AllLife Bank

Contents

- Business Problem Overview and Approach
- Objective
- Data Overview
- Exploratory Data Analysis (EDA) – Univariate Analysis
- EDA – Bivariate Analysis Between Two Attributes
- Model Performance Summary
- Conclusions

Business Problem Overview and Solution Approach

- **Core Business Idea**
  AllLife Bank wants to focus on its credit card customer base in the next financial year. Advised by their marketing research team that the penetration in the market can be improved, the marketing team proposes to run personalized campaigns to target new customers as well as upsell to existing customers.

- **Problem to Tackle**
  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.

- **Financial Implications**
  The company wants to harness the available data of their existing credit card customers to make the marketing expenditure more efficient in running personalized campaigns to target new customers as well as upsell to existing customers.

- **Machine Learning to Solve the Problem**
  By applying unsupervised learning techniques will be able to determine what customer profiles should be targeted more and predict the potential customer who is going to purchase the newly introduced travel package.

Objective

To identify different segments within the existing customer base, based on their spending patterns as well as past interaction with the bank, using clustering algorithms

Based upon the provided data set, conduct an analysis to complete the following:

- Explore and visualize the dataset
- Apply the clustering algorithms to discover hidden or interesting patterns in unlabeled data
- As needed, apply dimensionality reduction techniques to reduce the dimension of data
- Generate a set of insights and recommendations that will help the bank to better market to and service these customers.

Data Overview

- Dataset contains information in relation to their personal and customer banking profile

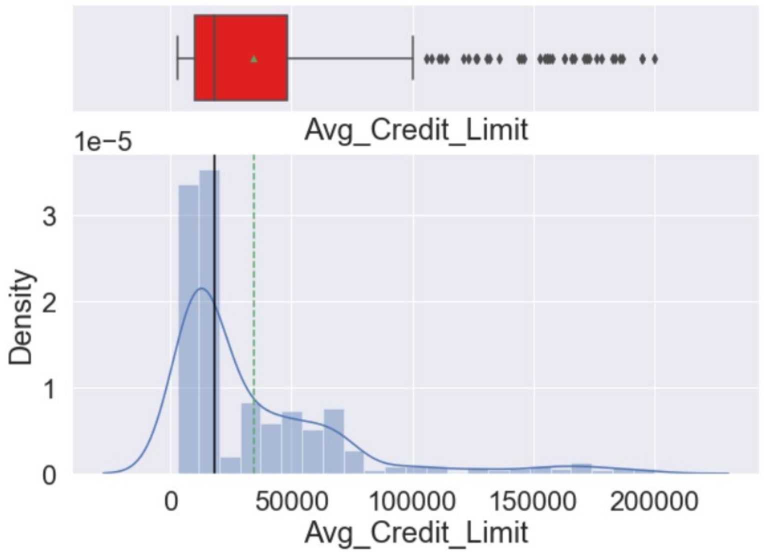
Variable	Description
I_No	Primary key of the records
Customer Key	Customer identification number
Average Credit Limit	Average credit limit of each customer for all credit cards
Total credit cards	Total number of credit cards possessed by the customer
Total visits bank	Total number of Visits that customer 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 number of calls made by the customer to the bank or its customer service department (yearly)

Total Entries	Total Variables	Missing Data	Duplicate Observations
660	7	0	5 pairs

Data Prep Required:

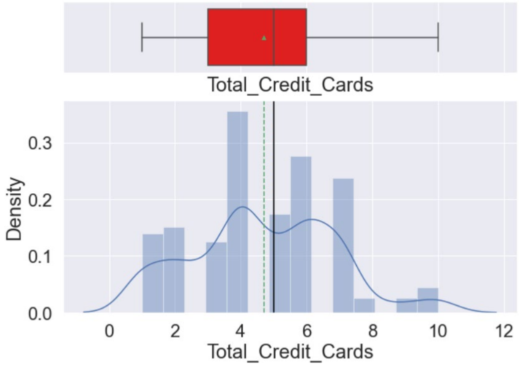
- I_No attribute immediately removed from the dataset, based on not providing value to any type of analysis
- Analysis of duplicate customer identification observations required

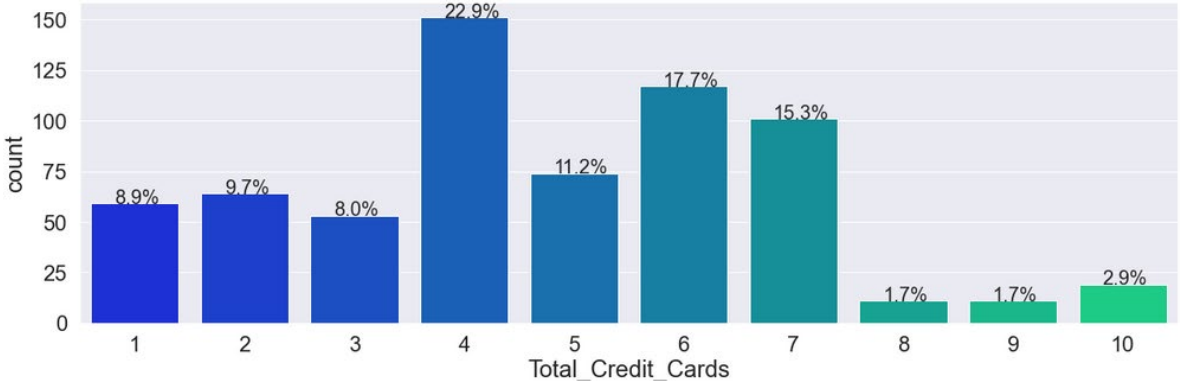
EDA – Univariate Analysis: Average Credit Limit



- Contains 110 unique observations and will be handled as a continuous type of numerical value
- Displays a range of 3k to 200k maximum, with skewing since the mean is ~34.6k, while the IRQ 50% is 18k
- The unique counts revealed that the top unique entry counts provided credit limits in a 5k - 20k range, with unique counts ranging from 18 to 35
- Displays possibly three peaks based on the KDE curve in the 10K, 50K, and the 170K ranges, with the majority of the customers in the 10k to 48k range based on the IRQ boxplot

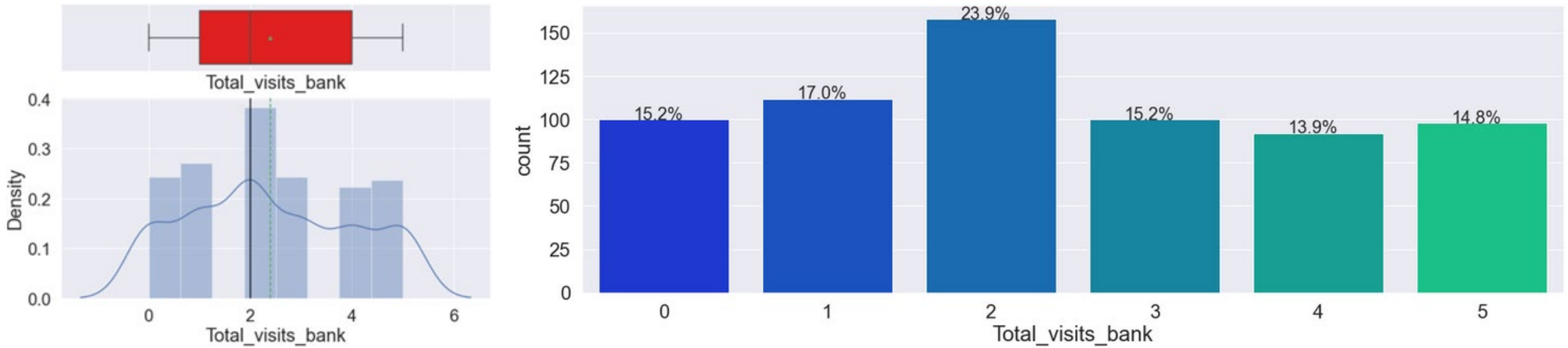
EDA – Univariate Analysis: Total Credit Cards





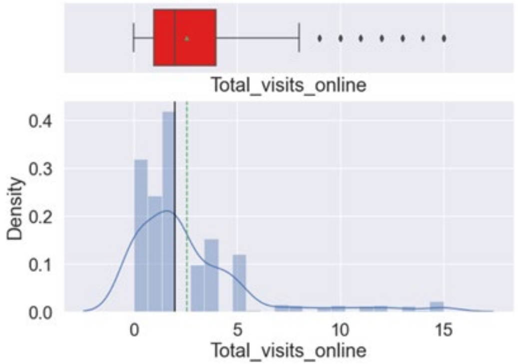
- Displays a range of 1 to 10 credit cards held by the bank's customers, with an average of ~5 credit cards per customer
- Though the average is ~5 credit cards, customers with 4, 6, or 7 credit cards is much more common
- Displays 3 to a possible 5 groups of credit card holders, with 8, 9, 10, and 5 consisting of one group, 8, 9, and 10 as a second group, and 4, 6, and 7, being a third group

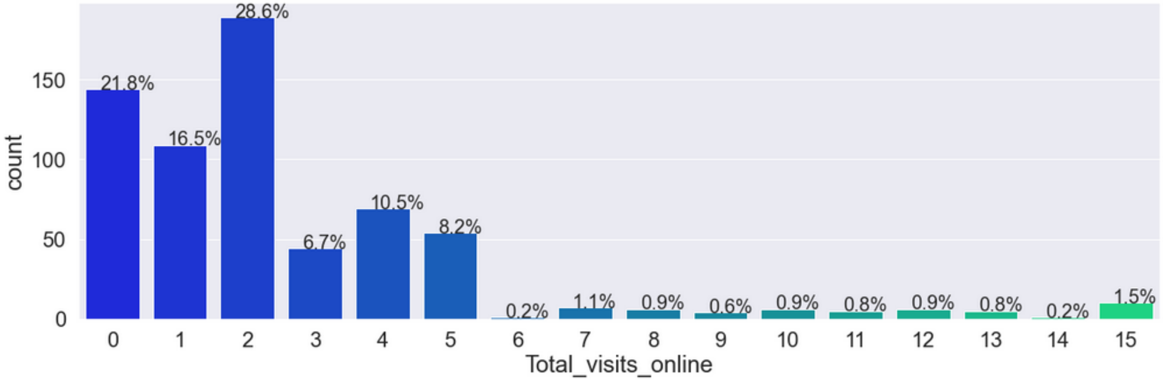
EDA – Univariate Analysis: Total Visits to Bank



- Displays a range of 0 to 5 total visits to the bank, with an average of ~2.5 visits per customer
- 2 total visits is the highest number 23.9% (158) of visits to the bank, while all other entries have 112 or less entry counts
- Displays two groups of visits to the bank, with one group consisting of 0, 1, 3, 4, and 5 visits, and 2 visits being a separate group

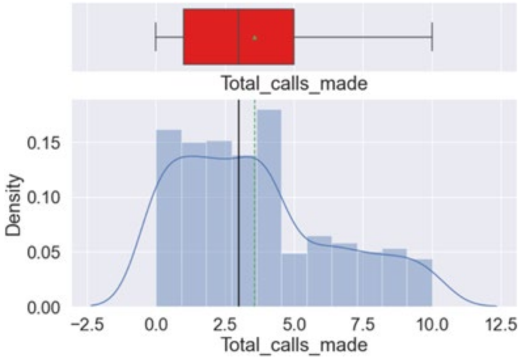
EDA – Univariate Analysis: Total Visits Online

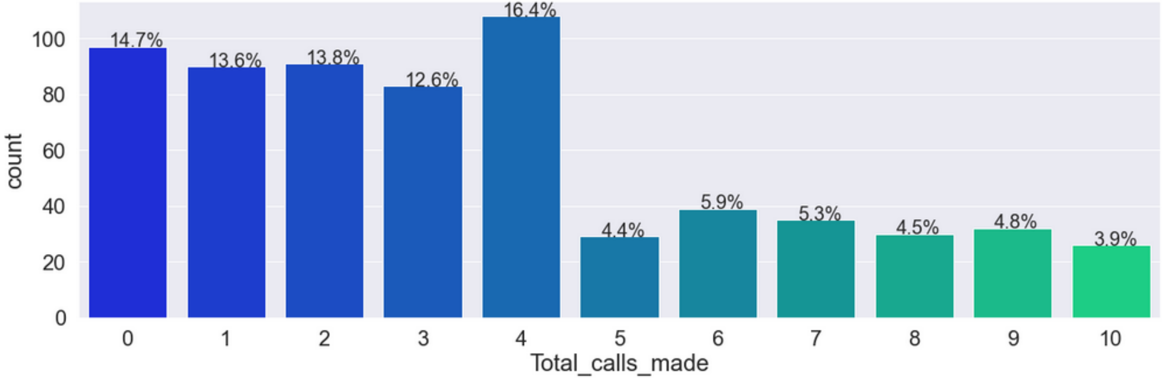




- Displays a range of 0 to 15 total online visits to the bank, with an average of ~2.6 visits per customer
- 2 total online visits is the highest number (28.6%) of online visits to the bank, while 0 (21.8%)and 1 (16.5%) have the next highest counts respectively
- Displays 3 to a possible 4 groups of visits online, with one group consisting of 0, 2, and possibly 1 visit, a second group consisting of 3, 4, and 5 visits, and a third group consisting of 6 - 15 total visits

EDA – Univariate Analysis: Total Calls Made

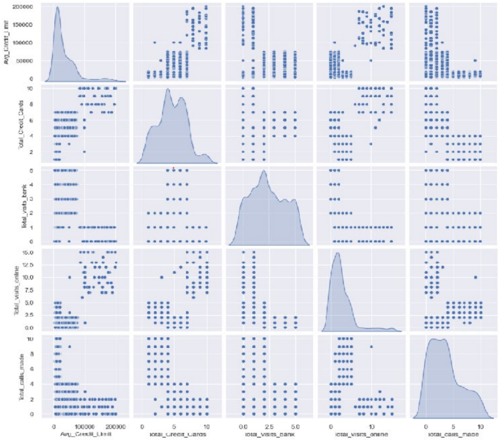




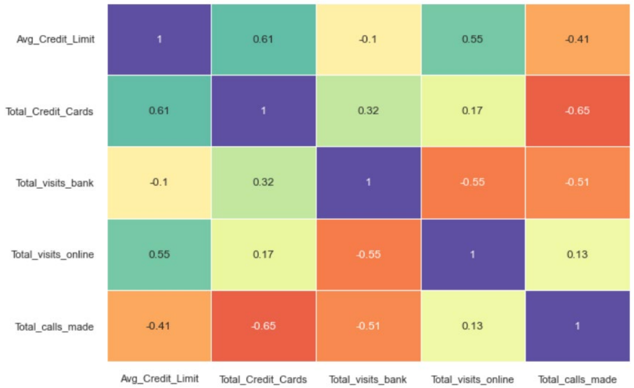
- Displays a range of 0 to 10 total calls made to the bank, with an average of ~3.6 calls to the bank per customer
- 4 total calls made is the highest number (16.4%) of calls to the bank followed by 0 (14.7%), then 2 (13.8%) and 1 (13.6%) respectively
- Displays two groups of calls made, with one group consisting of 0 - 4 calls, and the second group consisting of 5 - 10 calls

Exploratory Data Analysis – Bivariate Analysis

Covariance Plotted Using Pairplot



Correlation Based on Labels

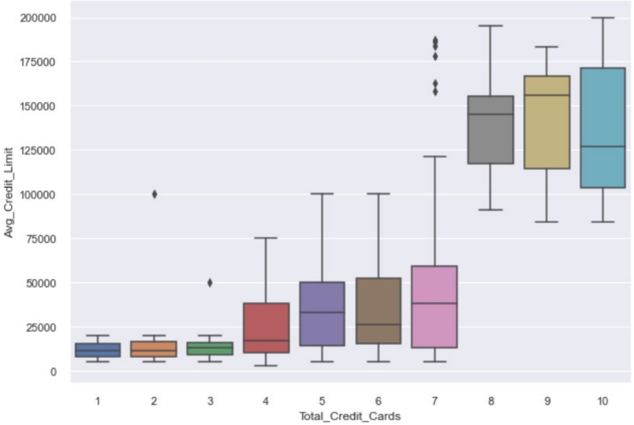


- Overall, the plots do not provide any direct correlation patterns between attributes, which may be due to four of the five attributes containing category type numerical data
- Possible groupings can be estimated for each of the attributes based on the peaks of the KDE curve for each attribute

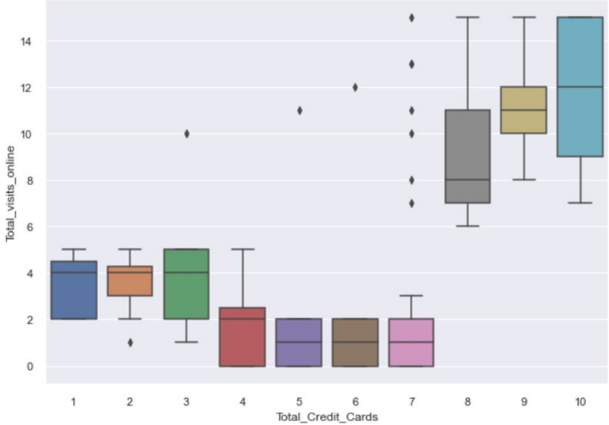
- The results of the correlation heatmap and the calculated correlation values display within the covariance table that the higher number of credit cards issued to customers is likely to lead to a much higher usage of total visits online as compared to visits or calls to the bank, as compared to calls and visits to the bank that are more common with customers having fewer credit cards

Exploratory Data Analysis – Bivariate Analysis

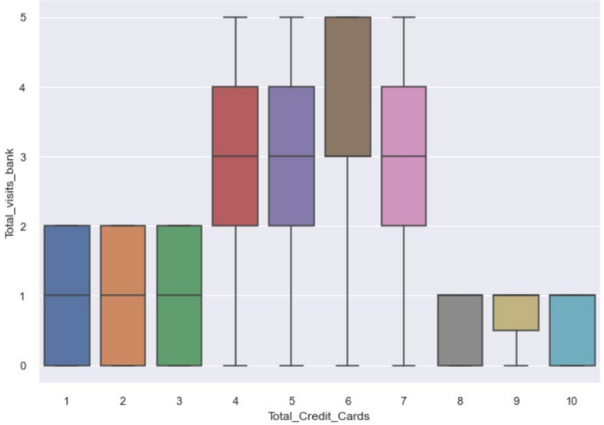
"Total_Credit_Cards" vs "Avg_Credit_Limit"



"Total_Credit_Cards" vs “Total_Visits_Online"



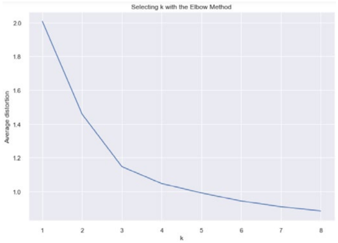
"Total_Credit_Cards" vs " Total_Visits_ Bank"

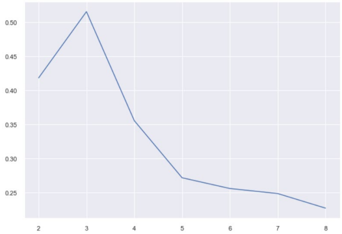


- The overall observation from the bivariate analysis displays strong grouping of observations based on the attributes being compare against. The grouping ranges from 1 - 3, with groups of 3 was more prevalent amount the groupings
- The bivariate groupings can provide the banks Operations team the critical information that is needed to upgrade their service delivery model, to ensure that customer queries are resolved faster. Grouping based on the two attributes of Total_Credit_Cards held by a customer and the Avg_Credit_Limit provide the bast way for the marketing team to run personalized campaigns to target new customers as well as upsell to existing customers.

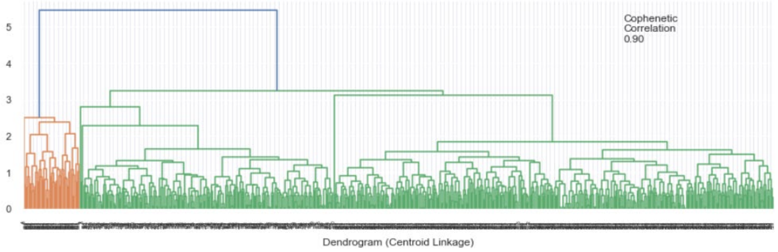
Model Overview

- Two techniques were applied to determine the groupings that the customers should be separated into
  - **K-Mean Clustering** – this centroid based method allows us to group points in such a fashion that similar points are in the same group and to determine the best number of clusters was 3 by utilizing tools such as:
    - Elbow Curves
    - Silhouette Scores
    - Cluster Profiling



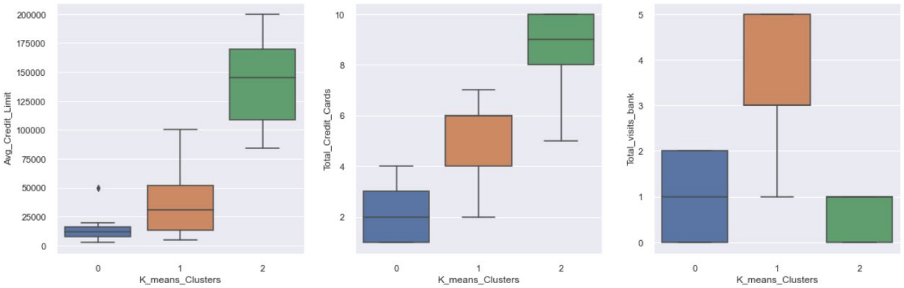


  - **Hierarchical Clustering** – Which allowed the data scientist to decide ultimately that 3 clusters were wanted after the analysis, by using tools such as
    - Dendrogram Trees for different linkage methods
    - Cophenetic correlation calculations and metrics to choose the best model
    - Cluster Profiling



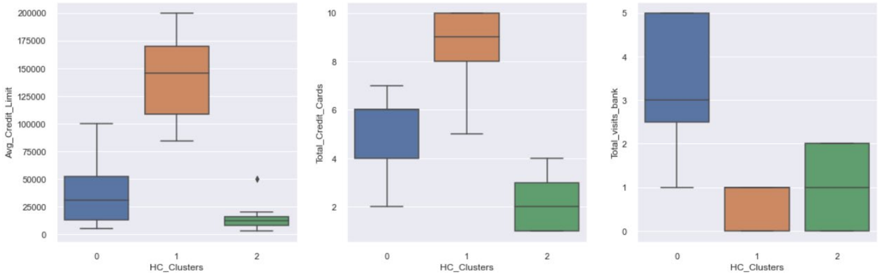
Model Performance Summary: K-means vs. Hierarchical Clusters

Boxplot of Scaled Numerical Variables for Each K-Means Cluster



neans_Clusters	Customer_Key	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	k=3 Group	count_in_each_segment
0	55239.830357	12174.107143	2.410714	0.933036	3.553571	6.870536	0.000000	224
1	54881.329016	33782.383420	5.515544	3.489637	0.981865	2.000000	1.000000	386
2	56708.760000	141040.000000	8.740000	0.600000	10.900000	1.080000	2.000000	50

Boxplot of Numerical Variables for each Hierarchical Clustering



HC_Clusters	Customer_Key	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	count_in_each_segments
0	54925.966408	33713.178295	5.511628	3.485788	0.984496	2.005168	387
1	56708.760000	141040.000000	8.740000	0.600000	10.900000	1.080000	50
2	55163.973094	12197.309417	2.403587	0.928251	3.560538	6.883408	223

- Overall results between the two methods provided very similar results, with only 1 observation difference
- Each method provided trade-offs to the other method
  - K-means Cluster – Took more time and steps to complete, but yielded a more accurate result the first time based on Elbow Curves and the Silhouette scores
  - Hierarchical Clustering – Was able to be completed in less time with fewer steps, but the clustering yielded a poor initial result
- In the end, the variable used for clustering was both the Avg_Credit_Limit and Total_Credit_Cards attributes

Business Insights and Recommendations

- Recommendations based on EDA and interpretation of the model input variables
  - AllLife Bank should make a strong effort to expand marketing based on the Average Credit Limits and Total Credit Cards held by their customer base
  - Based upon the Total Credit Cards attribute the total credit cards held by a customer and the Avg_Credit_Limit provide the bast way for the marketing team to run personalized campaigns to target new customers as well as upsell to existing customers
  - The bivariate groupings can provide the banks Operations team the critical information that is needed to upgrade their service delivery model, to ensure that customer queries are resolved faster.

- Insights based on EDA and interpretation of the model input variables
  - The overall observation from the bivariate analysis displays strong grouping of observations based on the attributes being compare against. The grouping ranges from 1 - 3, with groups of 3 was more prevalent amount the groupings
  - Univariate analysis can provide much insight into the dataset prior to conducting any unsupervised clustering technique especially though the use of count plots
  - Once the dataset has been clustered this new attribute can be added back into the original table to enable other regression type testing

greatlearning
Power Ahead

Happy Learning !