# UNSUPERVISED LEARNING

Coded Project Report

AllLife Bank Customer Segmentation Project

Submitted to

**greatlearning**
*Learning for Life*

by

## Subhadeep Seal

In Partial Fulfillment of PGP-DSBA

**TEXAS McCombs**
The University of Texas at Austin
McCombs School of Business

# Table of Contents

## 1.1 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.

## 1.2 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.

## 1.3 Data Description

The data provided is of various customers of a bank and their financial attributes like credit limit, the total number of credit cards the customer has, and different channels through which customers have contacted the bank for any queries (including visiting the bank, online, and through a call center).

### 1.3.1 Data Dictionary:

- 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 the 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)

## 1.3.2 Understanding the structure of data:

| | Sl_No | Customer Key | Avg_Credit_Limit | Total_Credit_Cards | Total_visits_bank | Total_visits_online | Total_calls_made |
|---|---|---|---|---|---|---|---|
| **120** | 121 | 16577 | 10000 | 4 | 2 | 4 | 6 |
| **573** | 574 | 96534 | 66000 | 7 | 2 | 2 | 3 |
| **70** | 71 | 21976 | 14000 | 1 | 2 | 4 | 7 |
| **19** | 20 | 14309 | 19000 | 4 | 1 | 5 | 8 |
| **385** | 386 | 16418 | 8000 | 7 | 4 | 2 | 0 |
| **405** | 406 | 90189 | 51000 | 7 | 3 | 0 | 4 |
| **647** | 648 | 79953 | 183000 | 9 | 0 | 9 | 2 |
| **596** | 597 | 74083 | 42000 | 6 | 4 | 2 | 0 |
| **424** | 425 | 61122 | 33000 | 5 | 4 | 2 | 0 |
| **642** | 643 | 91575 | 97000 | 10 | 1 | 9 | 2 |

# 1.4  Steps:

- Performing univariate analysis on the data to better understand the variables and to get an idea about the number of clusters.
- Performing EDA and get insights.
- Creating visualizations to explore data and present the insights.
- Executing K-means clustering use elbow plot and analyse clusters using boxplot.
- Executing hierarchical clustering (with different linkages) with the help of dendrogram and cophenetic coefficient.
- Analysing clusters formed using boxplot.
- Calculating average silhouette score for both methods.
- Comparing K-means clusters with Hierarchical clusters.
- Analysis of the clusters formed, explain how one cluster different from another.

## 2. Exploratory Data Analysis

## 2.1 Pair-plot
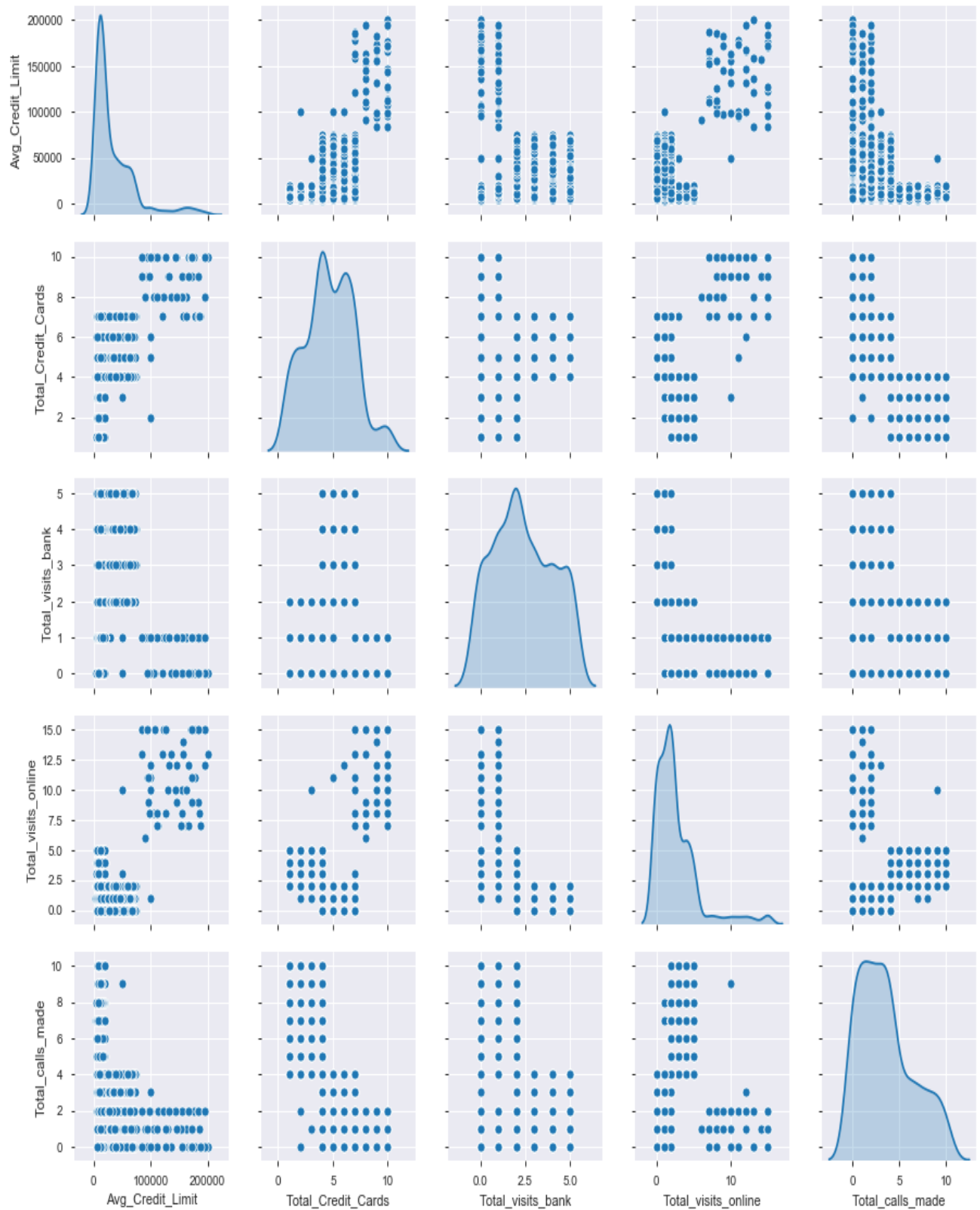
## 2.2 Heatmap for correlation
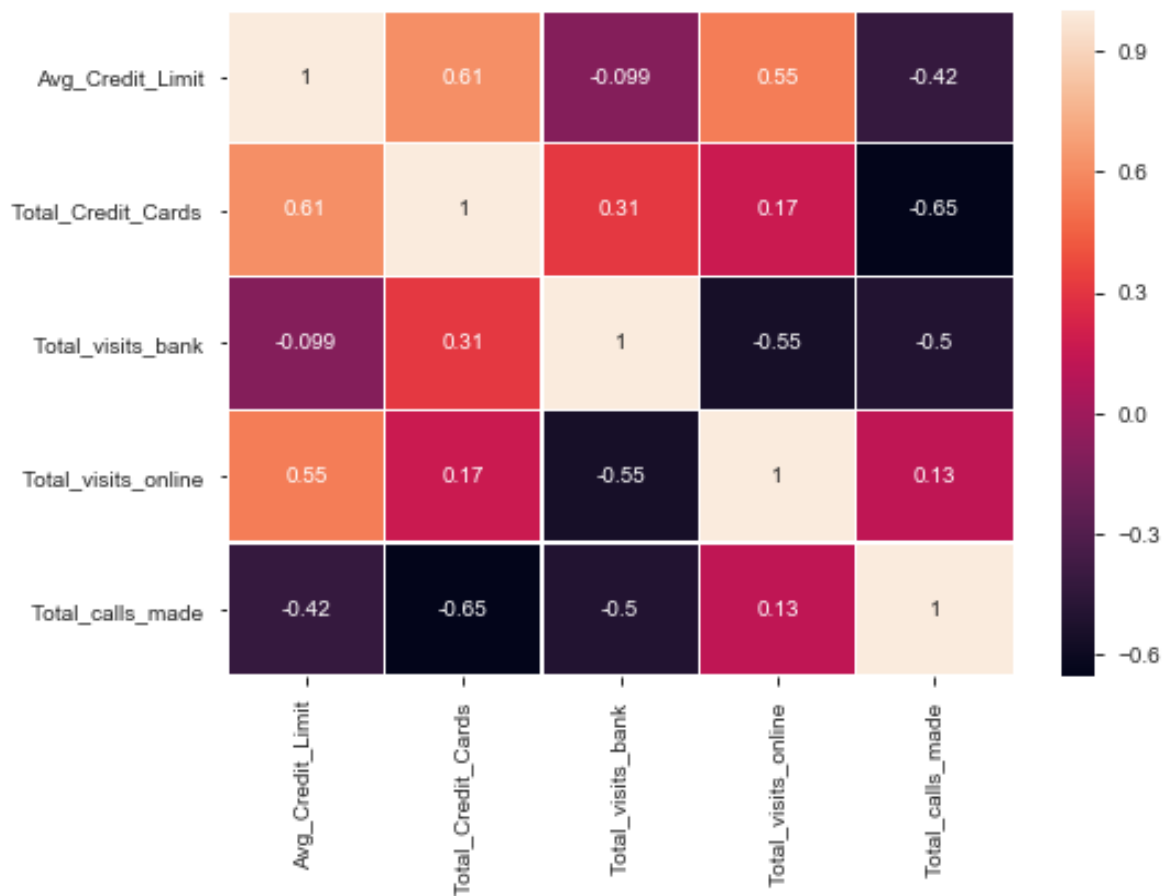


**Comments:**

- Total credit cards and total visit online has medium positive correlation with average credit limit : 0.61,0.55 respectively.

- Total credit cards and total visit bank has medium negative correlation with total calls made : -0.65,-0.5 respectively.

- Total visit online and medium negative correlation with total visit bank : -0.55.

# 3.1 K-means Clustering - Boxplot for analysing clusters
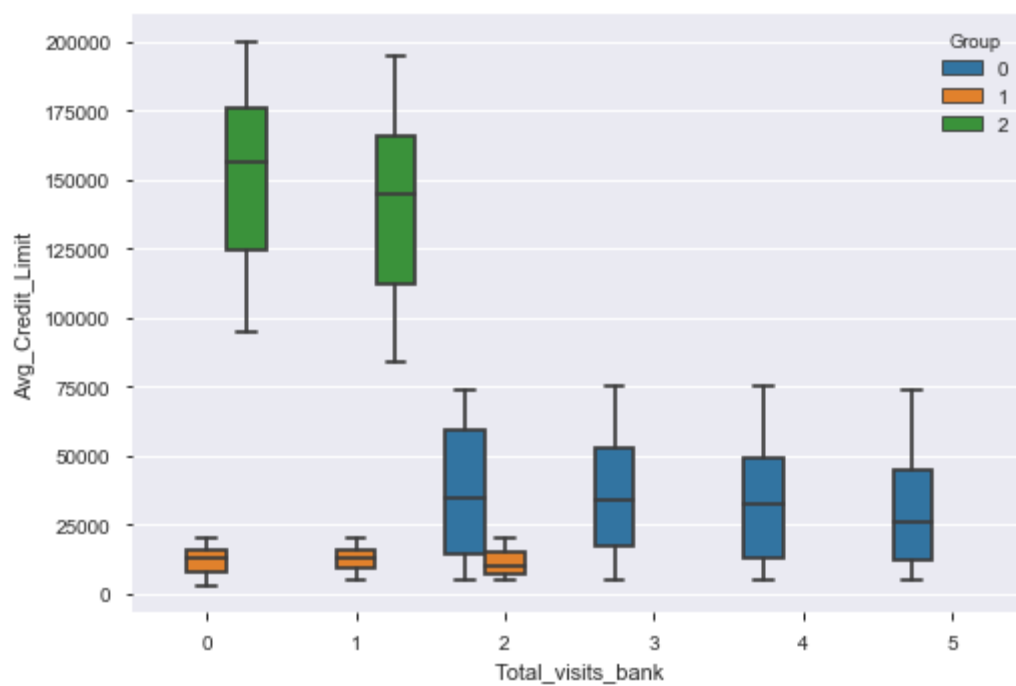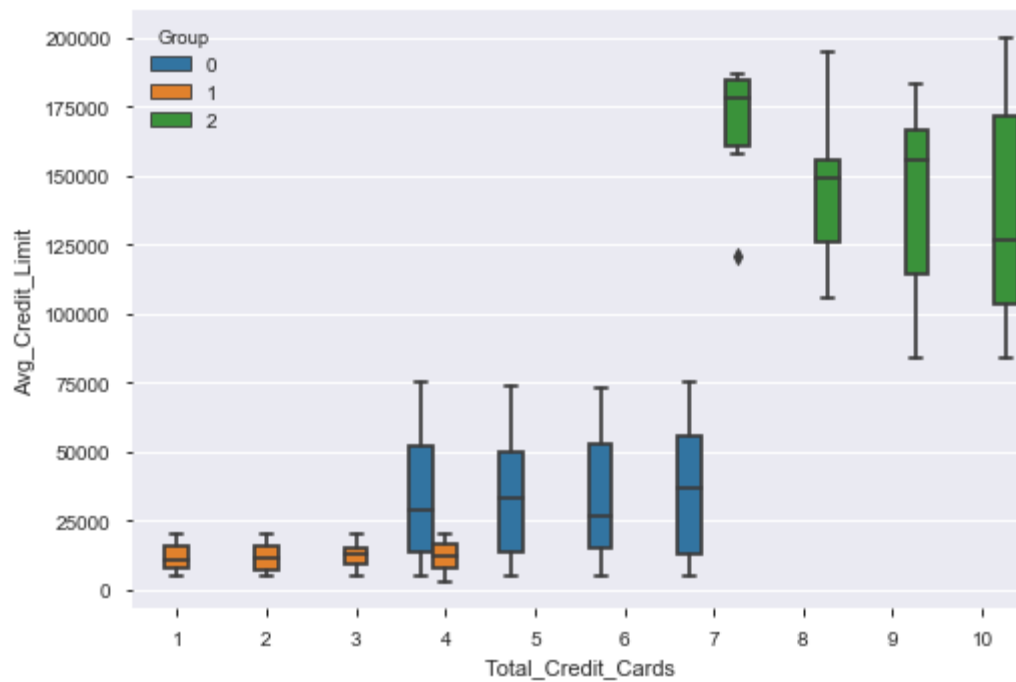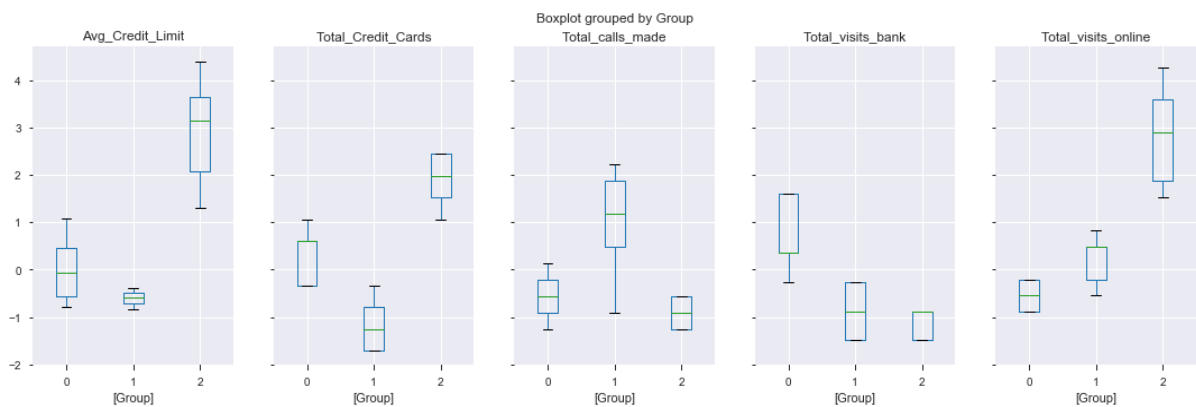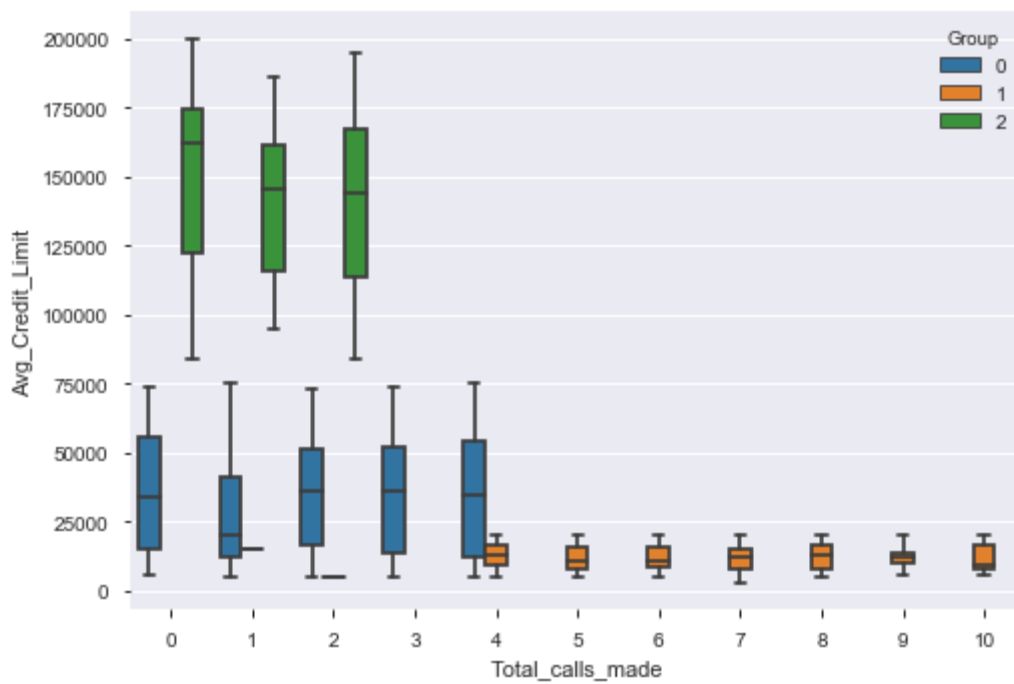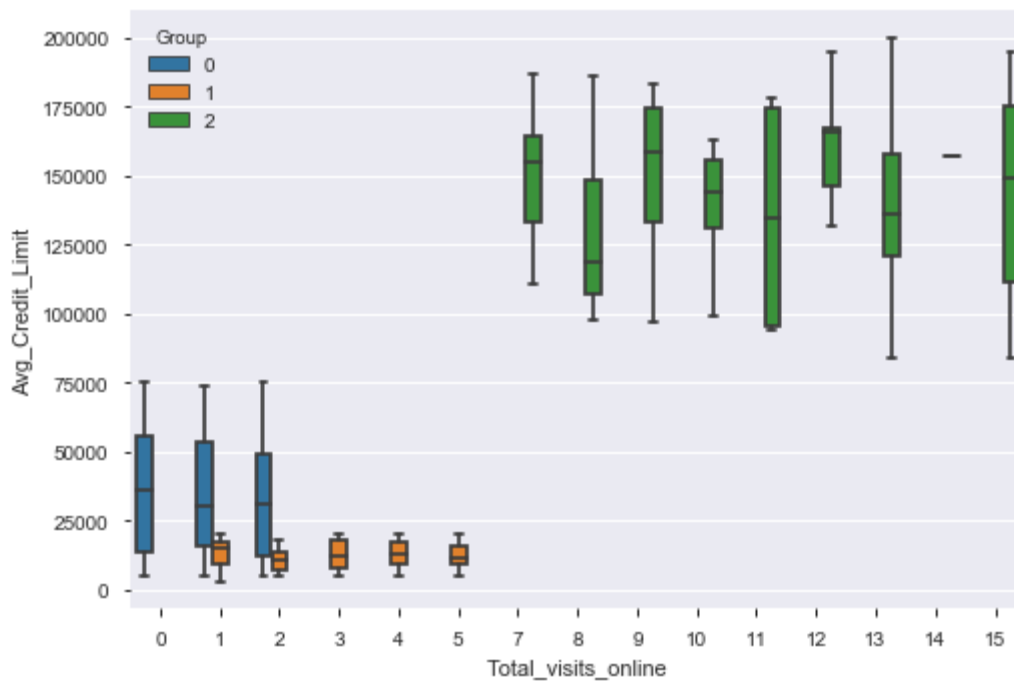
**Comments:**

- Group 2 has an obvious distinction compare to group 0 and 1.

  - Customers who have average credit limit of 75k and above are in group 2.

  - Customers who have more than 7 credit cards are in group 2.

  - Customers who have visited online more than 6 times are in group 2.

- Group 1 has less credit cards compare to other groups - 1-4 credit cards.

- Group 1 has one attribute that is distinct from other group namely, customers who make calls more than 4 times.

- Group 0 visited bank more than other groups - more than 2 times and up to 5 times.

- Group 0 also less visited bank online compare to other groups.

## 3.2 k-means cluster center compare with features

| | Avg_Credit_Limit | Total_Credit_Cards | Total_visits_bank | Total_visits_online | Total_calls_made |
|---|---|---|---|---|---|
| 0 | -0.015485 | 0.377806 | 0.670462 | -0.553460 | -0.558444 |
| 1 | -0.594379 | -1.055726 | -0.905492 | 0.326169 | 1.142877 |
| 2 | 2.905755 | 1.927283 | -1.110958 | 2.889150 | -0.893968 |

**Comments:**

- Group 2 has the highest value for Avg_Credit_Limit,Total_Credit_Cards & Total_visits_online.

- Group 2 has the lowest value for Total_visits_bank.

- Group 1 has the highest value for Total_calls_made.

- Group 1 has the lowest value for Avg_Credit_Limit & Total_Credit_Cards.

- Group 0 has the highest value for Total_visits_bank.

- Group 0 has the lowest value for Total_visits_online.

## 4.1 Key Questions:

- How many different segments of customers are there?

  * There are 3 different segments of customers in AllLife Bank credit card customer base.

- How are these segments different from each other?

  Using k-means clustering labels as reference: (Group 0-2)

Group 1
  * Customers who have average credit limit below 25k and own credit cards 1-4 max.
  * Customers who seldom visit bank 0-2 times.
  * Customers who visit bank online moderately (1-5 times)
  * Customers who make most phone calls (4-10 times)
  * 34% of customers are in this group.

    Group 2
  * Customers who have average credit limit above 75k and own most credit cards 7-10 max.
  * Customers who least visited bank 0-1 times.
  * Customers who most visit bank online . (7-15 times)
  * Customers who make least phone calls (0-2 times)
  * Only 7.3% of customers are in this group.

    Group 0
  * Customers who have average credit limit between 25k and 75k and own credit cards 4-7 max.
  * Customers who most visited bank  2-5 times.
  * Customers who least visit bank online. (0-2 times)
  * Customers who make moderate phone calls (0-4 times)
  * Majority of customers are in this group - 58.4%.

## 4.2 Actionable insights and Recommendations

* Group 1 own less credit card than others, bank should target group 1 to upsell credit cards services.

 * Besides, bank should provide higher credit limit to target group 0 where most of the customers are. With higher credit limit, group 0 would be able to spend more.

 * Since group 0 use the online banking the least, bank should promote more to group 0 in order for them to use it.

 * Assuming group 1 who make most phone calls are the customers perceive the support services of the bank poorly. Bank should target group 1 and provide better customers service by conducting feedback survey through phone.