**Factor Analysis and Solutions to**

**Retain credit card customers**

**Github Link To Project:**

**https://github.com/sathyareddy25/Project-STU**

**Team-STU**

**(**Umanaveena MekalaVY58857,

Thanmayee Pola-FJ06779

Sathya ThoutiReddy-JB23975**)**

[1.Abstract 2](#_Toc122732716)

[2.Background 2](#_Toc122732717)

[Dataset Attribute Dictionary 2](#_Toc122732718)

[3. Exploratory Data Analysis 4](#_Toc122732719)

[4. Experimental Results 11](#_Toc122732720)

[5.](#_Toc122732721) **[Conclusion](#_Toc122732721)** [12](#_Toc122732721)

[6. References 14](#_Toc122732722)

# 1.Abstract

The use of credit cards is crucial in the banking industry. We may scratch credit cards as buyers with the most advantageous deals and financial stability. We use credit cards to gain airline miles, hotel discounts, grocery shopping savings, and sign-up incentives on different apps. For each purchase, users receive extra prizes and their credit score. Users that use several credit cards and do not utilize just one are despised by banks. Churners are the name given to these users. Closing credit cards means doing so after the bonus has been credited to your account and before the subsequent annual fee is assessed. To gain advance cash, yearly fees, interest charges, and in-hand money, banks, corporate finance departments, and business finance managers give rewards on credit cards. For a credit card company, retaining existing customers is much cost efficient than gaining new customers. We find the reason of the customers leaving the credit card services, identify who might leave and somehow convince them to stay by providing them with better services and turn their decision of leaving. A team of Data Analyst’s at a Credit Card Company is concerned about the high attrition rate of 16.06% of customers. Our aim is to find the reason of the customers leaving the credit card services, identify who might leave and somehow convince them to stay by providing them with better services and turn their decision of leaving. We used various data mining techniques to classify whether customer is willing to stay or leave. To help with decision-making, machine learning techniques can be useful to accurately predict the customers who are leaving. Utilizing supervised learning algorithms such as support vector machine (SVM), decision tree, random forest, and logistic regression.

# 2.Background

The dataset can also be found on Kaggle which is where the csv for the dataset was accessed. The dataset that is available for public use is already cleaned and uses the 10 attributes. Through further investigative research there is no way to tell how the dataset was narrow downed and cleaned. The attributes for the dataset are listed in the dataset attribute dictionary below. Using this data our group wanted to find the best classification technique to predict whether a customer is leaving the company or not through various supervised learning algorithms. Attrition Flag is the target variable.

## Dataset Attribute Dictionary

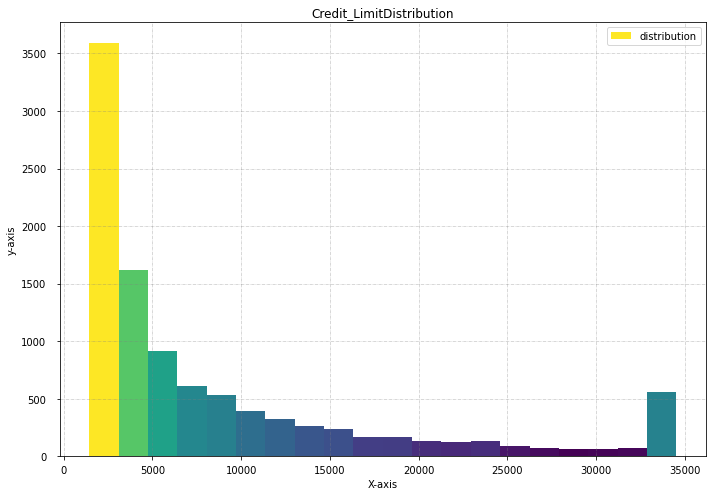
**The dataset represents a credit card company, and it has data factors affecting users such as:**

|  |  |  |
| --- | --- | --- |
| S.No | Variable | Description |
| 1 | Income Category | Annual Income Category of the account holder |
| 2 | Marital Status | Married, Single, Divorced, Unknown |
| 3 | Customer Age | Customer's Age in Years |
| 4 | Gender | M=Male, F=Female |
| 5 | Average Utilization Ratio | Average Card Utilization Ratio |
| 6 | Credit limit | Credit Limit on the Credit Card |
| 7 | Card Category | Type of Card (Blue, Silver, Gold, Platinum) |
| 8 | **Months Inactive** | No. of months inactive in the last 12 months |
| 9 | Dependent Count | No. of the dependents |
| 10 | Attrition Flag | Attrited Customer = 1, Existing customer =0 |

Converting attrited customers, existing customers, male customers, female customers to 0 and 1. Attrited customers - 1, Existing customers - 0 and Male customers - 0, Female customers – 1.

# Exploratory Data Analysis

In order to better understand our dataset, we conducted some preliminary data exploration before diving into running machine learning algorithms. This included using python to perform data visualization to see how various attributes correlated with each other and the output class. Figure 1 is a visualization of credit limit distribution among customers. Additional visualizations were programmed in python to gain more insights into the dataset. Here we are continuous features and creating histograms for all columns.



To gain more insights from figure 1 we did more exploratory data analysis using Python .

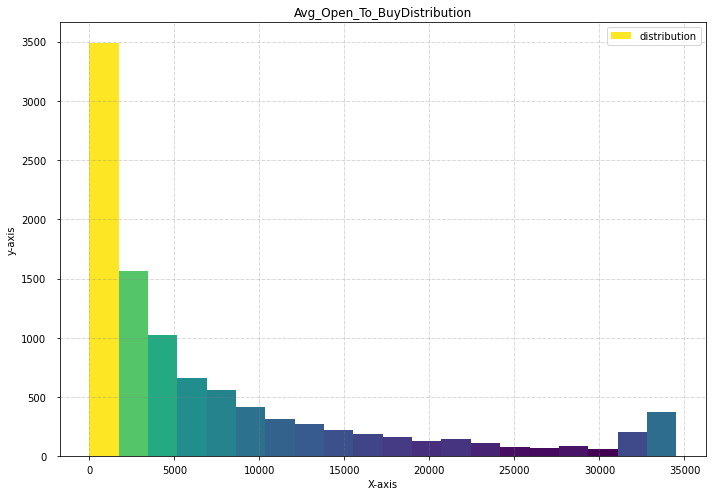


Fig2 ,average open to buy distribution

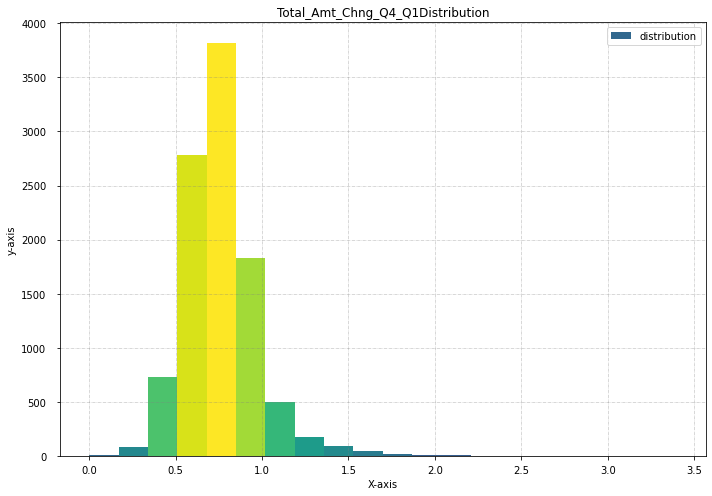


Fig 3,Total amount change distribution

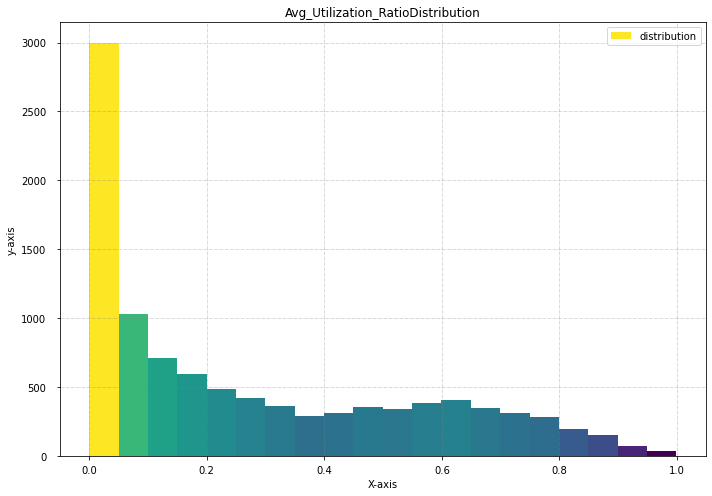


Fig 5.Average utilization ratio distribution.

. *Fig 5* shows that utilization of credit card by customers.. Additionally, This piece of information is important in determining the ground truth for our model.

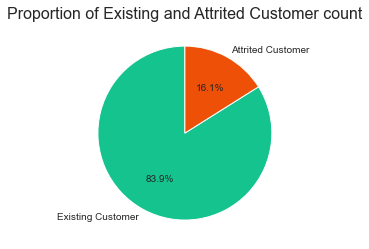
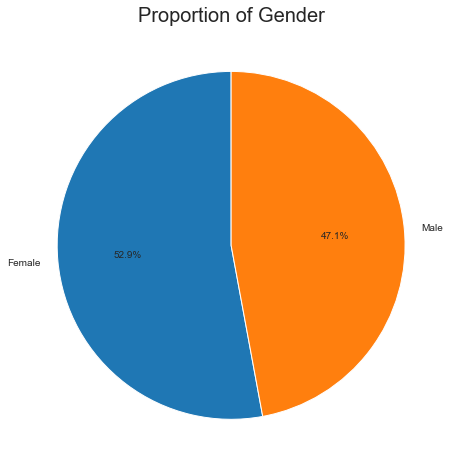


Fig6. Customers proportion based on attirition status.(y variable proportion)



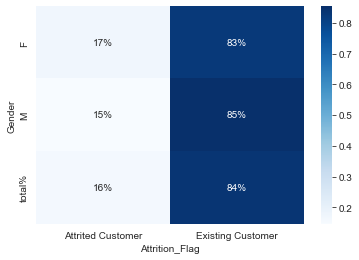
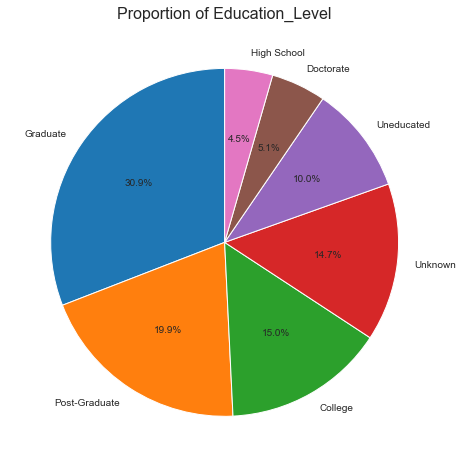


Fig 7&8. Customer distribution based on gender. Attirition proportion between gender to see the probability of churning in each gender category.



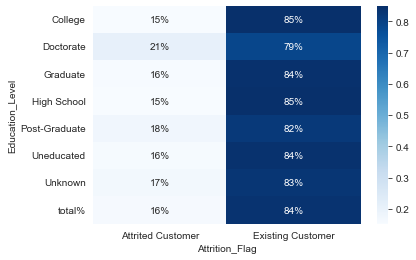
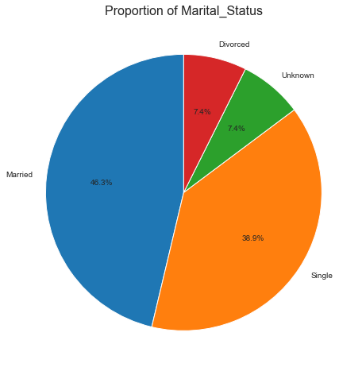


Fig 9&10.Proportion of education level of customers.



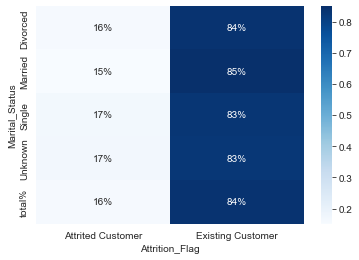
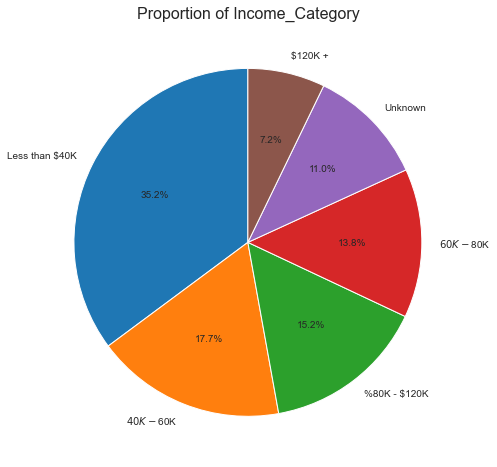


Fig 11&12.Attirition proportion of marital status of customers.



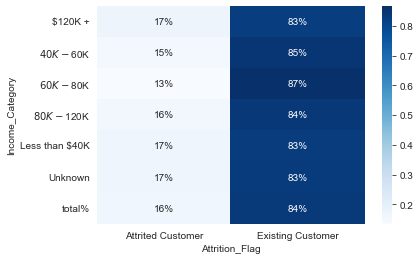


Fig 13.Attirition proportion of customers based on income category.

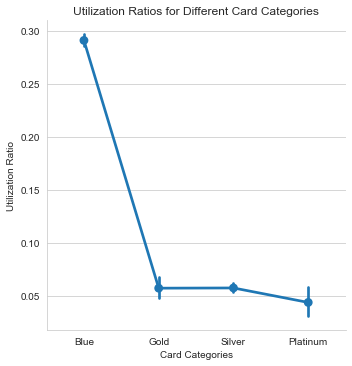


Fig 14.The graphs shows the categories of cards utilised by the customers.so that we can understand which kind of card is most used.

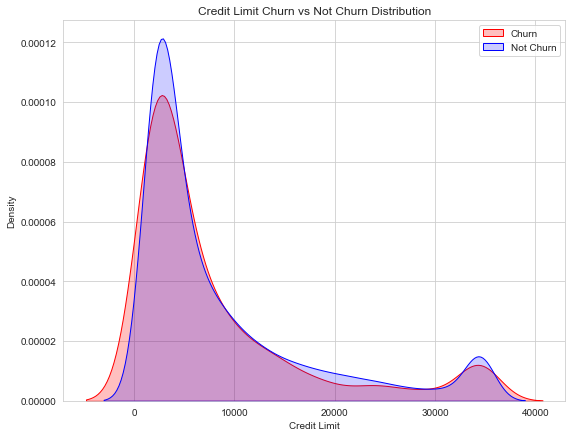


Fig 15.Distribution of credit limit churn and not churn.

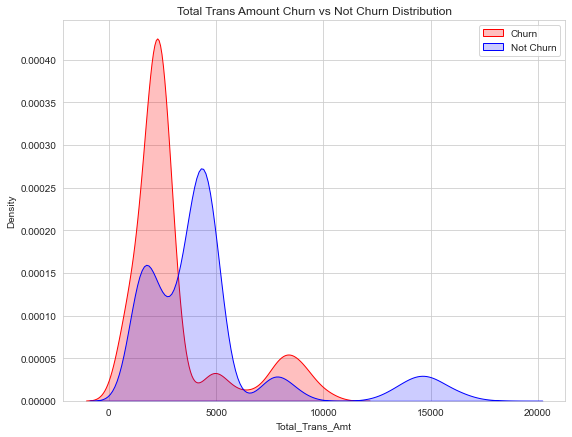


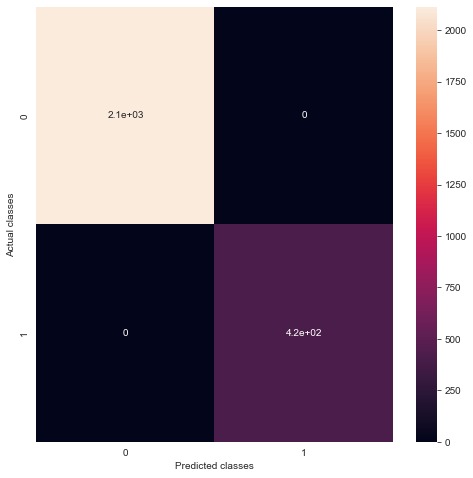
Fig 16.Distribution of transaction total amount helping to understand chur and not churn ratio.

# Experimental Results

Our experimental results proved to be very promising in being able to correctly classify whether or not individuals leaving the credit card company. For each machine learning model, we recorded the percentage of instances correctly classified (accuracy) Our results can be seen in the table below.

|  |  |  |
| --- | --- | --- |
| Algorithm | Logistic Regression | Navie Bayes Classifyer |
| Training  accuracy score | 1.0 | 1.0 |
| Training  precision score | 1.0 | 1.0 |
| Training  recall score | 1.0 | 0.97 |
| Accuracy score | 1.0 | 0.99 |
| Precision score | 1.0 | 1.0 |
| Recall score | 1.0 | 0.96 |

The logistic regression model, which obtained an accuracy of almost 100% and ultimately showed the greatest potential. The performance of the naive Bayes classification model, which attained a 99% accuracy rate and an Since naive bayes classifier is somewhat higher, suggesting that the model may be overfitting, one may contend that logistic regression may provide a superior fit.

Below is a confusion, matrix from our best model. 

# **Conclusion**

 We can infer that the logistic regression is the best-performing algorithm for our data among the other investigated models to research and forecast customer churn behaviour after taking into account various assessment criteria like accuracy and precision. The bank may use this model to anticipate which clients would think about leaving. With this insight, the bank may take proactive measures to keep the client by resolving any problems they may be having that might be motivating them to think about leaving.

**6.Future Work**

Despite identifying whether or not someone is leaving the credit card company is very challenging problem to solve, we are able to relatively accurately predict whether or not someone would leave the credit card company. Now knowing that various companies can now accurately assess whether or not someone will leave the credit card company using machine learning, it is important to consider how results may be further improved. With a larger dataset models could be better fine-tuned and deliver more accurate results.

# References

Predict Churning customers Kaggle. https://www.kaggle.com/datasets/sakshigoyal7/credit-card-customers

# Customer Churn Prediction:A Survey

<https://www.researchgate.net/publication/343787983_Customer_Churn_PredictionA>

# *Altinisik, F.: Predicting Customers Intending to Cancel Credit Card Subscriptions Using Machine Learning Algorithms: A Case Study. IEEE (2020)*