Predicting the Likelihood of Customer Churn

Sima Nikzad
Data Science Capstone project

The Problem

Customer Churn(attrition) Rate: rate at which customers, stop doing service with an entity. It is expressed as percentage of service subscriber, who discontinue their subscription within a given period of time.

XXX Bank is trying to identify the contributing factor(s) for customer churn and if this can this be avoided or predicted?
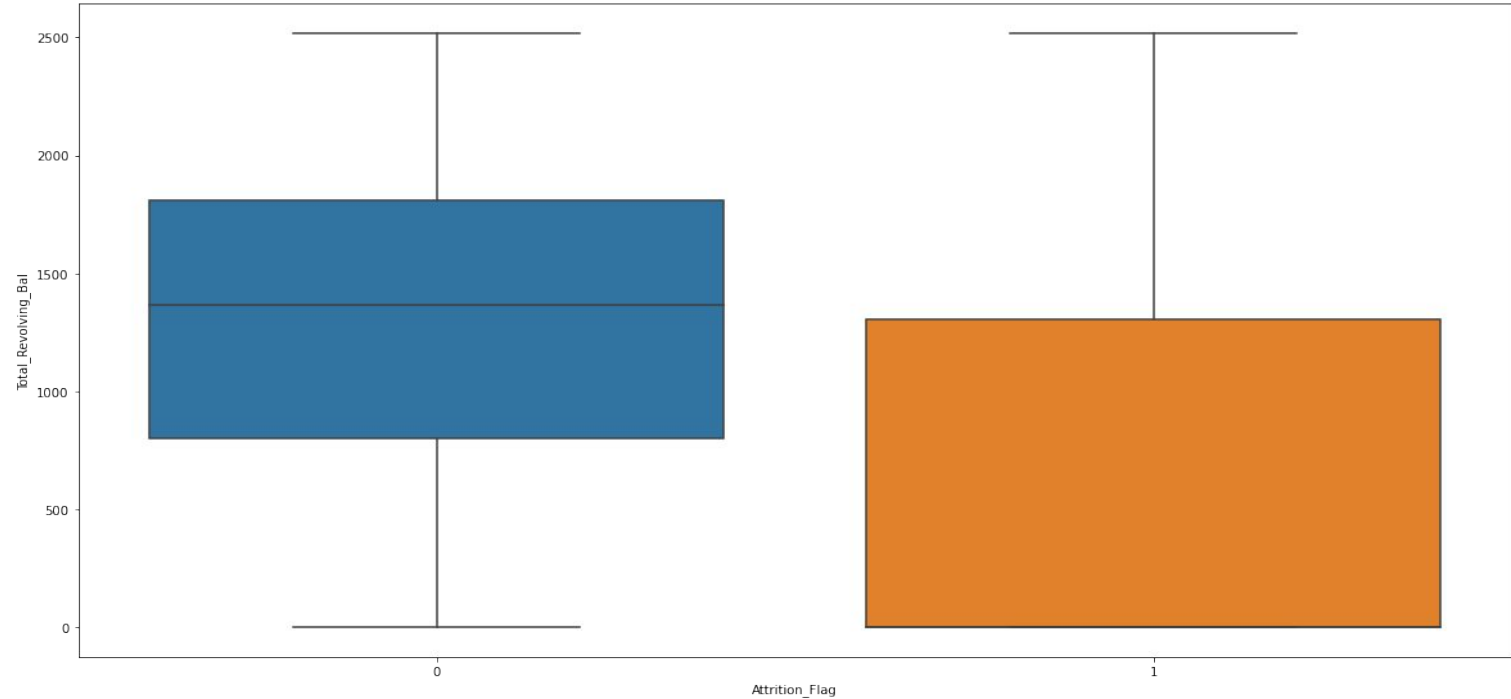
Contributing Factors
Highly Correlated

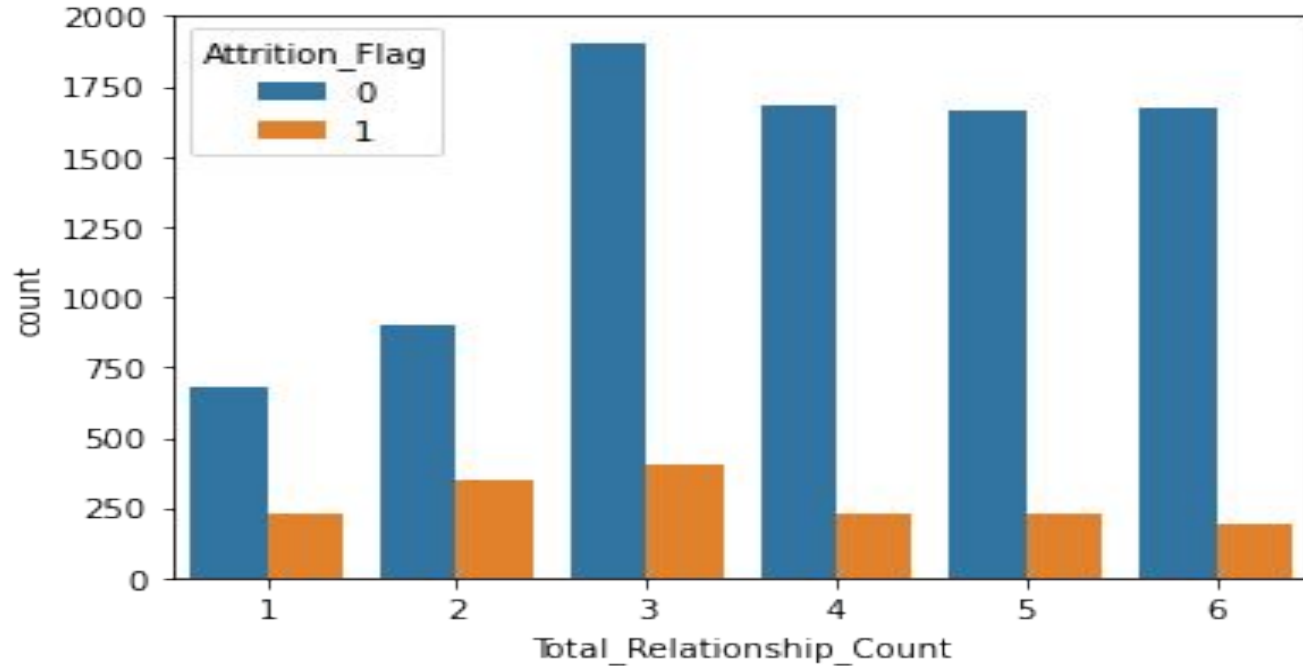## 1.   Total Number of Transaction
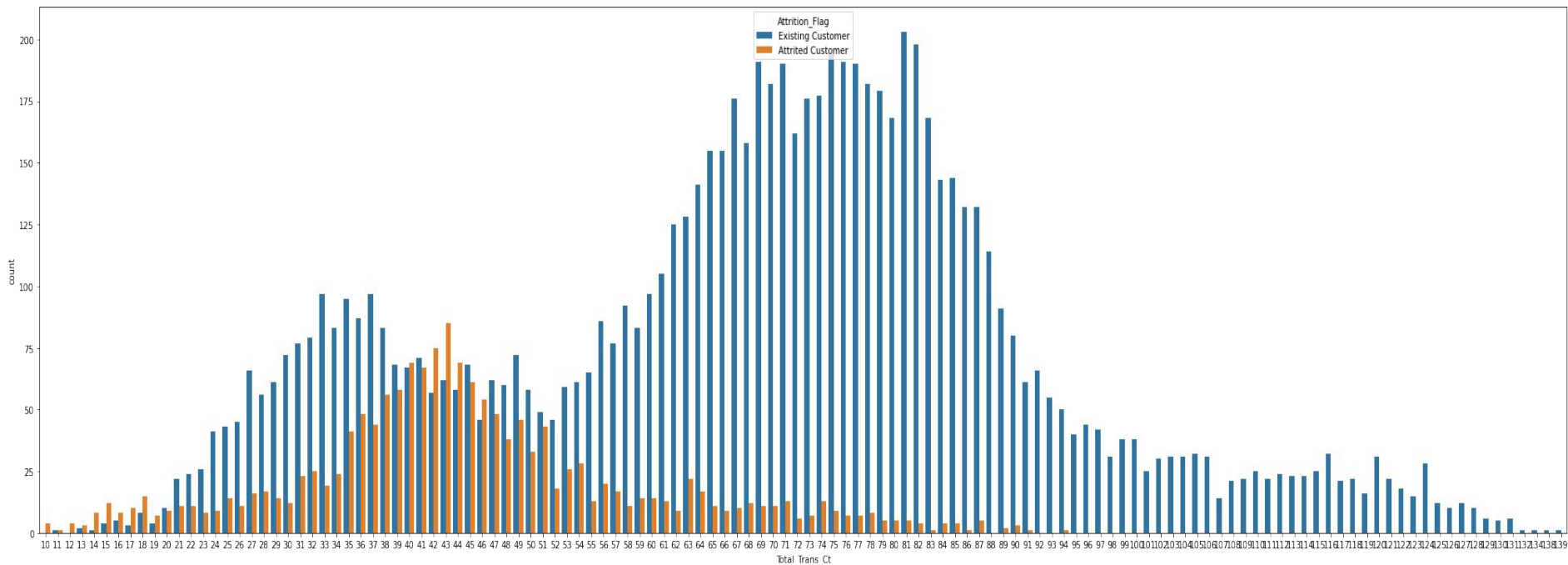
As we can clearly see, individual who use their credit card more often, are more likely to stay with our bank.

2. Total Revolving Balance

Customers with higher revolving balance are more likely to stay with our bank
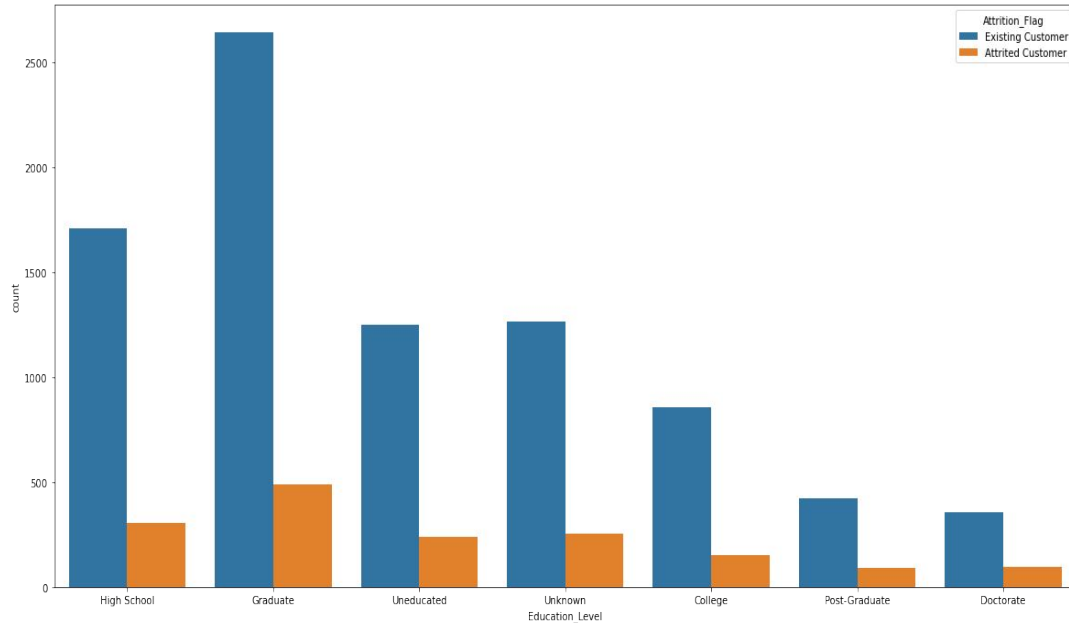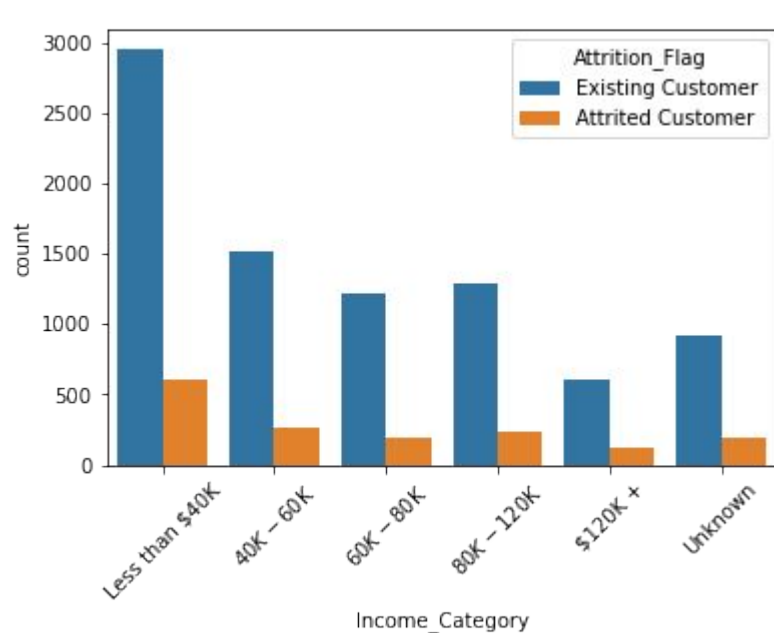
3. Total Relationship Count

Customers who are using more than two of the banks products are more likely to stay with our bank
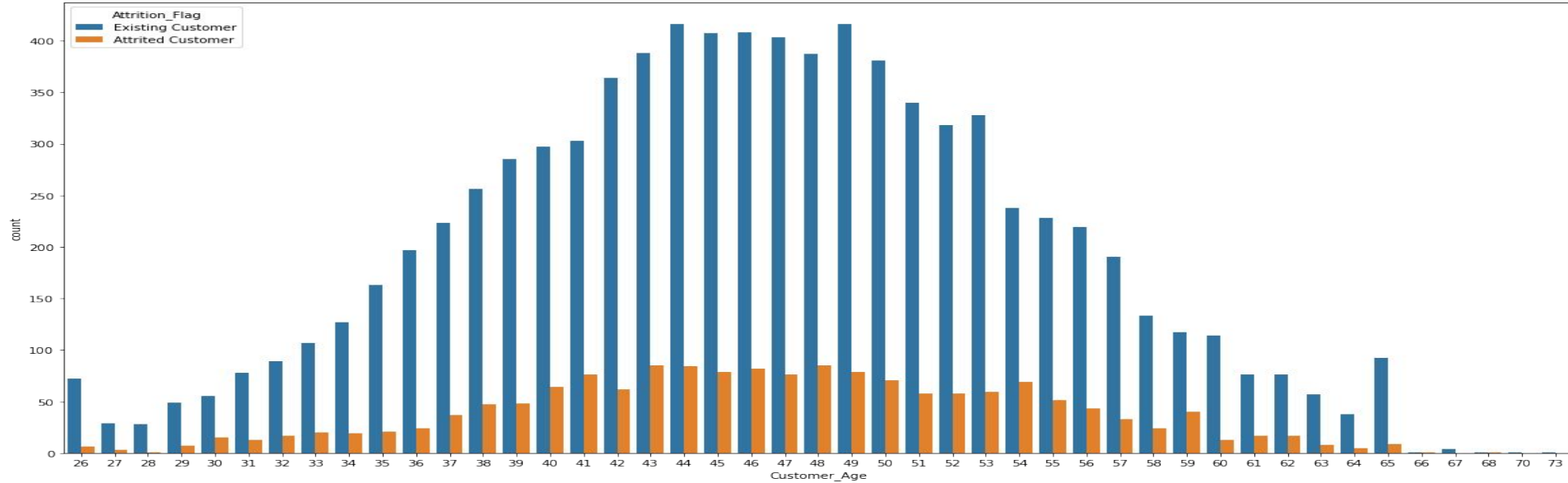
Non-Contributing Factors
Very little (if any) Correlated

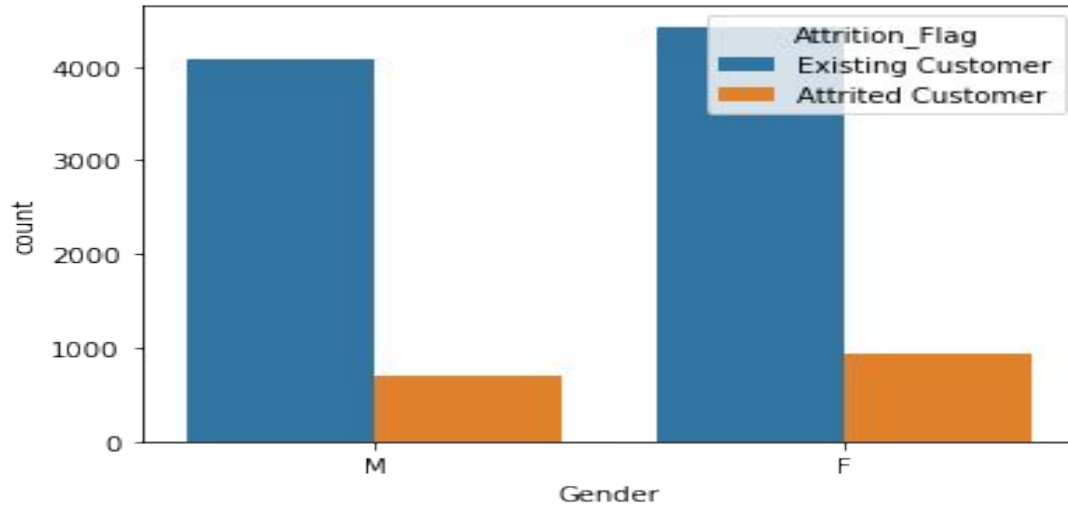### *Education and Income level*

The majority of the bank's customers are the ones make less than 40K and or have graduate degree

Age

Customer's age has very little (if any) correlation with churn rate

Gender

Customer Gender has no correlation with churn rate

Marital Status and Type of Credit Card
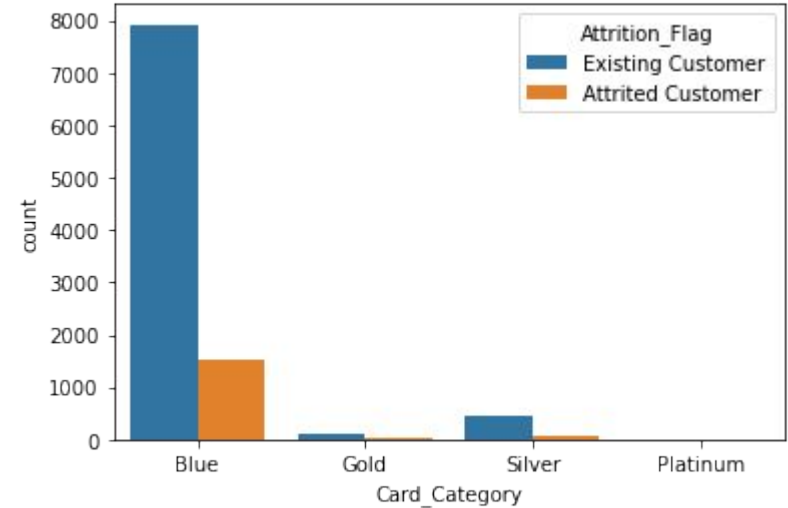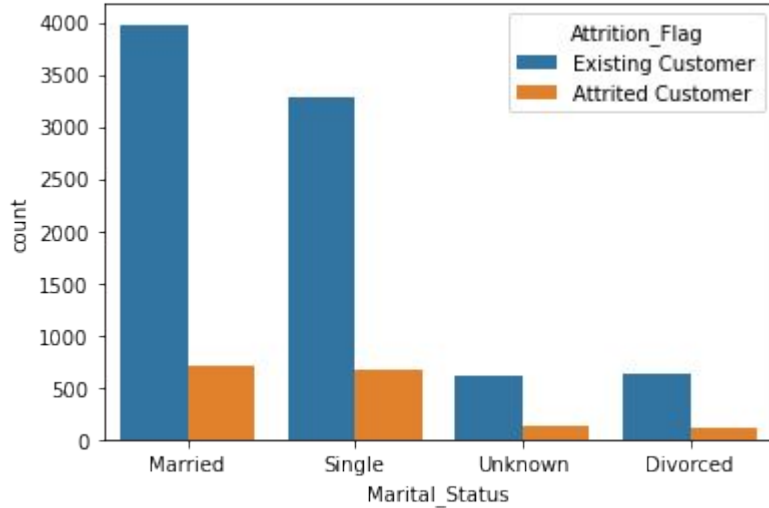
The rate of customer churn is slightly higher in single individual than the married ones and

The majority of bank's customer are the Blue card holder and since the number of all the other type of card holders are much smaller, it is unclear if type of card has any affect on churn rate

Machine Learning Modeling

Type : Supervised Learning

Binary Classification: 0 for Existing customer and 1 for Churned customer

Imbalance data : 15% data tagged with class 1

Tools : Python scikit Learn

Modeling Steps

1. Label Encoding
2. Data splitting into training and test sets (20% - 80%)
3. Use 3 different Classification Algorithms
4. Apply classification report to identify accuracy score
5. Create confusion matrix to find Precision and recall
6. Cross validation for finding best hyperparameter tuning

Classification Algorithms

1. Random Forest Classifier (Accuracy score of 96%)
2. Decision Tree Classifier (Accuracy score of 93%)
3. XGBClassifier (Accuracy score of 97%)

Since the last one has the best score, we will pick this model

Confusion Matrix for XGBClassifier

Confusion matrix, without normalization
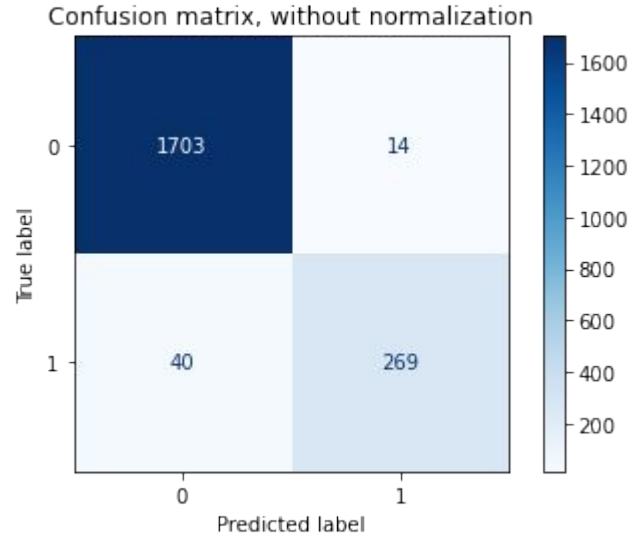
1. The actual number of 'existing' that was labeled correctly by our model (dark Blue)

2. The actual number of 'churned' that was labeled correctly by our model (medium blue)

3. The number of 'churned' that was predicted falsely as 'existing' by our model, (light blue at bottom)

4. The number of 'existing' that was predicted falsely as 'churned' by our model, (light blue on top)

Most influential features predicted by our model
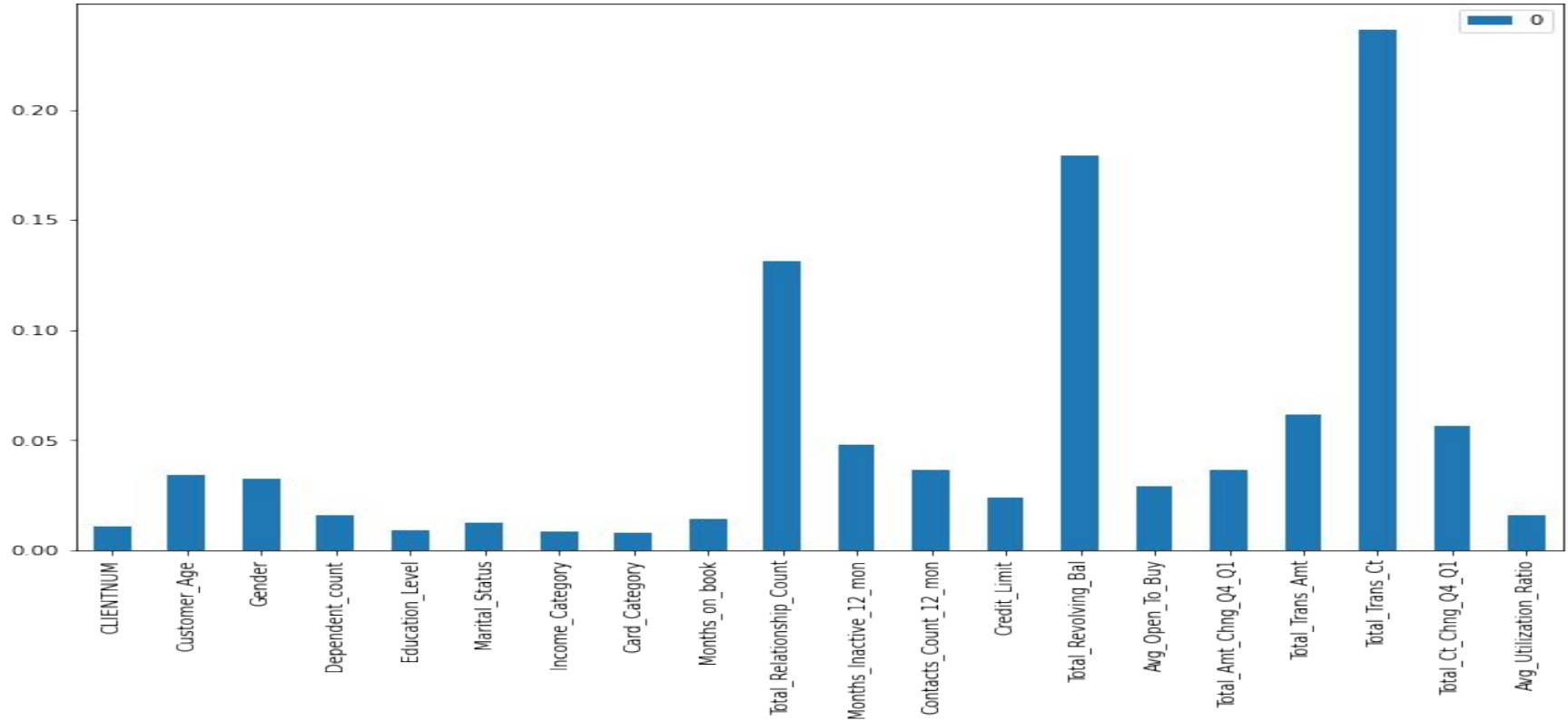
Conclusion: As we observed earlier in the visualization section, our model confirms the most important features affecting our churn rate

Recommendations

Based on the observed data and our model prediction, three features play the most important role in churn rate and those are:

1. **Total number of transactions,** management should find a way to motivate customers to use their credit card more often, it can be in a form of rewards, cash-back, contests ...

2. **Total Revolving Balance,** our data shows the customers who have more revolving balance are more likely to stay with our bank, in comparison with the ones with zero or low revolving balance. It is recommended that management find a way for customers to keep their revolving balance, the solution might be to offer zero interest on first x amount of revolving balance.

3. **Total Relationship Count,** this refers to the total number of bank's products each customer is using, it is suggested that management offer ways to encourage our customers in doing more business with our bank, for example opening checking/saving account, credit cards, loans by offering them rewards, cash back, interest free credit cards.

Further research

There are many factors that can contribute to the churn rate, which are missing in this data, it is recommended that the future data include some/all those features as well. Those features are;

1. Credit card annual fee (if any)
2. Credit card interest Rate
3. Detail customer satisfaction survey, which contains all areas of customer service
4. Cash back/ Rewards / sign-ins bonuses (if any)
5. Any other data that management is confident will affect the churn rate

Data source

https://www.kaggle.com/sakshigoyal7/credit-card-customers