are obtained from both algorithms

BUSINESS RECOMMENDATIONS

- **For In-person users:**

These customers visit the bank around 3 times per year. These visits could be used to talk to the customers to understand what their needs are in terms of credit cards.

The bank could also use mail as a medium to target this customer group with ads.

Count of customers in each cluster	K-Means clustering	Hierarchical Clustering
First Cluster (In-person users)	386	387
Second Cluster (Online users)	50	50
Third Cluster (Phone users)	224	223

BUSINESS RECOMMENDATIONS

- **For Online users:**

Since these customers visit the bank website a lot, the bank could target this customer base by online marketing techniques.

The bank could also use email as a medium to target this customer group with ads.

- **For Phone users:**

These customers could be offered a higher credit limit.

Since there is a direct positive correlation between Avg_Credit_Limit and Total_Credit_Cards, it could lead to increasing the number of credit cards a customer has.

The bank could call these customers and use phone as a medium to target this customer group with ads.