

# Bank Customer card Churn Prediction and Analysis

Padma Parameshwaran

# The Problem

Bank is concerned about more customers leaving their bank.

Customer leaving bank:  
**16.07 %**

Finding who might churn will give the bank to attract them with additional incentives.



**Can we Predict which customer is going to churn next?**  
**What factors affect customer churn?**

# Data Information

Details : Bank customers who own the credit card.

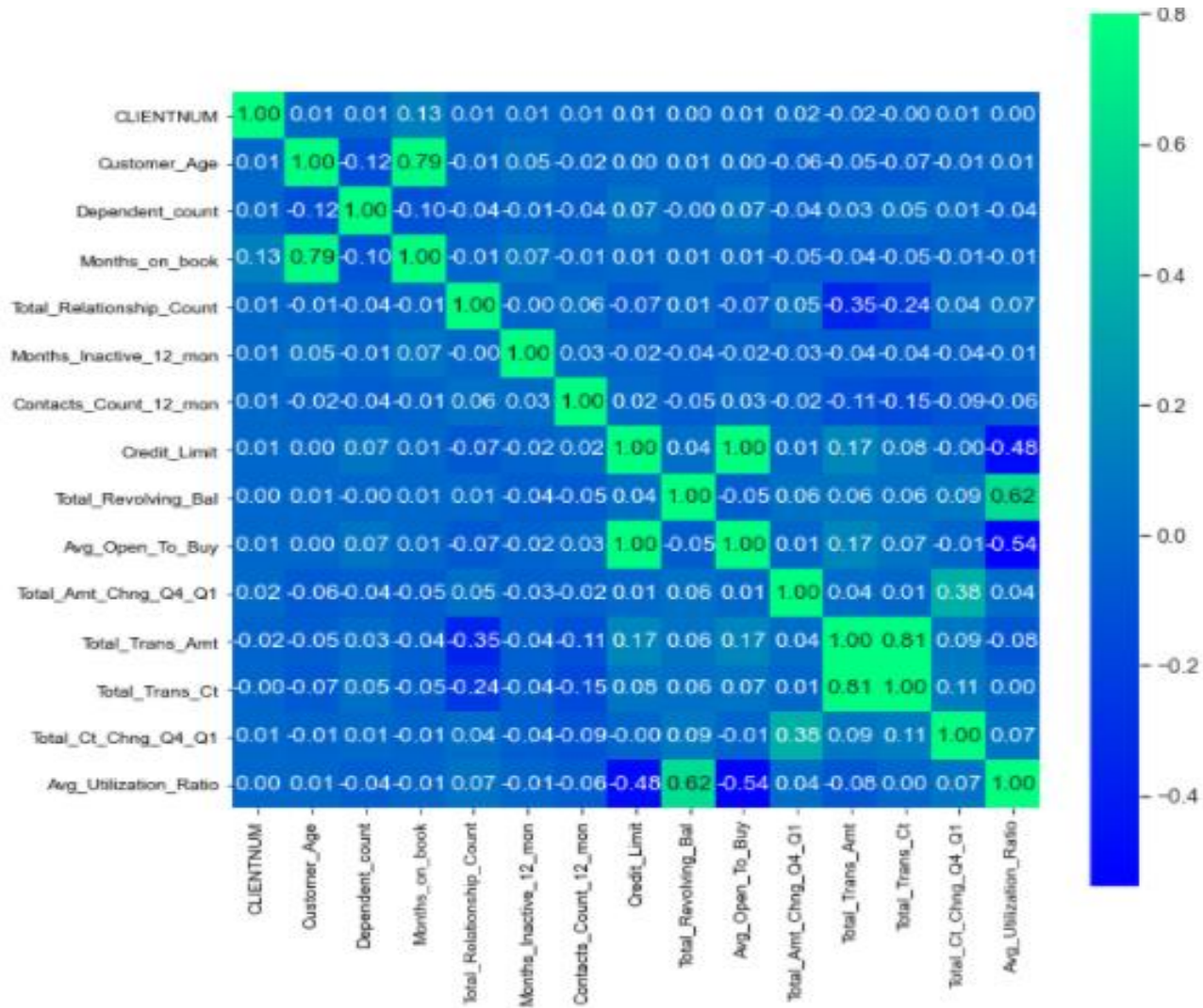
Number of Records: 10127

Number of attributes:23

Sample Attributes : Client number, Attrition Flag, Customer Age, Gender, Dependent count, Education Level, Marital Status, Income Category,...

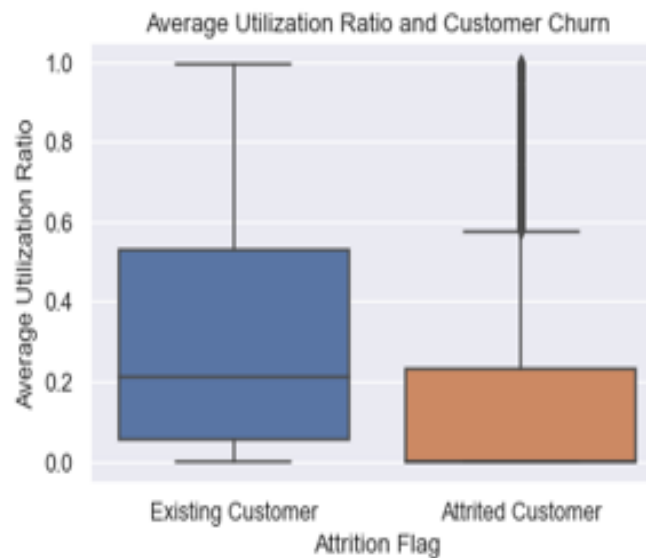
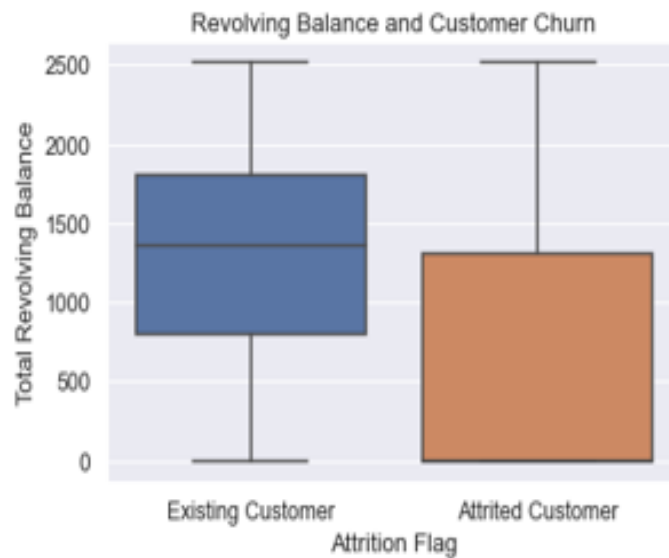
Acquired from: <https://www.kaggle.com/sakshigoyal7/credit-card-customers>

# Correlation of attributes

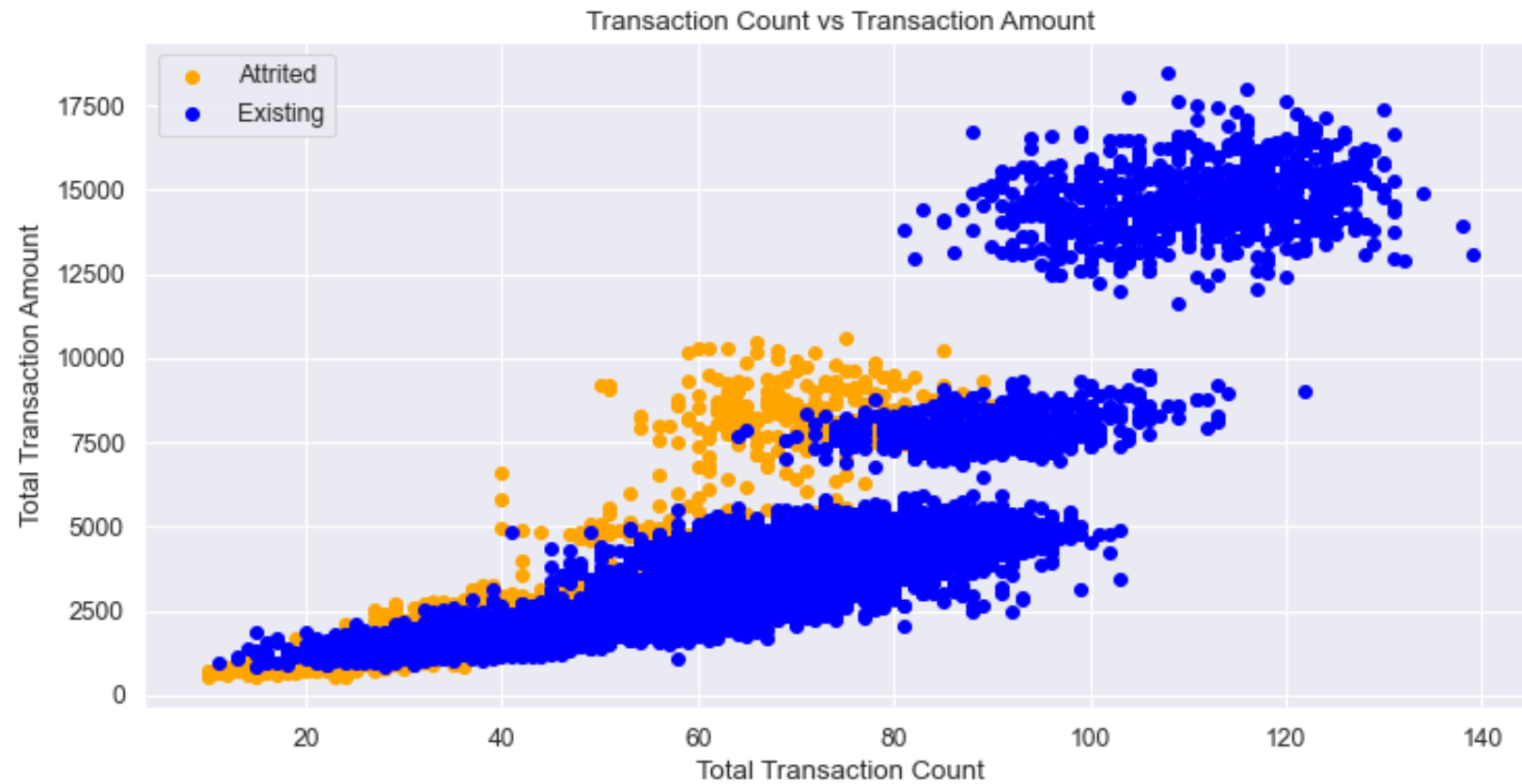


# Key Findings

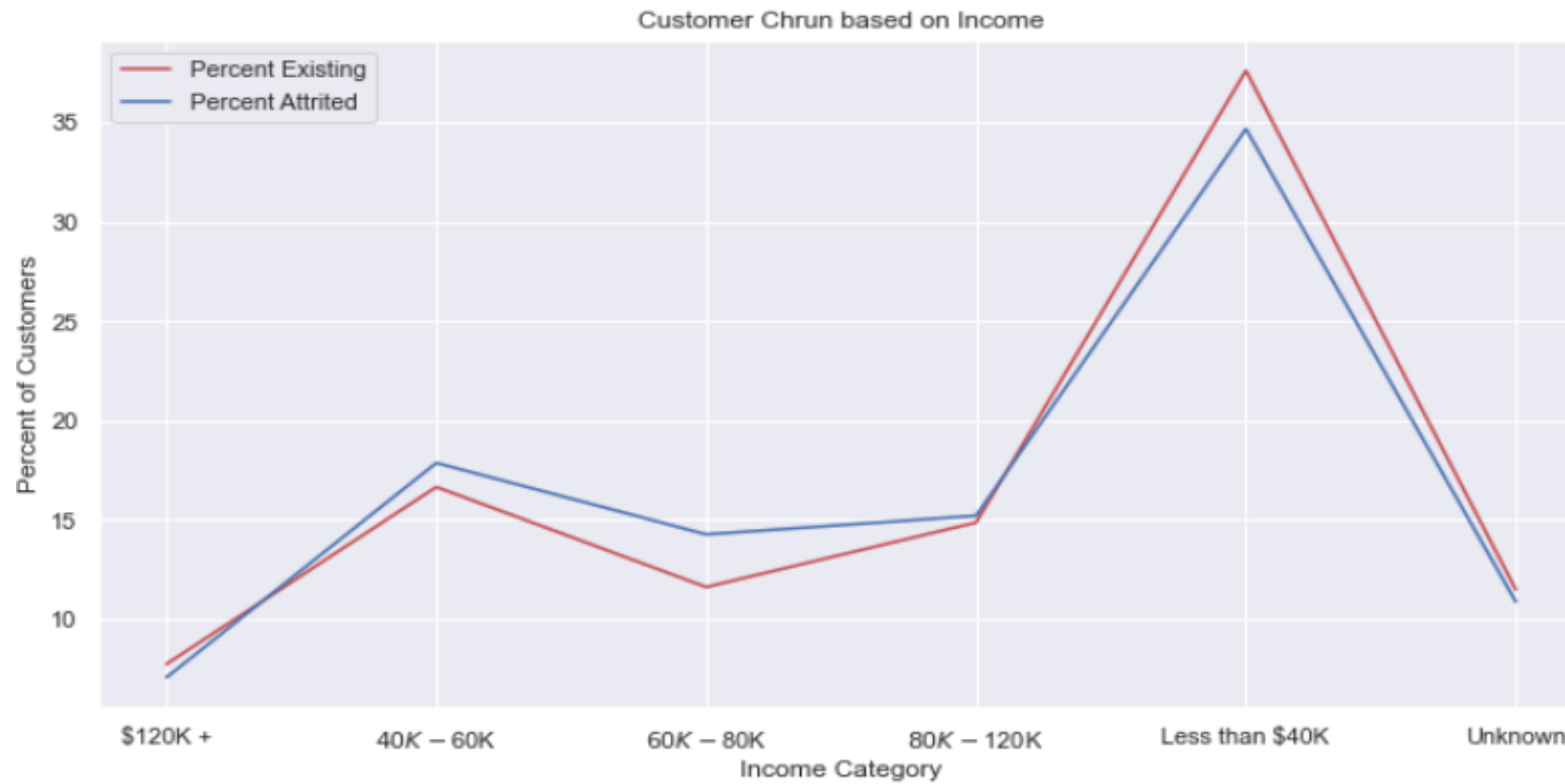
## Revolving Balance and Average Utilization Ratio



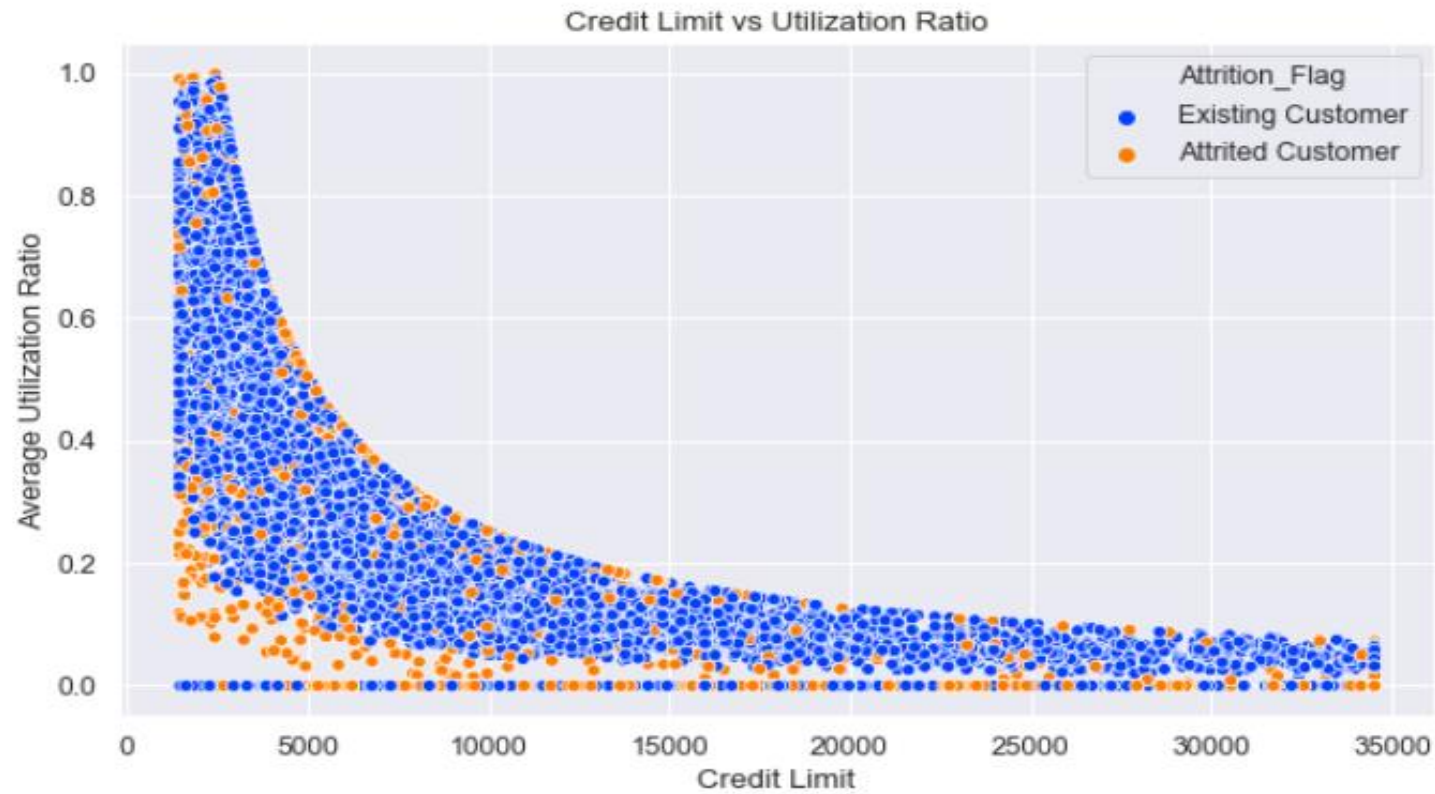
Existing customers have at least \$500 as their minimum balance and their utilization is also exceptionally low.



Customers who left the bank had fewer transaction counts than existing customers.



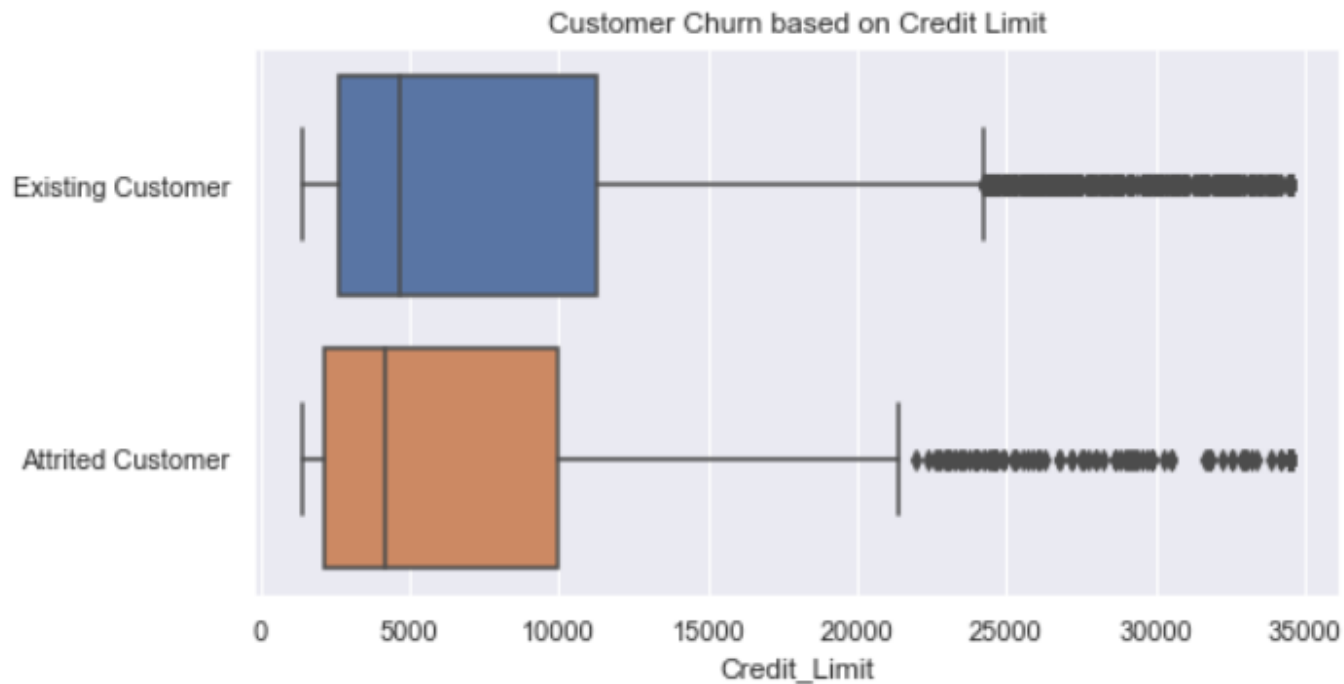
Customers whose income is in the range 40K-80K have higher percentage of staying with bank than other income



Most of the customers who are churned have lower credit limit and utilization ratio



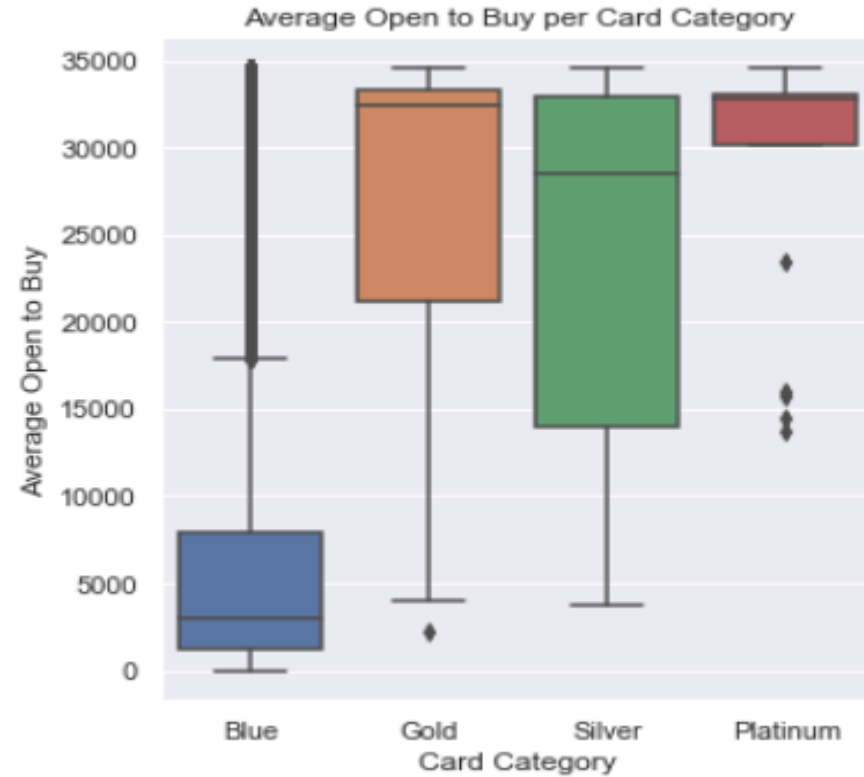
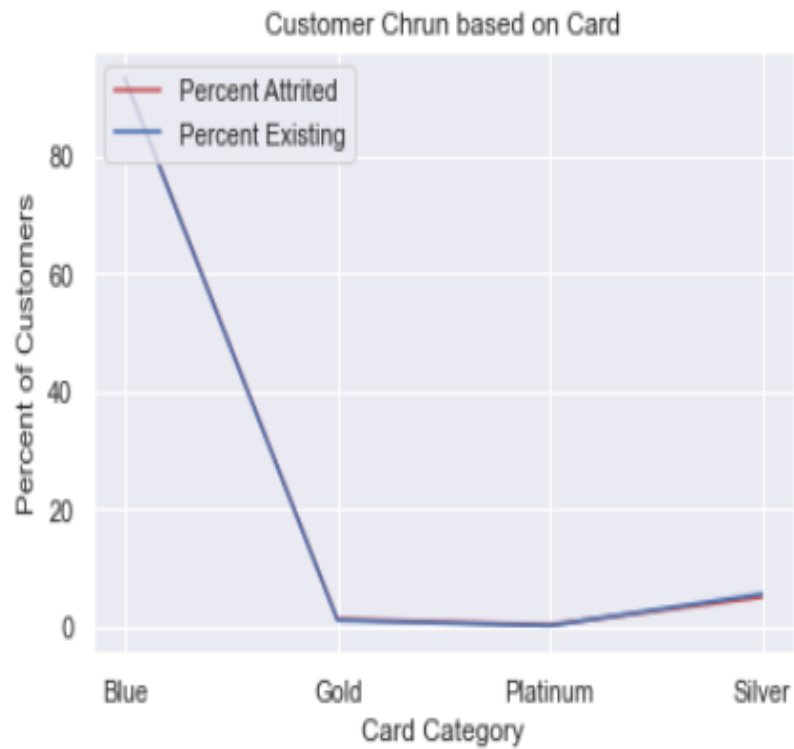
# Effect of Credit Limit on churning



Null Hypothesis: Mean of credit limit of churned customers equal to mean of credit limit of existing customers

P-value = 0.016

# Card Category and Open to Buy

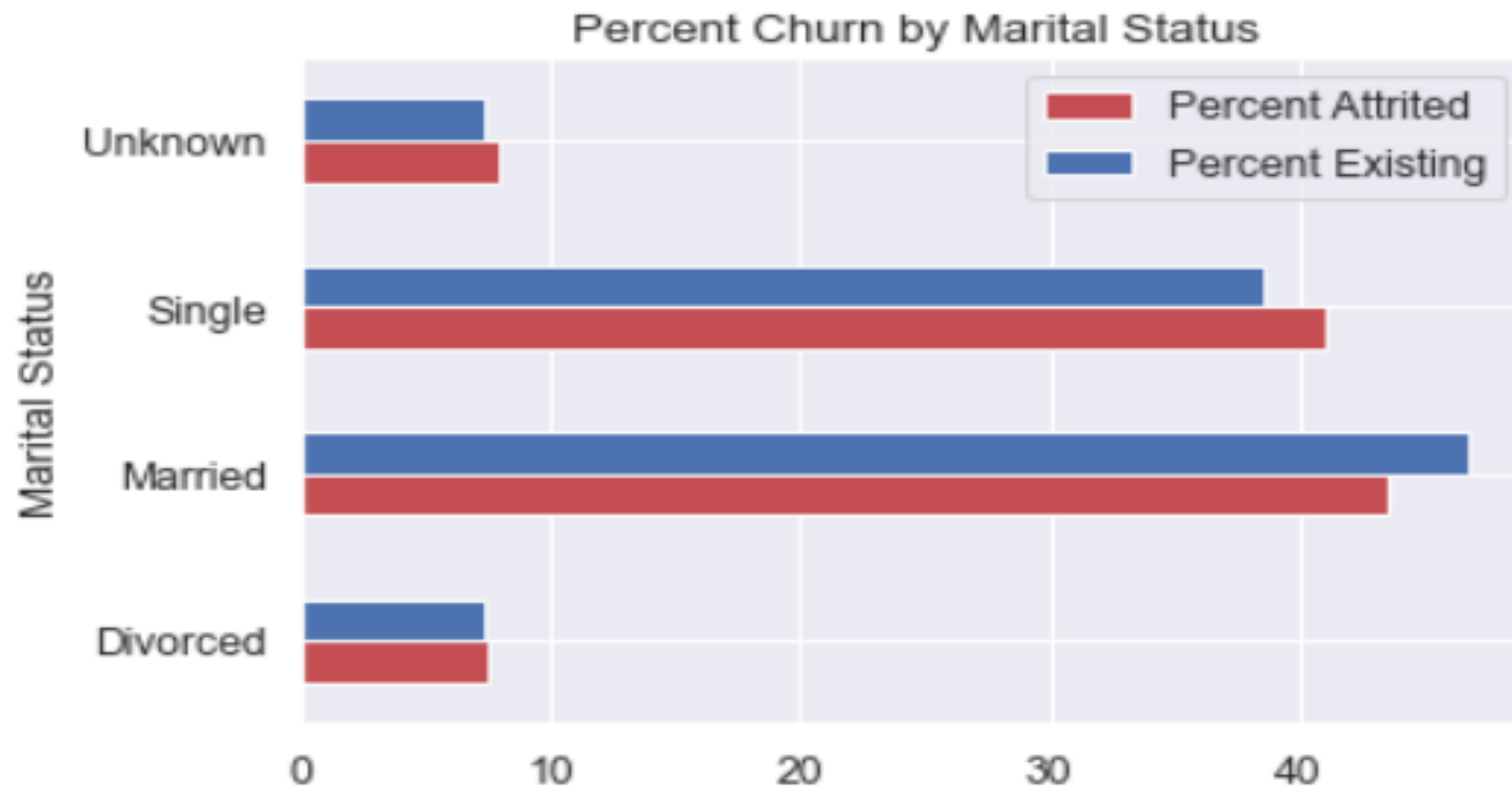


More activity on silver card holders though more customers are blue card holders

# Inactivity and Contacts with the bank

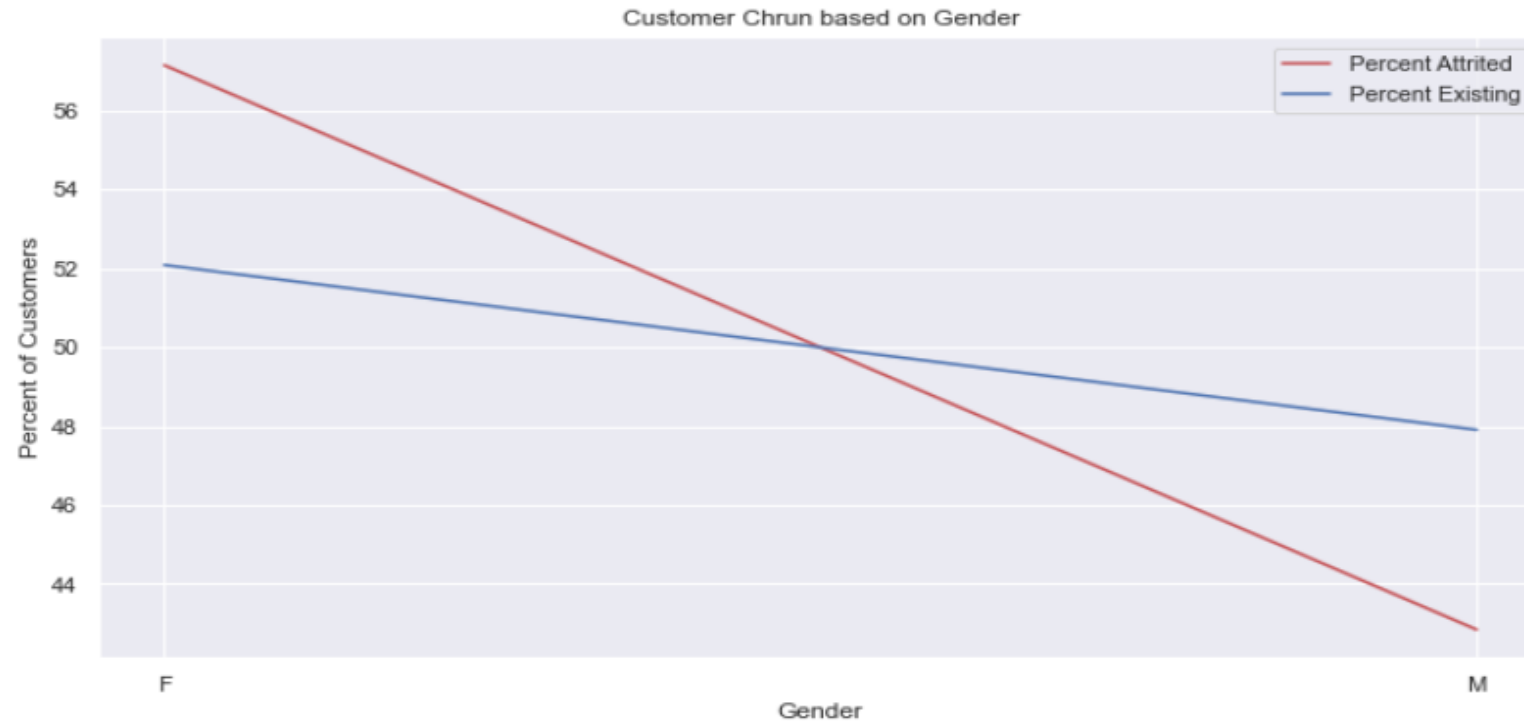


If the number of contacts or Months Inactive exceed 2 the bank must be cautious and look for any churning signal



Large percentage of married customers are loyal to the bank.

# Gender



There is higher percentage of female customers who churn compared to male customers

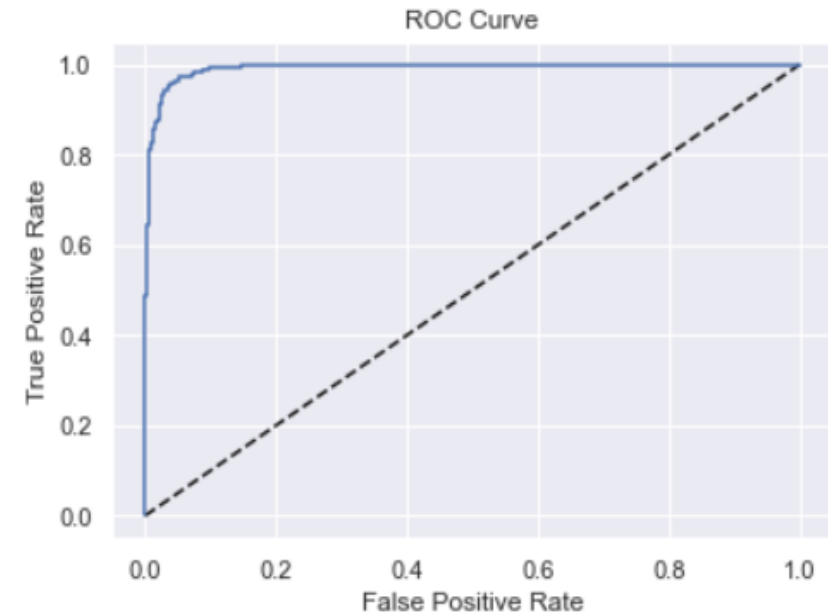
```
F    5358
M    4769
Name: Gender, dtype: int64
```

# Modelling

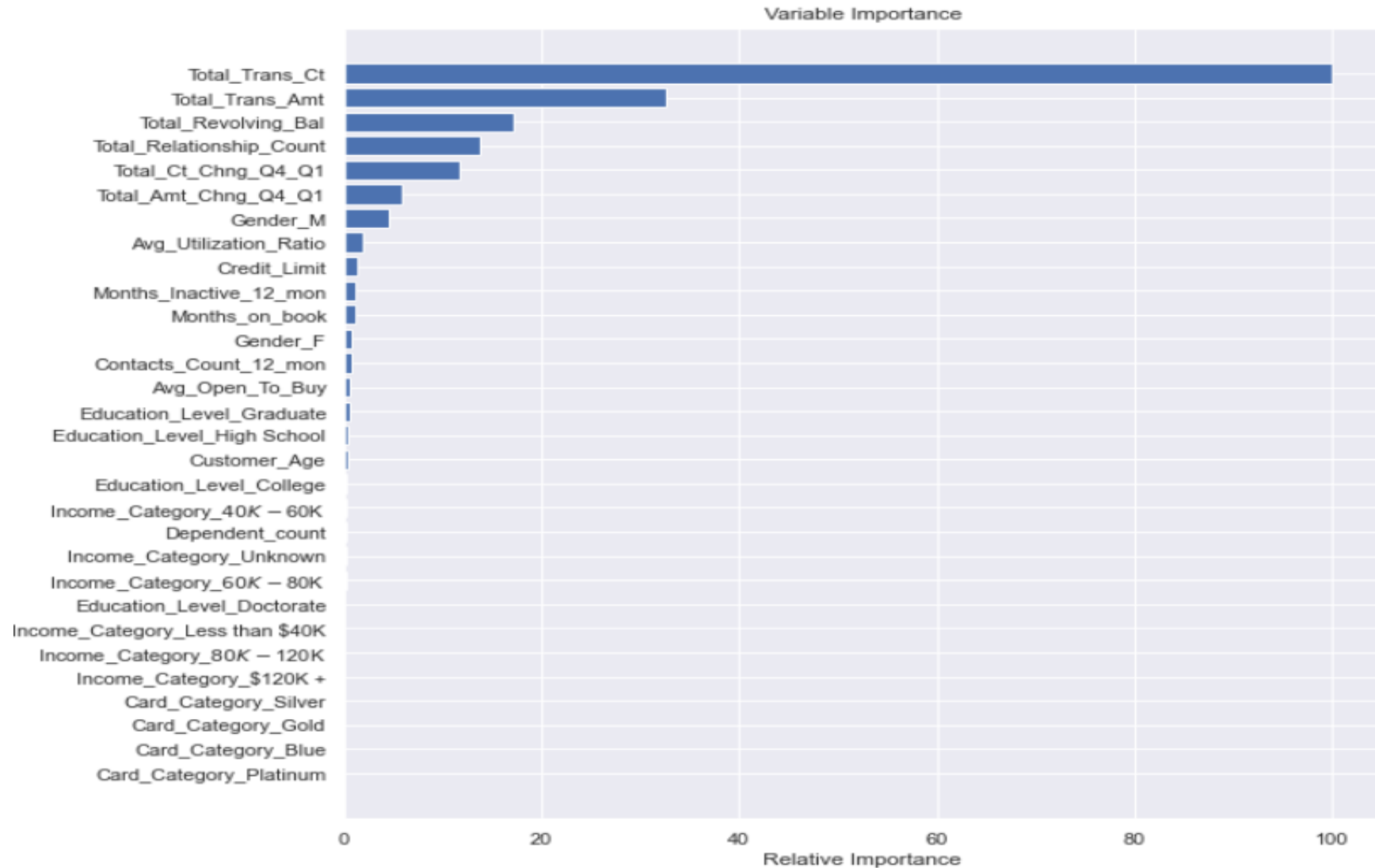
| Model                                | Accuracy | Precision | Recall      | F1-score |
|--------------------------------------|----------|-----------|-------------|----------|
| Logistic Regression                  | 0.89     | 0.81      | 0.48        | 0.60     |
| Random Forest                        | 0.96     | 0.92      | 0.86        | 0.88     |
| SVM                                  | 0.91     | 0.78      | 0.68        | 0.72     |
| KNN                                  | 0.86     | 0.70      | 0.26        | 0.37     |
| Gradient Boosting                    | 0.97     | 0.93      | 0.89        | 0.91     |
| Over Sampling with Gradient Boosting | 0.96     | 0.89      | <b>0.90</b> | 0.89     |

Best Model:  
Gradient Boosting

The area under the ROC curve is **0.99246**



# Feature Importance



## Recommendations

These are the recommendations to attract new customers of these categories and to retain existing customers

- Age of the customer between 45 and 55
- Status Married
- Income \$40K - \$80K
- Gender Male



# Future Improvements

- Some outliers and columns can be removed to check if the performance improves further.
- This can be combined with additional data of other products that the customers use in the same bank and cluster their behavior for further marketing.