Our Bank’s consumer credit card business manager is concerned about their customer attrition:

OUR BANK CHURN - Technical Report

**Problem Statement:**

Customers keep moving to other banks and leaving Our Bank in their dust. This is a big problem,

because they know it costs a lot more money to get new customers than it does to retain existing

customers.

They want to analyse the data to find out the reason behind this, predict future customers who

might leave, and to determine whether or not there are strategies they can use to encourage

existing customers to stay. They also want to know if they can improve the data, they collect in

any way.

**Goal:** To find why customers are leaving Our Bank, and to predict future customers who might leave. To also determine strategies, they can use to encourage the existing customers to stay.

**Audience:** Our Bank’s Consumer Credit Business Manager, Stakeholders.

My success criteria: I would be successful if I would be able to find the reason for customers to leave Our Bank. Also, to predict the future customers who would leave and to find if there are strategies to encourage the existing customers to stay.

**Data Sources:** Our Bank Customer Data.csv

**Data Dictionary:**

The data provided by Our Bank was clean with no null values. The file had a total of 10,127 records(rows), each record uniquely recognized by the client id (which is named as CLIENTNUM here).

|  |  |
| --- | --- |
| Column Name | Description |
| Attrition\_Flag | If the account is closed, Attrited Customer.  If not, Existing Customer. |
| Customer\_Age | Customer's Age in Years |
| Gender | M=Male, F=Female |
| Dependent\_count | Number of dependents |
| Education\_Level | Educational Qualification of the account holder |
| Marital\_Status | Married, Single, Divorced, Unknown |
| Income\_Category | Annual Income Category of the account holder (< $40K, $40K -  60K, $60K - $80K, $80K-$120K, >$120K |
| Card\_Category | Type of Card (Blue, Silver, Gold, Platinum) |
| Months\_on\_book | Period of relationship with bank |
| Total\_Relationship\_Count | Total no. of products held by the customer |
| Months\_Inactive\_12\_mon | No. of months inactive in the last 12 months |
| Contacts\_Count\_12\_mon | No. of Contacts in the last 12 months |
| Credit\_Limit | Credit Limit on the Credit Card |
| Total\_Revolving\_Bal | Total Revolving Balance on the Credit Card |
| Avg\_Open\_To\_Buy | Open to Buy Credit Line (Average of last 12 months) |
| Total\_Amt\_Chng\_Q4\_Q1 | Change in Transaction Amount (Q4 over Q1) |
| Total\_Trans\_Amt | Total Transaction Amount (Last 12 months) |
| Total\_Trans\_Ct | Total Transaction Count (Last 12 months) |
| Total\_Ct\_Chng\_Q4\_Q1 | Change in Transaction Count (Q4 over Q1) |
| Avg\_Utilization\_Ratio | Average Card Utilization Ratio (Amount Used/Credit Limit) |
| Credit Score (Calculated Field): FICO (Fair Isaac Corporation, is a data analytics company focused on credit scoring services. It’s FICO score, a measure of consumer credit risk, has become a fixture of consumer lending in the United States. New credit is a variable that we do not have in our dataset, hence except this variable, the rest 4 variables have been considered to calculate the credit score. | (0.35\*Total\_Trans\_Amt) +(0.3\*Avg\_Open\_To\_Buy) +(0.15\*Months\_on\_book) +˘(0.1\*Total\_Relationship\_Count) |
| Predicted\_Churn\_code (Calculated field):  A logistic regression modelling was performed as part of the analysis to predict the future customers who would leave. This is the predicted output of the logistic regression model. | 1 – indicates the customer will stay with the bank  0 – indicates the customer will leave the bank |
| Predicted churn (Same as above) | 1 – is named as Existing Customer  2- is named as Attrited Customer |
|  |  |
|  |  |

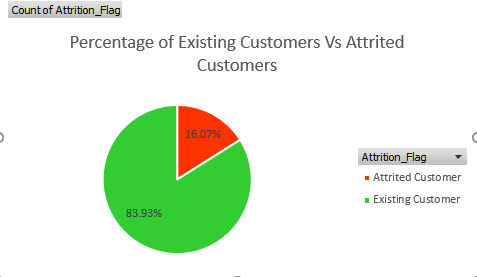
Following is the analysis of the dataset before applying the Logistic regression prediction model:

Patterns, Trends and Insights:

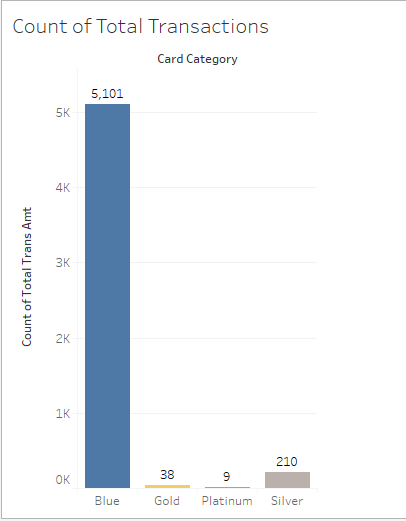
Analysis of the dataset given:

As of now, 16.07 % of the customers have left the bank and 83.93% of the customers are continuing with the bank. 16% is a big chunk. Let’s find out why. But before that lets understand more of the data.

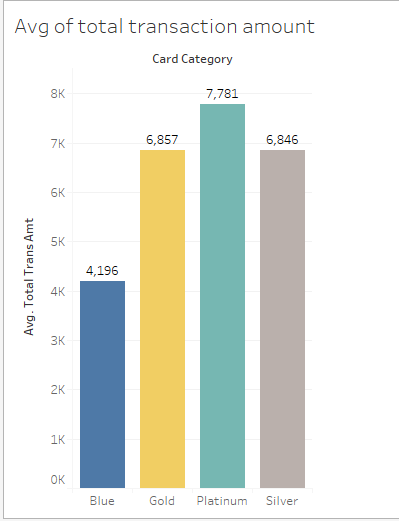
Throughout, attrited customers will be indicated in red and existing customers will be indicated in green.



The bank has more customers owning a blue credit card as when compared to the other credit cards.



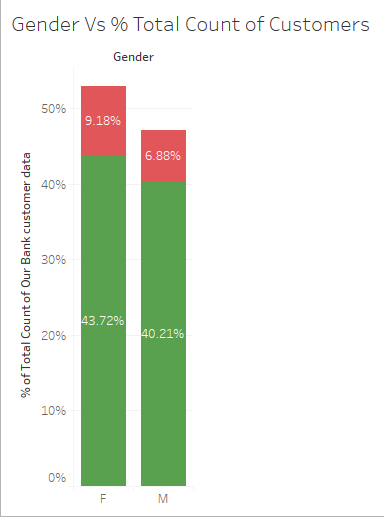
On the contrary, when average transaction amount is considered, platinum is at the highest at $7781 and blue being the least at $4,196.



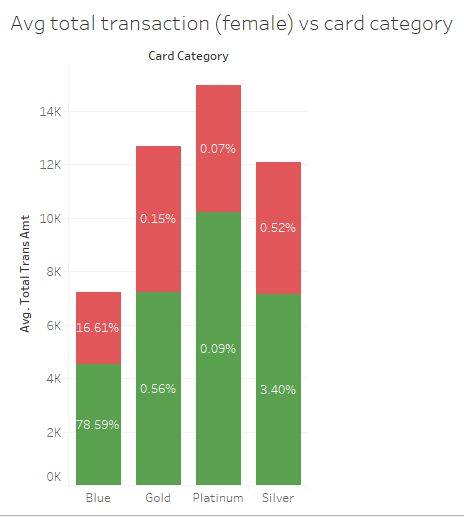
Now, considering the males and females:

Currently, the bank has more female customers with 57.16% and males at 42.84%.

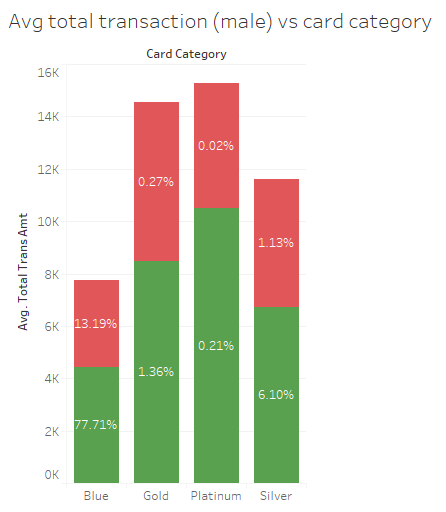
And a larger portion of female customers are attritioning at 9.18% as to compared to their male counterparts at 6.88%. This can be justified, as there are more female customers, we expect attritioning to be happening in the females. This also means that we got to look at the products being offered under each card category are female friendly or not.



Below graph shows, 16% of females owning a blue credit card have left the bank as compared to the percentages which are trivial across other category of credit cards.



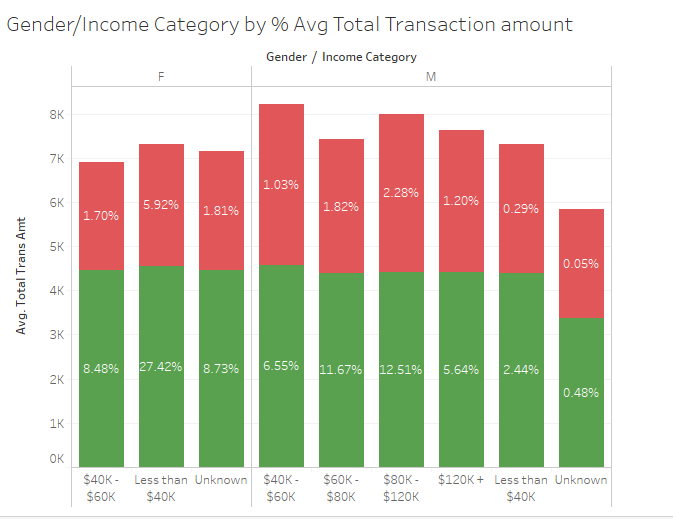
And below shows 13.19% of males who have left were also owning a blue credit card.



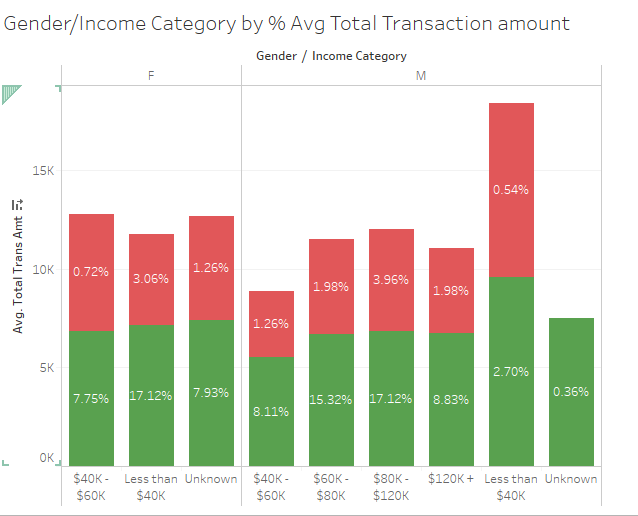
So, we understand here, that the problem was in the blue credit card holders. Lets delve deeper to understand why this happened?

Following is the card category wise break up of attrition in males and females, also considering their income categories.

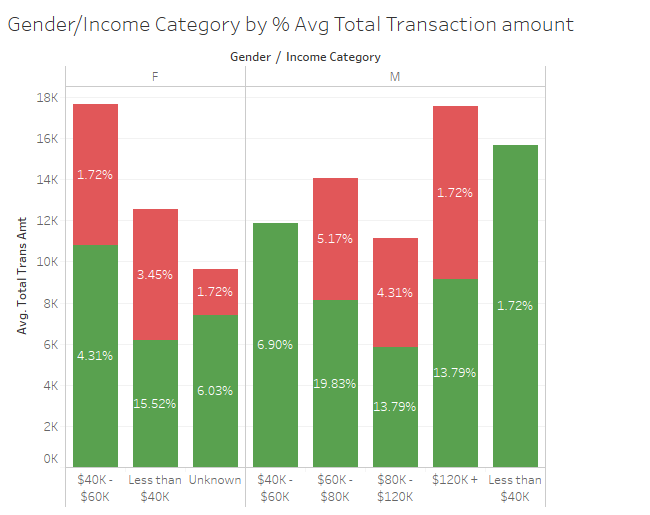
***BLUE:***



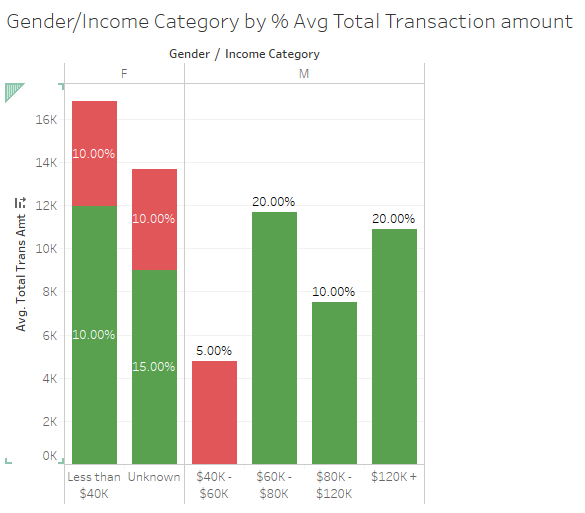
***SILVER:***



***GOLD:***



***PLATINUM:***



Analysis from the above card category wise graphs:

Key takeaways:

**FEMALES:**

***BLUE:***

1. We see attrition in all three income categories mainly, $40K to $60K, less than $40K and unknows. Majority of 5.92% attrition in less than $40K income category and 1.81% at unknow income category and 1.70% in the income category of $40K to $60K.

***SILVER:***

1. The same pattern follows with silver credit card holders as well. A major chunk of 3.06% attrited are in the income category group less than $40K and 1.26% are unknown income category.

***GOLD:***

1. Here also, the same pattern continues with less than $40K income category females attritioning at 3.45%, 1.72 % falling in the $40K-$60 K income category and another 1.72 % unknown income category.

***PLATINUM:***

1. 10% of customers under the income category of less than $40K have attritioned. And, 10% of the rest of customers who are females fall under the unknown income category.

Now, analysing the males:

**MALES:**

***BLUE:***

1. We see attritioning across all five income categories . Majorly, at 2.28% in the $80K-$120K income group and 1.82% in $60K-$80K income category.

***SILVER:***

1. Same pattern follows, 3.96% under the income category $80-$120K,

1.98% under the income category $60-$80K and , 1.98% under the income category of more than $120K.

***GOLD:***

1. Same pattern follows, 4.31% under the income category $80-$120K,

5.17% under the income category $60-$80K and , 1.72% under the income category of more than $120K

***PLATINUM:***

1. Luckily, here we see just 5% attrition under only one income category that is, $40K-$60K.

Interpretation to why they have left:

In males, we see that attrition is spread across all income categories. Particularly adding to weightage is the 5% attrition in Platinum card holders under the income category of $40K-$60K. Whereas no attrition in the higher income card categories. The lower income categories of people would have given up their Platinum cards due to the higher interest rates that are usually incurred on the higher end credit cards.

Meanwhile, other card categories of males would have left due to non-payment of debts.

In females, we see that a major portion of the attrition is coming from the 10% of the customers who are owning a higher end credit card (platinum where as their income categories are at less than $40K.

The same pattern observed in male and females is a clear indication that low income category groups were marketed into going for higher end cards such as platinum. These people defaulted later due to the higher credit rates associated with the higher end credit cards.

***Solution/Suggestion:***

Clearly, the solution to the problem is to cap the kind of credit cards issued to people based on their income category. So, this kind of attrition could be avoided.

Also, there is not just one reason for the customers who would have left Our Bank. A better way to keep customers is to periodically send surveys asking questions on if they are happy with the services offered, if there are pain points for the customers. This is basically knowing your customer periodically and getting into effect the changes if need to be made in terms of customer satisfaction.

Also, max credit limit across all the four categories of credit cards are at $34,516. This should be changed to a lower max credit limit for blue categories and step wise increase across the other credit cards with platinum getting the highest maximum credit limit.

This should be done to avoid customers who would end up using a max credit limit and later not being able to pay for the large debt followed.

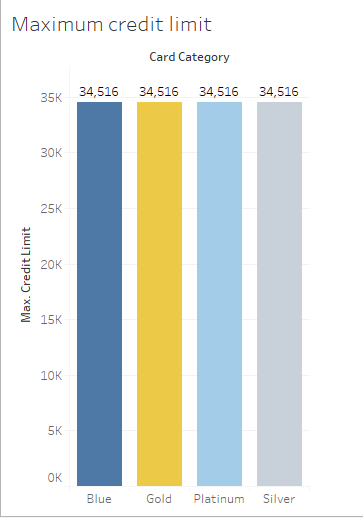
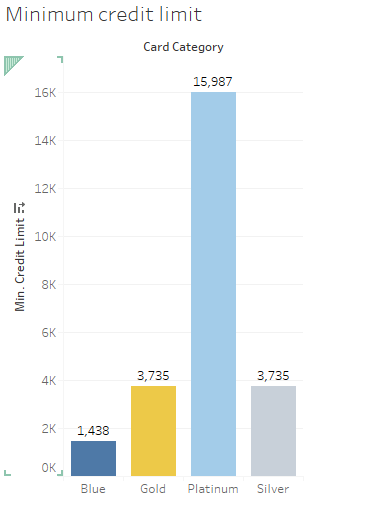
The below graphs would support this argument.

Ideally as the Credit limit increases the avg transaction amount increases. As the average transaction amount increases, the avg utilization ratio increases and as the avg utilization ratio increases the total revolving balance increases leading to more interest being levied by the bank on the customers. And this leads them to default or leave the bank.

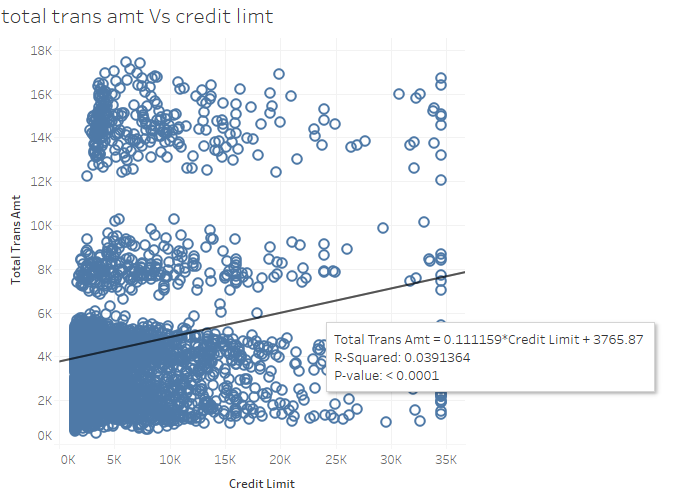
This is one angle to view the problem. The other could be customer satisfaction issues with the bank. This is where we need to do more analysis on the types of products/ services given by the bank to the customers.

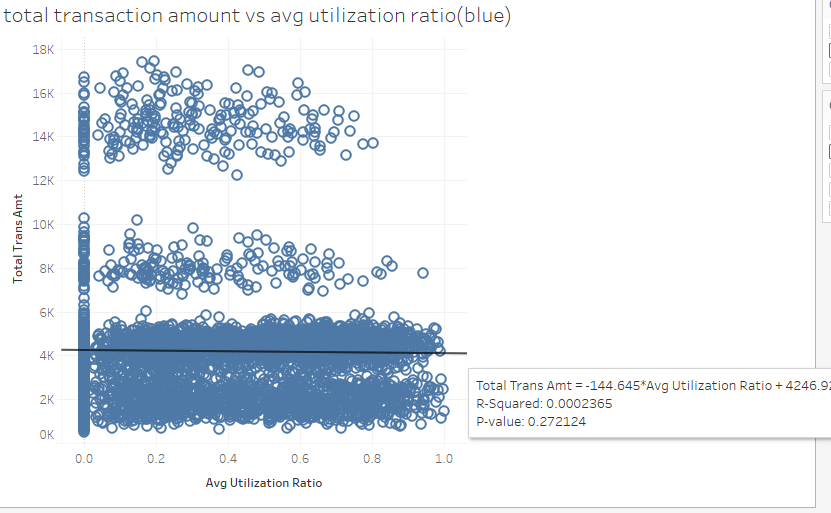
So, the solution to this is to keep a cap the max credit limit offered to customers according to their income category especially the platinum credit card holders (because that is where the major 10% attritioning is happening).

Higher income categories such as 60K and above are allowed to go for any of the credit cards. But lower income categories below 40K and 40K to 60K can only go for blue credit cards and blue or silver credit cards respectively and gold credit cards can be taken by all customers except under the income category of 60K and less.

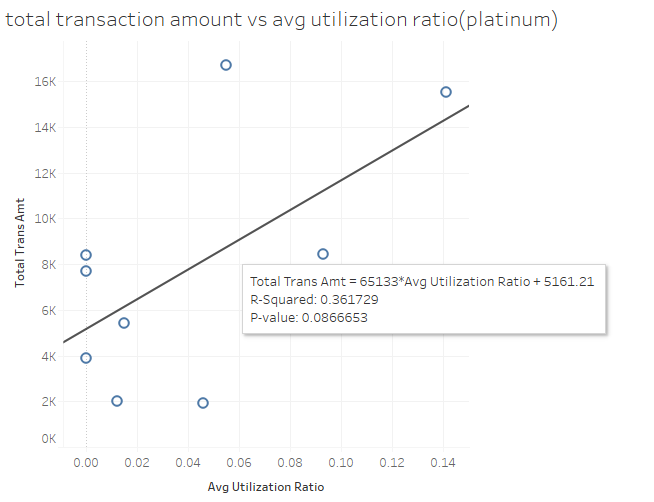
 

Trends and Patterns Observed:

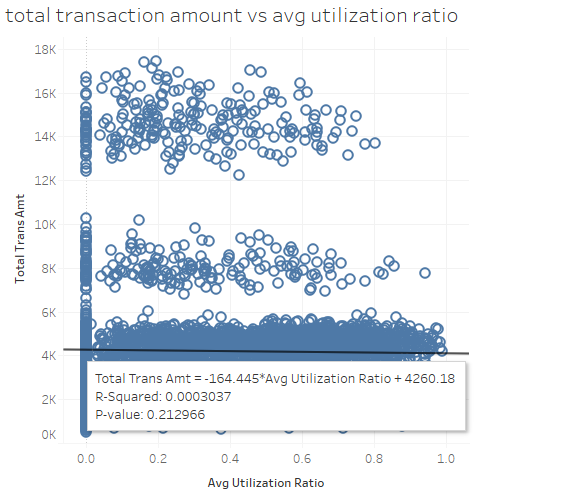




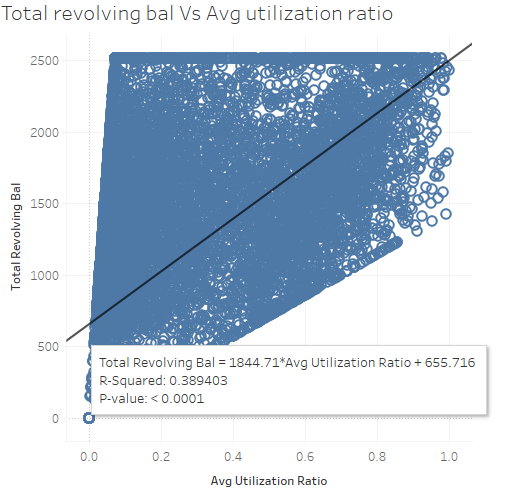
This the average total transaction amount vs average utilization ratio for blue credit card holders.



This is where the total transaction amount increases for the platinum customers. These platinum customers are females under the income category of less than $40K

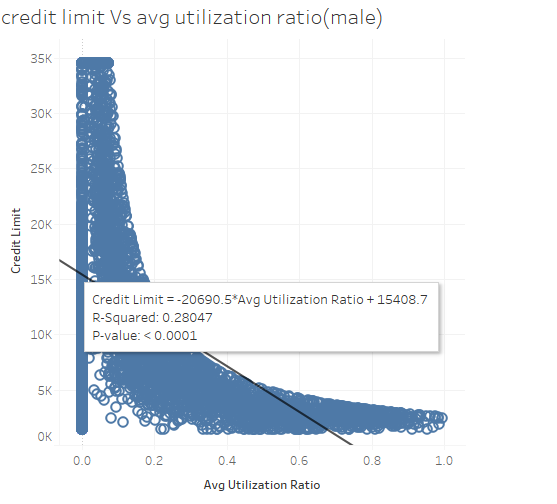


This the average total transaction amount vs average utilization ratio across females of all category of credit card.

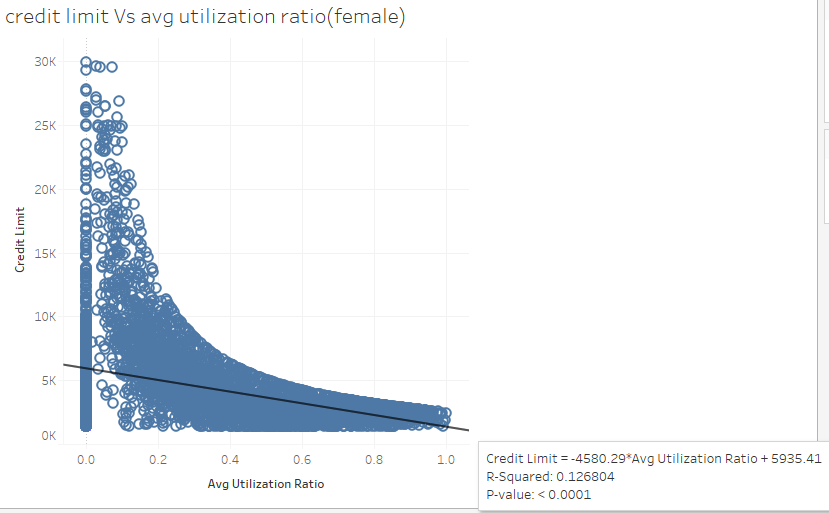


As the total revolving balance increases avg utilization ratio increases.

Avg Open to buy and the Avg Utilization ratio have a negative slope. As the average open to buy credit line increases, the average utilization decreases.

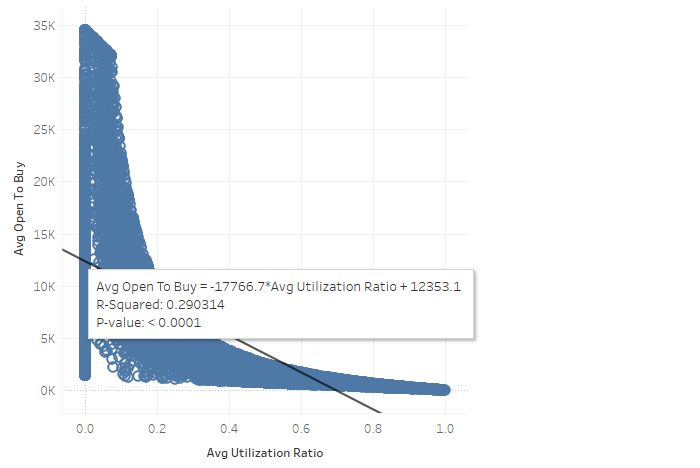


Avg Open to buy and the Avg Utilization ratio have a negative slope. As the average open to buy credit line increases, the average utilization decreases.

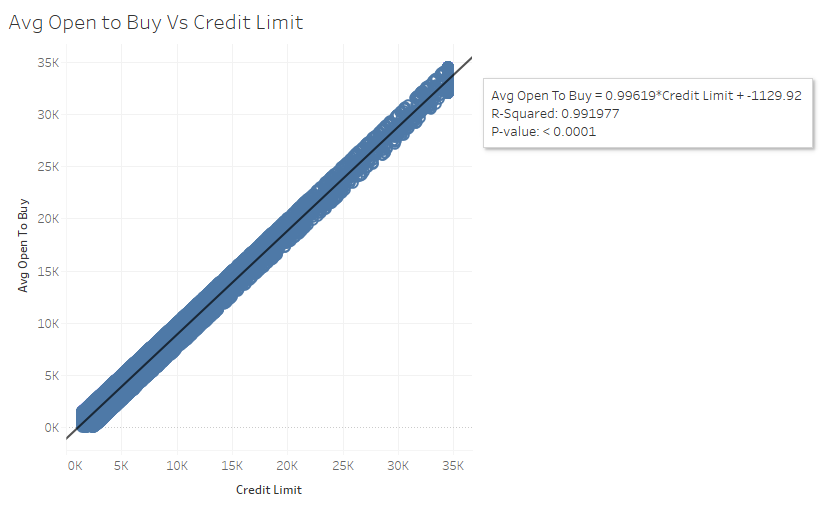


Avg Open to buy and the Avg Utilization ratio have a negative slope. As the average open to buy credit line increases, the average utilization decreases.

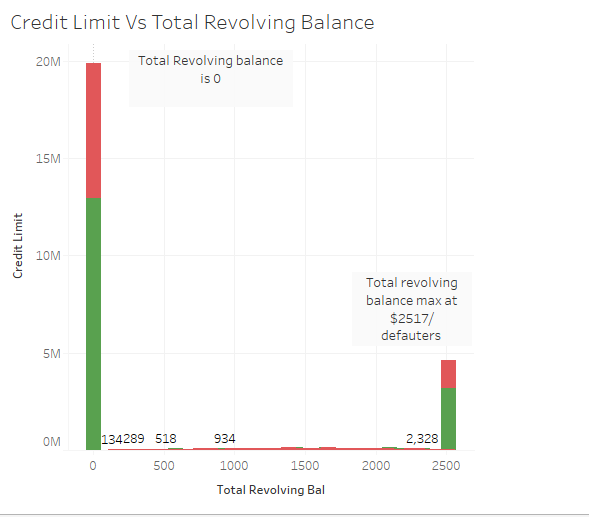
Below is the same graph as above, with female and males considered together.



Below is a perfectly corelated graph showing as the credit limit increases the avg open to buy also increases.

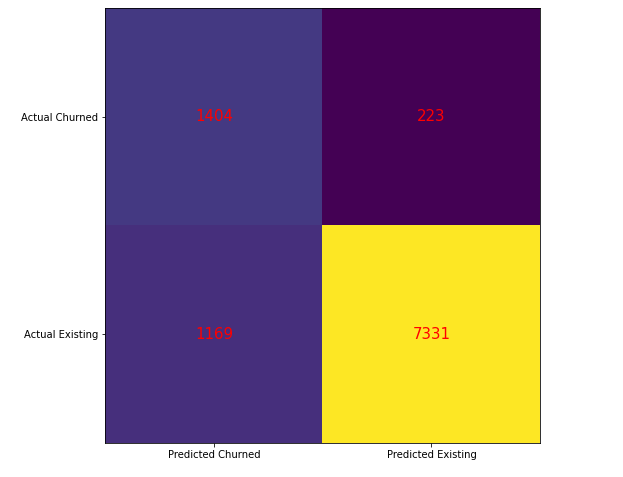


Below graph shows, even though the total revolving balance was 0, customers have left. Here the reason could be varied, either not satisfied with the products/services offered or might have gone for other competitive banks.



***To answer Our Bank’s Problem Statement to predict customers who are going to leave the bank, we performed a Logistic regression model as this suit best to look for customers who would leave the bank.***

The logistic model predicts the following:

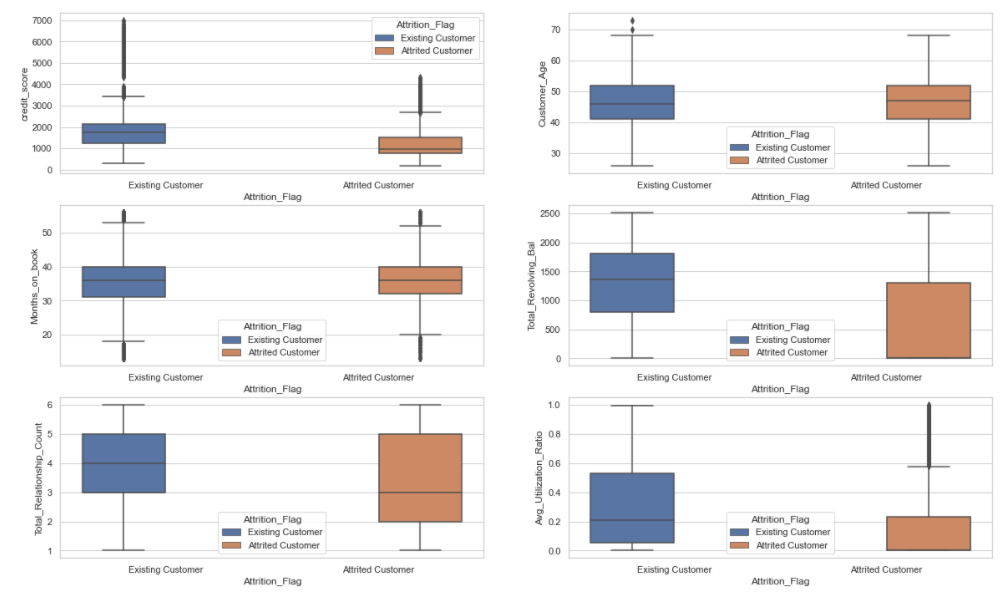


* 1404 customers who churned, the model predicted correctly that that the same number of customers have likely to have left the bank.
* 223 customers who churned, the model predicted incorrectly that they would still be with the bank.
* 169 customers who existed, the model predicted that these customers would churn.
* 7331 customers who are still existing with the bank, the model predicted correctly that these customers would remain with the bank.

So, now we as a bank need to find strategies need to find strategies to

* Retain these 1404 customers who have churned- and who we predicted will churn.
* And re-engage and retain these 7331 customers who are still loyal to us as the cost of retaining a customer is much lower than cost of acquiring new customers.

Now, another exploratory data analysis is as follows:



1. Existing customers have higher credit score rating than the attrited customers, but the difference is not very evident.

2. There is no significant difference in the customer age distribution between existing and attrited customers.

3. The months on books boxplot is slightly thicker for existing customers than when compared to attrited customers which means existing customers have had longer relationship with the bank as compared to the attrited ones.

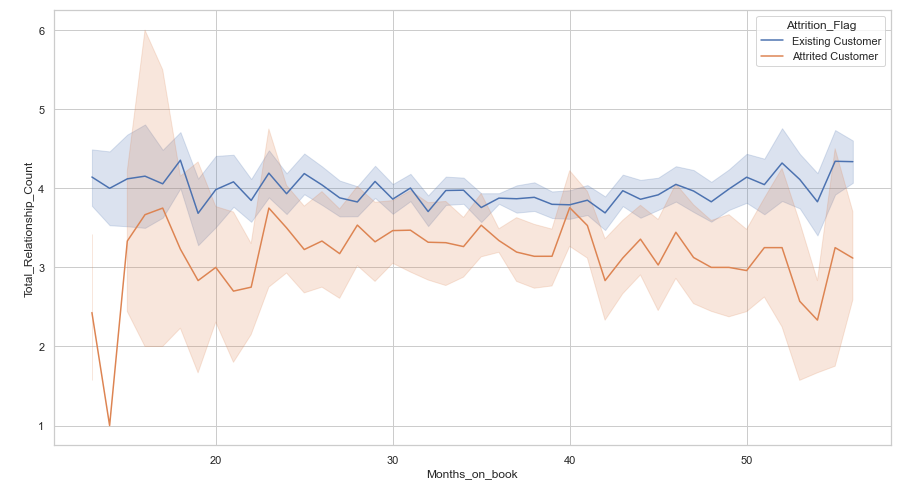
4. The customers who attrited the bank had a maximum of about $1400 as their balance. This means customers who have left the bank did not did not default with the bank but must have been unsatisfied with the Our Bank's Customer Services.

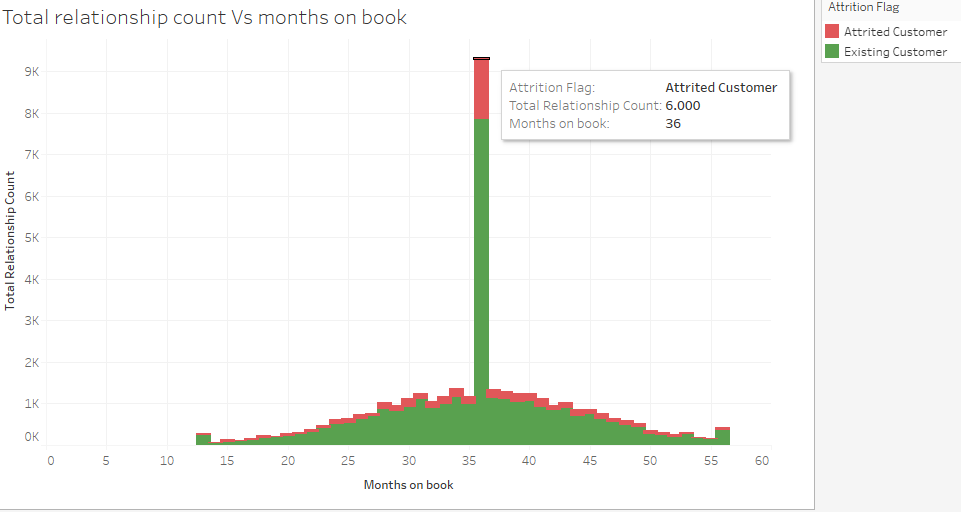
5. Surprisingly, customers owning more products with the bank have attrited. The bank may need to review their products here.

6. The average utilization ratio of existing customers is almost double that of attrited customers. This means attrited customers had less transaction ratio (less amount transacted with the bank)

Now doing a exploratory data analysis from the above two box plots and the below line graph, we find that at a period of 36 months, customers have had lower relationship count than when compared to the rest of the months.

This could be some scheme that the bank had come up with 36 months earlier, that in order to attract customers had come up like say low credit card interest rates or a customer attractive scheme that was valid for a period of duration 36 months, that made the customers to sign up for credit cards with the bank but later after the scheme ended the customers also attrited from the bank. This is one analysis.





General Trends and Patterns observed:

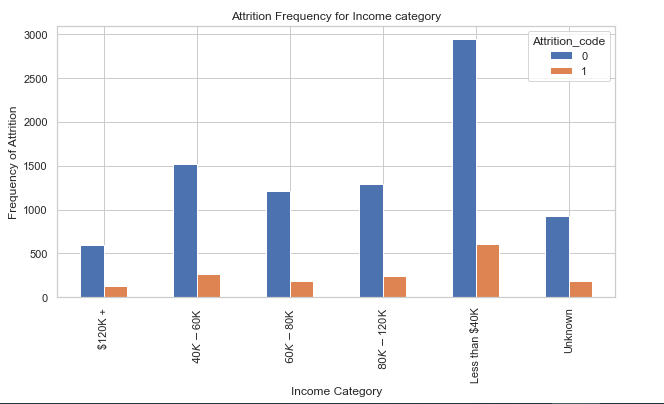
Frequency of attrition per income category:

More customers are in the income category of less than $40K and attrition is also found higher in this income category.

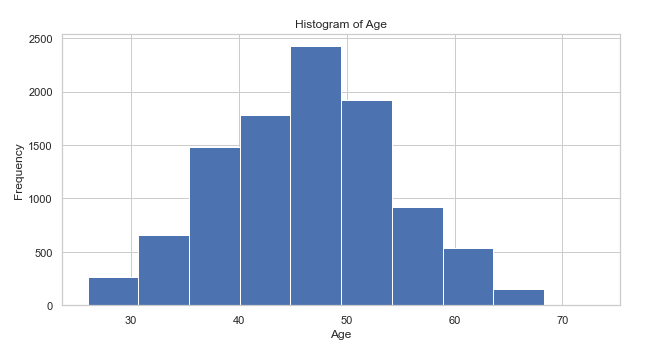
Legend:

Blue (0) means existing customers

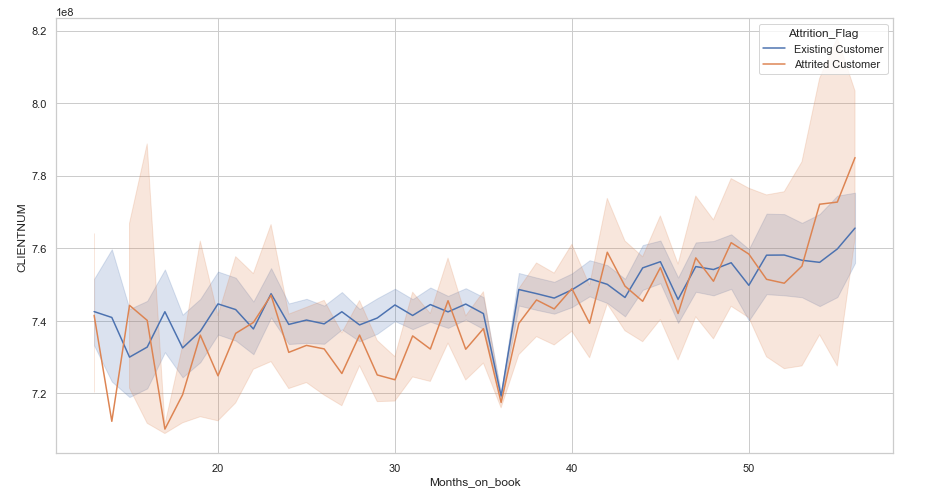
Orange (1) means attrited customers



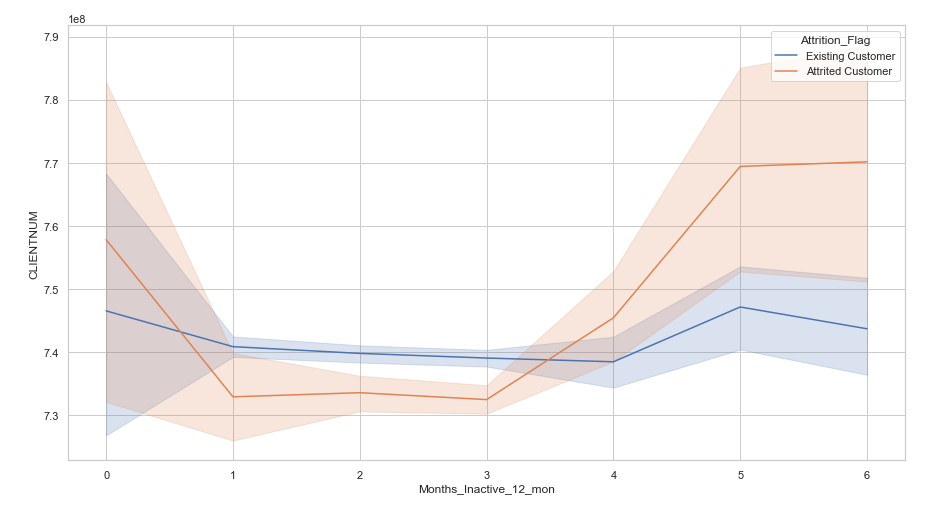
Below, shows the distribution of age of customers follow a normal bell curve.



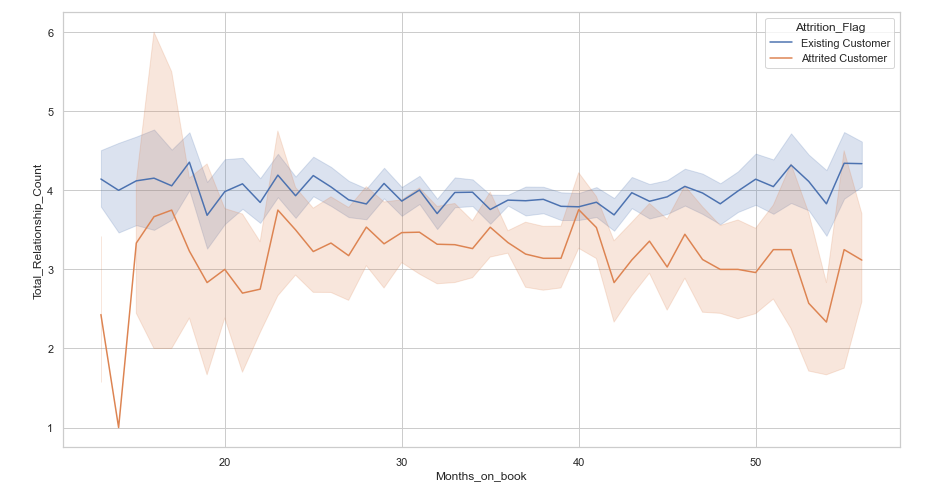
Below graph shows that, months on books of the existing and attrited customers follow a similar pattern.



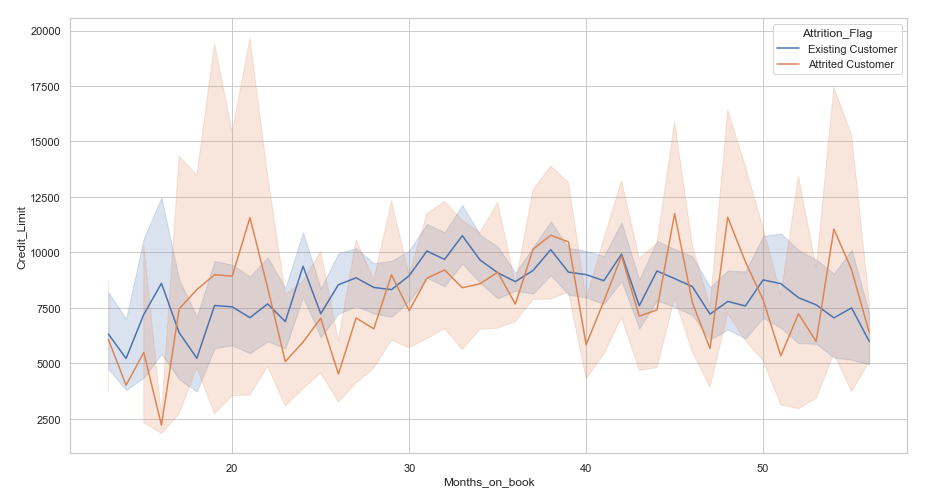
Similar pattern follows for both categories(attrited and existing) for the months inactive in the past 12 months.



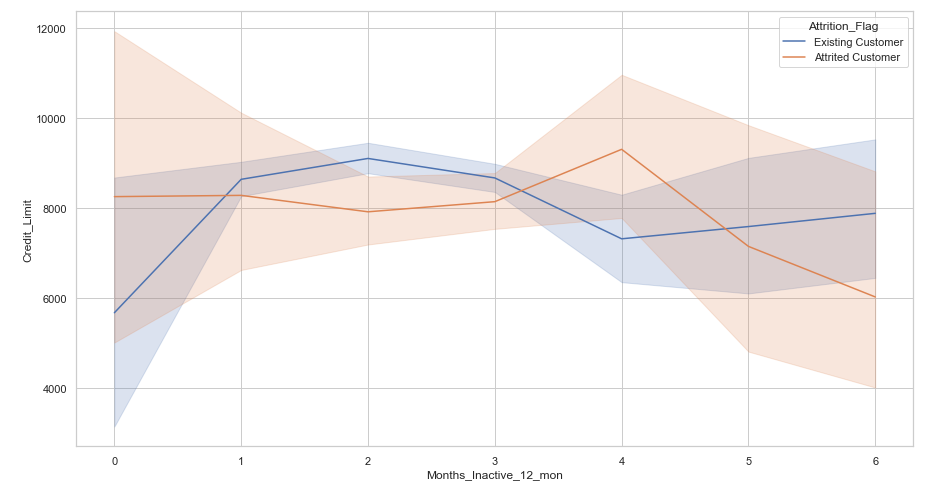
Existing customers had a higher relationship count (that is, held more number of products) than the the attrited customers.



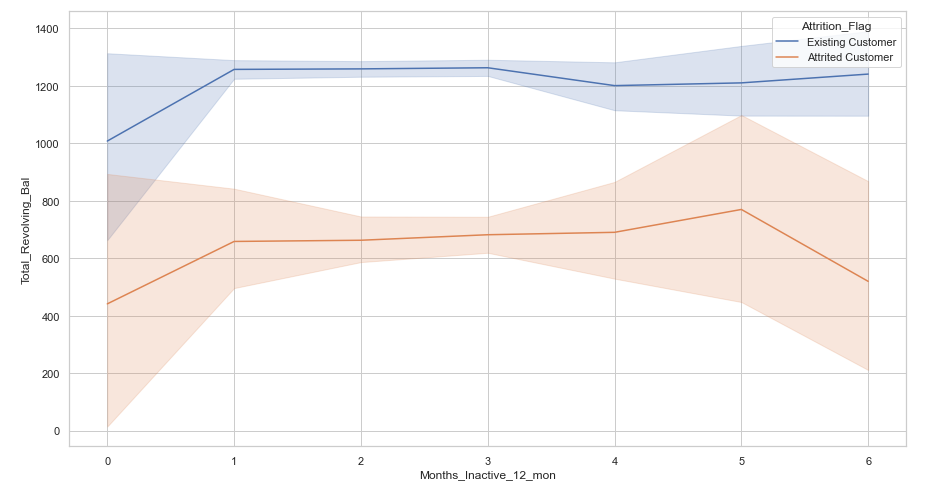
Customers who had more credit limit in the initial months and also likely in the later months are the one’s attritioned.

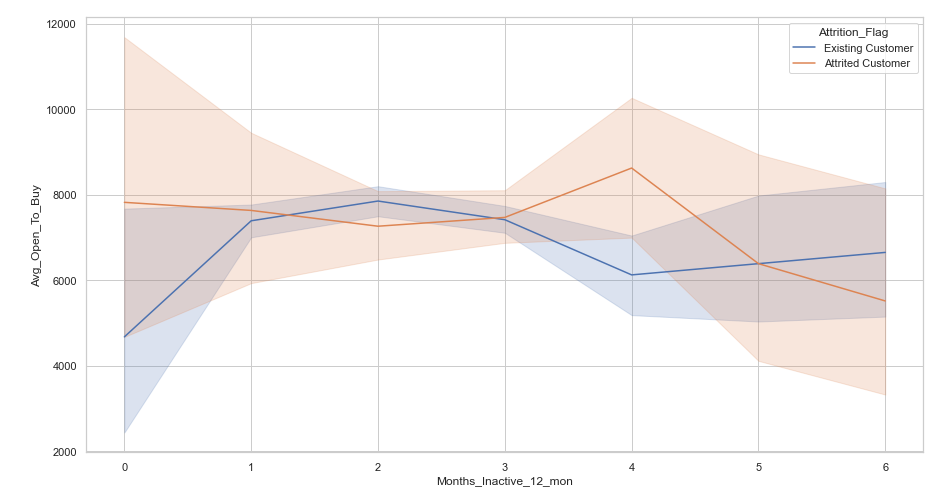


Credit Limit Vs Months inactive(in the past 12 months)

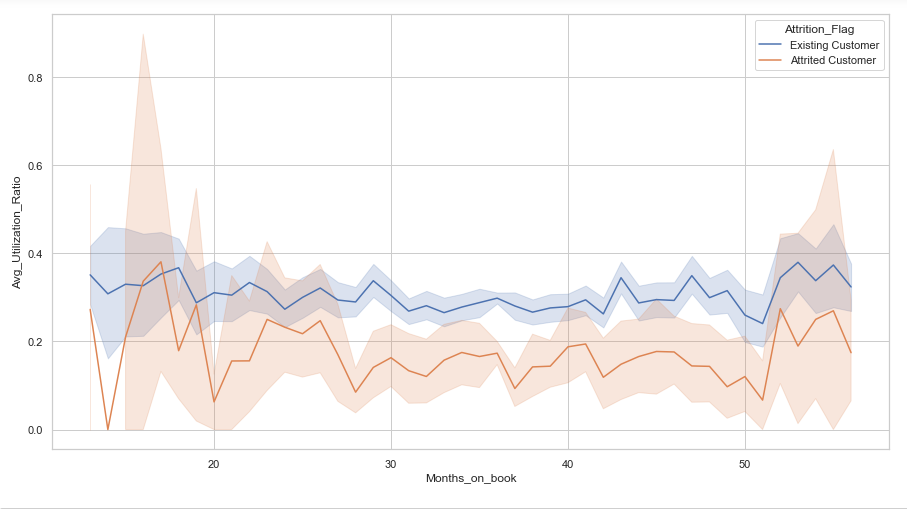
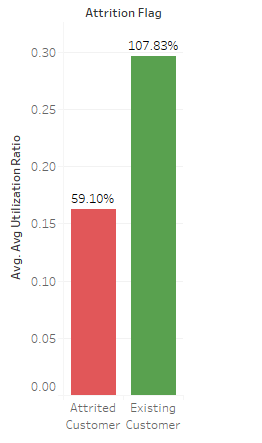


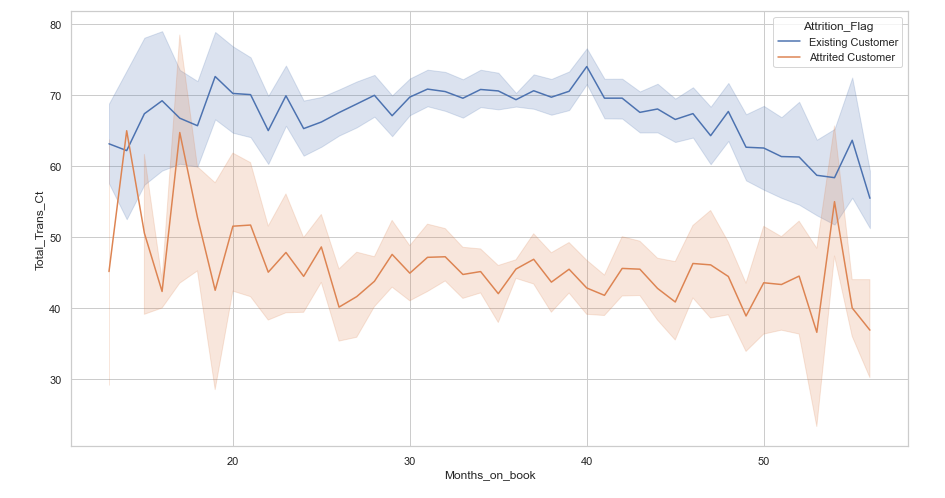
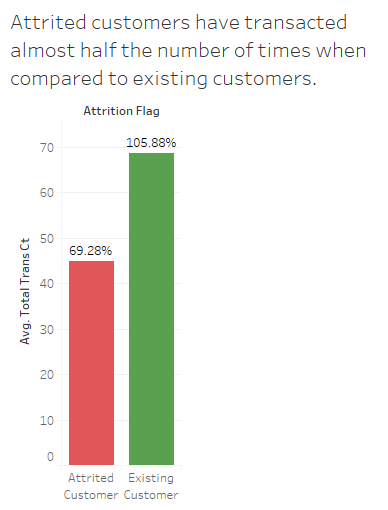
Attrited customers had less revolving balance in the past 12 months.





Average utilization ratio Vs Credit Limit: The graph shows attrited customers had lower utilization ratio(almost half) than existing customers. The bar chart shows more accurately the percentage.

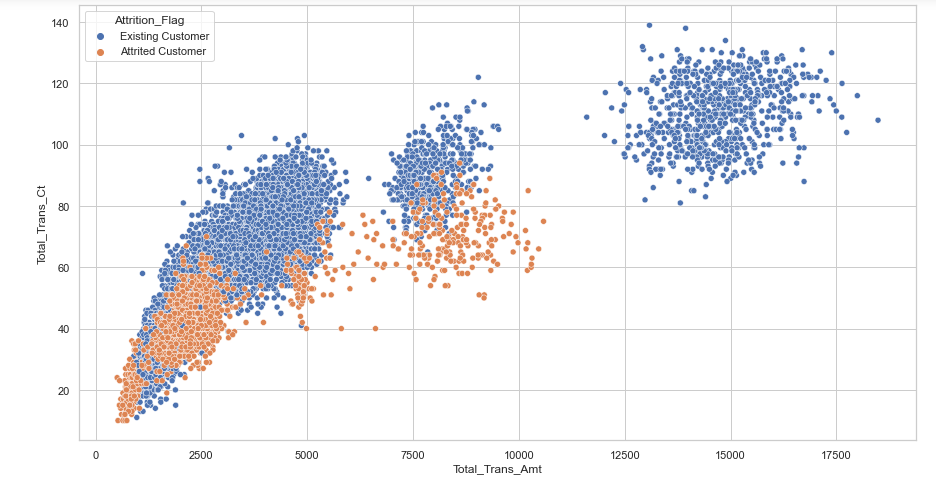
The same pattern follows in case of total transaction amount as well between the existing and the attrited customers.

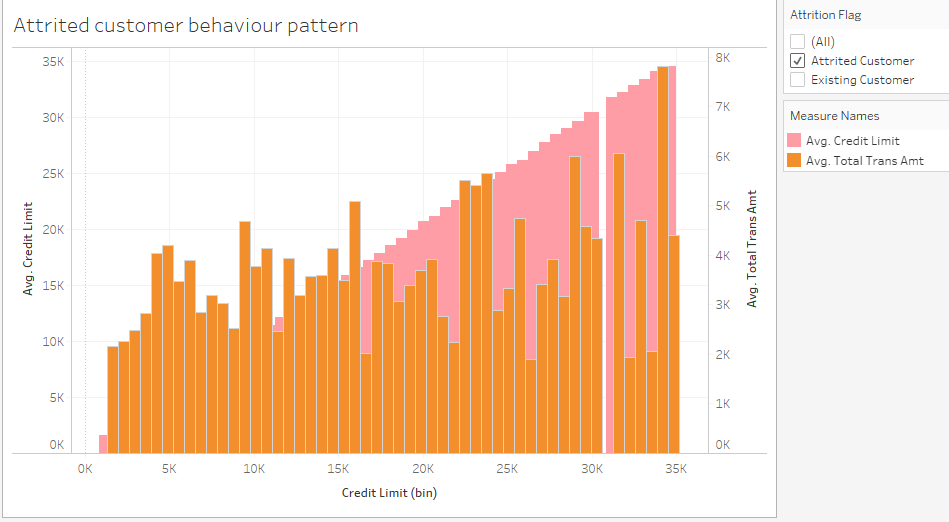
The below graph shows the pattern of attrited customers and the existing customers.

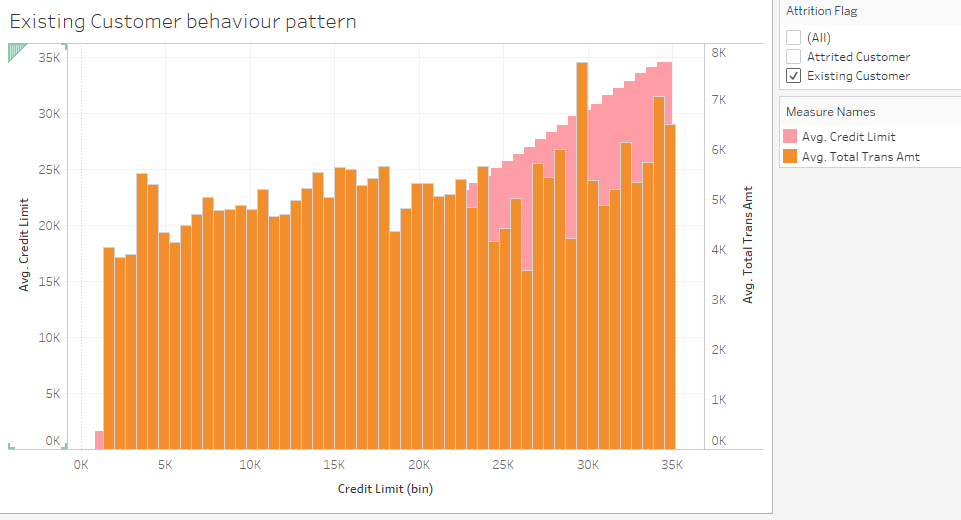
Each dot on this catterplot represents a customer.

We can see a pattern that attrited customers have had less transactions with the bank and also less transaction amount. And a major chunk of them left the bank after 70-80 transactions.

This is a good indicator to check on customers after their 60-70 transactions.

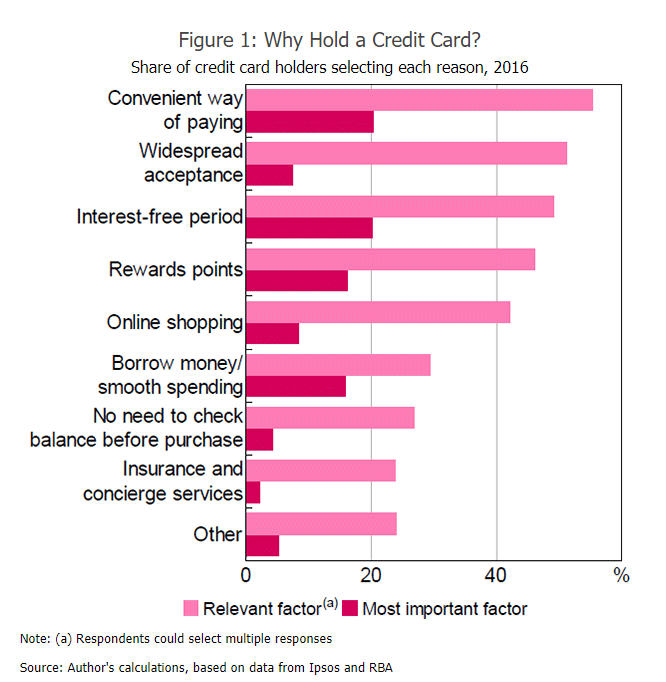






To answer the problem statement if there are recommendations on collecting the data, following are the suggestion/recommendations:

Recomendation 1: The graph has been taken from rba.gov.au.



The graph /survey shows the most important factors a customer considers while using a credit card.

Had there been data on the convenient way of payment, widespread acceptance and interest free period (which are also the most important factors according to the survey) offered by the bank, our analysis would be more accurate in finding reasons for customers to have left the bank.

Recommendation 2: Information on time as to when this data was collected is important.

The data does not mention about the time period during which this data was collected.

Time plays an important factor in determining the social conditions of people. For example, had the data been collected sometime during 2020, we know for a fact that the events in the year had played a major role for a customer to keep his credit card or not.

Also, had there been data about each customer’s transaction over a certain time period, our prediction would have been more accurate to understand the customer behaviour over the time period and as to when the customer had to leave the bank.

Next Steps: My next steps would be to consider doing analysis on customer behaviour and pattern as this would be accurate in assessing why have customers left the bank and also stayed with the bank.

Further reading:

The tableau work book has the 3 interactive dashboards along with other visualizations, dashboard 3 being the main.