Client Classification

By Pablo Nguema

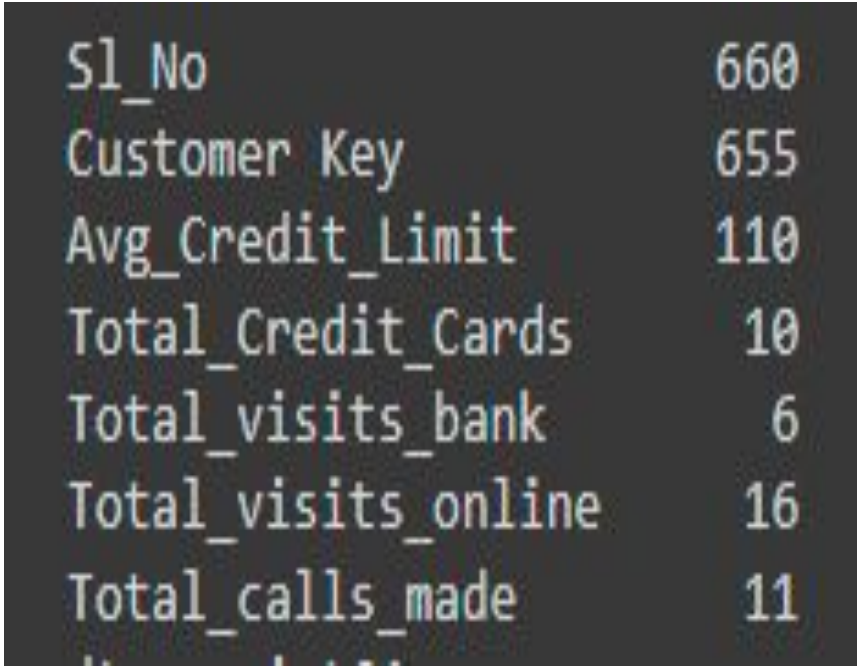
Objectives of Project

1)Analyse the distribution of the clients depending on the categories of 'Avg_Credit_Limit', 'Total_Credit_Cards', 'Total_visits_bank', 'Total_visits_online', 'Total_calls_made'

2)Create a model that would classify the client in groups in accordance to the aglomerate of their properties

3) Plan on how to advertise to them.

Analysis I



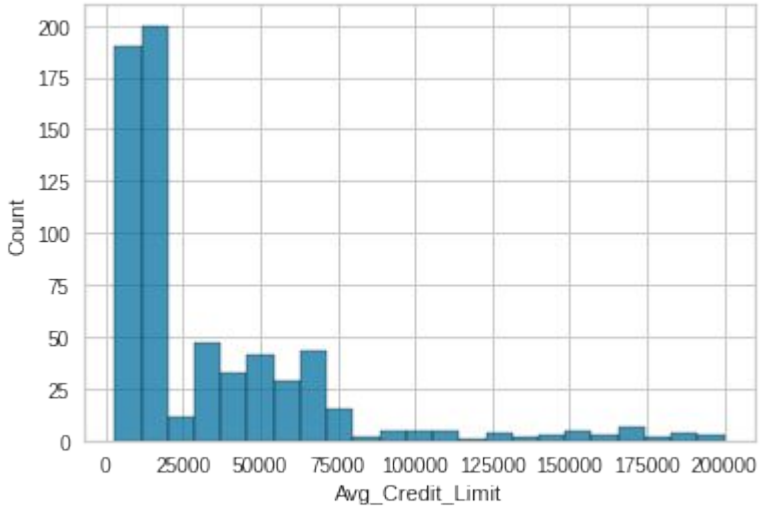
There are almost no replicas between Sl_No, adn Customer Key:

This suggest that there were at most 5 people coming back for different transactions.

The amount of credit limits are 110,

the max Credit Cards are 10,Calls made 11, Visits Online 16, Visits to the Bank 11

Analysis II(Credit Limit)

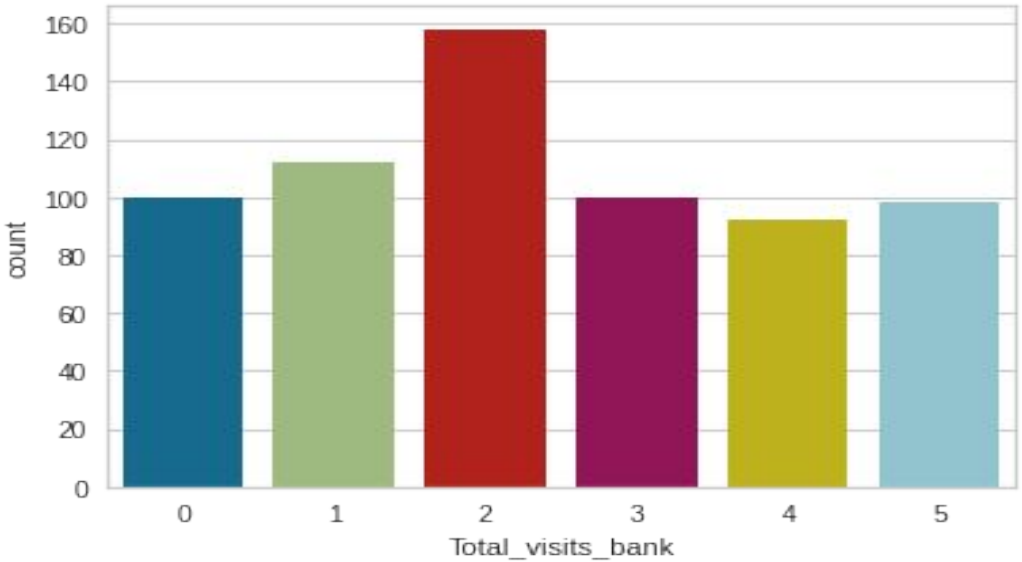


Most clients that we have signed with us are within the 0-25000 $ credit limit

It decreases at 25K$ while increases and remains around the same count between 25K-75K.

Most clients credit max are very much within 0- 75K $

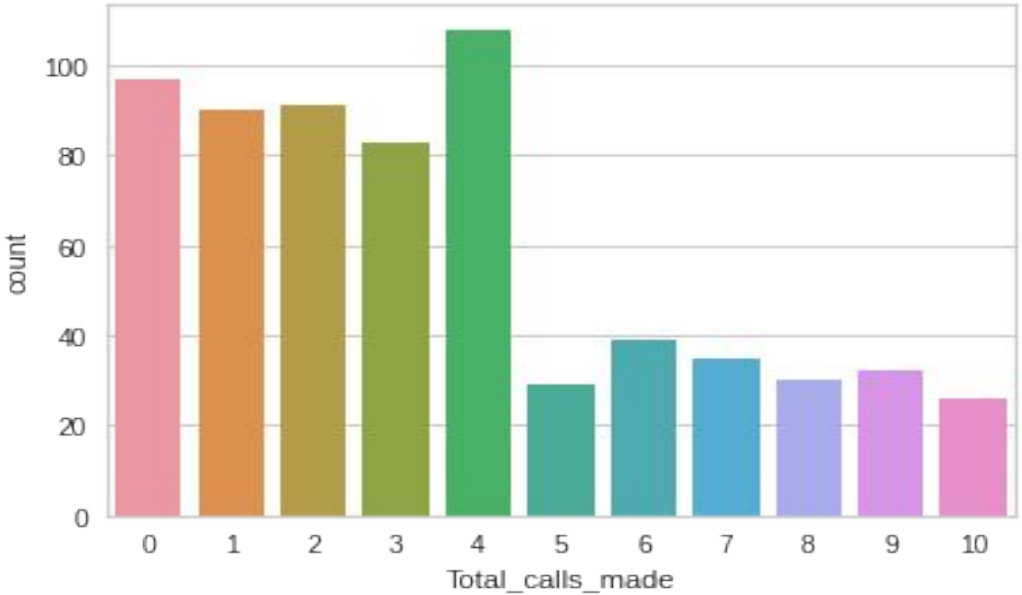
Analysis III(Total Visits)



Most people visit the banks on an mode of 2 times.

Could be due to several factors,but the one of the few could be age, phone line availability or lack of online understanding,which is linked to the age

Analysis IV(Total Calls made)

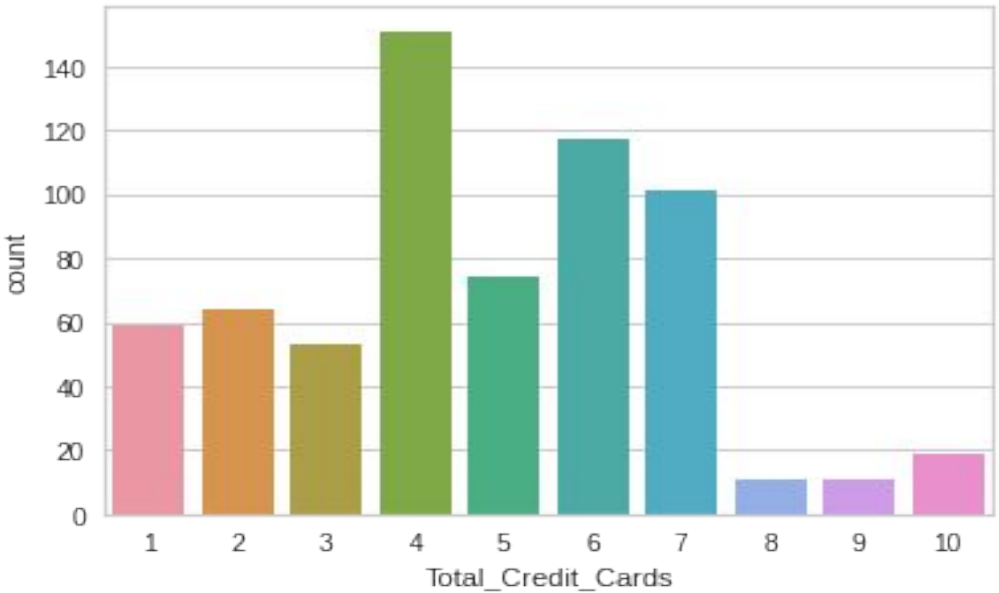


Most clients visited the called the bank 4 times,the second group is the 0 calls group.

The 0 calls groups might have come to the bank to get customer service

And those who call several times might have been interrupted on their first call

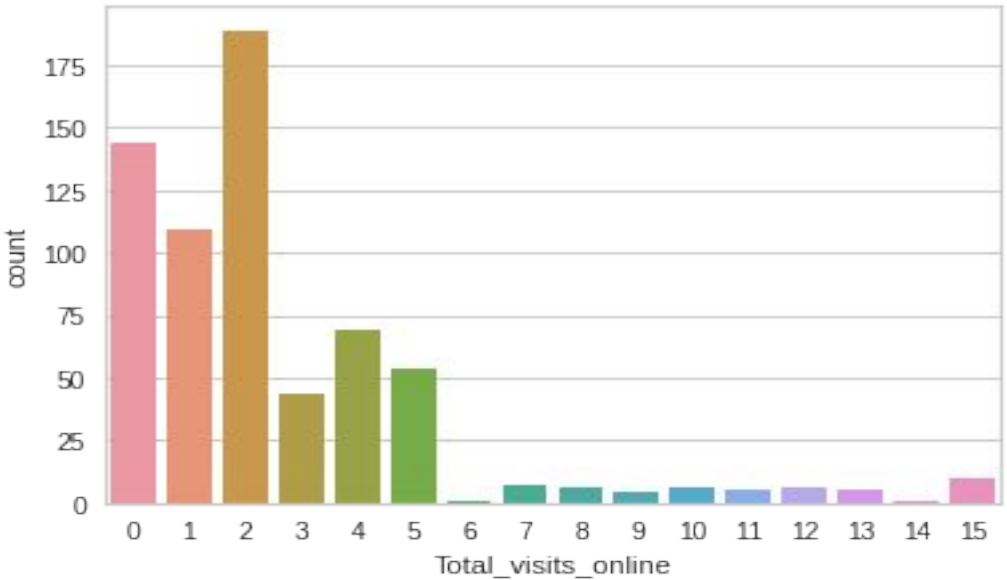
Analysis V(Total Credit Cards own)



Most of our clients have 4 credit cards with us, as far as we can tell.The second largest groups are 6 and 7.

Could be that they have different credit cards for different rewards type. While having credit cards for their children.

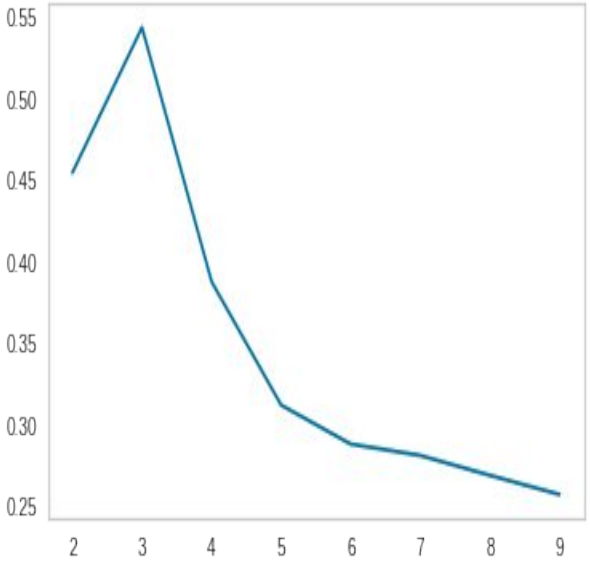
Analysis (Total Visits Online)

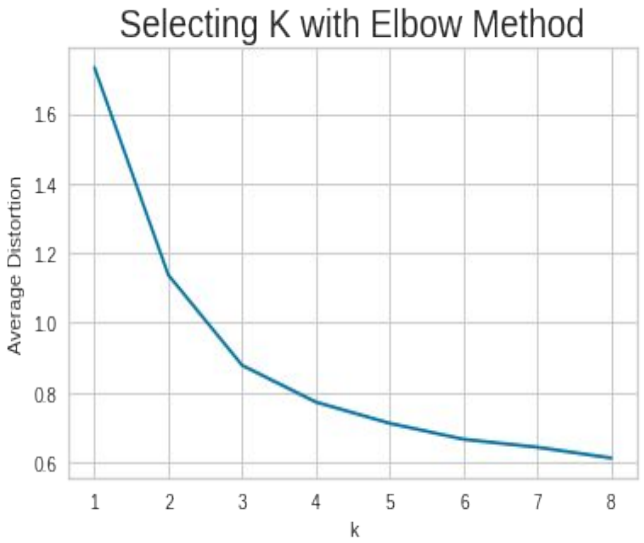


Most people check their bank account twice.

Could be to confirm transaction or check available credit,before and after falling a sleep.

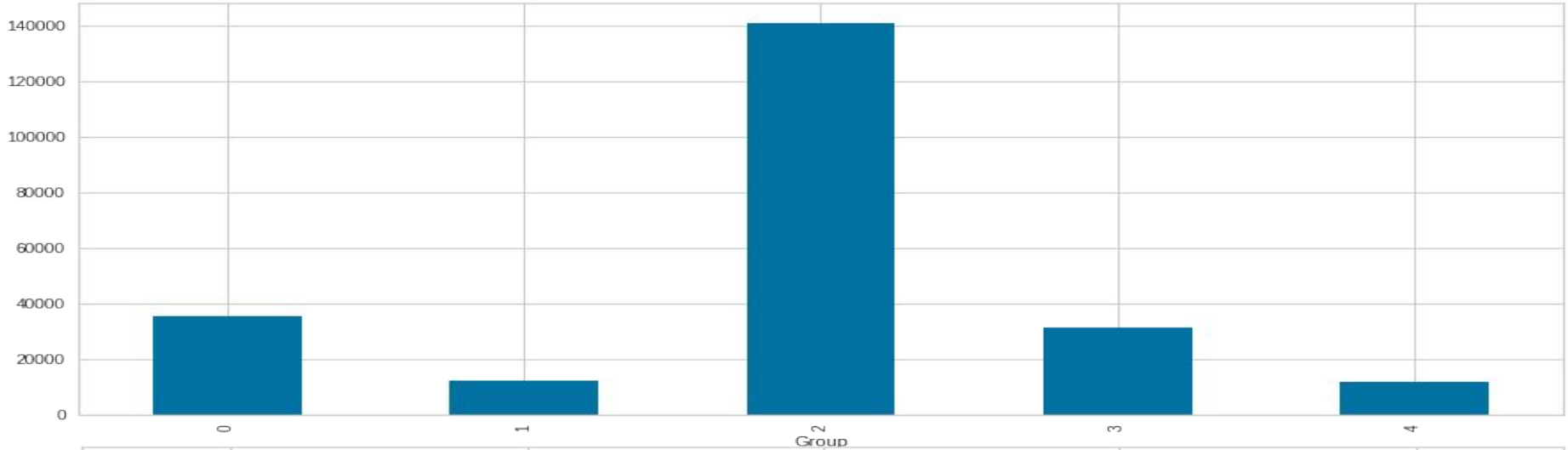
Model Analysis

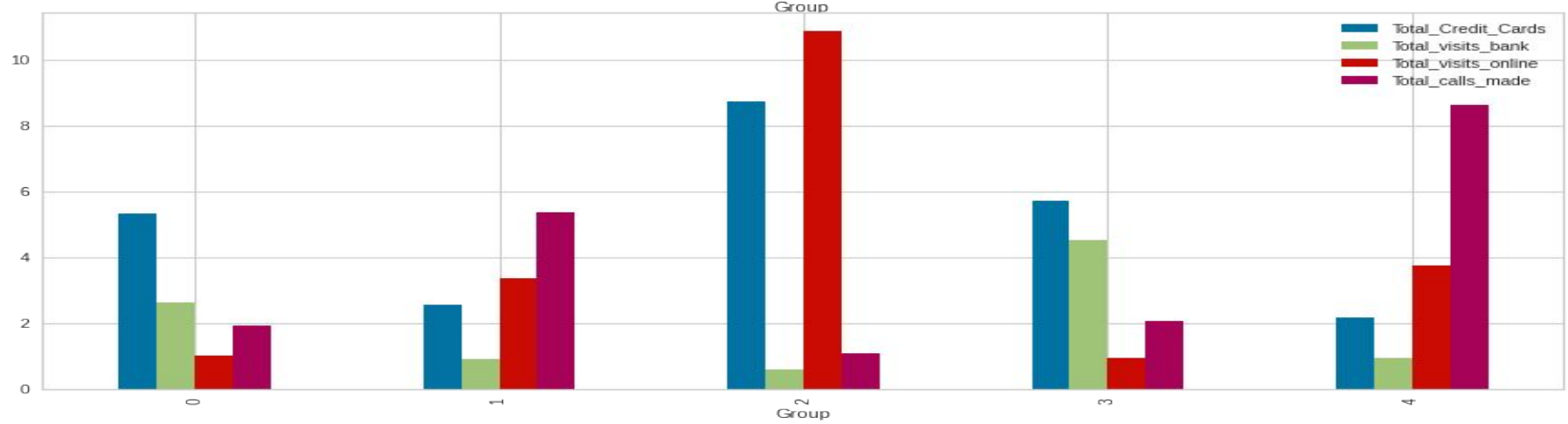




According to the silhouette and the elbow method graph.

Conclusion was drawn that the best amount of clusters would be 5


140000
120000
100000
80000
60000
40000
20000
0
0
1
2
3
4
Group


10
8
6
4
2
0
Total_Credit_Cards
Total_visits_bank
Total_visits_online
Total_calls_made
0
1
2
3
4
Group

Graphs Insights

1) Group 1:(190)
- Mean average credit limit of 35K
- 5 credit cards ownership on average
- Low visits online adn calls to the bank,with more than 2 visits to the bank

2) Group 2:(101)
- Less than 12K on credit limit
- 2 credit cards ownership
- Below 2 visits to the bank, with the second highest calls 5 times and over 3 visits online

3) Group 3:(50)
- Limit of 140K on average
- Over 8 credit cards ownership
- The lowest visits to the bank,highest online visits and below 1 call on average

Graphs Insights

1) Group 4:(196)
- 31K of credit limit on average
- 6 Credit cards ownership on average
- Average call made 2,1 visit on line and 5 visits to the bank(the highest)

2) Group 5:(123)
- 11.9K credit limit on average
- 2 cards ownership on average
- Averages of :1 visit,3 visits online,and 8 calls made,being the heights

Conclusion

Group 2 and 5,have the most phone interaction. This group could be our young adults trying to confirm certain transactions after having online visits,this could be our college students ,trying to avoid owning more than two credit cards. (The best way to market to them would be online, as they use mobile banking for check-ins)

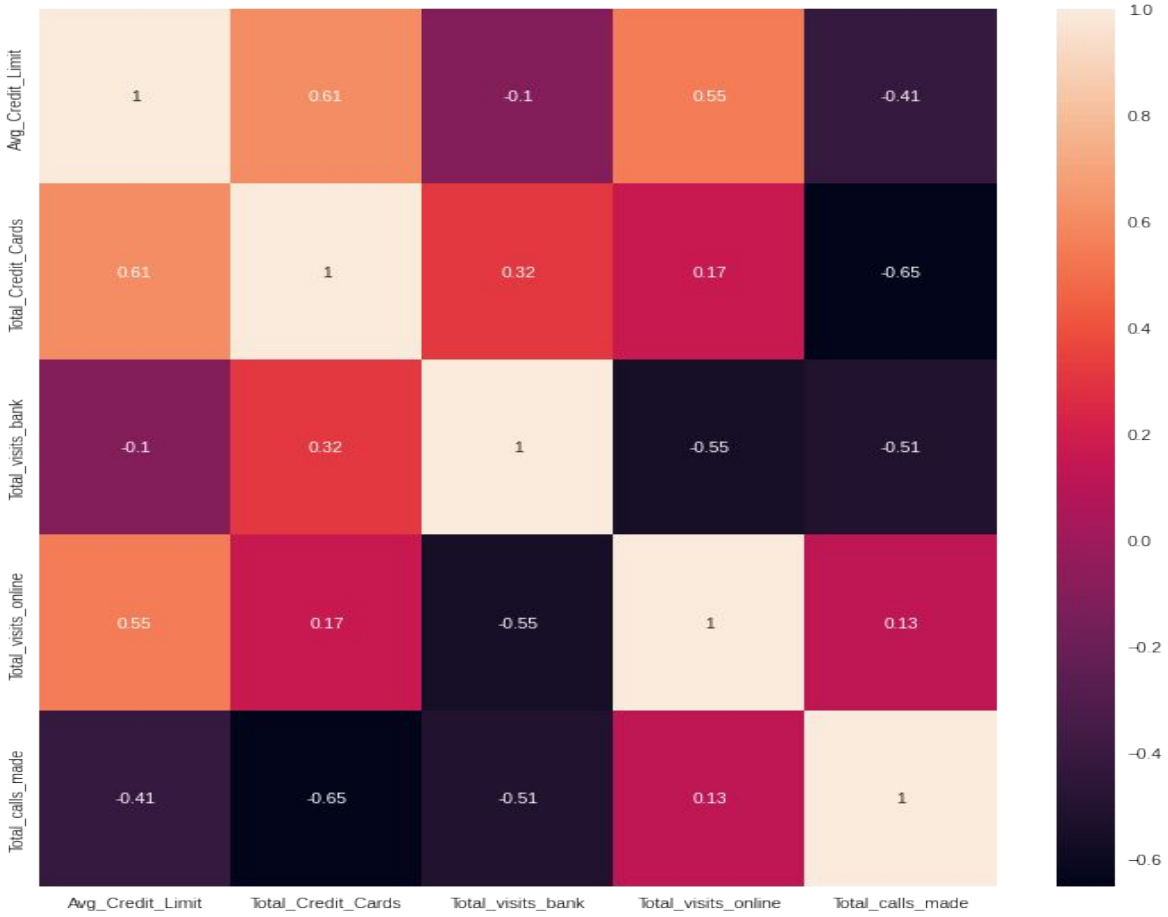
Group 1 could be young adults that might need to build their credit scores, as well travel and experience life. As their credit is low,however they have several credit cards, that could be linked to several packages on the credit cards(Contact the via phone call, try to convince them to start family planning, or to get experiences on the world)

Group 4 could be the elderly,which may explain the large amounts of visits to the bank to keep track of their transactions(contact them only in person)

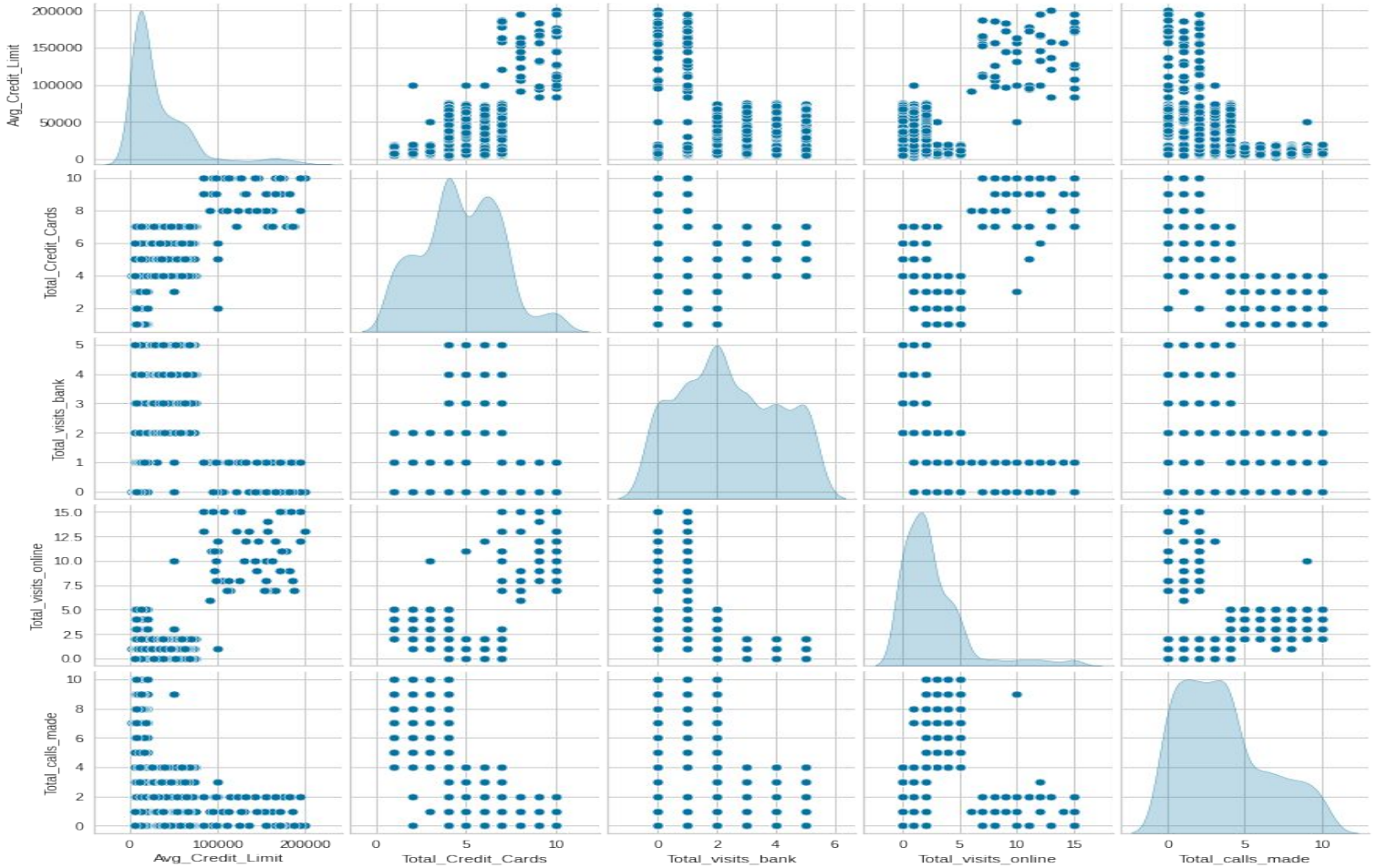
Group 3 could be our families or couples, which explains the consistent visits online,due to the lack of time,and use of different credit cards, some of which could be put to use for their children in case of emergencies (Can only be contacted via Email, and Message as they barely visit the bank)


APPENDIX

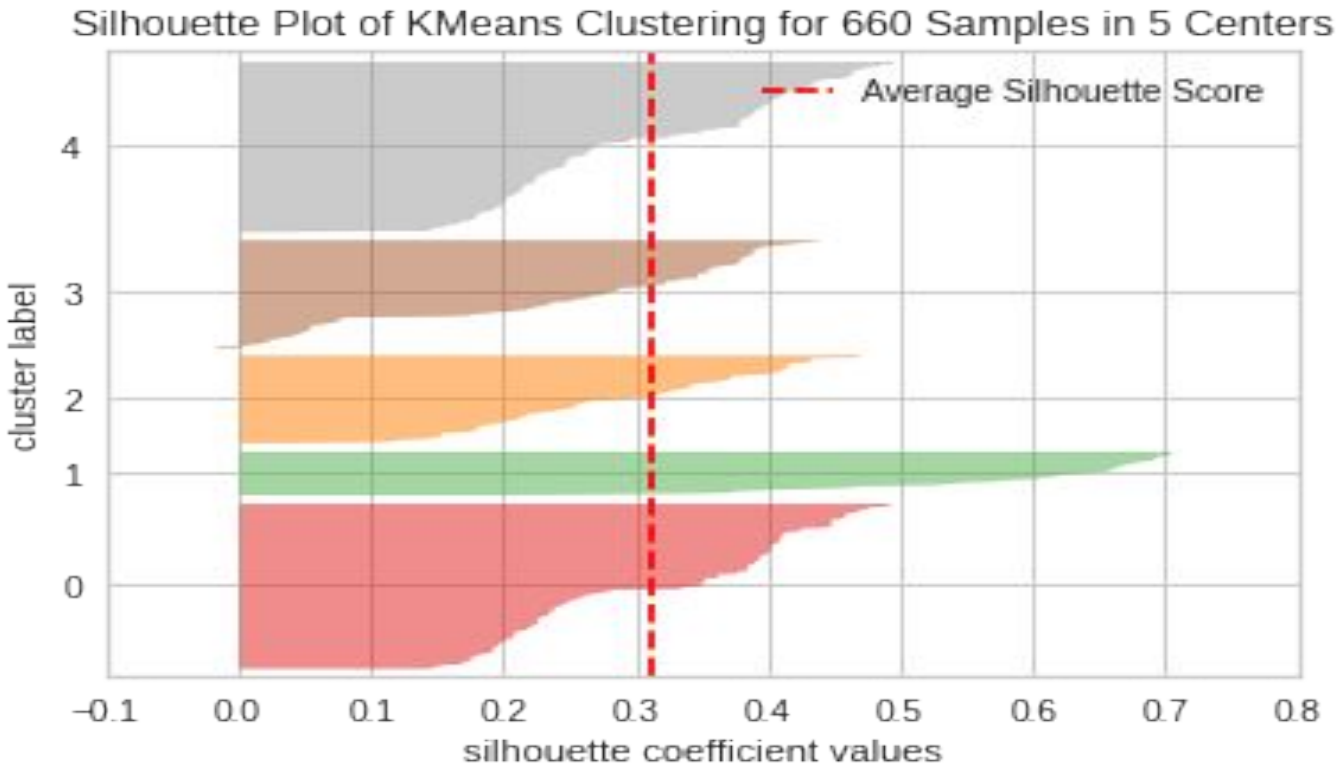
Correlation Heatmap



PairPlots Map



Distribution of the Silhouette



Summary of the Data

Group	Avg_Credit_Limit	Total_Credit_Cards	Total_visits_bank	Total_visits_online	Total_calls_made	K_means_segment	count_in_each_segment
0	35530.516432	5.347418	2.633803	1.023474	1.934272	2.680751	190
1	12083.333333	2.583333	0.908333	3.375000	5.366667	1.500000	101
2	141040.000000	8.740000	0.600000	10.900000	1.080000	2.000000	50
3	31488.505747	5.712644	4.528736	0.936782	2.091954	0.103448	196
4	12330.097087	2.194175	0.951456	3.776699	8.650485	4.000000	123

NoteBook Link

For more information go to the Notebook link,it will have all the required intel needed for the project.

Anything else can be asked to me directly

https://colab.research.google.com/drive/1hmTk3d4s3b3nBZgV7Ja4ugsduFM2WcWD#scrollTo=yMesyB20LzRZ