VI DATA ANALYTICS BOOTCAMP CAPSTONE PROJECT

By Yong Sook Prasit Attavit

PROBLEM STATEMENT

A business manager of a consumer credit card portfolio is confronted with the challenge of customer attrition. He would like to analyze the data to find out the underlying cause of this issue, and to leverage the findings to anticipate customers who are at risk of leaving. Furthermore, he would also like to know of any suggestions as to how to retain their customers.

The purpose of this study is to establish associations within the data to identify the demographic group most prone to experience credit card attrition. This approach will enable the credit card company to proactively engage with customers to deliver improved services and influence their decisions towards retaining the card.

DATASET BACKGROUND

The dataset used in this analysis can be found from the link below:
Credit Card Customers - Predict Churning Customers

It contains information for more than 10,000 customers mentioning their age, salary, credit limit, attrition status, etc. with 23 columns.

In the following slide, further elaboration into the fieldname, type of data associated with the field, as well as the definition of each field will be shown.

DATASET GLOSSARY [1/2]

Field Name	Type	Description
Clientnum	*Num*	Client number. Unique identifier for the customer holding the account
Attrition Flag	*Char*	Internal event *(customer activity)* variable - *if the account is closed then 1 else 0*
Customer Age	*Num*	Demographic variable - *Customer's Age in Years*
Gender	*Char*	Demographic variable - *M=Male, F=Female*
Dependent count	*Num*	Demographic variable - *Number of dependents*
Education Level	*Char*	Demographic variable - *Educational Qualification of the account holder (example: high school, college graduate, etc.)*
Marital Status	*Char*	Demographic variable - *Married, Single, Divorced, Unknown*
Income Category	*Char*	Demographic variable - *Annual Income Category of the account holder (< $40K, $40K - 60K, $60K - $80K, $80K-$120K, > $120K, Unknown)*
Card Category	*Char*	Product Variable - *Type of Card (Blue, Silver, Gold, Platinum)*

DATASET GLOSSARY [2/2]

Field Name	Type	Description
Months on book	*Num*	Period of relationship with bank
Total Relationship Count	*Num*	Total no. of products held by the customer
Months Inactive 12 mon	*Num*	No. of months inactive in the last 12 months
Contacts Count 12 mon	*Num*	No. of Contacts in the last 12 months
Credit Limit	*Num*	Credit Limit on the Credit Card
Total Revolving Bal	*Num*	Total Revolving Balance on the Credit Card
Avg Open To Buy	*Num*	Open to Buy Credit Line *(Average of last 12 months)*
Total Amt Chng Q4 Q1	*Num*	Change in Transaction Amount *(Q4 over Q1)*
Total Trans Amt	*Num*	Total Transaction Amount *(Last 12 months)*
Total Trans Ct	*Num*	Total Transaction Count *(Last 12 months)*
Total Ct Chng Q4 Q1	*Num*	Change in Transaction Count *(Q4 over Q1)*
Avg Utilization Ratio	*Num*	Average Card Utilization Ratio

DATASET PREPARATION

1. A filter was performed on the original .csv file to check for missing data.
2. Columns *F, G, H* corresponding to *Education_Level, Martial_Status, Income_Category* contained "Unknown" in their cells. These cells will be filtered out during data analysis.
3. Columns *V & W [Naïve Bayes Classifiers]* were not utilized for this report as we're not performing a superivised learning algorithm used for classification tasks.

F	G	H	V	W
Education_Level	Marital_Status	Income_Category	Naive_Bayes_Classifier_Attrition_Flag_Card_Cat	Naive_Bayes_Classifier_Attrition_Flag_Card_Cat
High School	Married	$60K - $80K	9.34E-05	0.99991
Graduate	Single	Less than $40K	5.69E-05	0.99994
Graduate	Married	$80K - $120K	2.11E-05	0.99998
High School	Unknown	Less than $40K	0.00013366	0.99987
Uneducated	Married	$60K - $80K	2.17E-05	0.99998
Graduate	Married	$40K - $60K	5.51E-05	0.99994
Unknown	Married	$120K +	0.00012303	0.99988
High School	Unknown	$60K - $80K	8.58E-05	0.99991
Uneducated	Single	$60K - $80K	4.48E-05	0.99996
Graduate	Single	$80K - $120K	0.00030251	0.9997
Uneducated	Unknown	$120K +	0.00019094	0.99981
Unknown	Married	$40K - $60K	0.00019751	0.9998
College	Single	$80K - $120K	4.78E-05	0.99995
Graduate	Unknown	$60K - $80K	9.61E-05	0.9999
Graduate	Married	Less than $40K	0.00011382	0.99989
Unknown	Unknown	$80K - $120K	6.35E-05	0.99994
Post-Graduate	Single	$80K - $120K	0.00023623	0.99976
Unknown	Married	$80K - $120K	0.00014953	0.99985
High School	Married	$40K - $60K	0.00017468	0.99983
Graduate	Married	Unknown	5.51E-05	0.99994

TABLEAU STORYBOARD SUMMARY

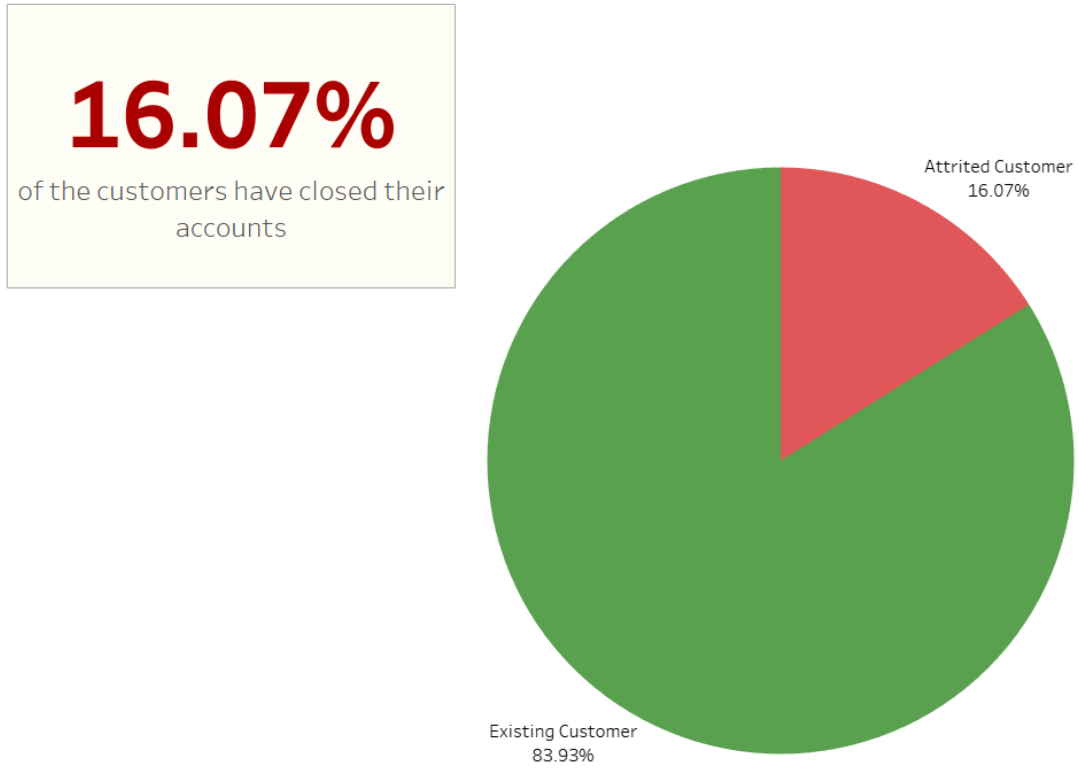
Below are the Contents for each Storyboard in Tableau:

1. Overview
2. Treemap
3. Analysis of Customer Dataset
4. Attrition Rate on Different Measures

Tableau Storyboard Link

STORYBOARD 1: OVERVIEW

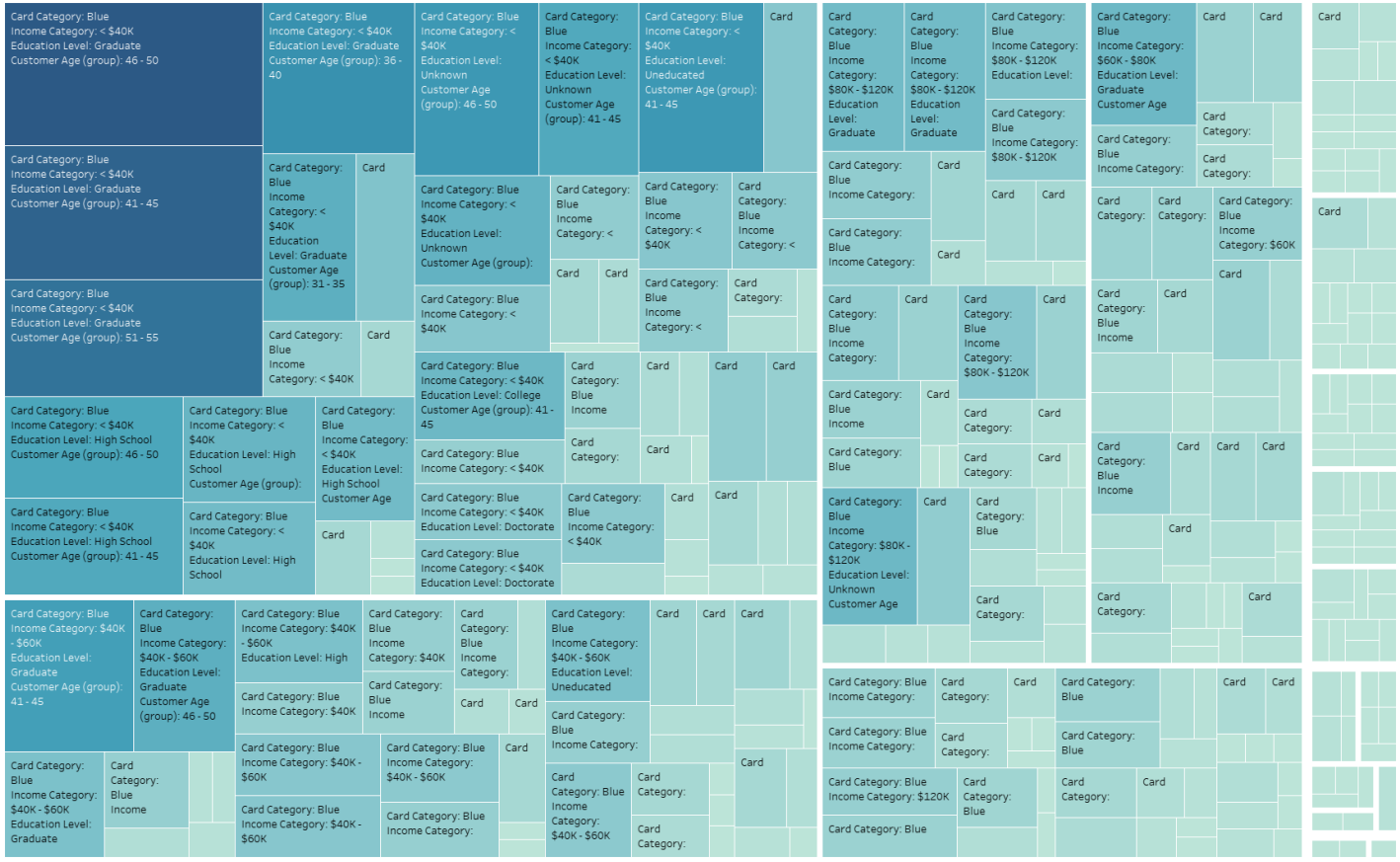
- In this dataset, we observe that **16.07%** of total customers have attritted over the past 12 month.
- In the next storyboard, a treemap is charted for a closer look into the category of customers that make up for the majority of attritted customers.



Storyboard 1: Overview

STORYBOARD 2: TREEMAP

- The treemap shown *only reflects the profile of the attritted customers.*
- The size + deepness of blue shade on the tile represents the largest demographic of attritted customers.
- From this treemap, we can infer that the majority of attritted customers are:
  1. Blue card holders
  2. Fall in the income category of <$40K
  3. Have a Graduate Degree
  4. Are middle-aged [Ages 41-55]



Storyboard 2: Treemap

STORYBOARD 3: ANALYSIS OF CUSTOMER DATASET [1/6]

- This dashboard shows a overview of the customer dataset. It contains an interactive pie chart in which users can interact to perform a comparison between attritted & existing customers.

- The distribution of attritted customers based on the below variables were examined:
  - Customer Age
  - Card Inactivity in Past 12 Months
  - Relationship with bank in Months

- A regression analysis (along with its accompanying boxplot) between Total Revolving Balance & Credit Limit on the different Income Categories is also performed in this dashboard.

Storyboard 3: Analysis of Customer Dataset

STORYBOARD 3: ANALYSIS OF CUSTOMER DATASET [2/6]

The dashboard was filtered to view solely attrite customers.

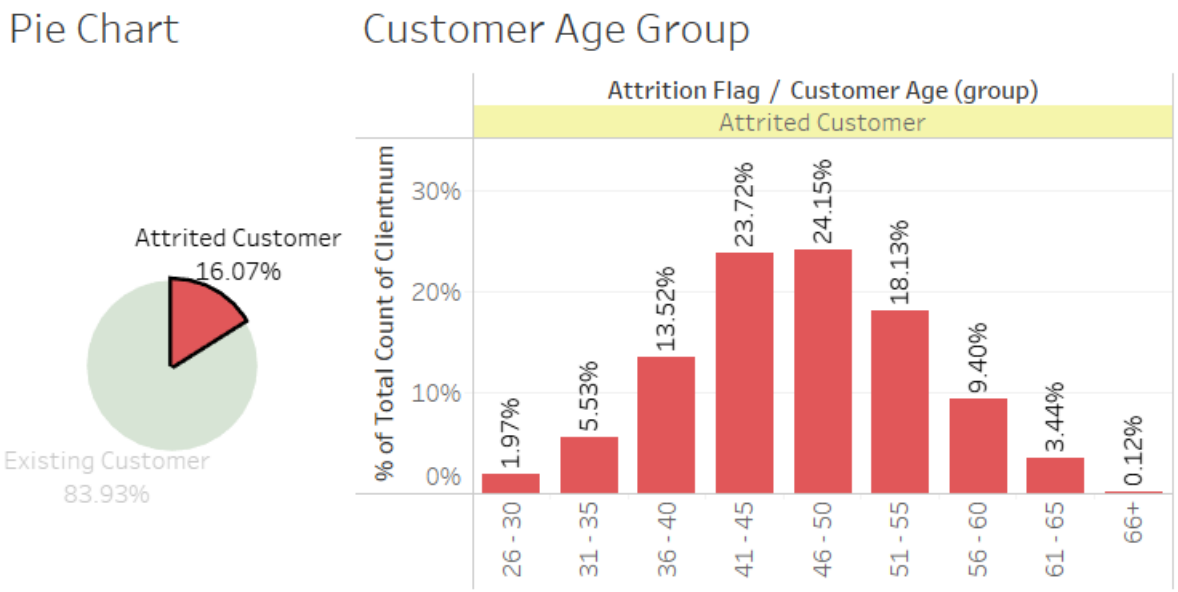
Observation:

Similar to what was previously shown in the treemap, a significant portion of attrite customers are between the ages 41-55.

Recommendation:

Based on the results, it seems that customers between the ages of 41 to 55 are not drawn to the products being offered.

As a result, it is recommended that the company examine its products and consider redesigning them to better appeal to this age group. Further surveys regarding product dissatisfaction could be conducted on this age group for better customer retention.



Storyboard 3: Analysis of Customer Dataset

STORYBOARD 3: ANALYSIS OF CUSTOMER DATASET [3/6]

The dashboard was filtered to view solely attrite customers.

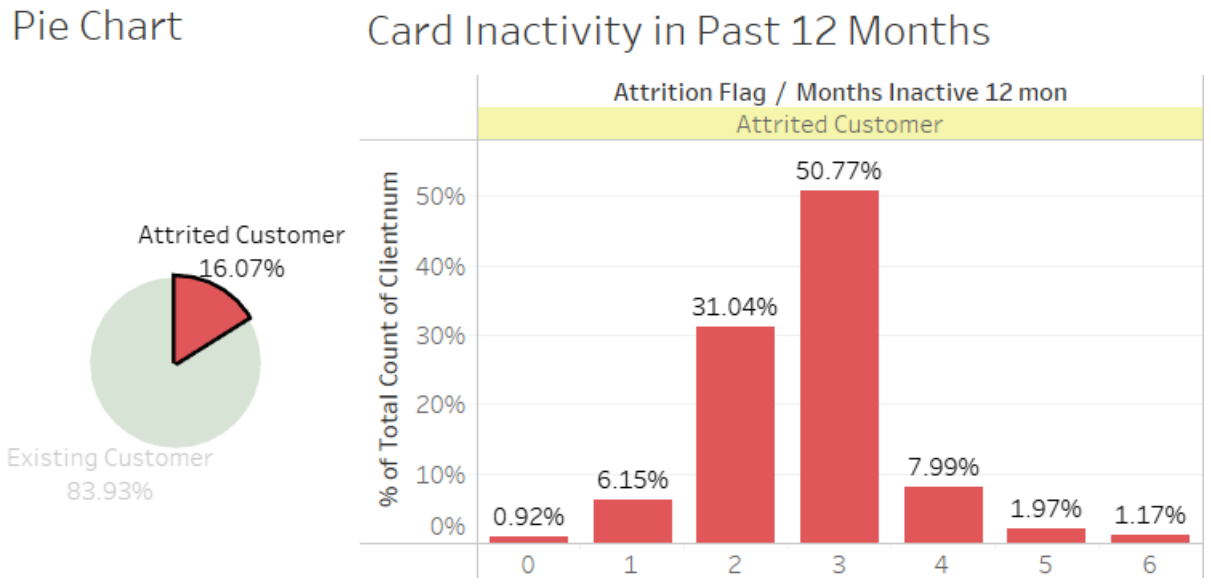
Observation:

A significant portion of attritted customers (>80% of attritted customers) began to attrite more aggressively at the 2nd month mark of card inactivity, then finally peaking at the 3rd month of card inactivity.

Recommendation:

It is recommended for the company to focus more resources on reaching out to such customers & to let them know of any credit card promotions to keep them engaged.

This can be achieved through regular engagement with such customers to inform them about incentives such as cashback, reward points or discounts for credit card usage.



Storyboard 3: Analysis of Customer Dataset

STORYBOARD 3: ANALYSIS OF CUSTOMER DATASET [4/6]

The dashboard was filtered to view solely attrite customers.

Observation:

The highest % of attritted customers come from customers who've stayed with the bank for 3 years [36 months].

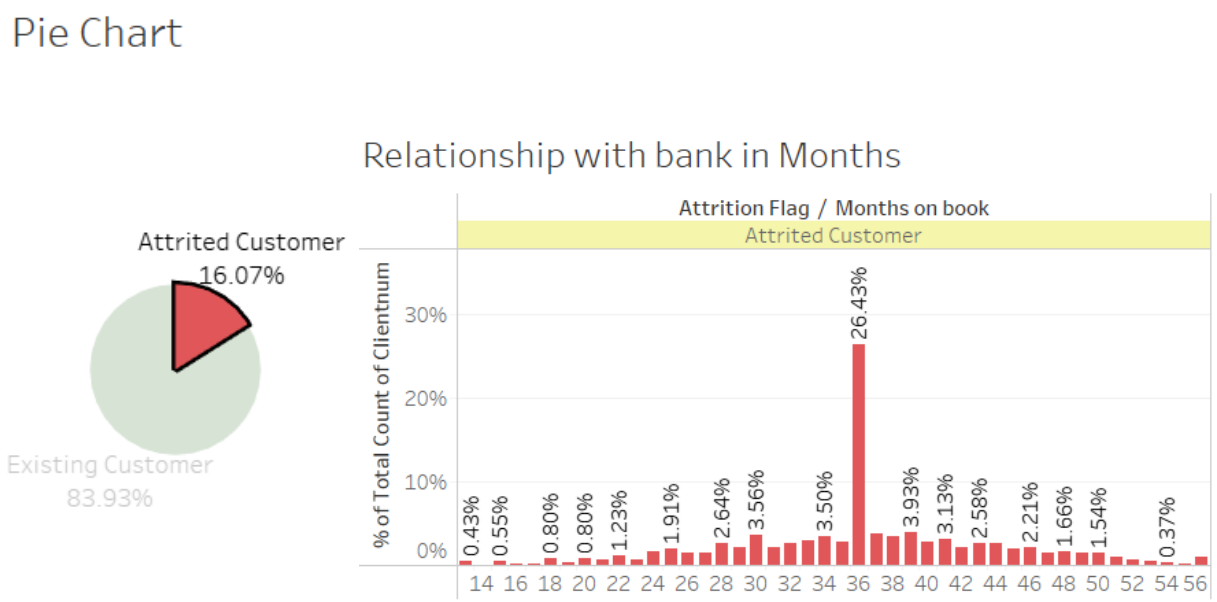
Recommendation:

It is recommended for the company to offer more exclusive benefits for long-time customers such as annual fee wavers or an enhanced rewards program for long-time customers.

Limitations of dataset:

Dataset does not contain a date as a continuous variable, hence it might be quite difficult to research if the customers leaving at the 3 year mark coincides with an economic downturn.

Further research can be done as to why a huge % of customer attrition occurs at the 3 year mark.



Storyboard 3: Analysis of Customer Dataset

STORYBOARD 3: ANALYSIS OF CUSTOMER DATASET [5/6]

<u>Definition of Measures</u>

Total revolving balance:

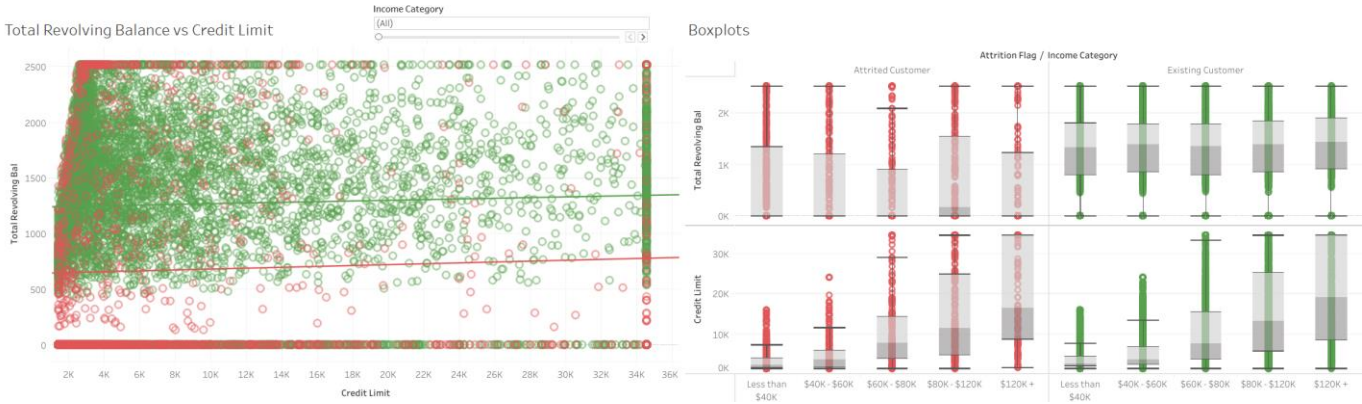
- Refers to the amount of money you owe on your credit card(s) that you haven't yet paid off in full.

Credit limit:

- The maximum amount of credit that you have been approved for by the credit card company. In simple terms, it's the maximum amount of money that a customer can spend using their credit card.

- Recalling that earlier on in the treemap plot, the income category with the highest attrition is in the <$40K income category range.

- An exploration into the total revolving balance of this income group was performed along with corresponding boxplots for 'Total Revolving Bal' & 'Credit Limit', which will be useful for determining the spread of the scatter plot.



Storyboard 3: Analysis of Customer Dataset

STORYBOARD 3: ANALYSIS OF CUSTOMER DATASET [6/6]

Observation:

Across all income groups, we noticed that there were individuals with no revolving balance who has attritted as well. From the corresponding boxplot, we also see that the median is "0" for most income groups.

This means that customers across all income groups with low revolving balance have a high attrition rate as compared to existing customers. This means that such customers are considered to be financially responsible and capable of managing their debt effectively.

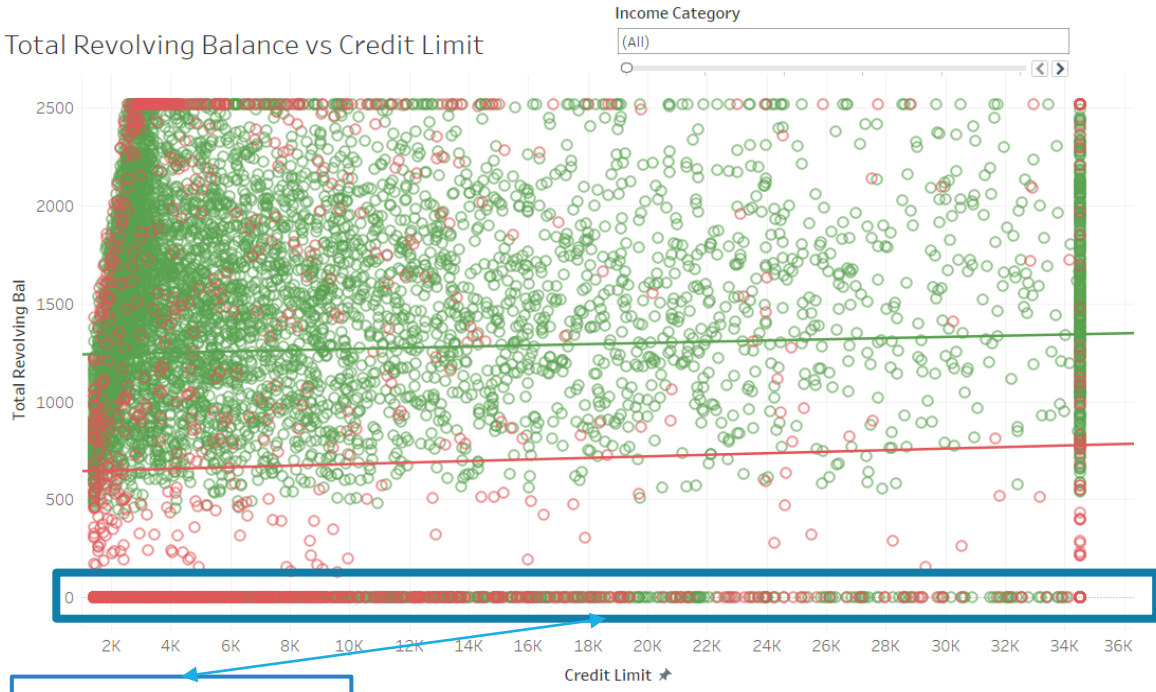
These are customers who chose to spend sparingly within their credit limits.
This demonstrates that they are not dependent on credit and have the means to pay off their balances in full every month, which reduces the credit risk for the credit card company.

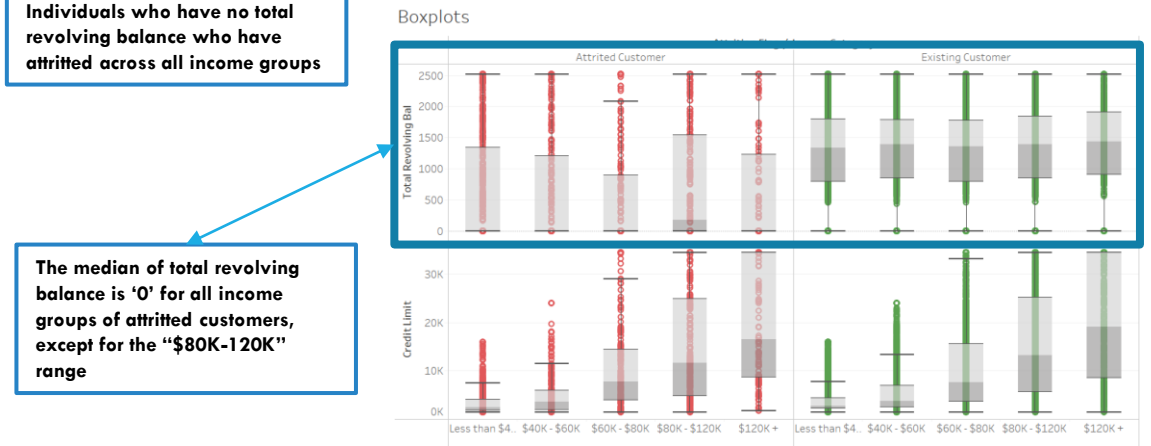
Recommendation:

For such credit card holders, the bank could look into methods to encourage such individuals to stay with the bank such as offering them a more premium credit card to further incentivize them to continue using their card with the bank.

Limitation:

There may be other underlying factors not shown here that resulting in such customers to attrite. Further analysis is required to give a definite suggestion.





STORYBOARD 4: ATTRITION RATES ON DIFFERENT MEASURES

- In this dashboard, the impact of attrition rates on various qualitative & quantitative measures will be determined via the dynamic toggling filter.
- 2 variables (Income group category & Card category) have been fixed as these were previously established in the treemap plot to be important variables linked to customer attrition.
- The corresponding line chart (labelled as plot) will serve as a visual aid for correlation between the different measures.

Storyboard 4: Attrition Rates on Different Measures

STORYBOARD 4: ATTRITION RATES ON DIFFERENT QUALITATIVE MEASURES

Observation:

When gender was selected via the toggle, we notice that:

1. The <$40K income group contains the highest number of existing & attritted credit card holders.

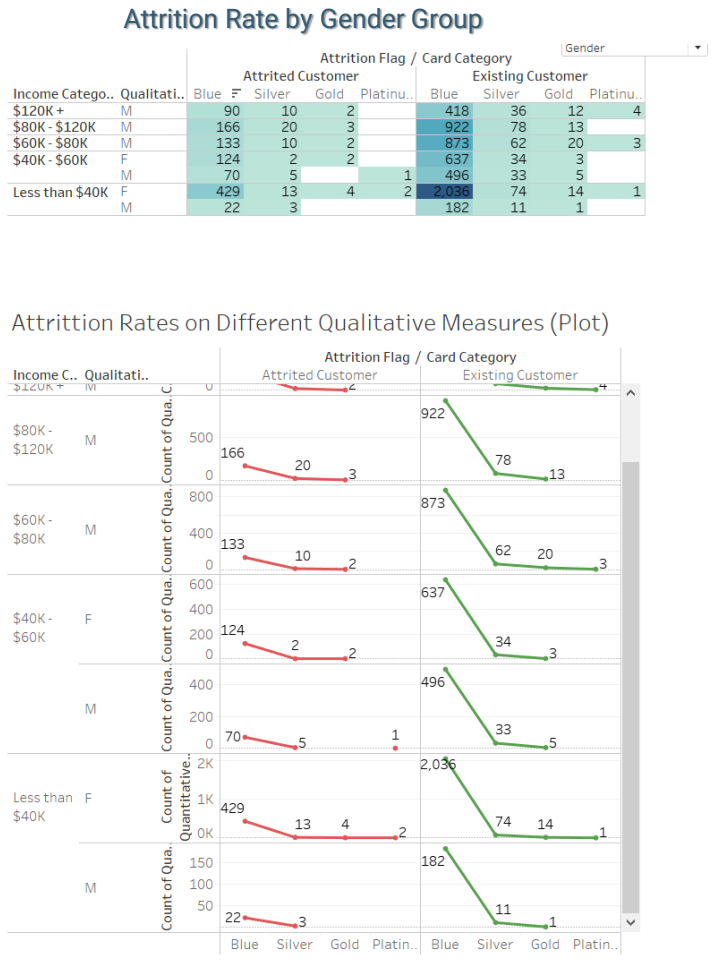
2. Female credit card holders in this income range accounts for the highest # of attrition.

Recommendation:

The bank could look into the blue card category and check if there are any credit cards catered towards women.

- If such credit cards already exists, a survey could be conducted to check if the current benefits provided are attractive to women.

- If the credit card does not exist, the bank could look into the creation of a woman-centered credit card.

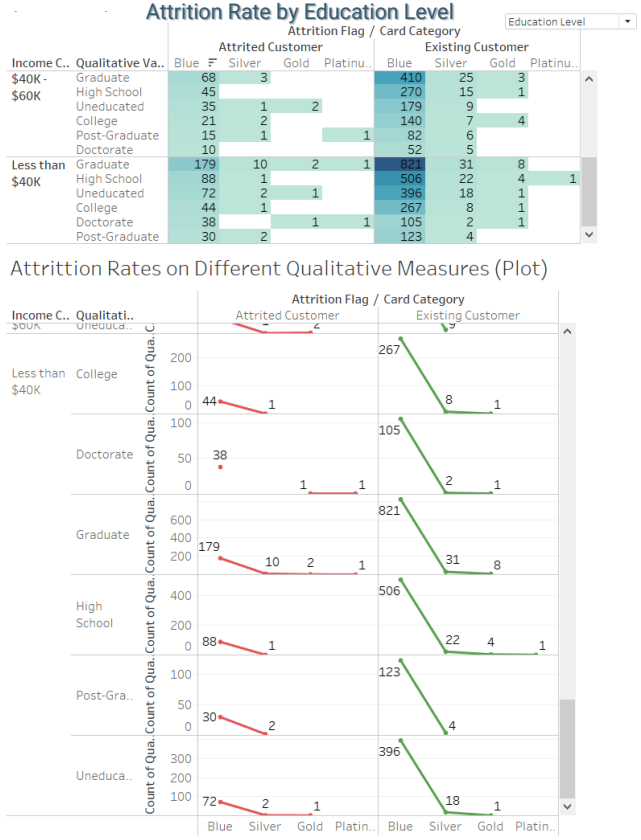


Storyboard 4: Attrition Rates on Different Measures

STORYBOARD 4: ATTRITION RATES ON DIFFERENT QUALITATIVE MEASURES

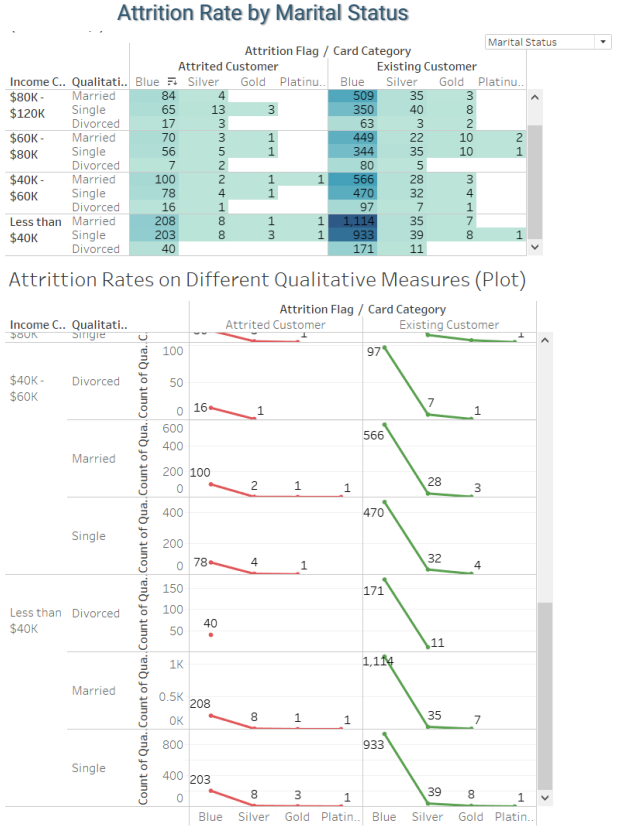
Observation:

In agreement with the findings of the treemap plot earlier, the group with the highest attrition comes from the <$40K income category holding graduate degree.



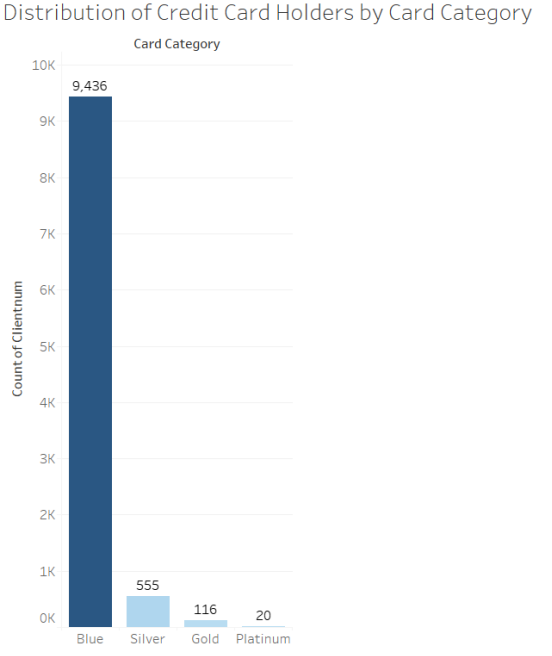
Observation:

We observe that the highest # of attritted customers come from the lowest income group from customers that are not divorced.



Conclusion:

For all 3 qualitative measures, it's difficult to observe significant trends of the impact on attrition rates based on credit card category as the dataset mainly consist of blue credit card holders.

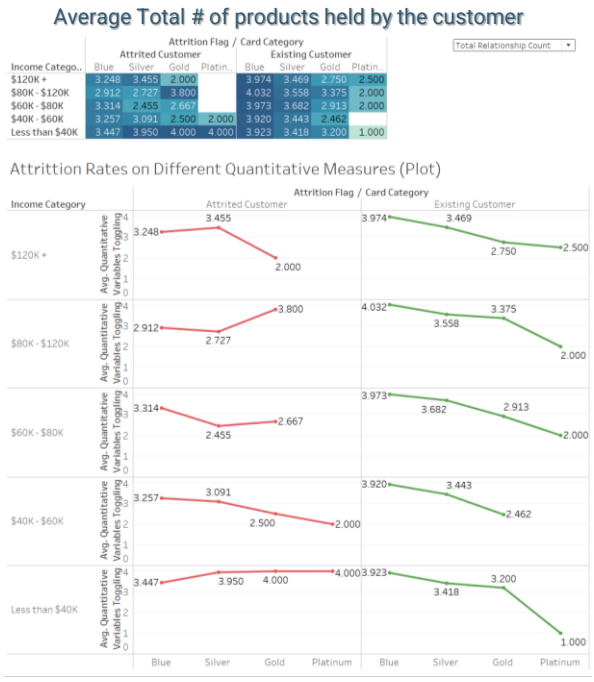


Storyboard 4: Attrition Rates on Different Measures

STORYBOARD 4: ATTRITION RATES ON DIFFERENT QUANTITATIVE MEASURES

Observation:

No clear impact on attrition rate can be determined based on the average total # of products that are being held by the customer.

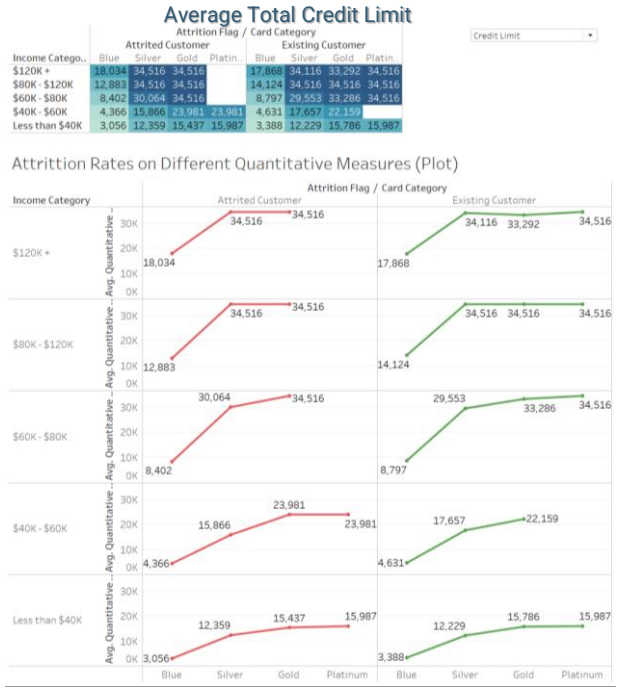


Storyboard 4: Attrition Rates on Different Measures

Observation:

We observe that there is generally a positive correlation between average total credit limit & increasing card tier list for all income groups.

The average credit card limit is higher for individuals owning a more premium card within the same income group. This may be because such credit card holders pay higher annual fees.



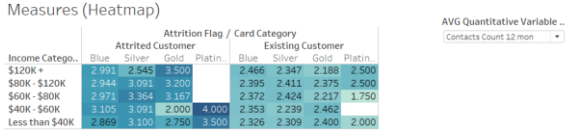
STORYBOARD 4: ATTRITION RATES ON DIFFERENT QUANTITATIVE MEASURES

Observation:

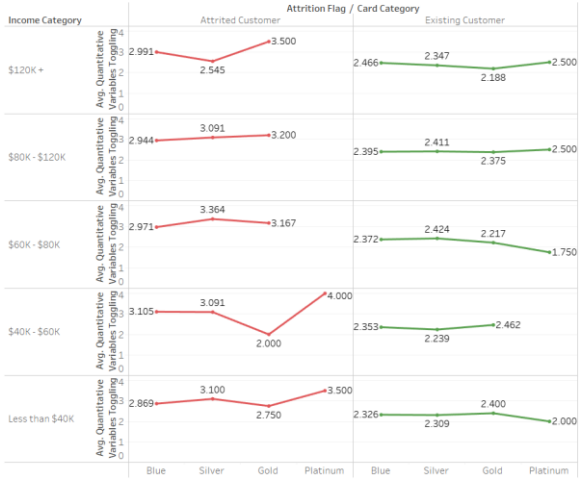
For attritting customers across all income groups, we notice that the average # of contact made to the customers in the past 12 months increases as the tier of the credit card increases. This shows good initiative from the bank as they are proactive in reaching out to loyal customers who own higher tier credit cards.

Average Total # of Contacts made to the Customer (past 12 months)

Attrition Rates on Different Quantitative Measures (Heatmap)



Attrition Rates on Different Quantitative Measures (Plot)



However, as seen from the boxplot in the <$40K income group, the average utilization ratio for platinum card holders are the least compared to all other credit card holders.

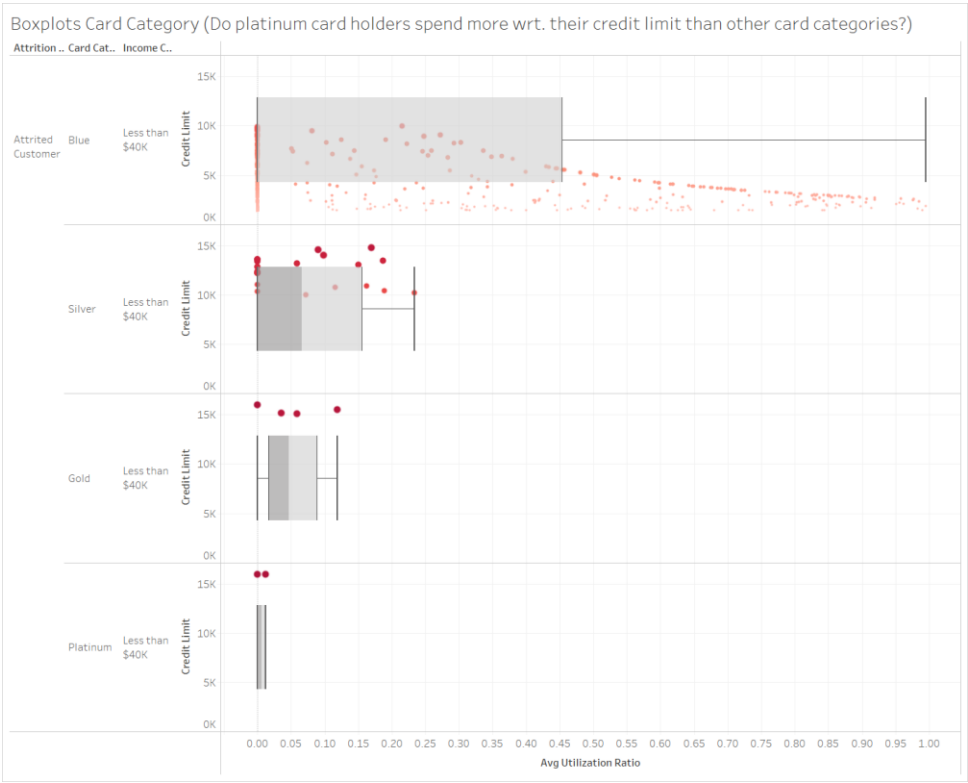
One may attribute this reason to be due to the high credit limit that is given to platinum card holders.

Recommendation:

It was previously established in the treemap plot that blue card holders in the <$40K income group have the highest attrition count.

Hence, more emphasis should be placed on contacting these category of customers particularly before their card becomes inactive for 2 months, as that is typically when customers start to attrite.

Boxplots Card Category (Do platinum card holders spend more wrt. their credit limit than other card categories?)



Storyboard 4: Attrition Rates on Different Measures

LIMITATIONS OF DATASET

1. Dataset only contains information about existing/attritted customers for the past one year. It's preferred to have a dataset across a longer time period for a more conclusive report.
2. The dataset provided may not offer a comprehensive understanding of the reasons why customers are leaving the bank. For instance, external elements such as alterations in the local economy or the emergence of new competitors in the market may not be recorded in the bank's data.

END OF PRESENTATION