Analyzing and Predicting Churn in a Multinational Financial Institution

Welcome to my presentation on customer churn analysis. Uncover insights and recommendations to tackle this challenge.

Drafted and presented by Faluta Opeyemi

Introduction to Customer Churn

Introduction to Customer Churn

Definition

Customer churn is the portion of customers who stop using a company's product or service within a specific preiod of time.

Costs

High customer churn rates can lead to significant financial losses in the form of reduced revenue and increased marketing and acquisition costs.

Note

In the course of reading through this analysis, please keep in my the following conventions:

With respect to churn;

“0” represents “customers that were retained ”

“1” represents “customers that churned”

In any other context;

“0” represents “No”

“1” represents “Yes”

Causes

Customer churn can be caused by a variety of factors, including poor customer service, poor customer satisfaction, ineffective marketing, and competition.

Solutions

Reducing customer churn requires a comprehensive approach, including but not limited to improving customer experience, increasing customer loyalty, and implementing effective retention strategies.

Crafting the Methodology

Crafting the Methodology

Data Source

The data was downloaded from Kaggle, a prominent data science community with numerous tools and datasets to choose from. According to Kaggle, the data set has a usability rating of 10.

Data Cleaning & Preprocessing

The dataset was well managed, documented and cleaned prior download, hence, little or no additional data cleansing was required.

Analytical Software and Libraries

MS Excel was initially utilized to eyeball the dataset before it was imported into Jupyter Notebook (Python) for further analysis. Libraries like Pandas, Numpy, Mathplotlib, Seaborn, Scikit-learn and Tkinter GUI toolkit were used.

Statistical Analysis

Key components of statistical techniques used in this project were descriptive statistics, correlation analysis, exploratory analysis and classification models for prediction.

Data Visualization

Transformed data into compelling visual narratives using charts. This approach provided valuable insights that served as the foundation for data-driven conclusions.

Limitations

Despite the usability score of the dataset, certain limitations such as the location of the bank, the currency used, the credit score of the country/ region and the documentation of the meaning of points earned which might be misleading.

Exploratory Data Analysis

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is the process of analysing a dataset to uncover insights, patterns and gain an understanding of the data's relationships. It involves a variety of statistical and visualization techniques to explore the data, understand the statistical data distribution, detect outliers and trends.
The following is the summary of the EDA activities that was performed on the data:

1 Correlation Analysis

This gave an idea of relationships or patterns to expect in the data. It served as the basis on which the data visualization, trends and data-based deductions were made. It is important to note that correlation does not always mean causation, and this was factored into the analysis as well.

2 Data Visualization

Python and several data analysis and visualization methods were used to explore and analyze the Bank's customer churn data. The visualizations were categorized into 3 different sections.

- Univariate visualization: visualizing a single variable
- Bivariate visualization: visualizing two variables
- Multivariate visualization: visualizing more than 3 variables

3 Exploring Relationships

The analysis focused on several key variables, including customer credit score, number of bank product, age group, service satisfaction score, points/rewards earned, active membership, geography and gender. All the aforementioned were explored to visualize their effect on customer churn.

View Appendix for more insight

EDA Insight Summary

EDA Insight Summary

The following are notable insights gotten from the exploratory analysis. These insights cuts across various variables that are important to the bank. In addition, all insights uncovered are solely data driven and as such results might change if additional data is given.
The following insights were explored in relation to churn;

1. **Churn Rate**

   There's a **20% churn rate** with middle aged adults (41-60) accounting for the majority of the Bank's churn.

2. **Customer Age Distribution**

   Fairly young customer base with young adults and adults (18-40) making up 64% of the bank's total customer population.

3. **Tenure Analysis**

   Based on the data, customers with **a year tenure** have the highest churn rate of 23%

4. **Number of Product Analysis**

   Customers who own 2 bank products churn the lowest rate (8%) with higher number of products resulting to an 80-100% churn rate. Note that all customers who owned 4 bank products all churned.

5. **Customer Geography Analysis**

   Germany has a 32% churn rate which is the highest compared to France (16%) and Spain (17%).

6. **Account Balance and Number of products**

   Analysis showed that the average account balance of a customer reduces as the number of product increases. This effect is more pronounced on the retained customer than their churned counterpart.

View Appendix for more insight

EDA Insight Summary

The following are notable insights gotten from the exploratory analysis. These insights cuts across various variables that are important to the bank. In addition, all recommendations made are solely data based and also in line with the financial sector's best practices.
The following insights were explored in relation to churn;

7 **Active Membership Analysis**

Based on the analysed data, an active customer of the Bank is less likely to churn.

8 **Credit Score Analysis**

Retained customers have a relatively higher credit score than their churned counterpart.

9 **Credit Card Ownership, Credit Score and Points Earned Analysis**

Based on the benchmarked filtering criteria, 1,176 customers qualify to own a credit card. The Bank should take a proactive approach to advise these customers as they are potentially at risk to churn.

In addition, further insight should be carried out on 337 churned customers who qualified to own a credit card. This will assist the Bank to identify factors that determines churn.

11 **Satisfaction Score**

The Bank needs to improve its satisfaction score across board as the average satisfaction score shows that customers are neutral (3).

View Appendix for more insight

Recommendations for the Bank

Recommendations for the Bank

1

Customer Onboarding

A recommended course of action is for the Bank to strengthen its onboarding programmes for new clients and measure the effectiveness on customer retention.

2

Offer Personalized Products

In addition to having a great customer onboarding, the Bank should review or strengthen their personalized products, incentivize bundle packages and services that better meet the needs of different customer segments.

3

Informative Newsletters

Based on the analysed data, it is important for the Bank to disseminate information to customers in an educative format through newsletters. Insights can be shared on spending habits, optimized credit card usage and ways to boost credit scores and the dangers of utilizing a high number of Banking products without fully understanding the product.

4

Know Your Customer (KYC)

The bank can improve their customer profiling and segmentation prediction which would boost the performance/ success of the recommended personalized product for the targeted customer. KYC is also useful for customer relationship management and credit risk assessment

Recommendations for the Bank

5

Proactive Credit Card Offerings

Based on the analysed data, there are customers who qualify to own a credit card based on their credit score and points earned. The Bank needs to be proactive by recommending their competitive credit card offerings to the qualified customers. The success of this recommendation should prove effective if the advantages of customer profiling is leveraged appropriately

6

Demographics Matter

It is important to take into identify the differences amongst various demographics. Whether it is the culture, spending habits or credit utilization among other factors. They all comes into play when considering customer experience, customer satisfaction and product offerings.

7

Continuous Improvement

Continuous improvement and actively acting on customers' feedback cannot be overemphasized. This will prove useful in improving the Bank's customer satisfaction score if implemented correctly.

8

What Matters

Based on the analysed data, the "Number of products" purchased by the customer has the most significant effect on customer churn. Other important features were credit card ownership and account balance. It is important for the Bank to pay attention to these features.

Predicting Churn with Machine Learning

Predicting Churn with Machine Learning

1

Data Preprocessing

Irrelevant columns and highly correlated features were dropped as to avoid overfitting the dataset. Categorical features were also encoded. Upon analysis the imbalanced target variable was handled using SMOTE which ensures between the target variable class. The train and test data were also scaled to ensure they are on the similar scale.

2

Model Building

Various classification models were trained and used in predicting customer churn. The dataset was splitted into test and train dataset. 25% of the dataset was set aside as the test variable and random state was set to 40 to ensure consistency in each iteration of the test.

3

Model Evaluation

The models were evaluated using Scikit learn metrics module. Evaluation was based on their accuracy score, precision score, recall score and F1 score. Based on the evaluation result, Random Forrest Classifier was selected as the model of choice with an F1 score of 87%.

Model Evaluation Metrics

Accuracy	This measured the total proportion of correctly classified customers
Precision	This measures how well the model could predict true positives (churned customers) out of all the predicted churned customers. **It answers the question:** Out of all the customers the model predicted as churned, how many actually churned?
Recall	This measures how well the model could predict true positives (churned customers) correctly. **It answers the question:** Out of all the customers who actually churned, how many did the model predict correctly?
F1 Score	A higher F1 score indicates that the model achieves a good balance between precision and recall. It's shows a comprehensive understanding of a model's effectiveness

Check appendix for Evaluation result for more insight

Predicting Churn with Machine Learning

4

Saving the Model and Scaler

The model of choice, Random Forrest Classifier and the scaler function used were saved. In production, actual data will be inputted with the expectation of a prediction. In this analysis, the actual data is stored in X_sam and y_sam.

GUI

5

A simple Graphical User Interface was created to allow users input customer details to predict if they will churn or not.

Conclusion

Conclusion

Reducing customer churn is critical for the long-term success of any business, particularly in the banking industry. It is common knowledge that it cost less to retain a customer than to onboard newer ones.

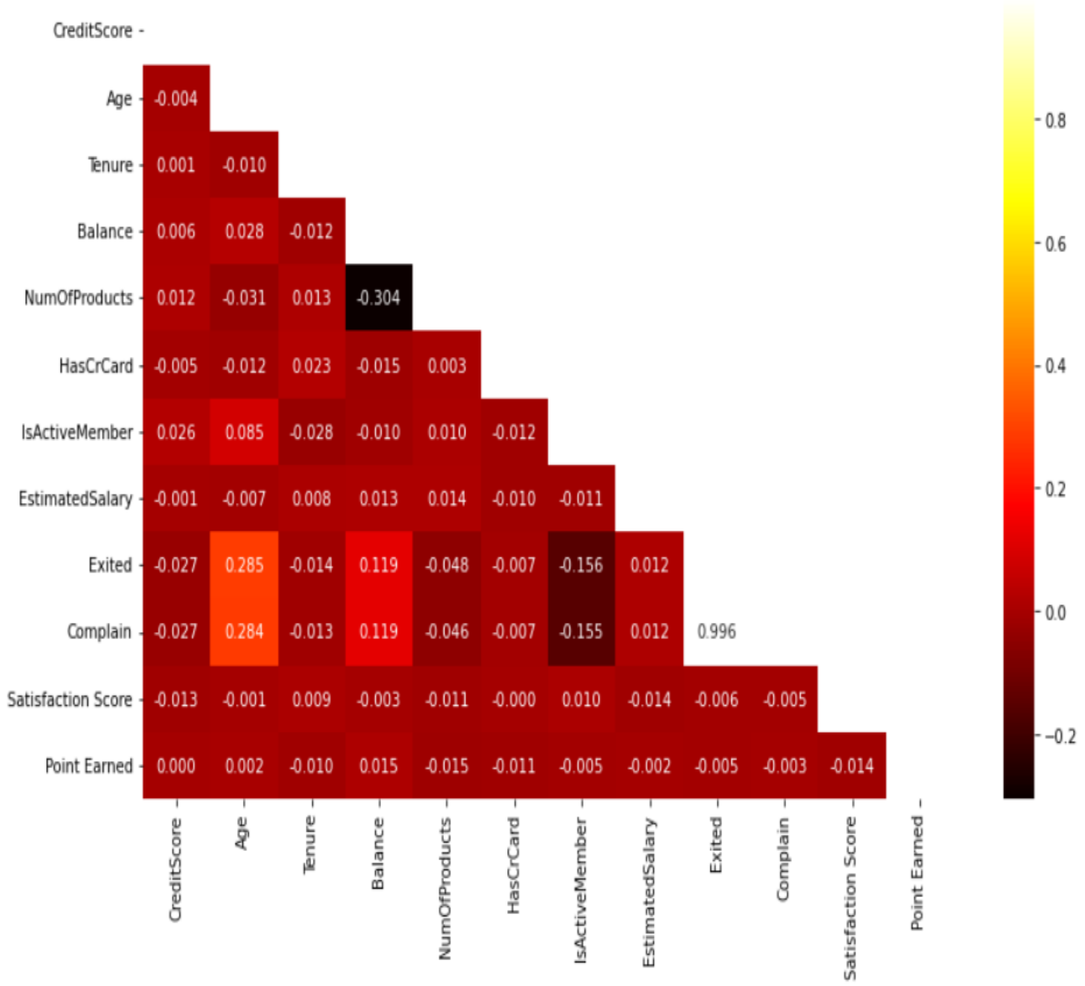
Feedbacks and questions on our findings and recommendations are welcomed. Kindly send a mail via link

- Impact of customer churn and its relevant contributing factors.
- The concept and importance of having a credit score.
- Advanced python visualisation skills. Improved on fully customised legends, ticks and labels amongst others.
- ML models, performance evaluation, ML libraries and GUI via tinker.

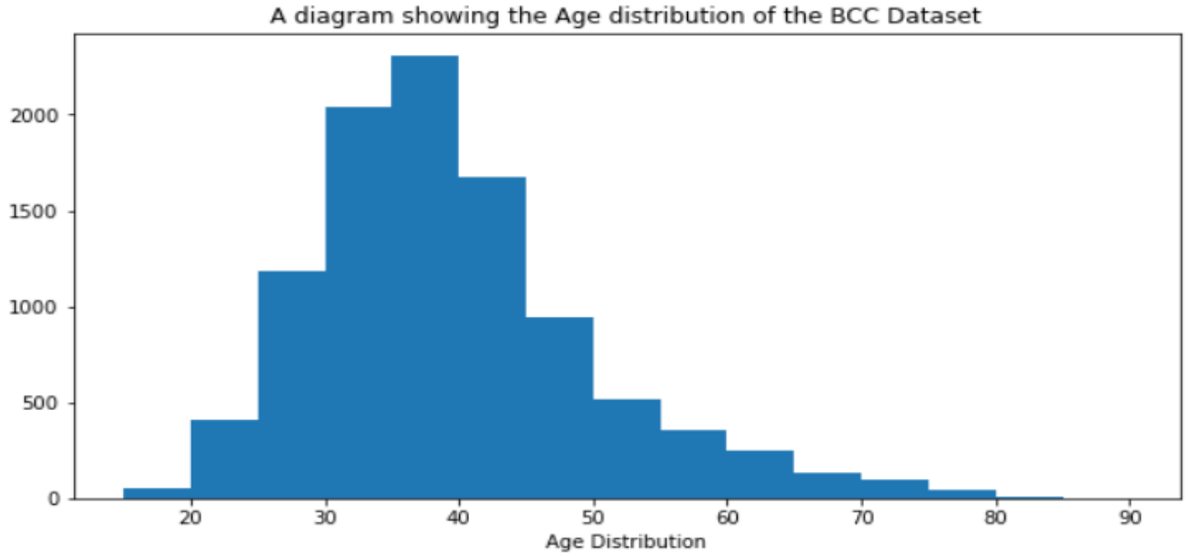
Appendix

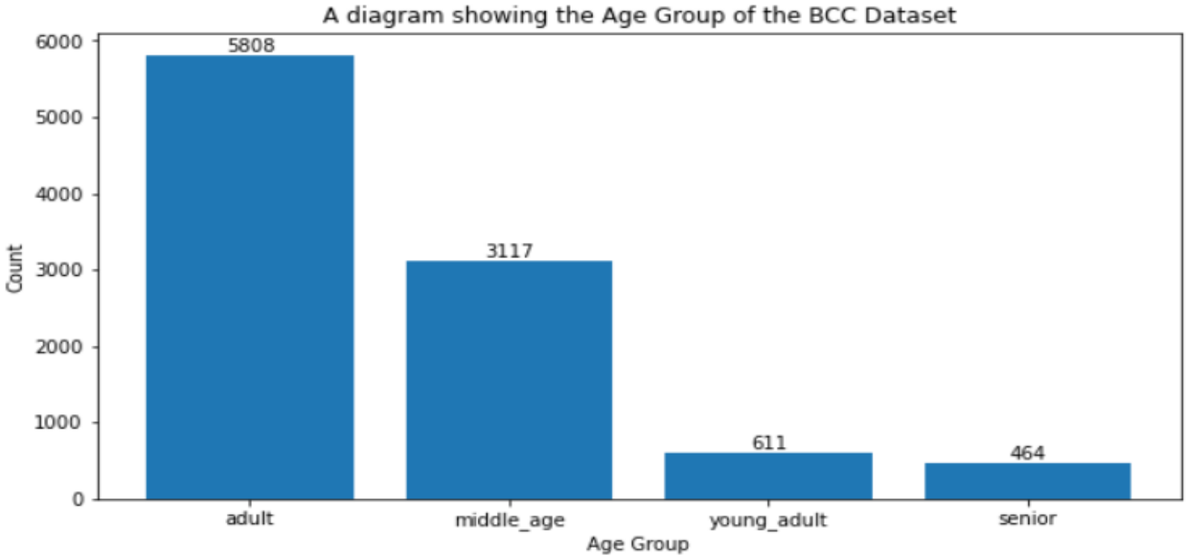
Appendix

Correlation Matrix



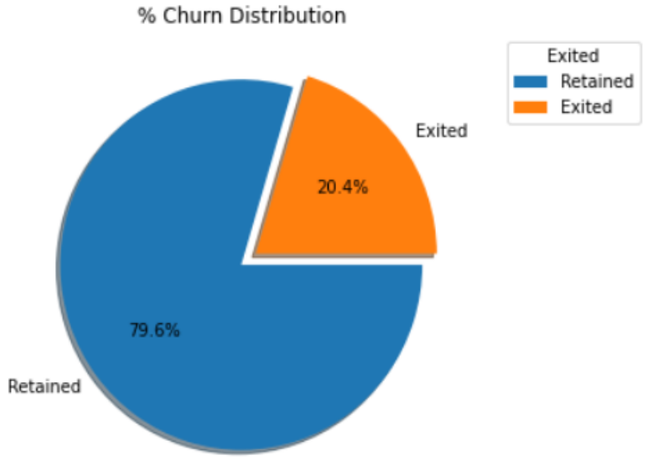
Age Distribution of the Data Set (Histogram and Bar Chart)





Appendix

Bank's Churn Rate Analysis

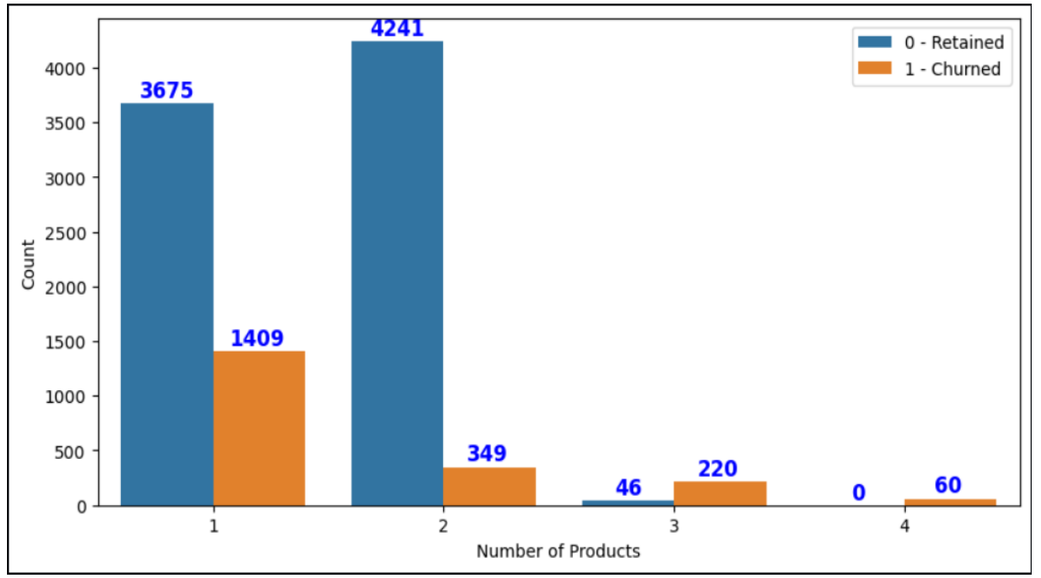


Observation: 20.4% of customers churned.

Churn rate by Tenure

Exited	0	1	% churn
Tenure			
0	318	95	23.0
1	803	232	22.0
2	847	201	19.0
3	796	213	21.0
4	786	203	21.0
5	803	209	21.0
6	771	196	20.0
7	851	177	17.0
8	828	197	19.0
9	770	214	22.0
10	389	101	21.0

Churn rate by Number of Product



Churn Rate by Geography

Exited	0	1	% churn
Geography			
France	4203	811	16.0
Germany	1695	814	32.0
Spain	2064	413	17.0

Churn by Products and Balance

Exited	0	1
NumOfProducts		
1	101053.0	92029.0
2	48721.0	90260.0
3	25744.0	85853.0
4	NaN	93733.0

F1 Score Performance Evaluation

	Models	F1Score
0	LR	0.800499
1	SVM	0.857070
2	DTC	0.832520
3	KNN	0.812059
4	RFC	0.880623
5	GBC	0.860512

Accuracy Performance Evaluation

	Models	ACC
0	LR	0.799045
1	SVM	0.859332
2	DTC	0.827681
3	KNN	0.805828
4	RFC	0.880683
5	GBC	0.860337

Appendix

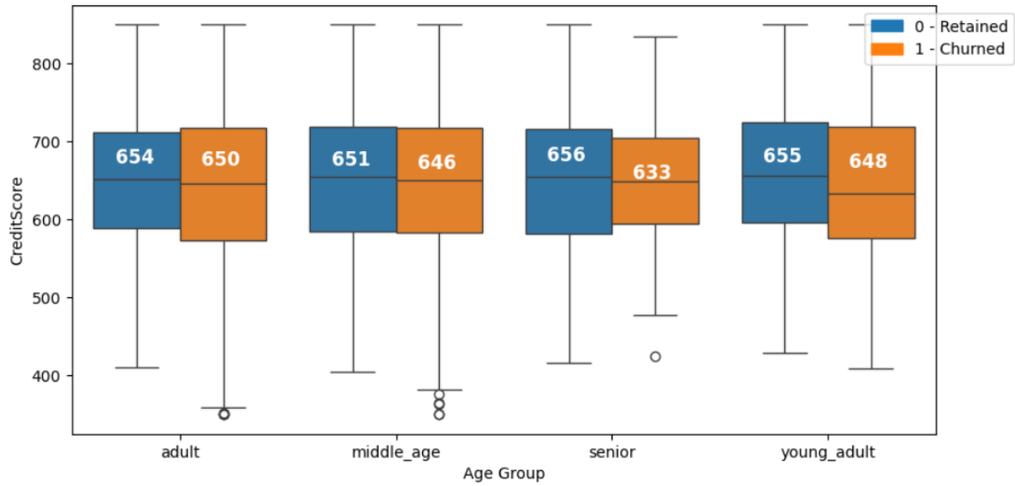
Churn Rate by Active Membership

```
IsActiveMember  Exited
0               0         35.0
                1         45.0
1               0         37.0
                1         45.0
Name: Age, dtype: float64
```

Churn by Credit Card Ownership, Credit Score and Points earned

HasCrCard	Exited	CreditScore	Point Earned
0	0	656.0	606.0
	1	646.0	634.0
1	0	653.0	603.0
	1	646.0	594.0

Churn by Credit Score and Age Group



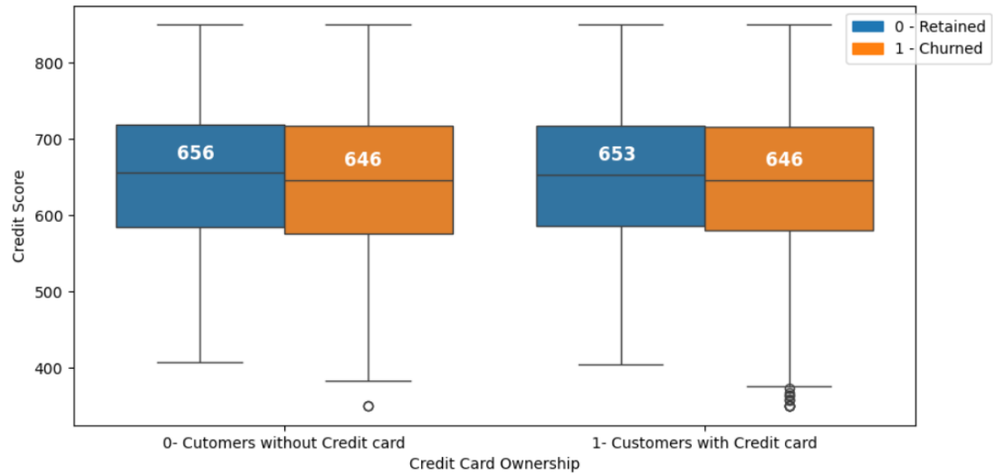
Churn by Products and Balance

Exited	0	1	% churn
NumOfProducts			
1	3675.0	1409.0	28.0
2	4241.0	349.0	8.0
3	46.0	220.0	83.0
4	NaN	60.0	100.0

Extension of the Analysis above including Satisfaction score

HasCrCard	Exited	CreditScore	Point Earned	Satisfaction Score
0	0	652.425815	608.141938	3.002144
	1	647.234910	618.872757	3.060359
1	0	651.594316	606.589343	3.024512
	1	644.631579	598.243509	2.970526

Churn by Credit Score and Credit Card Ownership



Thank You

Lagos, Nigeria

Opeyemi Faluta | LinkedIn

Falutaopeyemi@yahoo.com