

Capstone2 Final Report

The problem:

Customer attrition or (customer churn) is a tendency to abandon a brand and discontinue using the services or product of that business. The percentages of the customers to discontinue using the services of that particular brand is called churn(attrition) rate. This will be one of the health indicators of a business that is making profit based on their number of members using their products. In this section, I will propose some recommendations as to the reason, to what is the contributing factor to customer churn and the steps the bank can take in addressing those issues.

XXX bank is concerned about the fact that everyday, more and more of their customers leave their credit card services and have asked the Data scientists to help them find out the contributing reason(s) behind the churn rate and possible solution(s) to the problem.

Data:

The data that is provided by Kaggle, consists of 10,000 customers, including their age, salary, marital status, credit card limit, card category, etc. There are about 18 features. The data only include 16% of the customers who have churned , with limited data, it may not be an easy task to predict a good model.

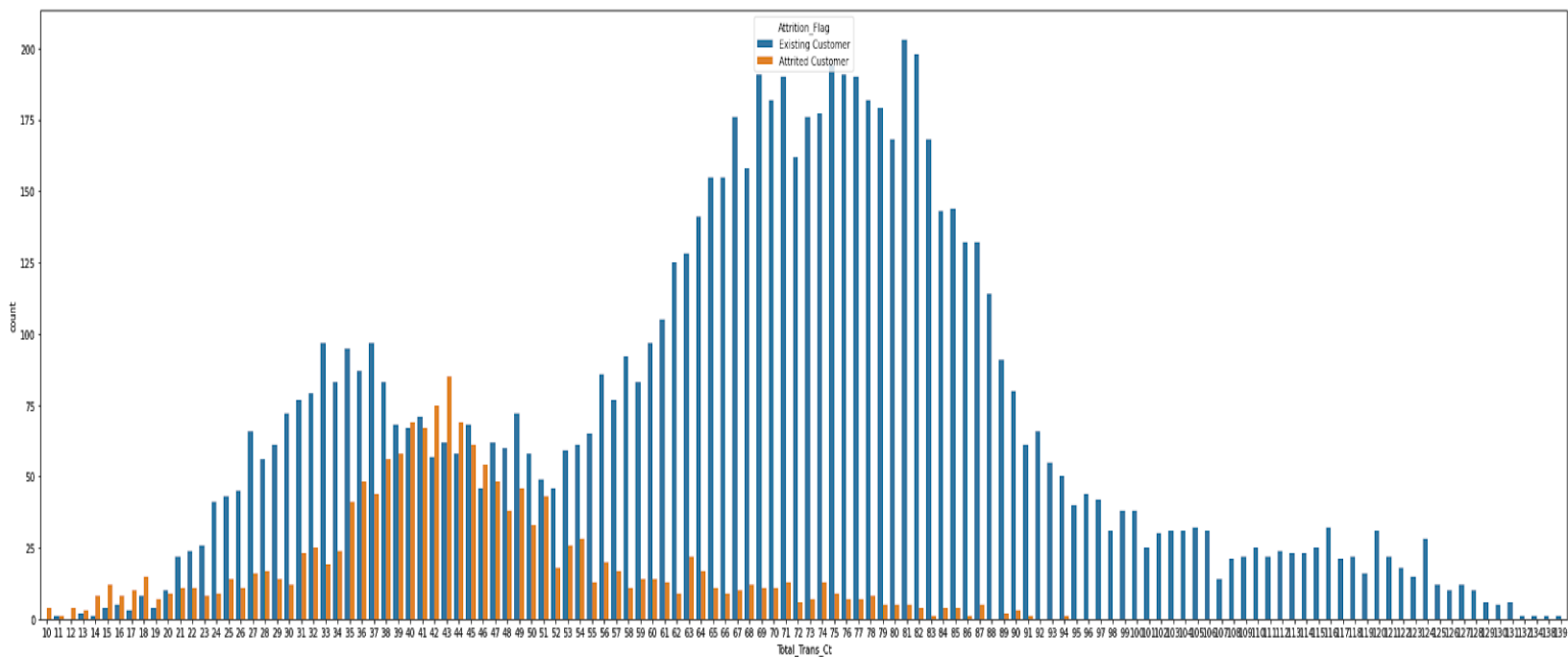
<https://www.kaggle.com/sakshigoyal7/credit-card-customers>

Findings:

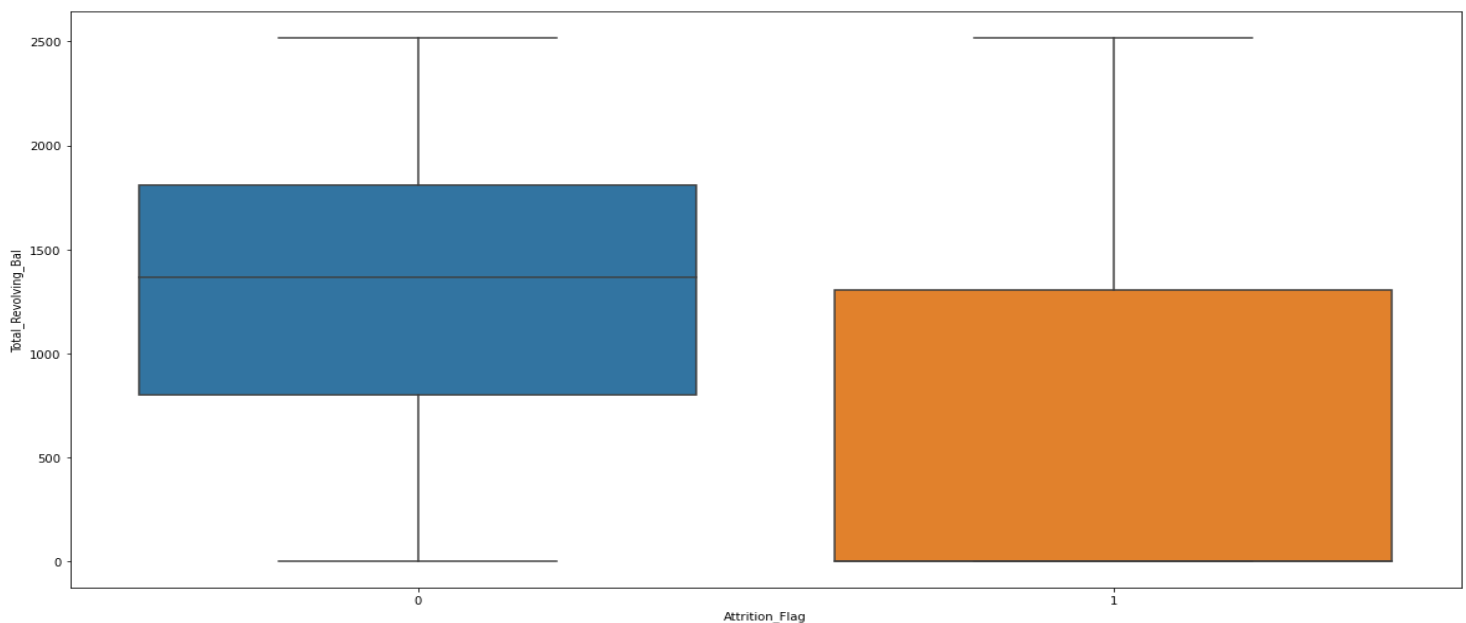
I have used many methods and different techniques to find the correlation between the data we have and the churn rate, from my findings, there are three features, that are mostly related to churning rate, and they are; total_tran_count , total_revolving_bal, total_relationship_count (total number of products held by customer). Here are all the data in charts and graphs, so the correlation can be easily seen.

1. Data Visualization- Features that are highly correlated

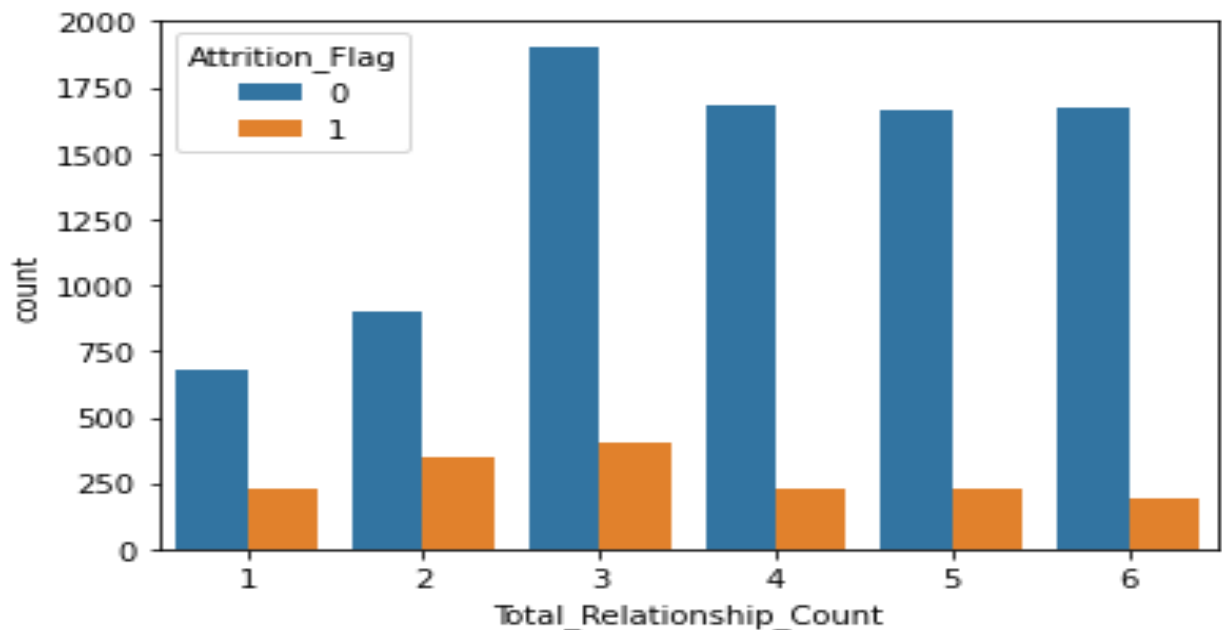
- **Total number of transactions** : in this section, we can see that the total number of transactions very correlated with churn rate, it shows that individuals who use their card more often , are more likely to stay with our bank, as opposed to the ones that use their card less often. (Blue color shows 'Existing Customer' and Orange color indicates 'Churned Customer')



- **Total Revolving Balance** : Individuals with higher revolving balance are more likely to stay with our bank, as opposed to, the ones with lower revolving balance

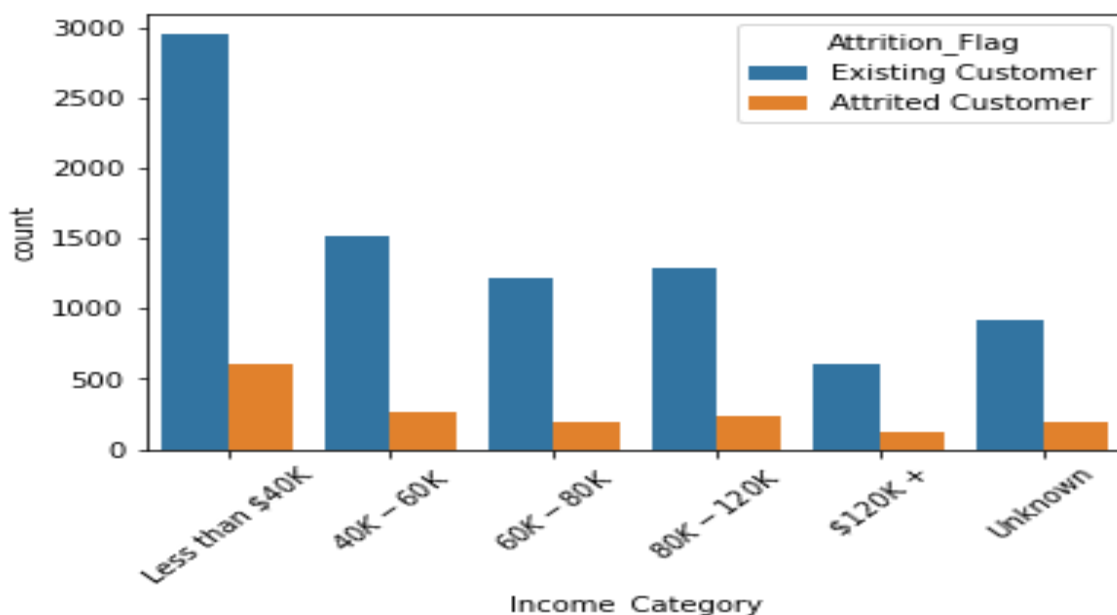


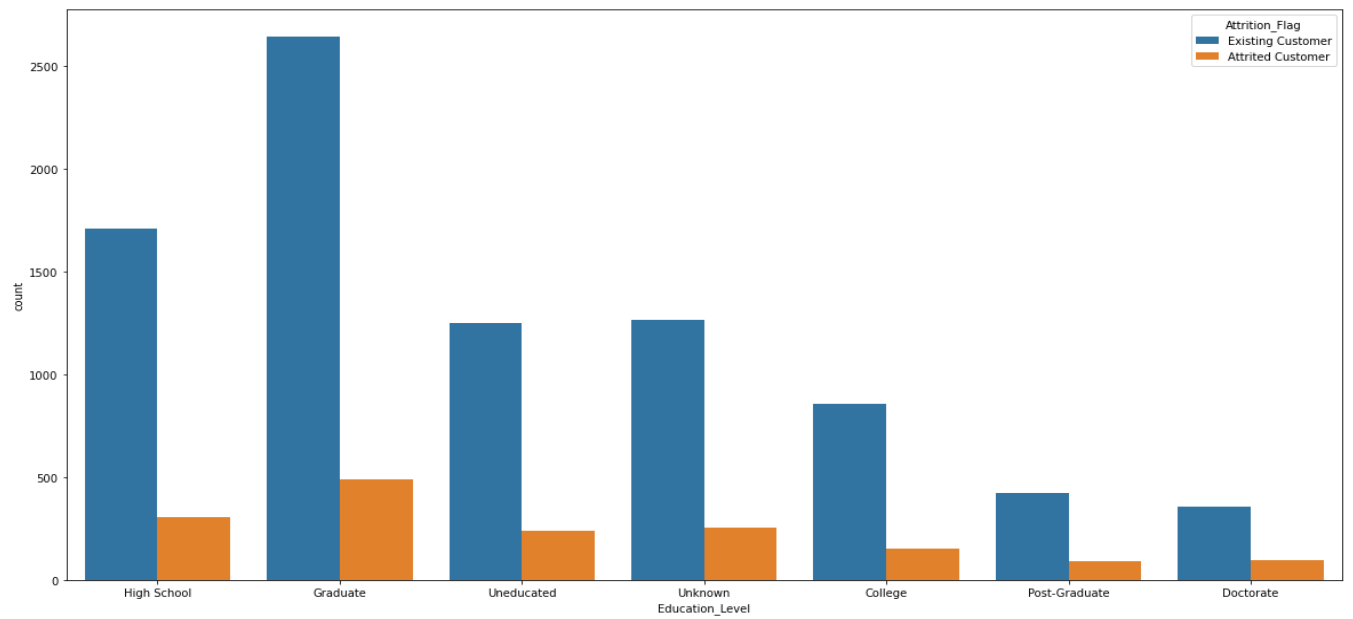
- **Total Relationship Count** : Customers who are using more than 2 of the bank's products (for example, 2 credit cards and/or checking/saving accounts) are more likely to stay with the bank, in comparison with the ones that are using only 1 or 2 of our products.



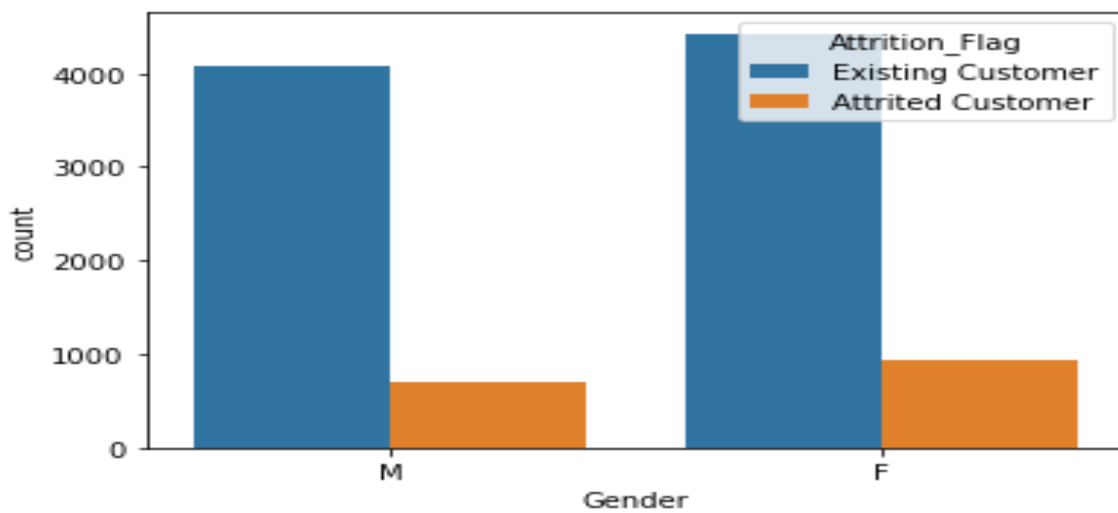
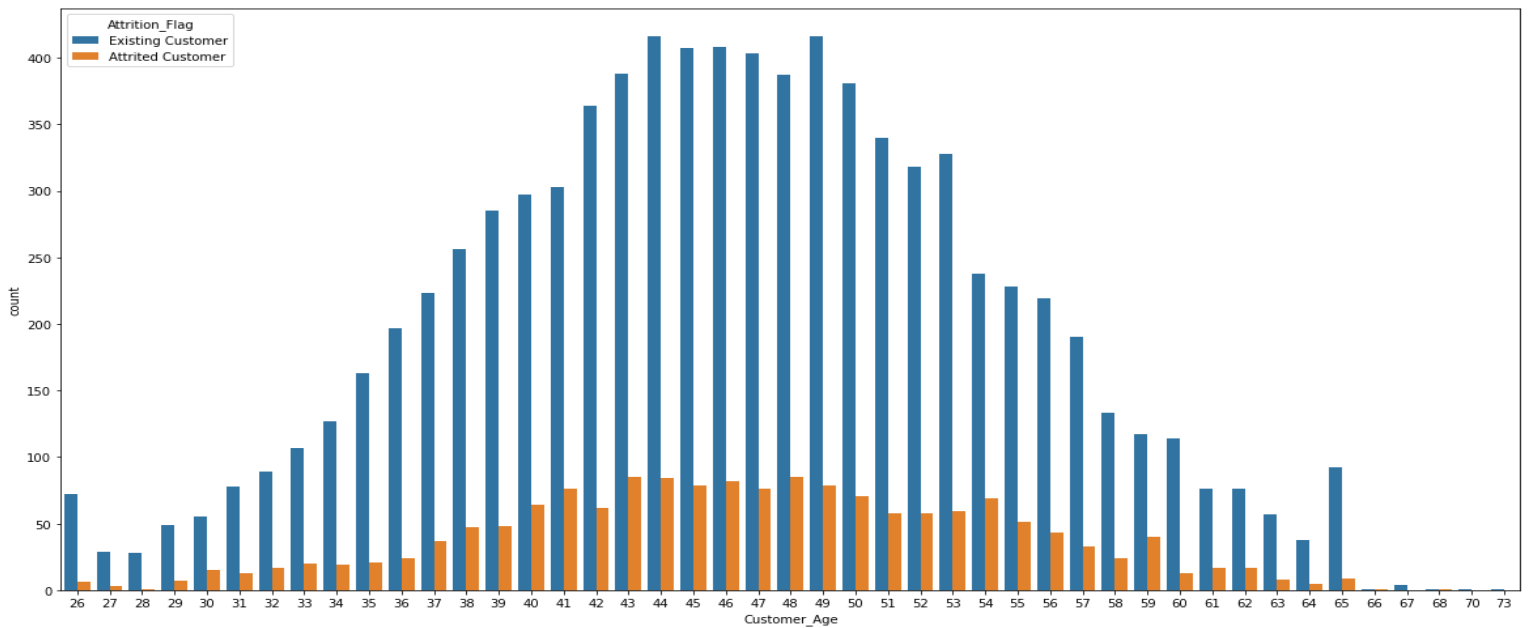
2. Data Visualization- Features that has very little, (if any) correlation

- **Income and Gender** : The majority number of bank's customers are the ones make less than 40K and/or have Graduate degree

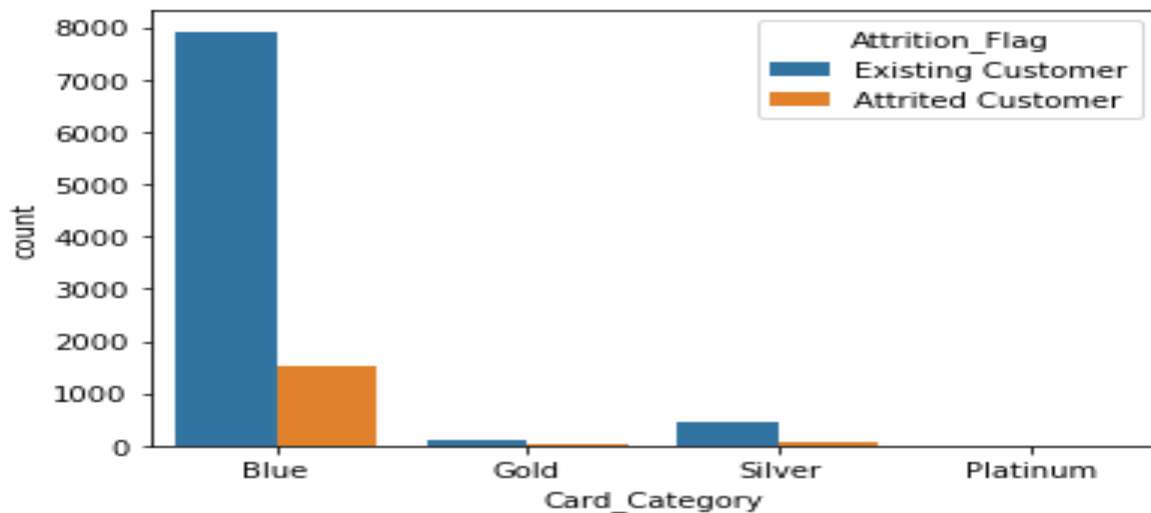




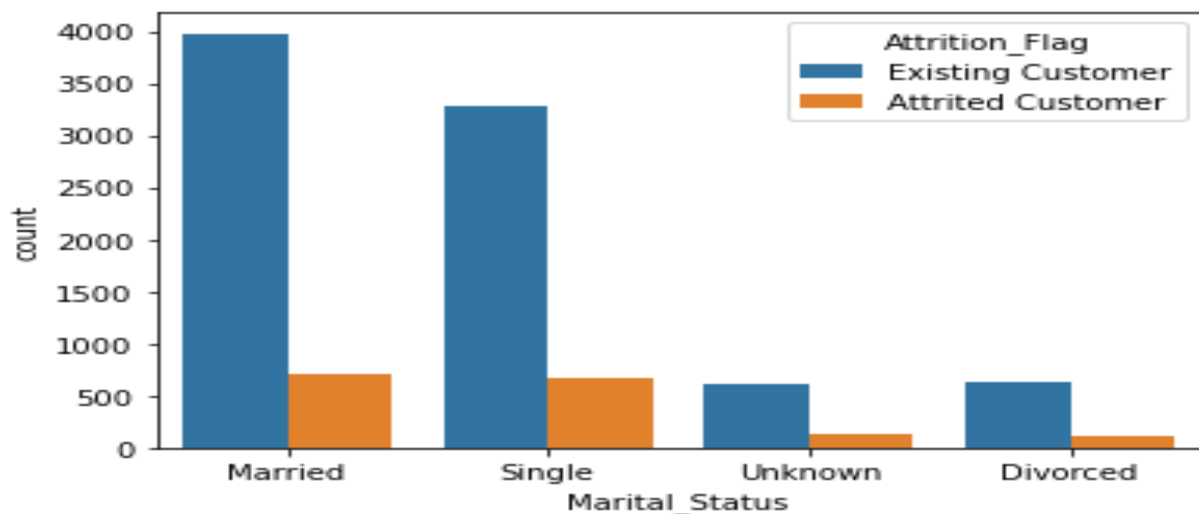
- **Age and Gender** : Customer age and gender has very little (if any) correlation with churning rate



- **Type of credit card** : This section shows that majority of bank's customers have the Blue card and since the percentage of the rest of the card holders are much smaller than the blue card holders, it is unclear that the type of card has any type of correlation with churn rate



- **Marital Status** :The rate of customer churn is slightly higher in single individuals than in married individuals, however, later in our modeling section, it does not show much correlation between the marital status and the churn rate



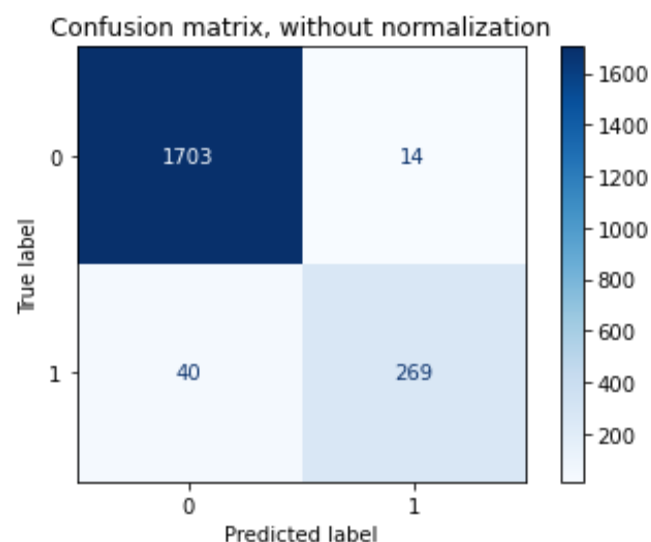
3. Data modeling

I have built different models to find a best performing one in regards to our data and pick the one with best accuracy score, our first model was built using 'RandomForestClassifier' with f1-score of 96% accuracy, which is a good model to predict our target. The second one is built with 'DecisionTreeClassifier' gives us the f1_score with 93% accuracy rate and the third one is built with 'XGBClassifier' gives us the f1-score of 97% which is our best model for predicting our target. So, I went ahead and picked that one.

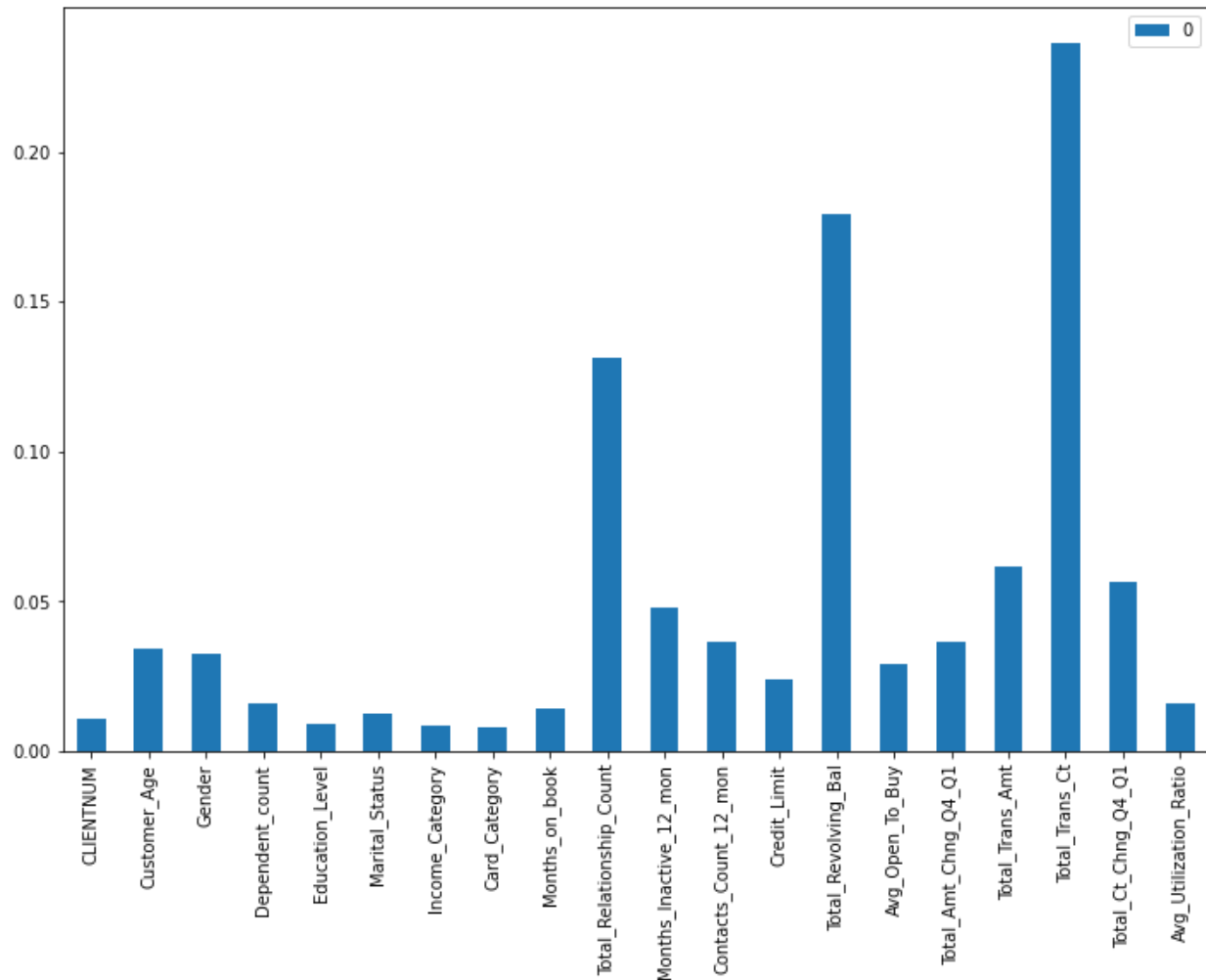
How does our model work? Based on the information that is provided in our dataset, the model will go through each and every individual, will predict who 'churned' and who 'stayed', then compares those results with the actual 'churned' and 'stayed' results and based on the comparison result, will calculate the accuracy score. These results will be shown in the confusion matrix. For our model, it shows 4 different numbers;

1. The actual number of 'stayed' 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 has been predicted 'stayed' by our model, (light blue at bottom)
4. The number of 'stayed' that was predicted as 'churned' by our model, (light blue on top)

The confusion matrix below shows, the model that was created, predicted 1703 people that stayed with the bank, actually labeled correctly and the 269 that actually churned, were correctly labeled as well. Other two numbers are the ones that we care the most about, it shows that 40 of the individuals who churned, were labeled as the ones that stayed with the bank, and 14 of the ones that stayed with the bank were labeled as churned.



Like we have predicted before from the charts in the previous section, our model confirms that the most important feature in our data analysis is Total_trans_count, Total-Revolving_bal and Total_relationship_count. Our model perfectly predicted that!



4. Recommendations:

As we can see, our data has many features, some have very little, if any, effect on our churn rate. Our model shows that the number one contributing factor in churn rate is 'Total number of transactions'. This means that the individuals who use our credit cards more often for purchases are more likely to stay with our bank. Thus the management must find a way to motivate their customers to use their credit cards more often, that can be in the form of rewards, cash-back, contests, etc,. This would require more extensive research on other datas and ultimately management discretion.

The second contributing factor in churn rate is, total revolving balance. That means customers who have more revolving balance are more likely to stay with the bank in comparison to the ones with lower or zero revolving balance. For this case, management must find a way for their customers to keep their revolving balance, even at a lower level. The solution can be to offer zero interest on first X amount of revolving balance, to motivate their customers to keep their revolving balance.

The third contributing factor is total relationship count, which is the total number of bank's products each customer is using. Management must motivate their customers in doing more business with them, like, opening checking/saving accounts, credit cards, loans, etc, by offering them rewards, cash-back, interest free new credit cards...

5. 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

