Data Visualization and Business Intelligence

Project report submitted to

Jamia Millia Islamia

New Delhi

In partial fulfillment of the requirements of the award of the Degree of

Master of Science in Banking and Financial Analytics

By

Saba Akram

Under the supervision of

External Supervisor

Nikhil Gupta

Branch Manager
Indian Bank, Noida Sector-61

Internal Supervisor

Dr. Saif Siddique

Associate Professor
Jamia Millia Islamia, New Delhi

Department of Economics

Jamia Millia Islamia

New Delhi

Submitted in June 2021

DECLARATION

I Saba Akram, student of Master of Science in Banking and Financial Analytics hereby declare that the thesis titled ***"Data Visualization and Business Intelligence"*** which is submitted by me to the Department of Economics, Jamia Millia Islamia, New Delhi, in partial fulfillment of the requirement for the award of the degree of Master of Science in Banking and Financial Analytics has not previously formed the basis for the award of any Degree, Diploma, Associateship, Fellowship or other similar title or recognition.

Place and Date

Saba Akram

CERTIFICATE

On the basis of declaration submitted by Saba Akram, student of Master of Science in Banking and Financial Analytics, I hereby certify that the thesis titled ***"Data Visualization and Business Intelligence"*** which is submitted in the department of Economics, Jamia Millia Islamia, New Delhi in partial fulfillment of the requirement for the award of the Degree of Master of Science in Bankiing and Financial Analytics is an original contribution with existing knowledge and faithful record of research carried out by them under my supervision and guidance.

To the best oof my knowledge this work has not been submitted in part or full for any Degree or Diploma to this University or elsewhere and that is free from plagiarism.

Signature of External Supervisor

Signature of Internal Supervisor

Countersigned by

Principal/Dean of the Faculty

ACKNOWLEDGEMENT

Through this Project Report I would like to thank both our Internal Supervisor ***Dr. Saif Siddique***, Associate Professor, Jamia Millia Islamia, New Delhi and our External Supervisor ***Nikhil Gupta***, Branch Manager, Indian Bank, Noida Sector-61, who has guided us a lot during the tenure of our research.

I would also like to thank our professors who were always available for us during the course of our research and helped a lot which proved very effective.

I would also like to thank our Head of Department, Dr. Halima Sadia Rizvi without the help of whom our research would be very hard to achieve. Her motivation played a key role in completing our research.

Thank You

Saba Akram

TABLE OF CONTENTS

1. INTRODUCTION

"A picture is worth a thousand words". We are all familiar with this expression. It especially applies when trying to explain the insight obtained from the analysis of increasingly large datasets. Data visualization plays an essential role in the representation of both small and large-scale data. One of the key skills of a data scientist is the ability to tell a compelling story, visualizing data and findings in an approachable and stimulating way. Learning how to leverage a software tool to visualize data will also enable you to extract information, better understand the data, and make more effective decisions.

Data visualization has been heavily utilized in the banking industry for years and has been inseparable part of the analytical processes that take place in the risk and reporting departments. Recently, the banking world is increasingly using Data Visualization and Business Intelligence that allow export, processing, modelling and analysis of large databases. Not only in the banking sector but almost in every business, the ability to easily drill down into the data enhances the operational efficiency of the business. Data Visualization brings Business Intelligence to life. It provides a new and more creative approach that improves drastically our ability to understand the information hiding in it, in a more constructive way. It also provides a better and faster way to identify patterns, trends, and correlation in the data sets that would remain undetected with a text or figure-based nature.

Like most industries, we as a banking and financial Analytics professionals love reports. As the scale of the organization grows, more reports are produced with increased levels of data points being considered and bigger amounts on the line. Data visualization can merge complex data sets and create actionable insights for a specified department, or the organization as a whole. Here we can sort out few areas where we can use data visualizations in banking and financial institutions where we can enhance the client experience, and save time/money along the way:

Client Portfolios/Statements – We should always be client focused, and there is no better way than to help our clients visualize their relationship rather than read about it. As more people access their account digitally (website, mobile app, etc), there is a need to provide a more dynamic visual for the client. When their statement is just static templates and reports with their name on it, we miss an opportunity to provide a differentiating experience. Based on conversations with the client, statements or annual reports can have specific visualizations customized for them.

Sales – Banks and Financial Institutions have considerable amounts of data on their clients/members. When an organization can harness that information and apply models to the data, it is easier to find which prospects or current customers need to be marketed to for specific product promotions.

Data Team – Mapping the data journey will be crucial to proper storage and governance of the invaluable data. As teams start to add roles like Chief Data Officers, Data Architects & Data Analysts, it is imperative for IT and data dependent teams to get aligned. The end users want to make sure they know where their information is coming from, and IT wants to ensure the security of their network and platform. One visualization showing all data sources (internal and external) ensures everyone is on the same page.

While working in the Indian Bank as an Intern I was given a topic to do a research on. "Data Visualization and Business Intelligence" was chosen as my topic. Data was arranged from the Kaggle.com. The dataset belonged to a bank which is facing a problem that the customers is leaving its Credit Card Service. The Dataset contained many variables along with the Attrition Flag as the target variable. The predictor variables are Client Number, Customer Age, Income Category, Education Level, Dependent Count, Gender, Marital Status, Card Category, Total Revolving Balance, Total Transfer Amount, Total Transaction Count, Average Open to Buy, Month on the book, Total Relationship Count, Months Inactive in the last 12 months, Contacts Count in the last 12 months, Credit Limit, Total Amount Change Q4 by Q1, Total Count Change Q4 by Q1 and Average Utilization Ratio.

2. LITERATURE REVIEW

For every industry to grow at the highest possible level, the data has been the driving factor for many a sector and organization. A sector that vests on data without a question is banking and finance. Finance is the key precursor for the banks to boost their value and growth thereby speeding up the process of wealth generation. As a matter of fact, banking and finance are two sectors which overlap with analytics as they inherently deal with data.

It must be understood by the aspirants that the banking sector has become a commoditized marketplace with almost every bank offering similar products and services. To stay competitive, data analytics led strategy based on customer insights can be a key differentiator to staying competitive – building bank market share and profitability and enhancing customer experience.

Today in the world of Banking and Finance, which deals with millions and millions of data of the customers, it is a big and challenging task for these institutions to deal with the data, such as storing them in such a manner so that they can be easily assessed when needed. Accessing the data is not only the thing but it should be kept in such a way that it can be easily analyzed. The behavior of the customers can be analyzed and hence certain loopholes in the policies can be amended by reading those behavior and many more.

Data visualization is the process of acquiring, interpreting and comparing data in order to clearly understand complex ideas, thereby facilitating the identification and analysis of meaningful patterns. Data visualization can be essential to strategic communication. It helps us interpret available data; detect patterns, trends, and anomalies; make decisions; and analyze inherent processes. All told, it can have a powerful impact on the business world.

The ways we structure and visualize information are changing rapidly and getting more complex day to day. Thanks to the rise of social media, the ubiquity of mobile devices, and service digitalization, data is available on any human activity that utilizes technology. The generated information is hugely valuable and makes it possible to analyze trends and patterns, and to use big data to draw connections between events. Thus, data visualization can be an effective mechanism for presenting the end user with understandable information in real time.

Every company has data, be it to communicate with clients and senior managers or to help manage the organization itself. It is only through research and interpretation that this data can acquire meaning and be transformed into knowledge.

Importance of Data Visualization

We live in the era of visual information, and visual content plays an important role in every moment of our lives. A study by SHIFT Disruptive Learning demonstrated that **we typically process images 60,000 times faster than a table or a text**, and that our brains typically do a better job remembering them in the long term. That same research detected that after three days, analyzed subjects retained between 10% and 20% of written or spoken information, compared with 65% of visual information.

The rationale behind the power of visuals

The human mind can see an image for just **13 milliseconds** and store the information, provided that it is associated with a concept. Our eyes can take in **36,000 visual messages per hour. 40%** of nerve fibers are connected to the retina.

All of this indicates that human beings are better at processing visual information, which is lodged in our long-term memory.

Consequently, for reports and statements, a visual representation that uses images is a much more effective way to communicate information than text or a table. It also takes up much less space.

This means that data visuals are more attractive, simpler to take in, and easier to remember.

The graph takes what the numbers cannot communicate on their own and conveys it in a visible, memorable way. This is the real strength of data visualization.

Role of Data visualization

Explaining: Visuals aim to lead the viewer down a path in order to describe situations, answer questions, support decisions, communicate information, or solve specific problems. When we attempt to explain something through data visualization, we start with a question, which interacts with the data set in such a way that enables viewers to make a decision and, subsequently, answer the question.

Exploring: Some visuals are designed to lend a dataset spatial dimensions, or to offer numerous subsets of data in order to raise questions, find answers, and discover opportunities. When the goal of a visual is to explore, the viewers start by familiarizing themselves with the dataset, then identifying an area of interest, asking questions, exploring, and finding several solutions or answers.

Analyzing: Other visuals prompt viewers to inspect, distill, and transform the most significant information in a dataset so that they can discover something new or predict upcoming situations.

So here through this major industrial project I completed my project on “Data Visualization and Business Intelligence”. I have taken the banking dataset from the Kaggle.com. That bank is facing a serious problem because its customers are leaving its Credit Card Service. The dataset contains 23 predictor variables and a target variable. The target variable is the “Attrition Flag”. The attrition percentage is 16.059%. We have used bar graphs in order to visualize the relationship between the “Attrition Flag” with the categorical variables. We have used histograms in order to visualize the relationship between the “Attrition Flag” and continuous variables. We have used scatterplots to visualize the multivariate relationships between “Attrition Flag” and other predictor variables both categorical and numerical. We have also used tables in order to understand the relation more accurately.

3. RESEARCH METHODOLOGY

A. Objective of the Research

The goal of data visualizations is to help us understand the object they represent. They are a medium for communicating stories and the results of research, as well as a platform for analyzing and exploring data. Therefore, having a sound understanding of how to create data visualizations will help us create meaningful and easy-to-remember reports, infographics, and dashboards. Creating suitable visuals helps us solve problems and analyze a study's objects in greater detail.

The main Objectives of our Research are:

- To visualize the data more effectively, efficiently, elegantly, accurately as well as meaningfully communicating information in such a manner so that our eyes can recognize, and our brain can comprehend.
- To detect the patterns hidden inside the variable, so that the information becomes clear and unambiguous.
- To help frame policies which can be beneficial for the banks and the customers and that would ultimately lead to the proper development of that bank and will lead to the welfare of the customers.
- To help find the customers who are mainly not satisfied with the current Credit Card Service provided by the Bank, and they are leaving their service.
- To find the reasons behind the customers leaving the Credit Card Service facility.
- To visualize other patterns which may not be related to the main objective of the study but can help find the loopholes in the banks other policies.

B. Source of Data Collection

The data is collected using mainly by secondary sources. The methods used to collect secondary data are:

- Related books
- Research Papers
- Online Articles
- Websites

C. Tools

The tools used in this study includes R Software, MS-EXCEL, MS-WORD. The major part of the "Data Visualization Techniques" of this study is done using the R Software. Data is imported using M.S. Excel. Final report of this study is prepared using M.S. Word.

4. DATA VISUALIZATION

The customers of the Bank are leaving the credit card services in large numbers. Therefore, in our research project we are going to apply the Data Visualization and Business Intelligence techniques in order to find the respectable solution. Here we have imported the data below. Let's have a little look.

```
BankChurners <- read.csv("D:/Research Materials/Credit Card Churners/BankChurners.csv", stringsAsFactors=TRUE)
library(ggplot2)
```

We see that there are 23 columns and 10127 records of the customers. Let's take a look at the variables: -

- CLIENTNUM - This is client identity number. Each customer has a unique identity number.
- Attrition_Flag - This variable shows the status of the customer whether the customer is existing or have attrited(Churned).
- Customer_Age- Demographic Variable. Customer's age in years.
- Gender- Demographic Variable M=Male and F=Female.
- Dependent_count - Demographic Variables. Number of Dependents.
- Education_Level- Demographic Variable."Unknown, Uneducated, High School, Graduted, Post-Graduate, Doctorate and College".
- Marital_Status - Demographic Variable. "Unknown, Married, Single and Divorced".
- Income_Category - Demographic Variable. Annual Income category of the account holder. "Unknown, Less than $40k, $40k-$60k, $60k-$80k, $80k-$120k and $120+".
- Card_Category - Product variable. Type of Card - Blue, Silver, GOld and Platinum.
- Month_on_the_book - Period of relationship with the bank.
- Total_Relationship_count - Total number of products held by customer.
- Months_Inactive_12_mon - Numbers of months inactive in last 12 months.
- Contacts_count_12_mon - Number of contacts in the last 12 months.
- Credit_Limit - Credit limit on the credit card.

- Total_Revolving_Balance - 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.

There are two more variables which is not required. Therefore, Let's omit these last two columns.

```
BankChurners<-BankChurners[,-c(22:23)]
colnames(BankChurners)
```

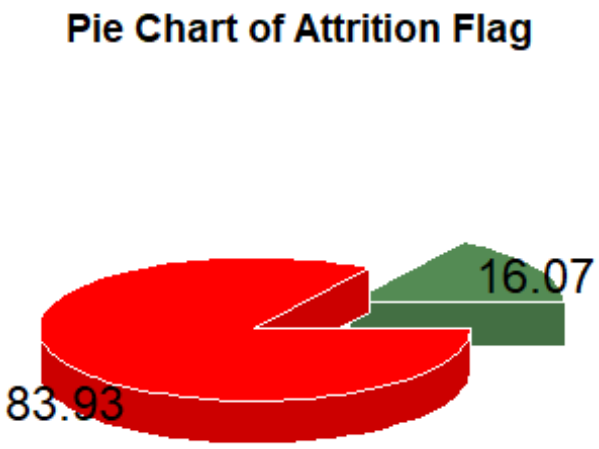
```
##  [1] "CLIENTNUM"                "Attrition_Flag"
##  [3] "Customer_Age"             "Gender"
##  [5] "Dependent_count"          "Education_Level"
##  [7] "Marital_Status"           "Income_Category"
##  [9] "Card_Category"            "Months_on_book"
## [11] "Total_Relationship_Count" "Months_Inactive_12_mon"
## [13] "Contacts_Count_12_mon"    "Credit_Limit"
## [15] "Total_Revolving_Bal"      "Avg_Open_To_Buy"
## [17] "Total_Amt_Chng_Q4_Q1"     "Total_Trans_Amt"
## [19] "Total_Trans_Ct"           "Total_Ct_Chng_Q4_Q1"
## [21] "Avg_Utilization_Ratio"
```

Now the data set contains 20 predictors worth of information of 10127 customers along with the target variable, "Attrition_Flag", an indication of whether the customer churned or not. We will try to visualize univariate relationships and then try to visualize multivariate relationships among the predictors and the target variable.

A. Target Variable - Attrition_Flag

```
# Counting the number of existing customers and attrited customers.
attrition_count<-table(BankChurners$Attrition_Flag)
attrition_count
```

```
##
## Attrited Customer Existing Customer
##              1627              8500

# Calculating the percentage of the existing customers and the attrited custo
mers.
attrition_percent<-round(prop.table(attrition_count)*100,2)

# Showing it through the pie chart

library(plotrix)
pie3D(attrition_percent,col = c("palegreen4","red"),labels = attrition_percen
t, explode = 0.25,labelcol = "black", border = "white",main="Pie Chart of Att
rition Flag")
```



After observing from the above findings, we get that 1627 customer out of 10127 customers which equals 16.07% of the customers attrited(churned) the credit card service provided by the bank.

One of the primary reasons for applying “Data Visualization and Business Intelligence” is to investigate the variables, examine the distributions of the categorical variables, look at the histograms of the numeric variables, and explore the relationships among sets of variables. In this way, we can visualize the data while keeping an eye on our overall goal. We begin by considering the categorical variables, and their relationship to the attrition.

B. EXPLORING CATEGORICAL VARABLES

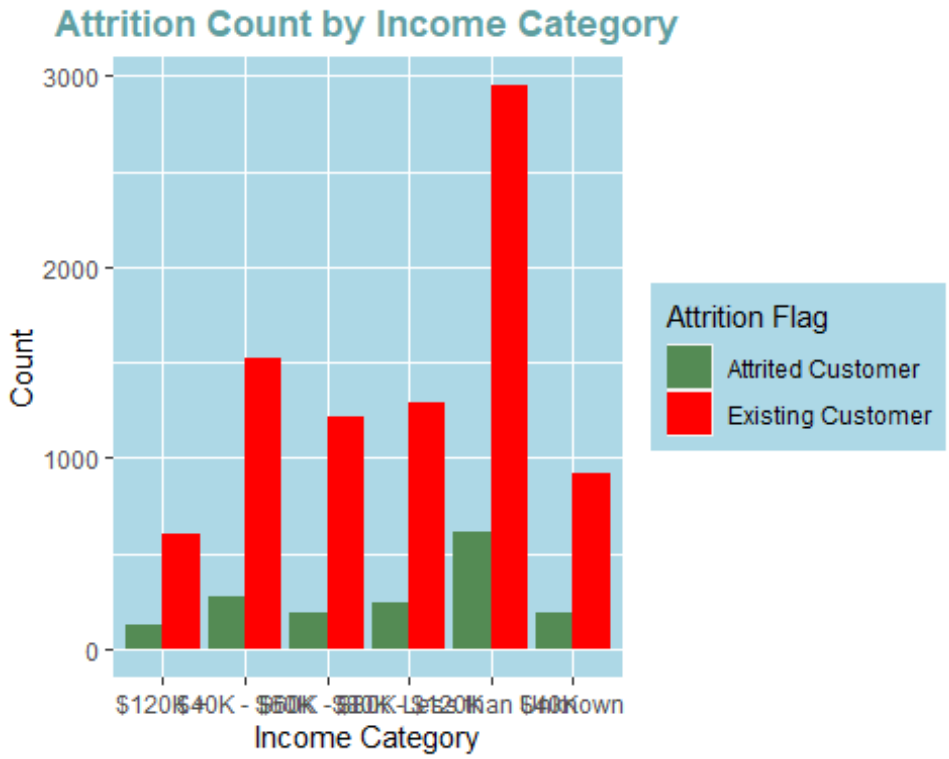
Here we have Income_Category, Card_category, Education_Level, Marital_Status and Gender as the categorical variables and we will look at them one by one by visualizing them.

Attrition_Flag - Income_Category

```
counts<-table(BankChurners$Attrition_Flag,BankChurners$Income_Category,dnn =
c("Attrition_Flag","Income_Category"))
counts
```

```
##                     Income_Category
## Attrition_Flag      $120K + $40K - $60K $60K - $80K $80K - $120K Less than
$40K
##   Attrited Customer     126         271         189          242
612
##   Existing Customer     601        1519        1213         1293
2949
##                     Income_Category
## Attrition_Flag      Unknown
##   Attrited Customer     187
##   Existing Customer     925
```

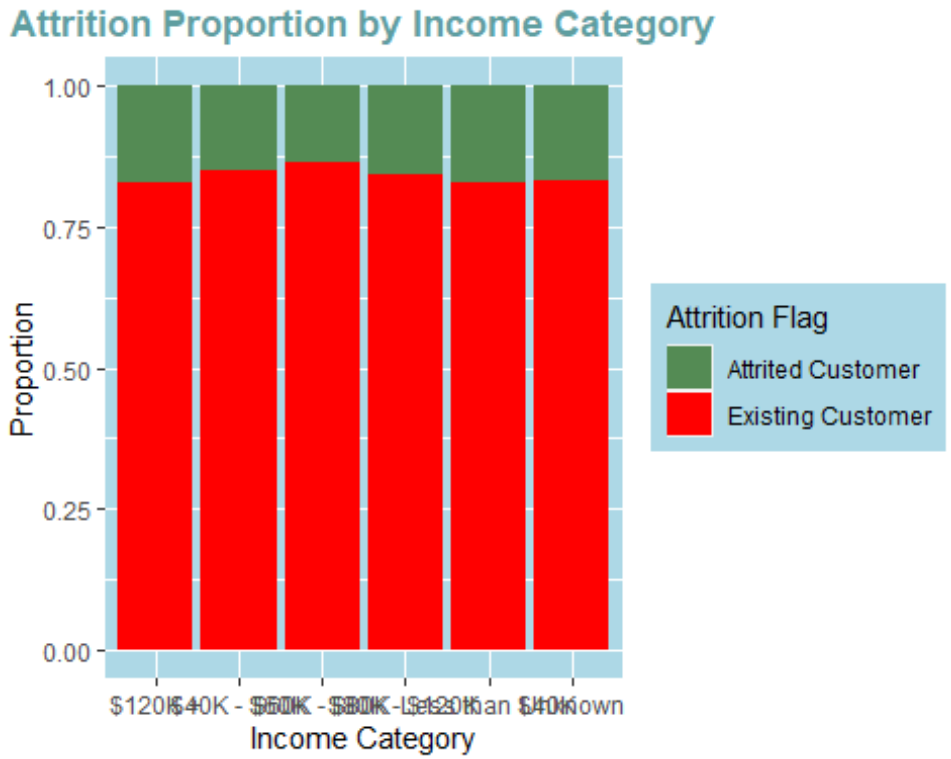
```
ggplot(data = BankChurners,aes(x=Income_Category,fill=Attrition_Flag))+geom_b
ar(position = "dodge")+labs(title = "Attrition Count by Income Category",x="I
ncome Category",fill="Attrition Flag")+theme(panel.background = element_rect(
fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold
",colour = "cadetblue"))+scale_x_discrete("Income Category")+scale_y_continuo
us("Count")+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.back
ground = element_rect(fill = "lightblue"))
```




```
# Income_Category wise proportion
row.margin<-round(prop.table(counts,margin = 1),4)*100
row.margin

##                    Income_Category
## Attrition_Flag      $120K + $40K - $60K $60K - $80K $80K - $120K Less than
$40K
##   Attrited Customer    7.74       16.66       11.62        14.87
37.62
##   Existing Customer    7.07       17.87       14.27        15.21
34.69
##                    Income_Category
## Attrition_Flag      Unknown
##   Attrited Customer   11.49
##   Existing Customer   10.88

#Attrition_Flag wise proportion
col.margin<-round(prop.table(counts,margin = 2),4)*100
col.margin

##                    Income_Category
## Attrition_Flag      $120K + $40K - $60K $60K - $80K $80K - $120K Less than
$40K
##   Attrited Customer   17.33       15.14       13.48        15.77
17.19
##   Existing Customer   82.67       84.86       86.52        84.23
82.81
```

```
##                     Income_Category
## Attrition_Flag       Unknown
##   Attrited Customer   16.82
##   Existing Customer   83.18
```

```
ggplot(data = BankChurners,aes(x=Income_Category,fill=Attrition_Flag))+geom_bar(position = "fill")+scale_x_discrete("Income Category")+scale_y_continuous("Proportion")+labs(title = "Attrition Proportion by Income Category",x="Income Category",fill="Attrition Flag")+theme(panel.background = element_rect(fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.background = element_rect(fill = "lightblue"))
```

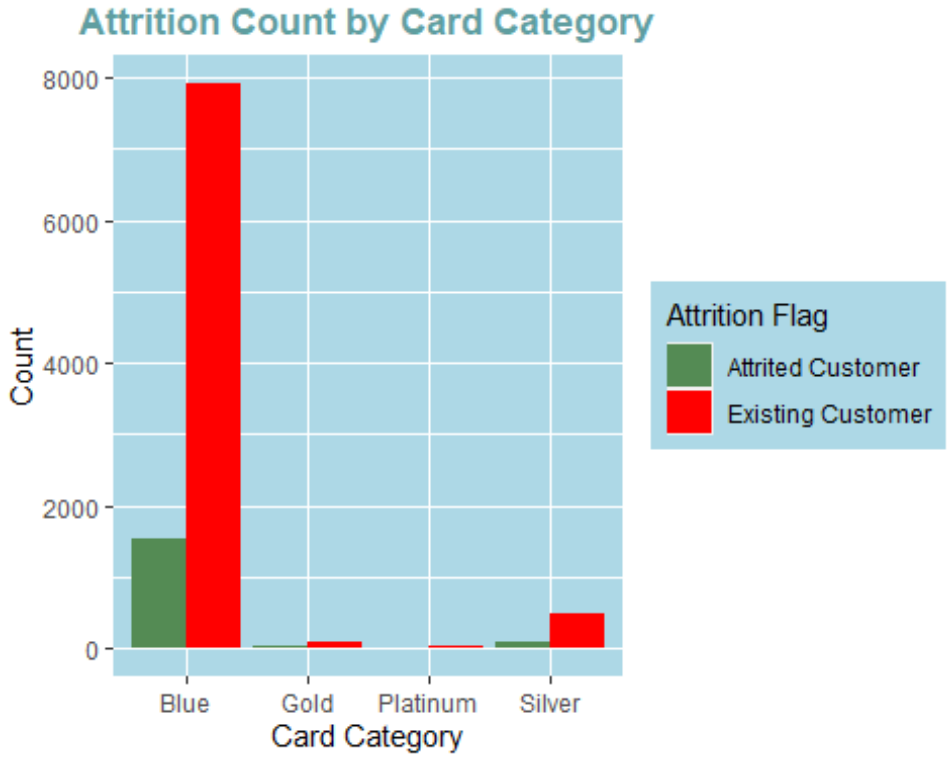


- From the above graphs and tables, we observe that the bank has maximum number of customers from low income category and goes on decreasing with the increase in their incomes.
- We see that the customers with the lowest and the most upper category income category, have maximum attrition.
- Our more focus should be on both types of customers but the customers with the lowest income category should be bank's main focus because it has maximum number of customers.

- The other income categories have attrition rate which are lower than the overall attrition rate.

Attrition_Flag - Card_Category

```
counts<-table(BankChurners$Attrition_Flag,BankChurners$Card_Category,dnn = c(
"Attrition_Flag","Card_Category"))
counts

##                     Card_Category
## Attrition_Flag      Blue Gold Platinum Silver
##   Attrited Customer 1519   21        5     82
##   Existing Customer 7917   95       15    473

ggplot(data = BankChurners,aes(x=Card_Category,fill=Attrition_Flag))+geom_bar
(position = "dodge")+labs(title = "Attrition Count by Card Category",x="Card
Category",fill="Attrition Flag")+theme(panel.background = element_rect(fill =
"lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colou
r = "cadetblue"))+scale_x_discrete("Card Category")+scale_y_continuous("Count
")+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.background =
element_rect(fill = "lightblue"))
```



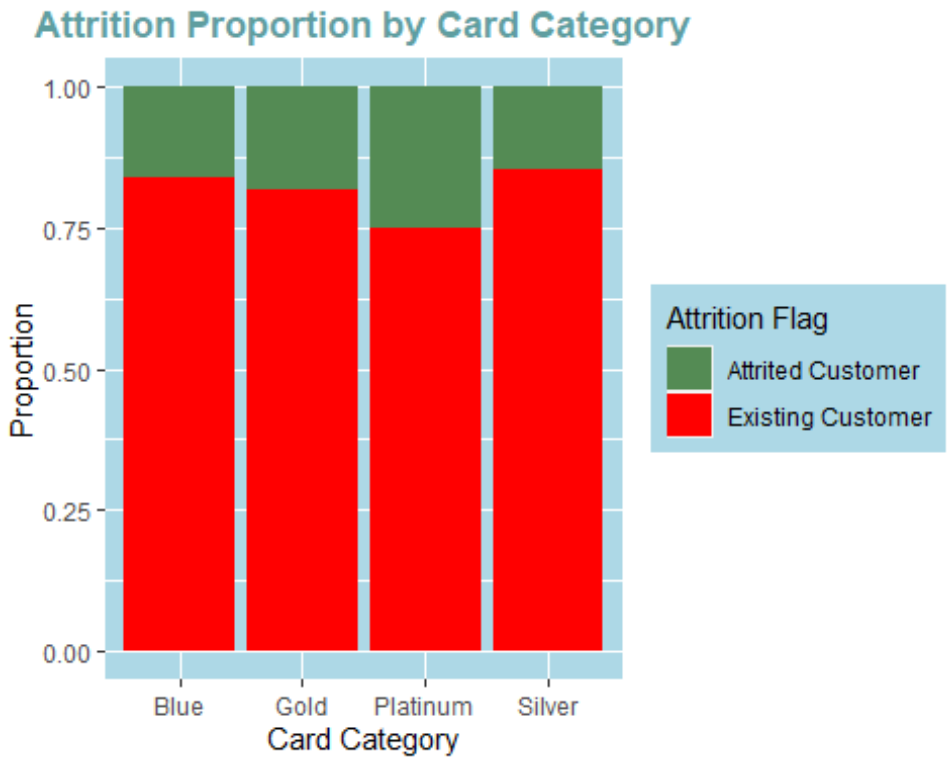
```
# Card_Category wise proportion
row.margin<-round(prop.table(counts,margin = 1),4)*100
row.margin
```

```
##                      Card_Category
## Attrition_Flag         Blue  Gold Platinum Silver
##   Attrited Customer 93.36  1.29     0.31   5.04
##   Existing Customer 93.14  1.12     0.18   5.56
```

```
#Attrition_Flag wise proportion
col.margin<-round(prop.table(counts,margin = 2),4)*100
col.margin
```

```
##                      Card_Category
## Attrition_Flag         Blue  Gold Platinum Silver
##   Attrited Customer 16.10 18.10    25.00  14.77
##   Existing Customer 83.90 81.90    75.00  85.23
```

```
ggplot(data = BankChurners,aes(x=Card_Category,fill=Attrition_Flag))+geom_bar
(position = "fill")+scale_x_discrete("Card Category")+scale_y_continuous("Pro
portion")+labs(title = "Attrition Proportion by Card Category",x="Card Catego
ry",fill="Attrition Flag")+theme(panel.background = element_rect(fill = "ligh
tblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "
cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.bac
kground = element_rect(fill = "lightblue"))
```



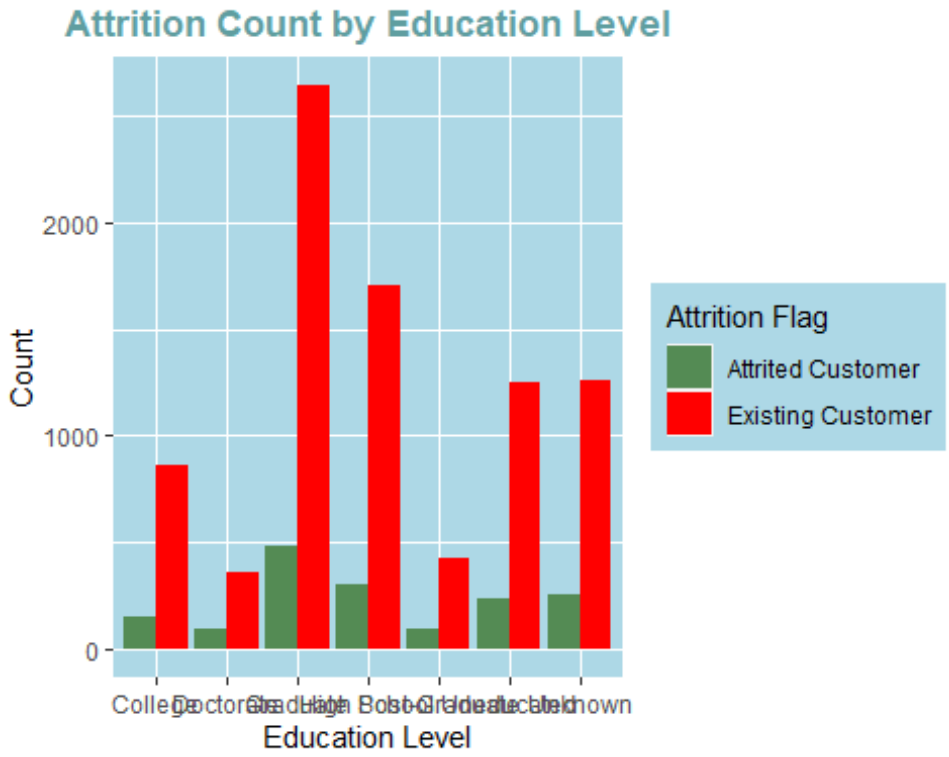
- From the above graphs and tables, we observe that the customers use blue card the most more than 93% followed by the silver card.
- With the minimum usage of the platinum and gold cards, the bank has maximum attrition rate in platinum (25%) followed by the gold card (18.1%). The bank should take this conclusion very seriously. The Bank may provide various incentives to customers to motivate them towards using the Platinum and Gold Cards.
- The Silver card customers consists of only 5%. The use of Silver Card should also be promoted.

Attrition_Flag - Education_Level

```
counts<-table(BankChurners$Attrition_Flag,BankChurners$Education_Level,dnn =
c("Attrition_Flag","Education_Level"))
counts
```

```
##                    Education_Level
## Attrition_Flag      College Doctorate Graduate High School Post-Graduate
##   Attrited Customer     154        95      487         306            92
##   Existing Customer     859       356     2641        1707           424
##                    Education_Level
## Attrition_Flag      Uneducated Unknown
##   Attrited Customer        237     256
##   Existing Customer       1250    1263
```

```
ggplot(data = BankChurners,aes(x=Education_Level,fill=Attrition_Flag))+geom_b
ar(position = "dodge")+labs(title = "Attrition Count by Education Level",x="E
ducation Level",fill="Attrition Flag")+theme(panel.background = element_rect(
fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold
",colour = "cadetblue"))+scale_x_discrete("Education Level")+scale_y_continuo
us("Count")+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.back
ground = element_rect(fill = "lightblue"))
```



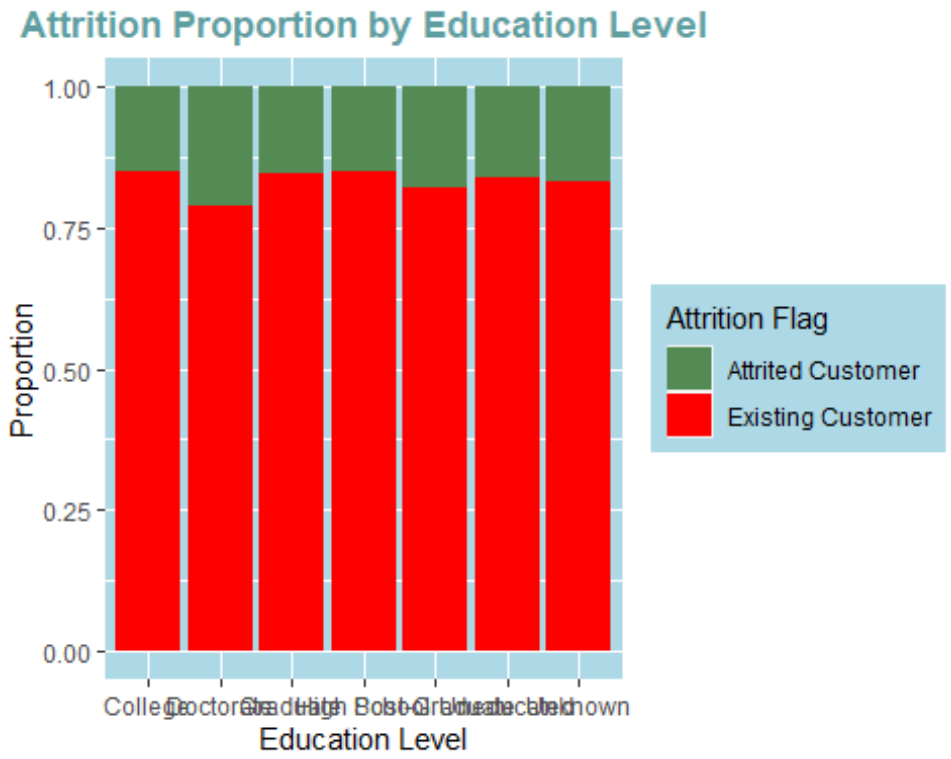

```
# Education_Level wise proportion
row.margin<-round(prop.table(counts,margin = 1),4)*100
row.margin
```

```
##                    Education_Level
## Attrition_Flag      College Doctorate Graduate High School Post-Graduate
##   Attrited Customer    9.47      5.84    29.93       18.81          5.65
##   Existing Customer   10.11      4.19    31.07       20.08          4.99
##                    Education_Level
## Attrition_Flag      Uneducated Unknown
##   Attrited Customer      14.57   15.73
##   Existing Customer      14.71   14.86
```

```
#Attrition_Flag wise proportion
col.margin<-round(prop.table(counts,margin = 2),4)*100
col.margin
```

```
##                    Education_Level
## Attrition_Flag      College Doctorate Graduate High School Post-Graduate
##   Attrited Customer   15.20     21.06    15.57       15.20         17.83
##   Existing Customer   84.80     78.94    84.43       84.80         82.17
##                    Education_Level
## Attrition_Flag      Uneducated Unknown
##   Attrited Customer      15.94   16.85
##   Existing Customer      84.06   83.15
```

```
ggplot(data = BankChurners,aes(x=Education_Level,fill=Attrition_Flag))+geom_b
ar(position = "fill")+scale_x_discrete("Education Level")+scale_y_continuous(
"Proportion")+labs(title = "Attrition Proportion by Education Level",x="Educa
tion Level",fill="Attrition Flag")+theme(panel.background = element_rect(fill
= "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",col
our = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+theme(leg
end.background = element_rect(fill = "lightblue"))
```



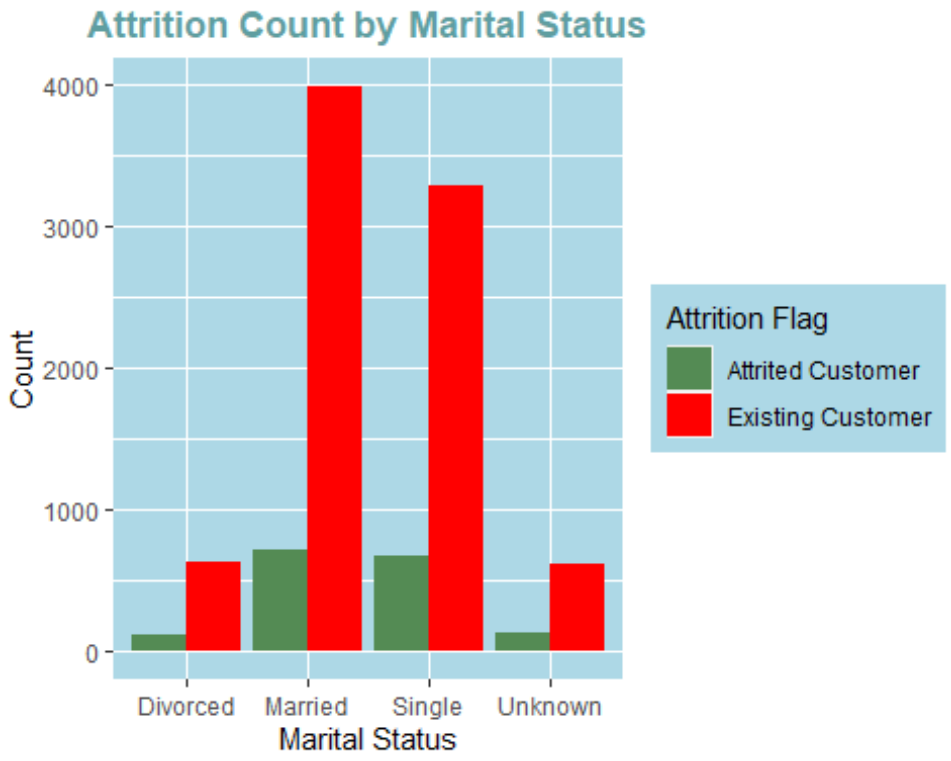
- From the above graphs and the tables, we observe that the society attached with the bank are civilized and educated but at the same time uneducated customers consists of approx 14%.
- We observe that the customers belong to the Post-Graduate and Doctorate category seems to churn the credit card facility at a higher rate than the others with 17.83% and 21.06% attrition rates respectively. Since the topmost educated category has the problem with the service, the bank should try to readjst its credit card facilities wisely.

Attrition_Flag - Marital_Status

```
counts<-table(BankChurners$Attrition_Flag,BankChurners$Marital_Status,dnn = c
("Attrition_Flag","Marital_Status"))
counts
```

```
##                    Marital_Status
## Attrition_Flag      Divorced Married Single Unknown
##   Attrited Customer      121     709    668     129
##   Existing Customer      627    3978   3275     620
```

```
ggplot(data = BankChurners,aes(x=Marital_Status,fill=Attrition_Flag))+geom_ba
r(position = "dodge")+labs(title = "Attrition Count by Marital Status",x="Mar
ital Status",fill="Attrition Flag")+theme(panel.background = element_rect(fil
l = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",c
olour = "cadetblue"))+scale_x_discrete("Marital Status")+scale_y_continuous("
Count")+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.backgrou
nd = element_rect(fill = "lightblue"))
```




```
# Marital_Status wise proportion
row.margin<-round(prop.table(counts,margin = 1),4)*100
row.margin
```

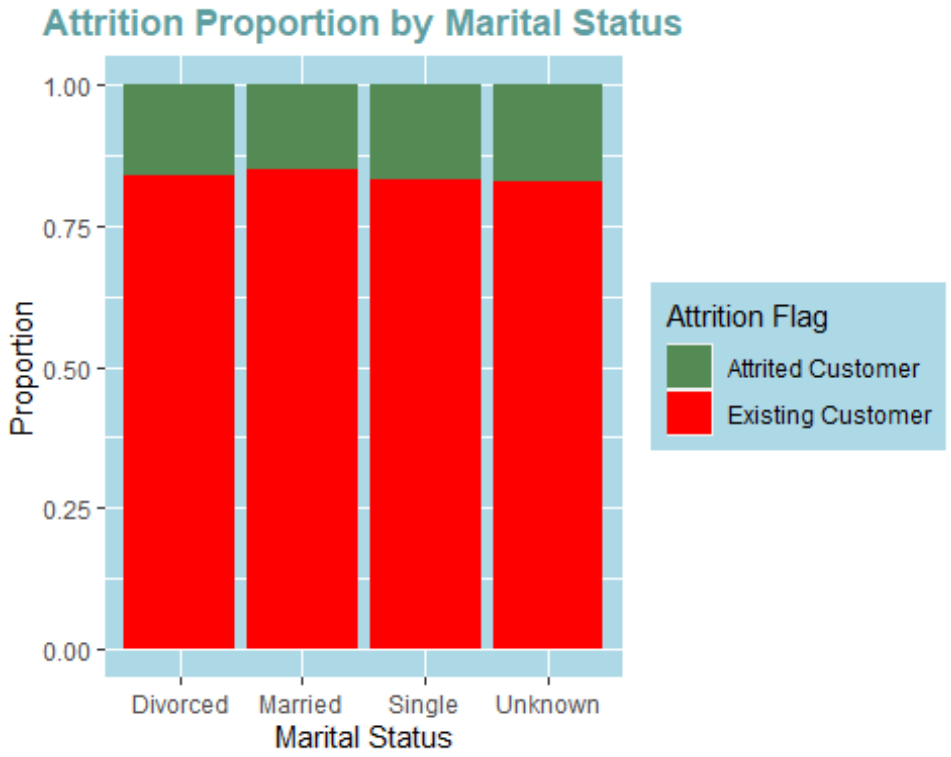
```
##                    Marital_Status
## Attrition_Flag      Divorced Married Single Unknown
```

```
##   Attrited Customer     7.44   43.58  41.06    7.93
##   Existing Customer     7.38   46.80  38.53    7.29
```

```
#Attrition_Flag wise proportion
col.margin<-round(prop.table(counts,margin = 2),4)*100
col.margin
```

```
##                     Marital_Status
## Attrition_Flag      Divorced Married Single Unknown
##   Attrited Customer    16.18   15.13  16.94   17.22
##   Existing Customer    83.82   84.87  83.06   82.78
```

```
ggplot(data = BankChurners,aes(x=Marital_Status,fill=Attrition_Flag))+geom_ba
r(position = "fill")+scale_x_discrete("Marital Status")+scale_y_continuous("P
roportion")+labs(title = "Attrition Proportion by Marital Status",x="Marital
Status",fill="Attrition Flag")+theme(panel.background = element_rect(fill = "
lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour
= "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.
background = element_rect(fill = "lightblue"))
```



- From the above graphs and tables, we observe that approximately 43% of the customers are married and 41% are single with the single category has a slightly higher attrition rate than the average.

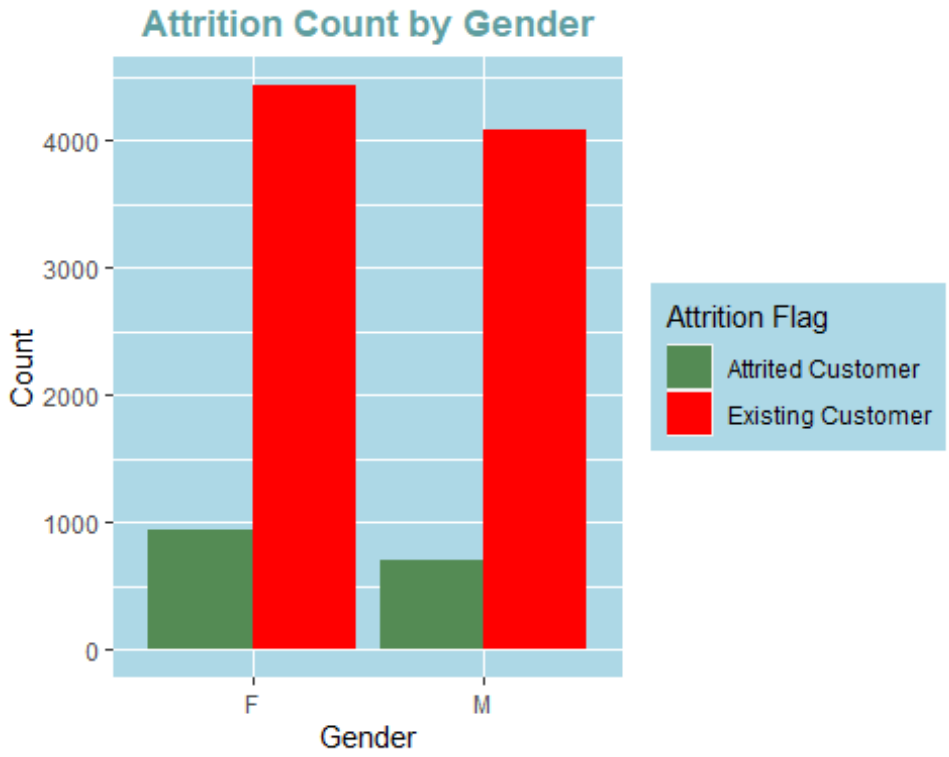
- The customers with the Unknown marital status consists of approx. 8% of the total customers but they have the highest attrition rate of 17.22%.
- Since the married and the single category consists of maximum customers with the reasonable attrition rate. The bank's focus should be on these two categories.

Attrition_Flag – Gender

```
counts<-table(BankChurners$Attrition_Flag,BankChurners$Gender,dnn = c("Attrit
ion_Flag","Gender"))
counts
```

```
##                     Gender
## Attrition_Flag         F    M
##   Attrited Customer  930  697
##   Existing Customer 4428 4072
```

```
ggplot(data = BankChurners,aes(x=Gender,fill=Attrition_Flag))+geom_bar(positi
on = "dodge")+labs(title = "Attrition Count by Gender",x="Gender",fill="Attri
tion Flag")+theme(panel.background = element_rect(fill = "lightblue"))+theme(
plot.title = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+sc
ale_x_discrete("Gender")+scale_y_continuous("Count")+scale_fill_manual(values
=c("palegreen4","red"))+theme(legend.background = element_rect(fill = "lightb
lue"))
```



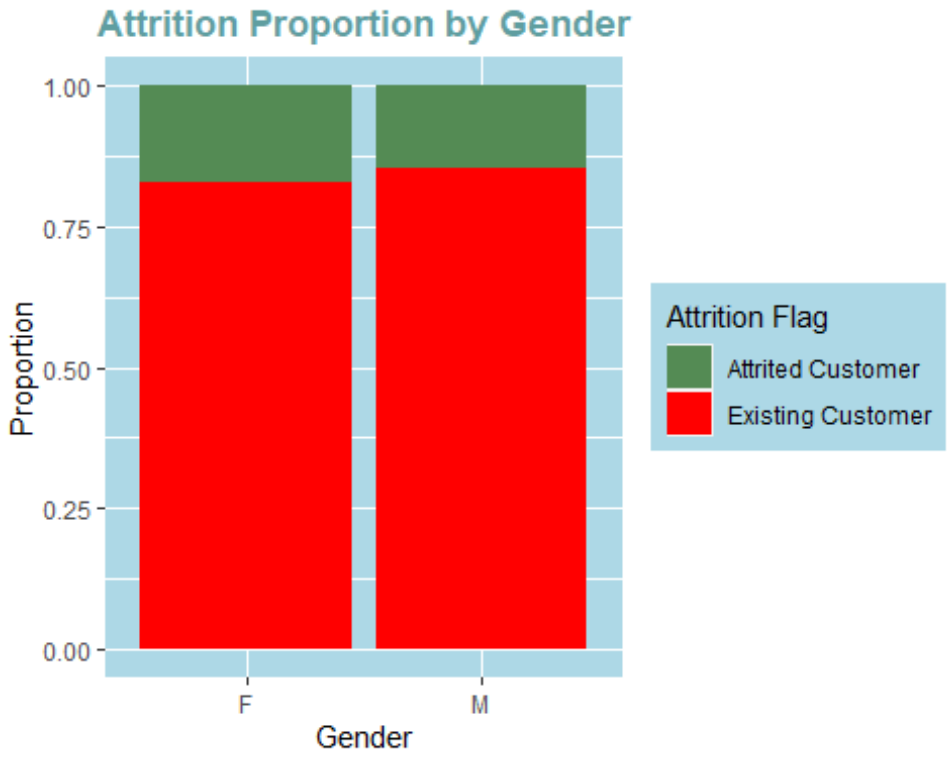

```
# Gender wise proportion
row.margin<-round(prop.table(counts,margin = 1),4)*100
row.margin
```

```
##                     Gender
## Attrition_Flag          F     M
##   Attrited Customer 57.16 42.84
##   Existing Customer 52.09 47.91
```

```
#Attrition_Flag wise proportion
col.margin<-round(prop.table(counts,margin = 2),4)*100
col.margin
```

```
##                     Gender
## Attrition_Flag          F     M
##   Attrited Customer 17.36 14.62
##   Existing Customer 82.64 85.38
```

```
ggplot(data = BankChurners,aes(x=Gender,fill=Attrition_Flag))+geom_bar(positi
on = "fill")+scale_x_discrete("Gender")+scale_y_continuous("Proportion")+labs
(title = "Attrition Proportion by Gender",x="Gender",fill="Attrition Flag")+t
heme(panel.background = element_rect(fill = "lightblue"))+theme(plot.title =
element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manu
al(values=c("palegreen4","red"))+theme(legend.background = element_rect(fill
= "lightblue"))
```



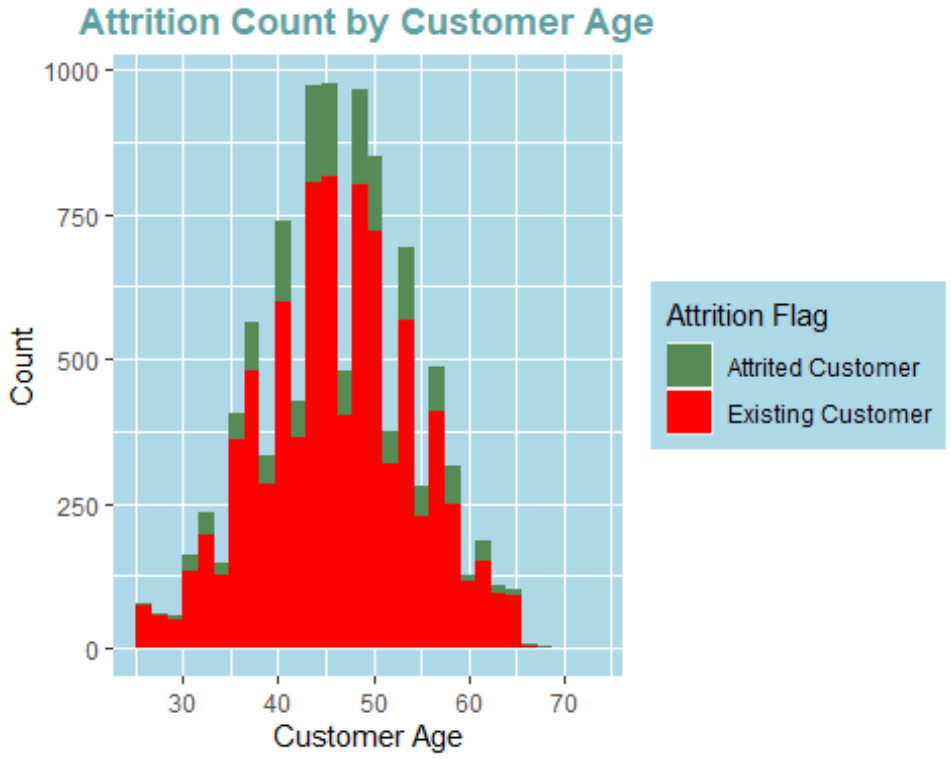
- More than 50% of the customers of the bank are females.
- Females also have more attrition rate, as is clear from the graph and the table with 17.36%.

C. EXPLORING NUMERICAL VARIABLES

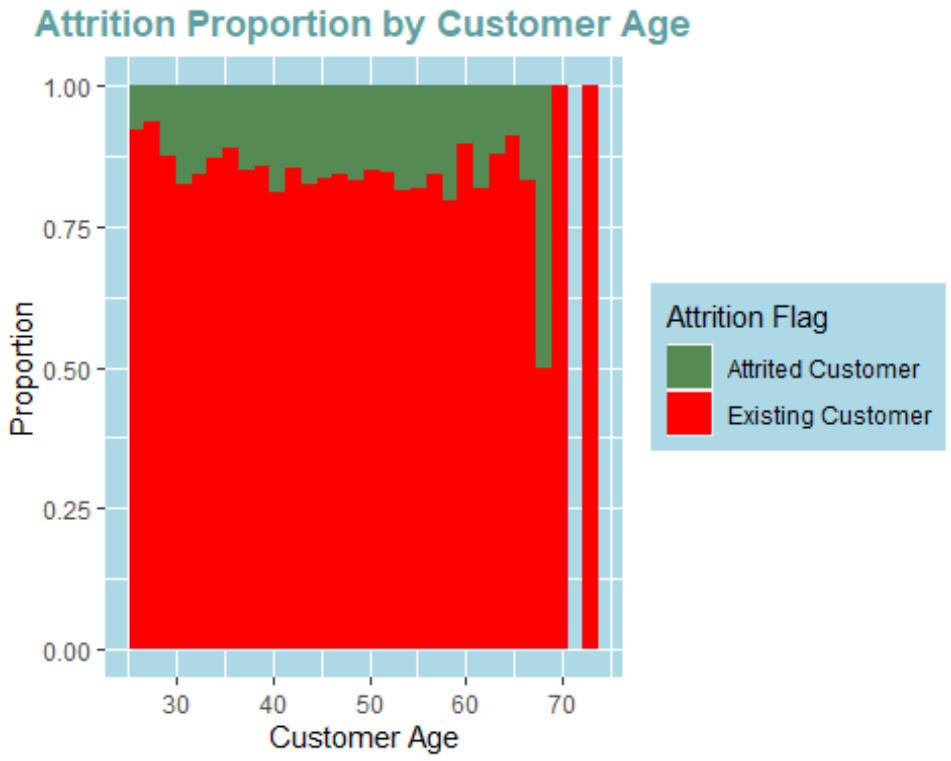
Now we have Customer_Age, Dependent_Count, Months_on_book, Total_relationship_count, months_inactive_12_mon, contacts_count_12_mon, credit_limit, Total_revolving_balance, Avg_open_to_buy, Total_amt_chng_Q4_Q1, Total_Trans_Amt, Total_Trans_count, Total_count_chng_Q4_Q1 and Avg_utilization_Ratio as the numerical variables and we will look at them one by one by visualizing them.

Attrition_Flag - Customer_Age

```
summary(BankChurners$Customer_Age)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   26.00   41.00   46.00   46.33   52.00   73.00

ggplot(data = BankChurners,aes(x=Customer_Age,fill=Attrition_Flag))+geom_hist
ogram(position = "stack",bins = 30)+scale_x_continuous("Customer Age")+scale_
y_continuous("Count")+labs(title = "Attrition Count by Customer Age",x="Custo
mer Age",fill="Attrition Flag")+theme(panel.background = element_rect(fill =
"lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colou
r = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+theme(legen
d.background = element_rect(fill = "lightblue"))
```



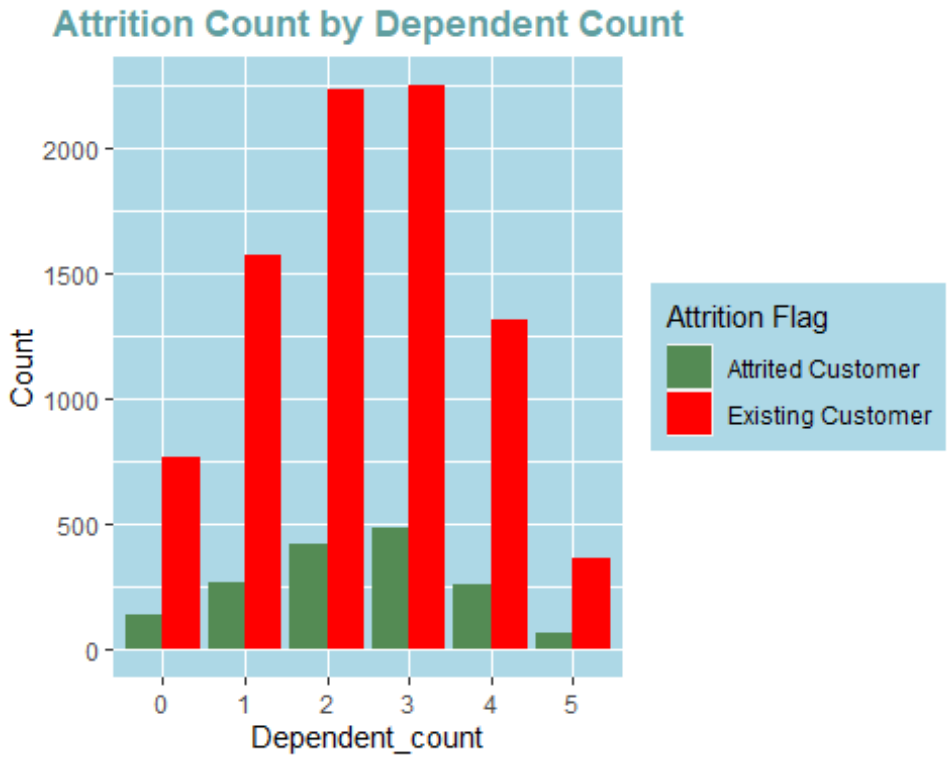

```
ggplot(data = BankChurners,aes(x=Customer_Age,fill=Attrition_Flag))+geom_hist
ogram(position = "fill",bins = 30)+scale_x_continuous("Customer Age")+scale_y
_continuous("Proportion")+labs(title = "Attrition Proportion by Customer Age"
,x="Customer Age",fill="Attrition Flag")+theme(panel.background = element_rec
t(fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bo
ld",colour = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+th
eme(legend.background = element_rect(fill = "lightblue"))
```



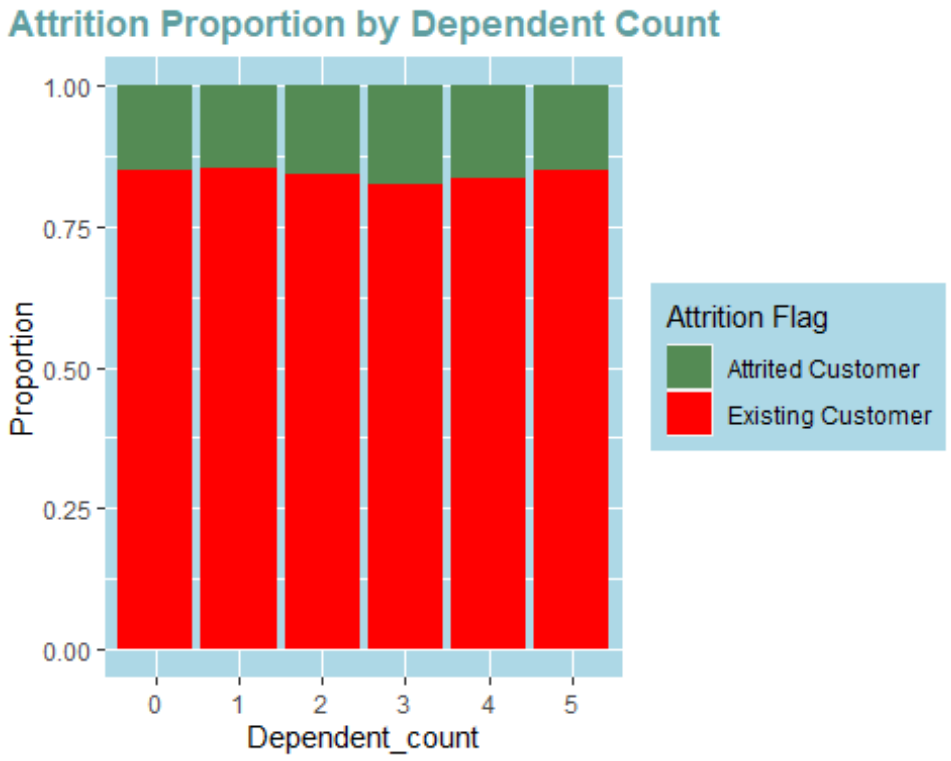
- From the above graphs and the summary, we observed that we have customers from age group 26 to 73.
- From the first graph we get that we have maximum number of customers from the middle age group 40-60 years.
- Also we observe that as the customer age increases, the attrition rate also increases. At the same time, we get that the customers who crossed 70 years of age have no attrition.

Attrition_Flag - Dependent_Count

```
ggplot()+geom_bar(data = BankChurners,aes(x=factor(Dependent_count),fill=fact
or(Attrition_Flag)),position="dodge")+scale_x_discrete("Dependent_count")+sca
le_y_continuous("Count")+guides(fill=guide_legend(title = "Attrition Flag"))+
scale_fill_manual(values = c("palegreen4","red"))+labs(title = "Attrition Cou
nt by Dependent Count",x="Dependent Count",fill="Attrition Flag")+theme(panel
.background = element_rect(fill = "lightblue"))+theme(plot.title = element_te
xt(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.background =
element_rect(fill = "lightblue"))
```



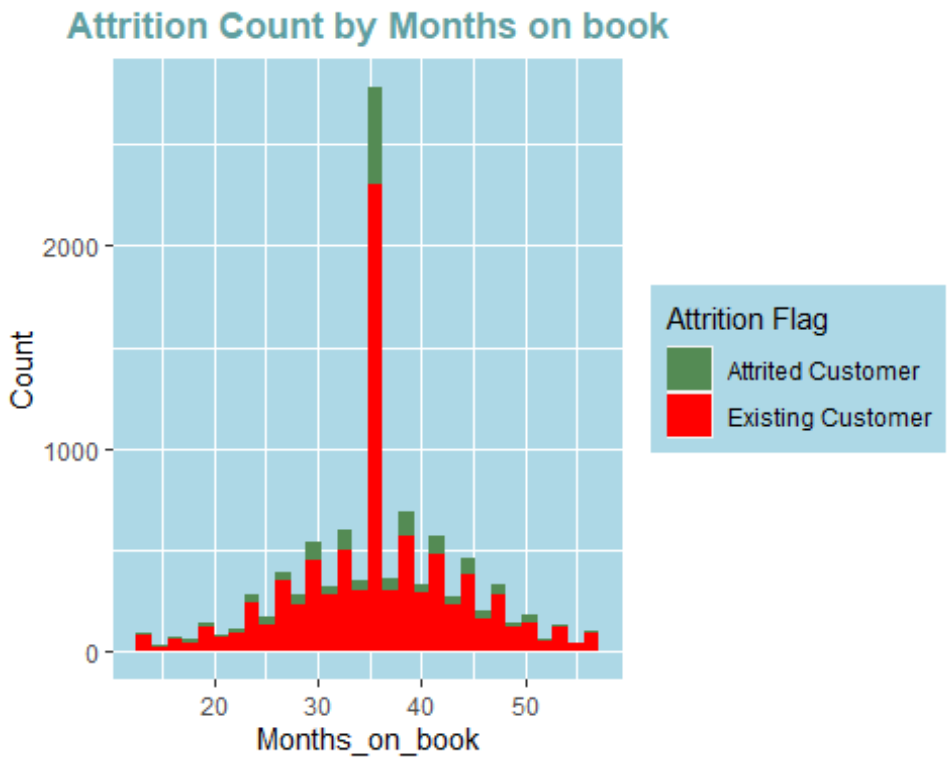

```
ggplot()+geom_bar(data = BankChurners,aes(x=factor(Dependent_count),fill=fact
or(Attrition_Flag)),position="fill")+scale_x_discrete("Dependent_count")+scal
e_y_continuous("Proportion")+guides(fill=guide_legend(title = "Attrition Flag
"))+scale_fill_manual(values = c("palegreen4","red"))+labs(title = "Attrition
Proportion by Dependent Count",x="Dependent Count",fill="Attrition Flag")+the
me(panel.background = element_rect(fill = "lightblue"))+theme(plot.title = el
ement_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.back
ground = element_rect(fill = "lightblue"))
```



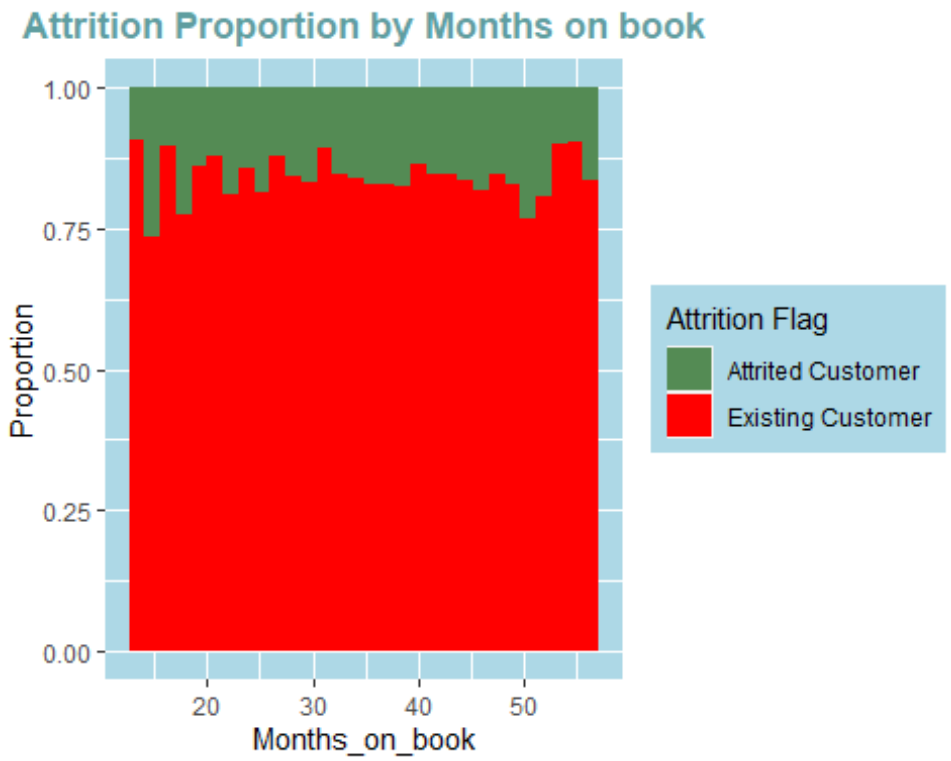
- From the first graph we observe that the customers with the dependent count are the highest.
- Also we observe that these two categories have the maximum attrition rate.
- Also, we have minimum attrition rate when dependend count is 0 and 5.

Attrition_Flag - Months_on_book (Period of relationship with the bank)

```
summary(BankChurners$Months_on_book)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   13.00   31.00   36.00   35.93   40.00   56.00

ggplot(data = BankChurners,aes(x=Months_on_book,fill=Attrition_Flag))+geom_hi
stogram(position = "stack",bins = 30)+scale_x_continuous("Months_on_book")+sc
ale_y_continuous("Count")+guides(fill=guide_legend(title = "Attrition Flag"))
+scale_fill_manual(values = c("palegreen4","red"))+labs(title = "Attrition Co
unt by Months on book",x="Months on book",fill="Attrition Flag")+theme(panel.
background = element_rect(fill = "lightblue"))+theme(plot.title = element_tex
t(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.background =
element_rect(fill = "lightblue"))
```



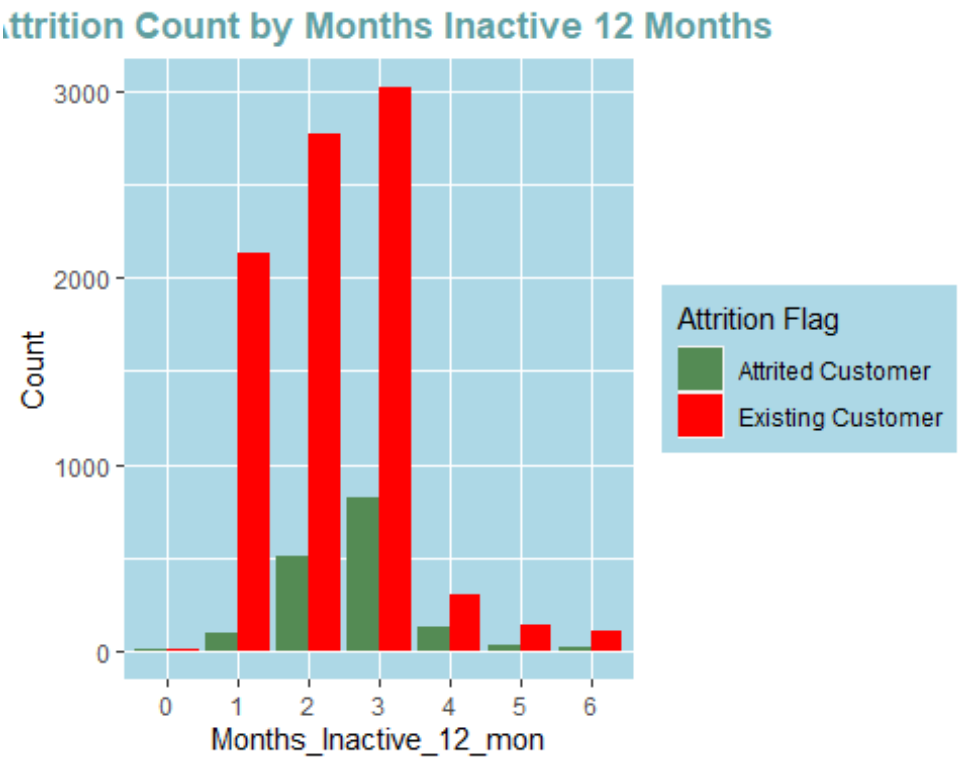

```
ggplot(data = BankChurners,aes(x=Months_on_book,fill=Attrition_Flag))+geom_hi
stogram(position = "fill",bins = 30)+scale_x_continuous("Months_on_book")+sca
le_y_continuous("Proportion")+guides(fill=guide_legend(title = "Attrition Fla
g"))+scale_fill_manual(values = c("palegreen4","red"))+labs(title = "Attritio
n Proportion by Months on book",x="Months on book",fill="Attrition Flag")+the
me(panel.background = element_rect(fill = "lightblue"))+theme(plot.title = el
ement_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.back
ground = element_rect(fill = "lightblue"))
```



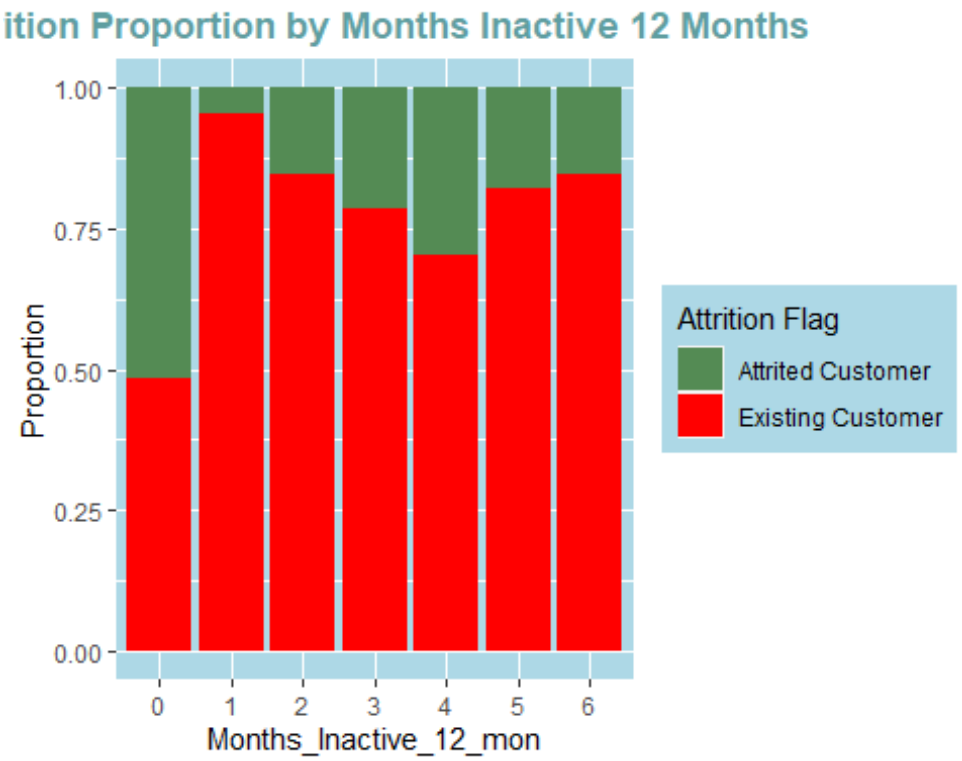
- From the above graph, we observe that most of customers of the bank consists of the customers with 30-40 Periods of relationship with the bank.
- However overall, we see that the attrition rate is constant throughout but if we observe closely, from the second proportionate graph we get that the customers with very small and very large period of the relationship with the bank seems to churn at a higher attrition rate.

Attrition_Flag -Months_Inactive_12_mon (Numbers of months inactive in last 12 months)

```
ggplot()+geom_bar(data = BankChurners,aes(x=factor(Months_Inactive_12_mon),fi
ll=factor(Attrition_Flag)),position="dodge")+scale_x_discrete("Months_Inactiv
e_12_mon")+scale_y_continuous("Count")+guides(fill=guide_legend(title = "Attr
ition Flag"))+scale_fill_manual(values = c("palegreen4","red"))+labs(title =
"Attrition Count by Months Inactive 12 Months",x="Months Inactive 12 months",
fill="Attrition Flag")+theme(panel.background = element_rect(fill = "lightblu
e"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "cade
tblue"))+theme(legend.background = element_rect(fill = "lightblue"))
```



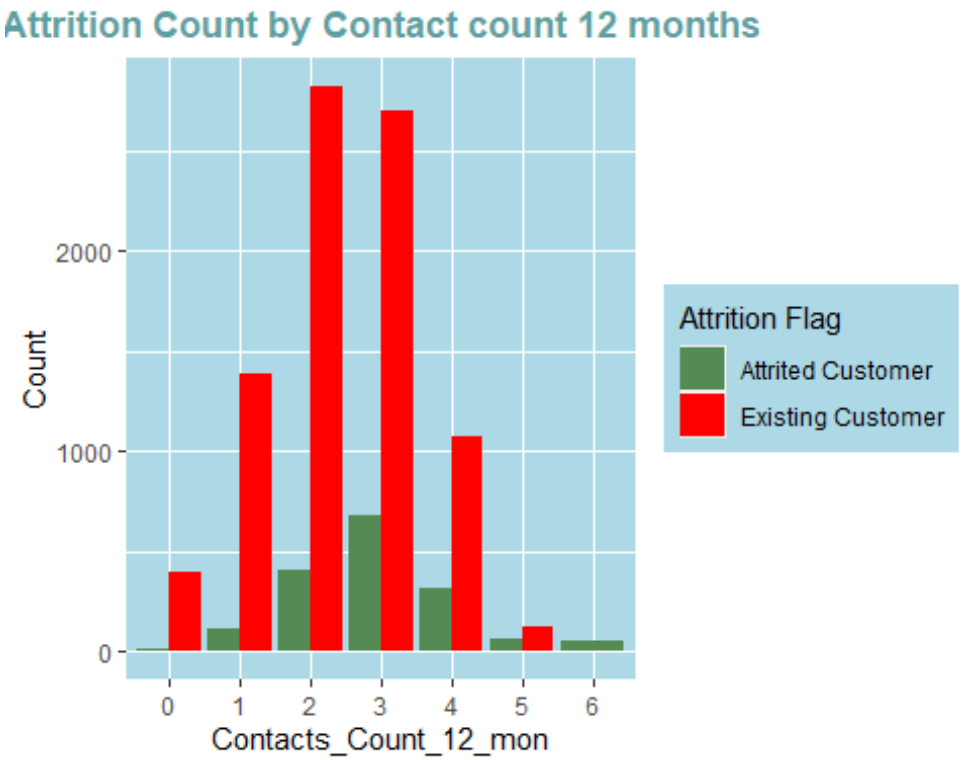

```
ggplot()+geom_bar(data = BankChurners,aes(x=factor(Months_Inactive_12_mon),fi
ll=factor(Attrition_Flag)),position="fill")+scale_x_discrete("Months_Inactive
_12_mon")+scale_y_continuous("Proportion")+guides(fill=guide_legend(title = "
Attrition Flag"))+scale_fill_manual(values = c("palegreen4","red"))+labs(titl
e = "Attrition Proportion by Months Inactive 12 Months",x="Customer Age",fill
="Attrition Flag")+theme(panel.background = element_rect(fill = "lightblue"))
+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "cadetblu
e"))+theme(legend.background = element_rect(fill = "lightblue"))
```



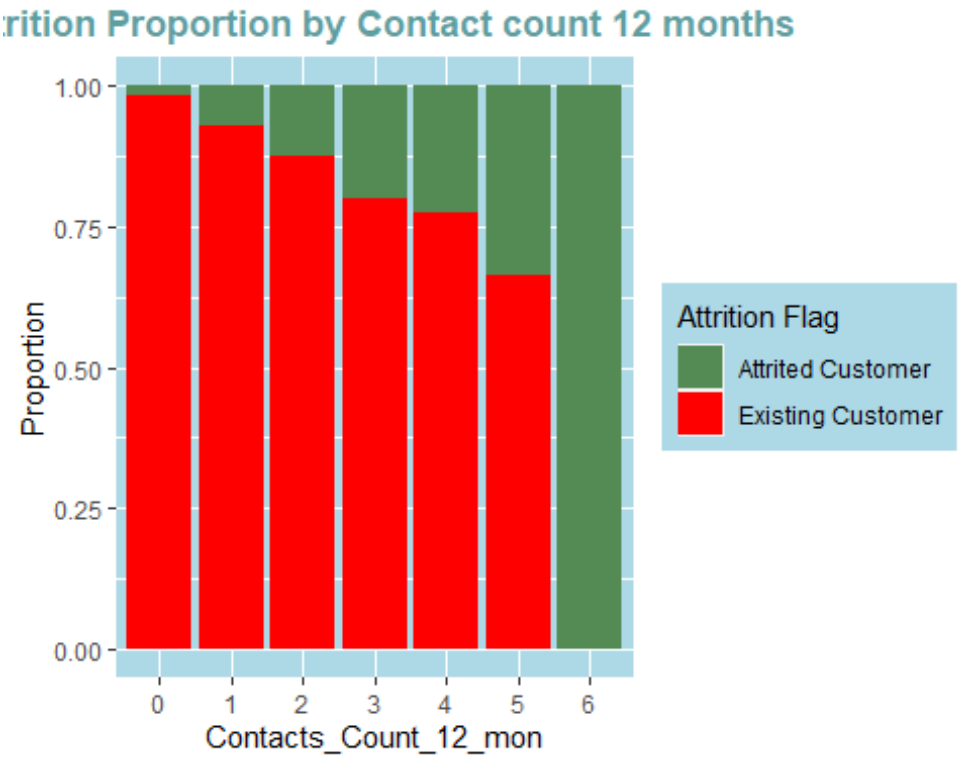
- From the first above graph, we observe that there are very less customers who are inactive, while in the second graph we see that these active and regular customers have the maximum attrition rate which is extremely alarming. Therefore the bank should make the polcies keeping in mind for the regular customers.
- Second thing we observe that the bank has almost all customers who remain inactive from 1 month minimum to 6 months minimum which is also not good for the healthy for the bank. We have maximum customers who are inactive for 3 months in last 12 months followed by 2 months followed by 1 month.
- Also we see that from 1 inactive month to the 4 inactive months in the last 12 months, we have constantly increasing attrition rate.
- The customers who were inactive 5 and 6 months in the last 12 months have the relatively low attrition rate.

Attrition_Flag - Contacts_count_12_mon (Number of contacts in the last 12 months)

```
ggplot()+geom_bar(data = BankChurners,aes(x=factor(Contacts_Count_12_mon),fil
l=factor(Attrition_Flag)),position="dodge")+scale_x_discrete("Contacts_Count_
12_mon")+scale_y_continuous("Count")+guides(fill=guide_legend(title = "Attrit
ion Flag"))+scale_fill_manual(values = c("palegreen4","red"))+labs(title = "A
ttrition Count by Contact count 12 months",x="Contact count 12 month",fill="A
ttrition Flag")+theme(panel.background = element_rect(fill = "lightblue"))+th
eme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "cadetblue")
)+theme(legend.background = element_rect(fill = "lightblue"))
```




```
ggplot()+geom_bar(data = BankChurners,aes(x=factor(Contacts_Count_12_mon),fil
l=factor(Attrition_Flag)),position="fill")+scale_x_discrete("Contacts_Count_1
2_mon")+scale_y_continuous("Proportion")+guides(fill=guide_legend(title = "At
trition Flag"))+scale_fill_manual(values = c("palegreen4","red"))+labs(title
= "Attrition Proportion by Contact count 12 months",x="Contact count 12 month
s",fill="Attrition Flag")+theme(panel.background = element_rect(fill = "light
blue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "c
adetblue"))+theme(legend.background = element_rect(fill = "lightblue"))
```

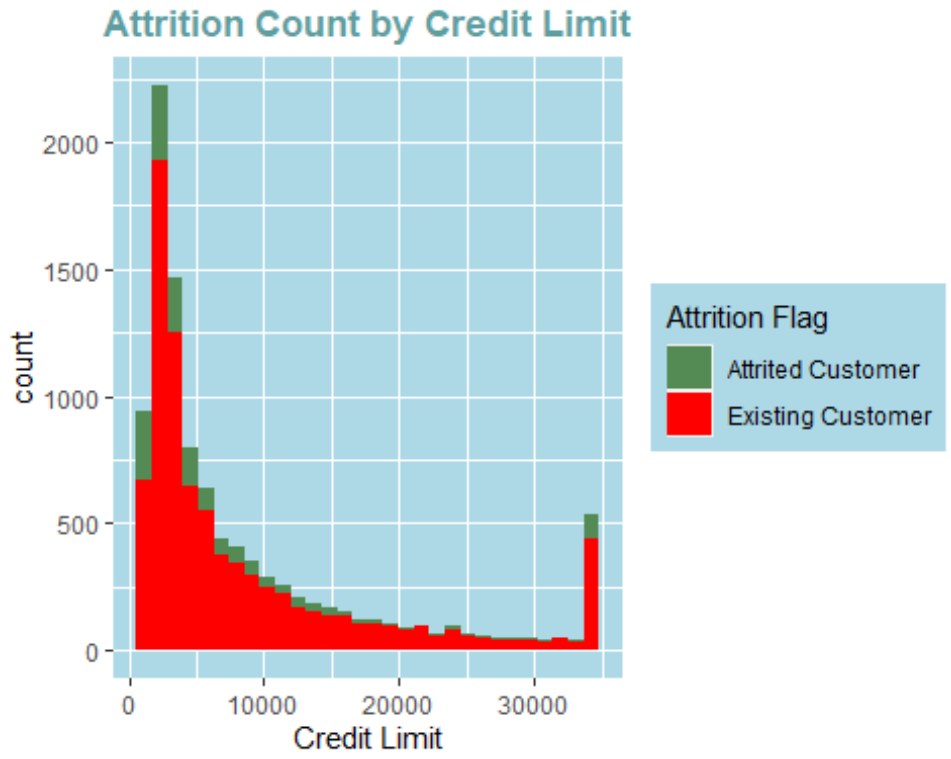


- From the above first graph we observe that there are maximum customers in the banks with 2 and 3 contact counts, followed by 1 and 4 whic is indeed a very large proportion of the customers.
- We also observe from the second graph that as the number of contact counts increases the attrition rate increases at a very fast rate indicating the the customers are not getting satisfied despite contacting.
- The Bank should observe carefully the complaints of the customers who contacted and try to identfy the problem and reframe the credit card policies in order to remove the problems faced by the credit card customers.

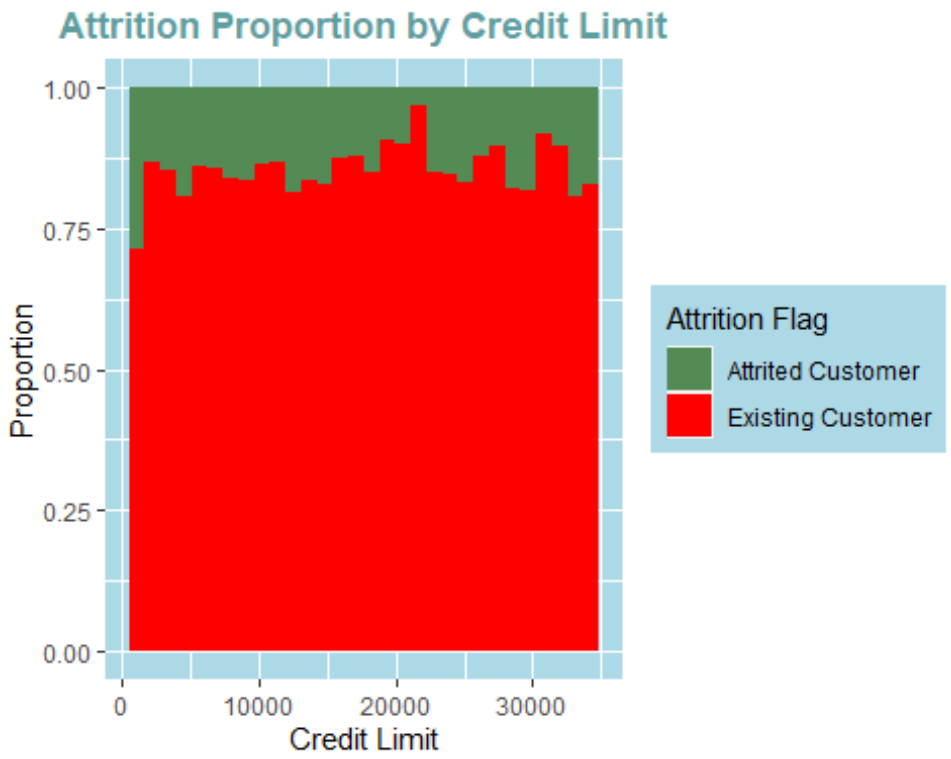
Attrition_Flag - Credit_Limit (Credit limit on the credit card)

```
summary(BankChurners$Credit_Limit)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    1438    2555    4549    8632   11068   34516
```

```
ggplot(data = BankChurners,aes(x=Credit_Limit,fill=Attrition_Flag))+geom_hist
ogram(position = "stack",bins = 30)+labs(title = "Attrition Count by Credit L
imit",x="Credit Limit",fill="Attrition Flag")+theme(panel.background = elemen
t_rect(fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face
= "bold",colour = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"
))+theme(legend.background = element_rect(fill = "lightblue"))
```




```
ggplot(data = BankChurners,aes(x=Credit_Limit,fill=Attrition_Flag))+geom_hist
ogram(position = "fill",bins = 30)+scale_x_continuous("Credit Limit")+scale_y
_continuous("Proportion")+labs(title = "Attrition Proportion by Credit Limit"
,x="Credit Limit",fill="Attrition Flag")+theme(panel.background = element_rec
t(fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bo
ld",colour = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+th
eme(legend.background = element_rect(fill = "lightblue"))
```

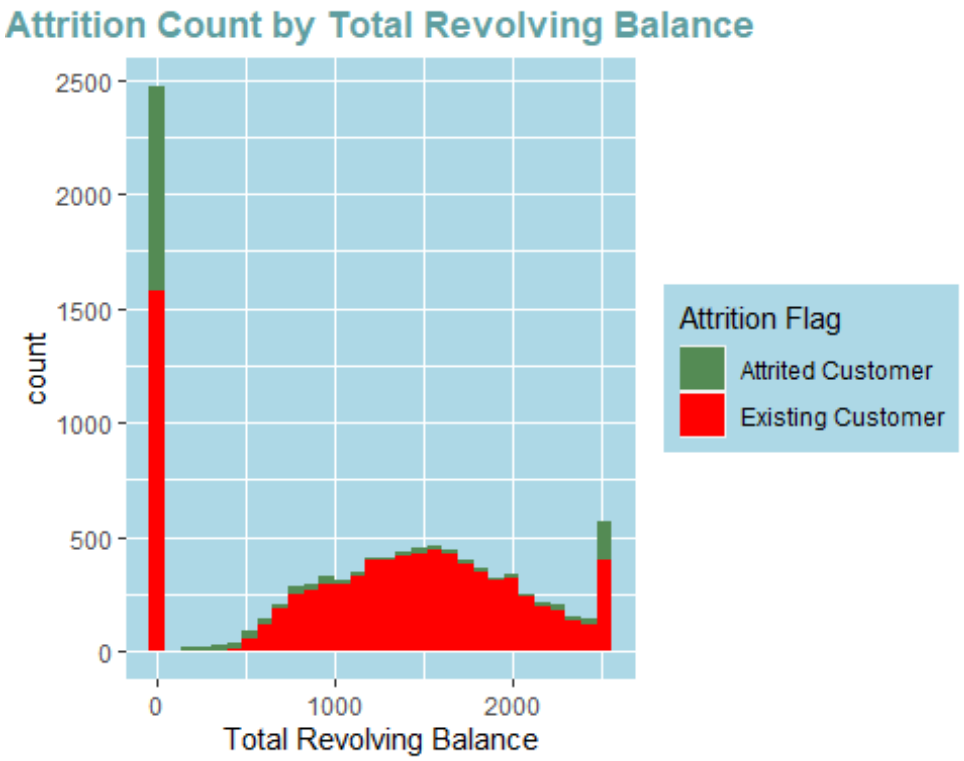


- From the above graphs, we get that there are maximum customers with low credit limit. As the credit limit of the customer increases, their count decreases.
- From the second graph above we observe that upto 2000 Credit Limit the attrition rate of customers decreases though slowly. As the Credit Limit goes beyond 2000 the attrition rate starts increasing.
- Therefore, the area of concern is the customer with low credit limit, say, less than 1000 and greater than 3000.

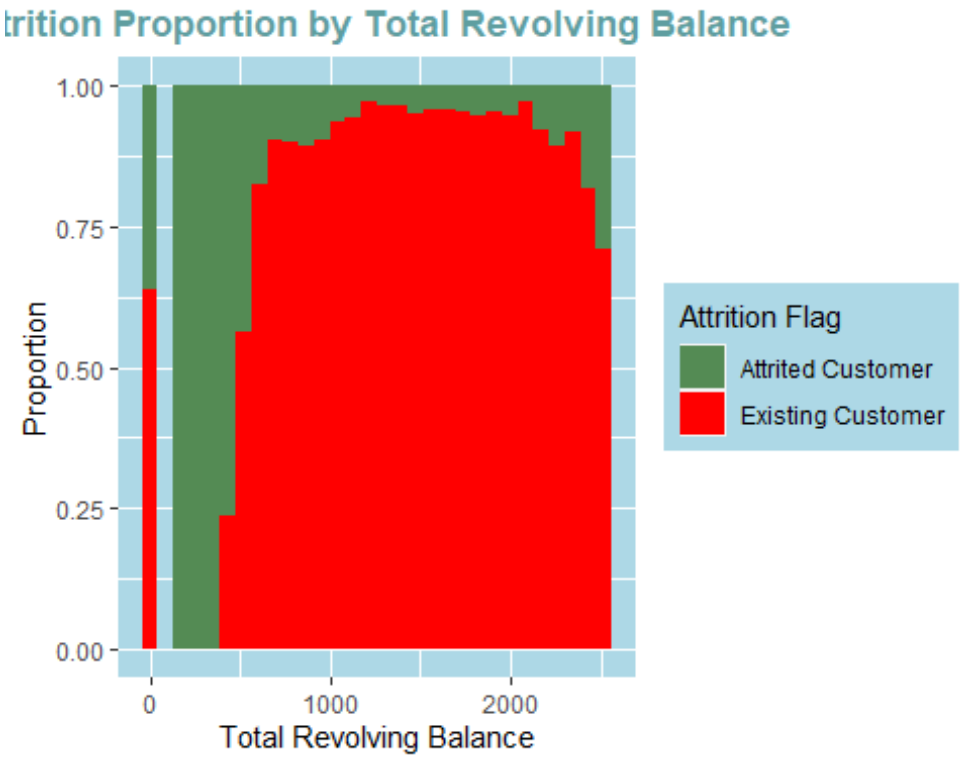
Attrition_Flag -Total_Revolving_Balance

```
summary(BankChurners$Total_Revolving_Bal)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       0     359    1276    1163    1784    2517

ggplot(data = BankChurners,aes(x=Total_Revolving_Bal,fill=Attrition_Flag))+ge
om_histogram(position = "stack",bins = 30)+labs(title = "Attrition Count by T
otal Revolving Balance",x="Total Revolving Balance",fill="Attrition Flag")+th
```

```
eme(panel.background = element_rect(fill = "lightblue"))+theme(plot.title = e
lement_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manua
l(values=c("palegreen4","red"))+theme(legend.background = element_rect(fill =
"lightblue"))
```



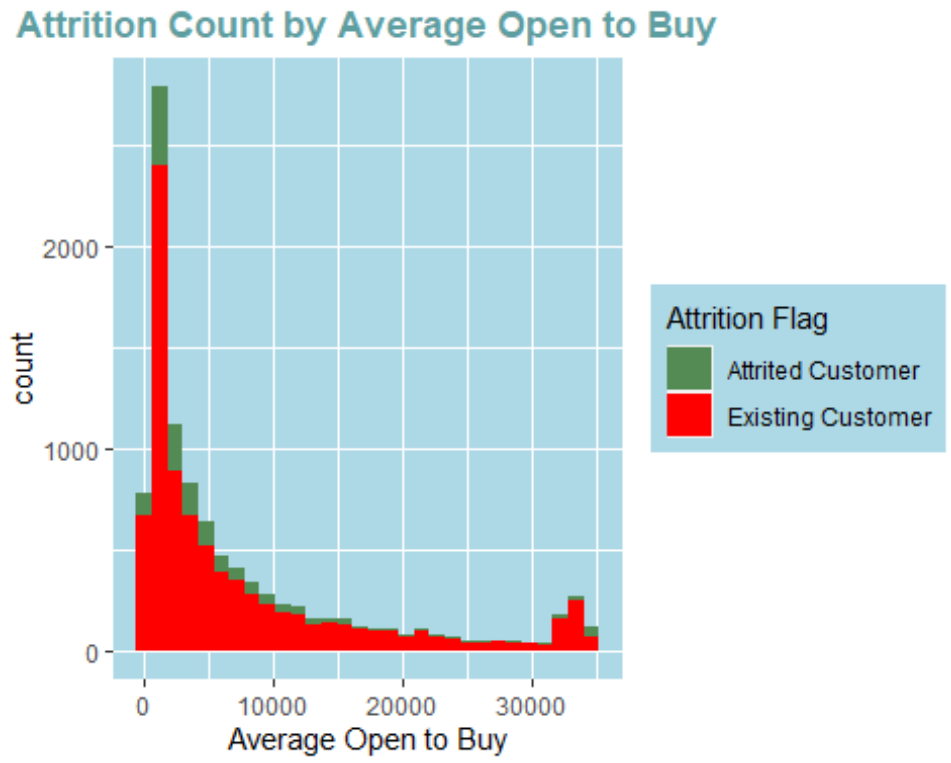

```
ggplot(data = BankChurners,aes(x=Total_Revolving_Bal,fill=Attrition_Flag))+ge
om_histogram(position = "fill",bins = 30)+labs(title = "Attrition Proportion
by Total Revolving Balance",x="Total Revolving Balance",fill="Attrition Flag"
)+theme(panel.background = element_rect(fill = "lightblue"))+theme(plot.title
= element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_ma
nual(values=c("palegreen4","red"))+theme(legend.background = element_rect(fil
l = "lightblue"))+scale_y_continuous("Proportion")
```



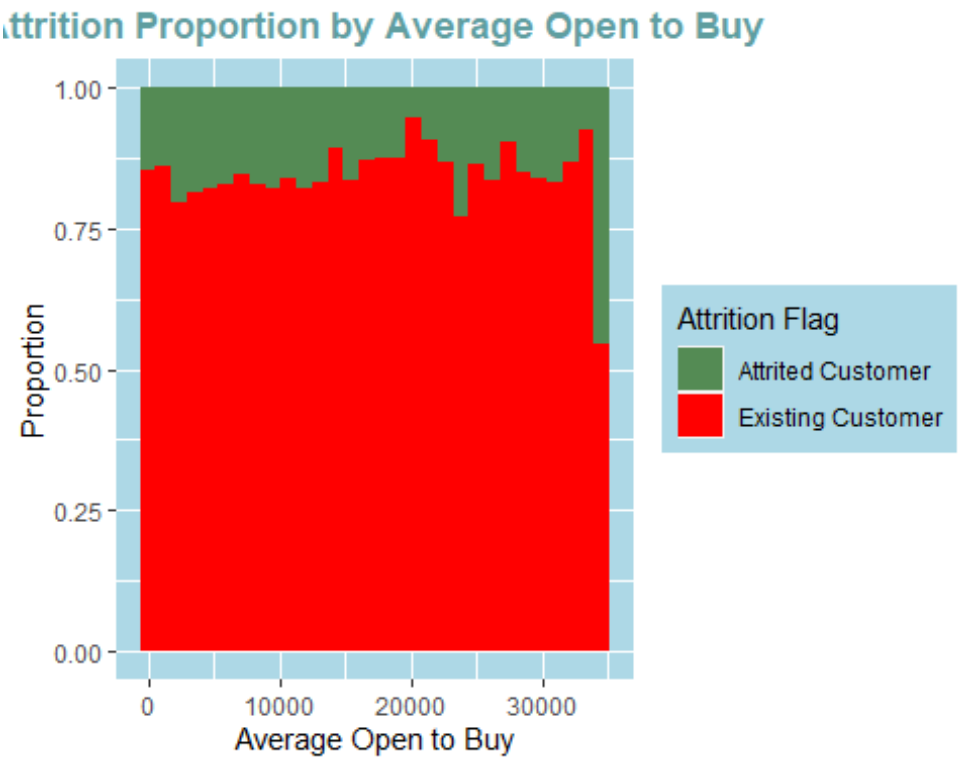
- From the above graphs we can say that there are approximately 2500 customers who have not any Revolving Balance which is itself a very large nuumber. Also thuus section has a very large attrition rate (approximately 35%). Therefore we can conclude that the customers with the) Total Revolving Balance should be the area of concern.
- Next thing we observe that there are more customers with the Total Revolving Balance from 1000 to 2000. The Total Revolving Balance starts decreasing after 1600 and continues to decrease till 2500 until we get a huge increase in the customers with approx 2600 Total Revolving Balance.
- From the proportionate graph above we get that, the Bank has maximum Attrition rate with the customers with the low Total Revolving Balance. As the Total Revolving Balance with the customers increases, the attrition rate starts decreases.
- Also we see that beyond 2000 Total Revolving Balance the attrition rate again starts increasing. Therefore, the bank's focus should be on the customers with the low and high Total Revolving Balance.

Attrition_Flag - Avg_open_to_buy (Open to buy credit line (Average of last 12 months))

```
summary(BankChurners$Avg_Open_To_Buy)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##       3    1324    3474    7469    9859   34516

ggplot(data = BankChurners,aes(x=Avg_Open_To_Buy,fill=Attrition_Flag))+geom_h
istogram(position = "stack",bins = 30)+labs(title = "Attrition Count by Avera
ge Open to Buy",x="Average Open to Buy",fill="Attrition Flag")+theme(panel.ba
ckground = element_rect(fill = "lightblue"))+theme(plot.title = element_text(
hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(values=c("
palegreen4","red"))+theme(legend.background = element_rect(fill = "lightblue"
))
```




```
ggplot(data = BankChurners,aes(x=Avg_Open_To_Buy,fill=Attrition_Flag))+geom_h
istogram(position = "fill",bins = 30)+labs(title = "Attrition Proportion by A
verage Open to Buy",x="Average Open to Buy",fill="Attrition Flag")+theme(pane
l.background = element_rect(fill = "lightblue"))+theme(plot.title = element_t
ext(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(values
=c("palegreen4","red"))+theme(legend.background = element_rect(fill = "lightb
lue"))+scale_y_continuous("Proportion")
```



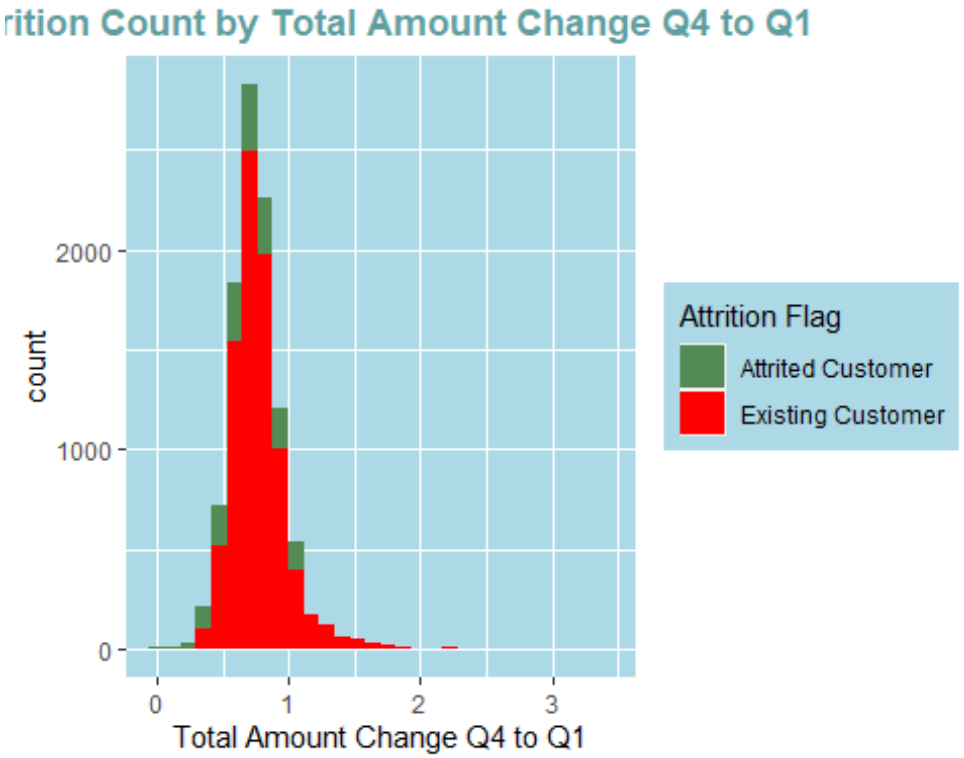

```
cor(BankChurners$Credit_Limit,BankChurners$Avg_Open_To_Buy) # Findinng correl
ation

## [1] 0.9959805
```

- From the above graph we observe that the pattern of this graph seems so similar to the Credit Limit, the details of which we have mentioned above.
- Also after finding the relationship between them we get that they have a very high correlation between them i.e. 0.9959. Therefore it is not required to keep both these variables. Anyone of them needs to be omitted.

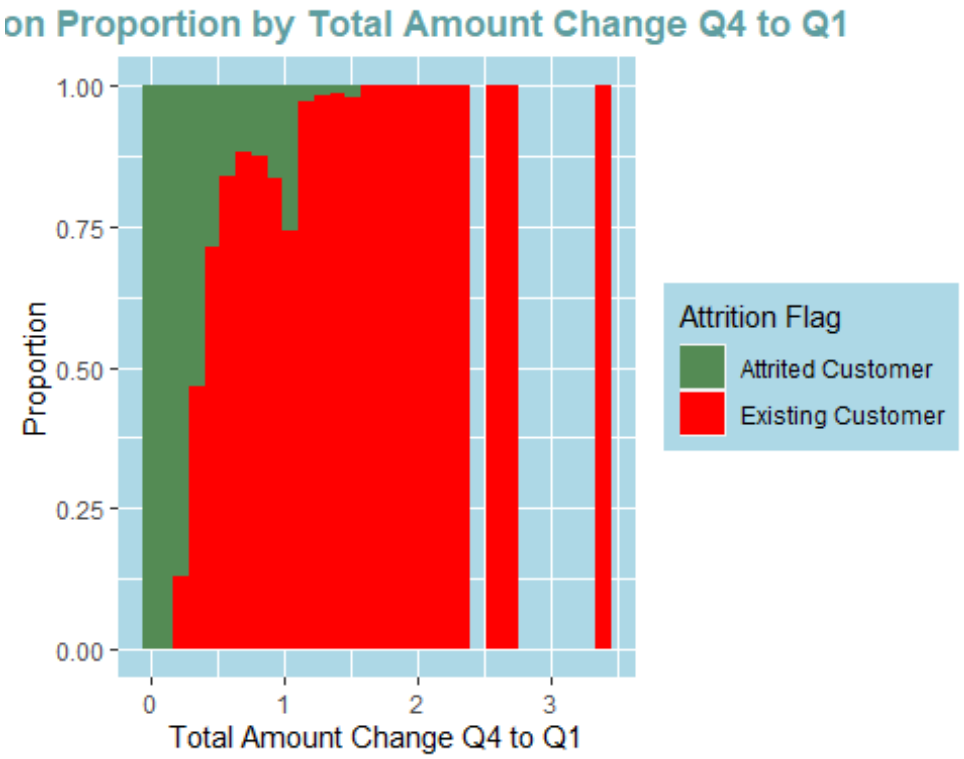
Attrition_Flag - Total_Amt_chng_Q4_Q1

```
summary(BankChurners$Total_Amt_Chng_Q4_Q1)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.6310  0.7360  0.7599  0.8590  3.3970

ggplot(data = BankChurners,aes(x=Total_Amt_Chng_Q4_Q1,fill=Attrition_Flag))+g
eom_histogram(position = "stack",bins = 30)+labs(title = "Attrition Count by
```

```
Total Amount Change Q4 to Q1",x="Total Amount Change Q4 to Q1",fill="Attritio
n Flag")+theme(panel.background = element_rect(fill = "lightblue"))+theme(plo
t.title = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale
_fill_manual(values=c("palegreen4","red"))+theme(legend.background = element_
rect(fill = "lightblue"))
```



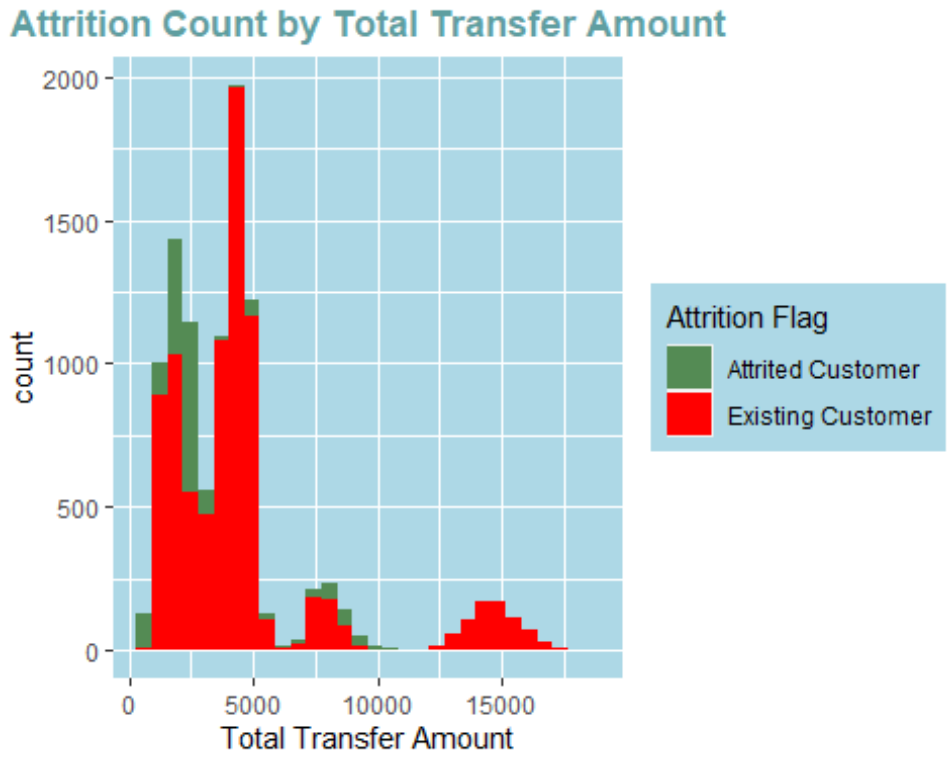

```
ggplot(data = BankChurners,aes(x=Total_Amt_Chng_Q4_Q1,fill=Attrition_Flag))+g
eom_histogram(position = "fill",bins = 30)+scale_y_continuous("Proportion")+l
abs(title = "Attrition Proportion by Total Amount Change Q4 to Q1",x="Total A
mount Change Q4 to Q1",fill="Attrition Flag")+theme(panel.background = elemen
t_rect(fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face
= "bold",colour = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"
))+theme(legend.background = element_rect(fill = "lightblue"))
```



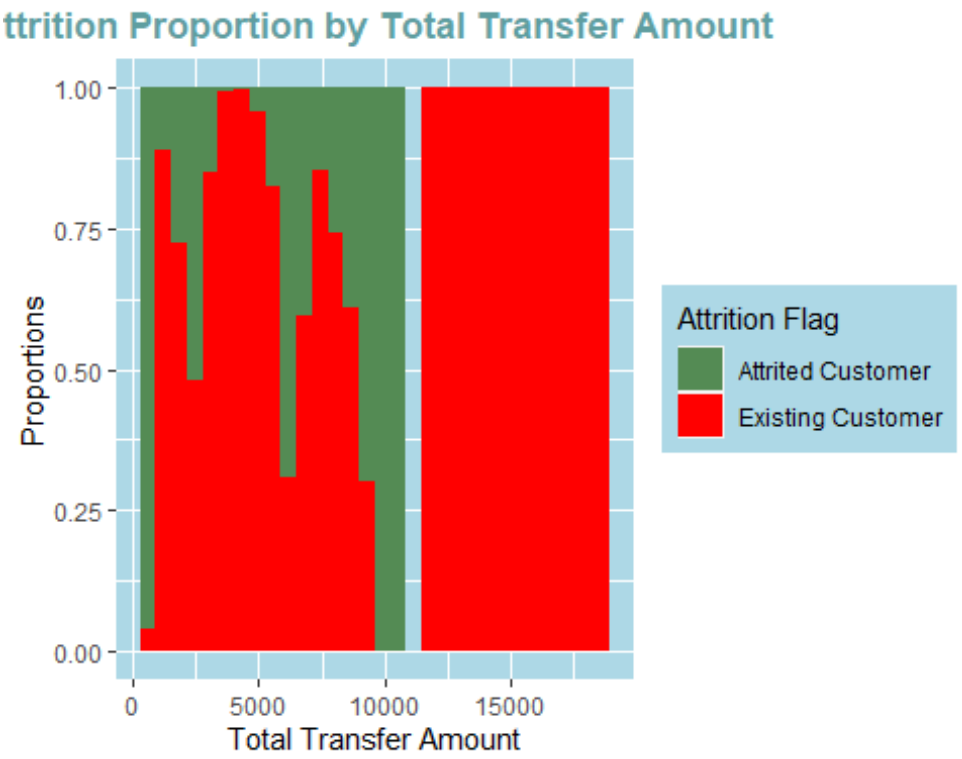
- From the above graph we get that there are maximum customers between 0.4-0.1 Total_Amt_Chng_Q4_Q1. This means that most of the customers got their amounts decreased from Q4 to Q1 which is infact a matter of concern for the bank.
- From the second graph we observe that when Total_Amt_Chng_Q4_Q1 is 0, the bank has attrition rate of 100% which is indeed an alarming stage for the bank.
- As the Total_Amt_Chng_Q4_Q1 incrases the attrition rate decreases.
- When Total_Amt_Chng_Q4_Q1 is greater than 1.5, we see that there is 0 attrition rate which is a good sign.
- The Bank should focus mainly on those customers who have Total_Amt_Chng_Q4_Q1 less than 1, because they are more prone to leaving the credit card service of the bank.
- Indeed, the Bank should try to know the reason behind their decreasing bank balance and frame the policies accordingly.

Attrition_Flag - Total_Trans_Amt

```
summary(BankChurners$Total_Trans_Amt)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##     510    2156    3899    4404    4741   18484

ggplot(data = BankChurners,aes(x=Total_Trans_Amt,fill=Attrition_Flag))+geom_h
istogram(position = "stack",bins = 30)+labs(title = "Attrition Count by Total
Transfer Amount",x="Total Transfer Amount",fill="Attrition Flag")+theme(panel
.background = element_rect(fill = "lightblue"))+theme(plot.title = element_te
xt(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(values=
c("palegreen4","red"))+theme(legend.background = element_rect(fill = "lightbl
ue"))
```



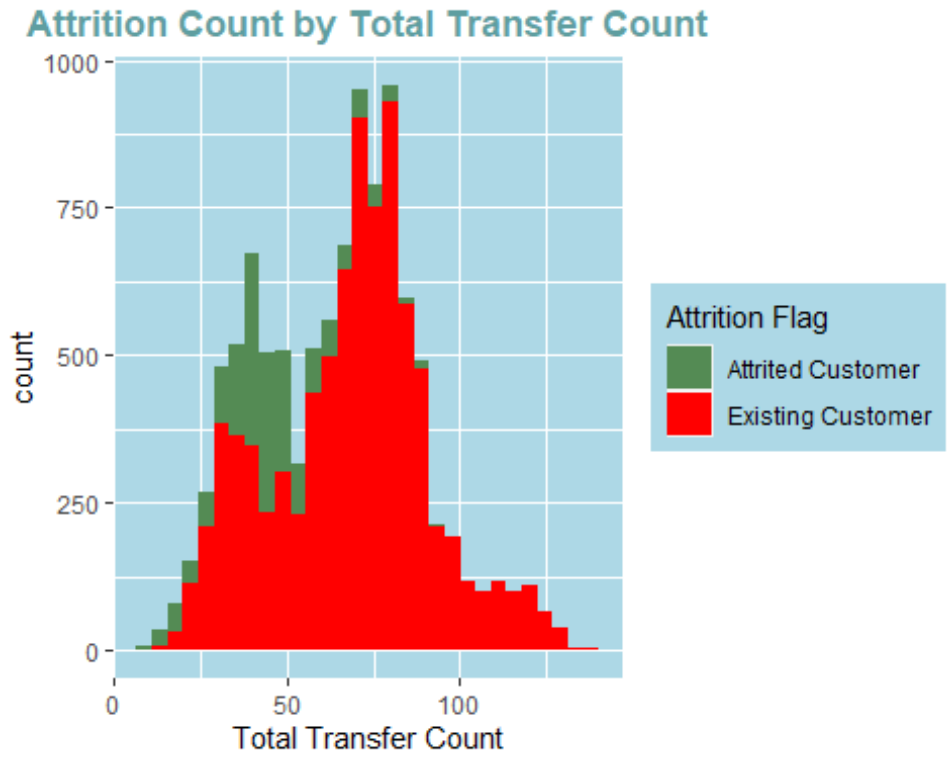

```
ggplot(data = BankChurners,aes(x=Total_Trans_Amt,fill=Attrition_Flag))+geom_h
istogram(position = "fill",bins = 30)+labs(title = "Attrition Proportion by T
otal Transfer Amount",x="Total Transfer Amount",fill="Attrition Flag")+theme(
panel.background = element_rect(fill = "lightblue"))+theme(plot.title = eleme
nt_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(va
lues=c("palegreen4","red"))+theme(legend.background = element_rect(fill = "li
ghtblue"))+scale_y_continuous("Proportions")
```



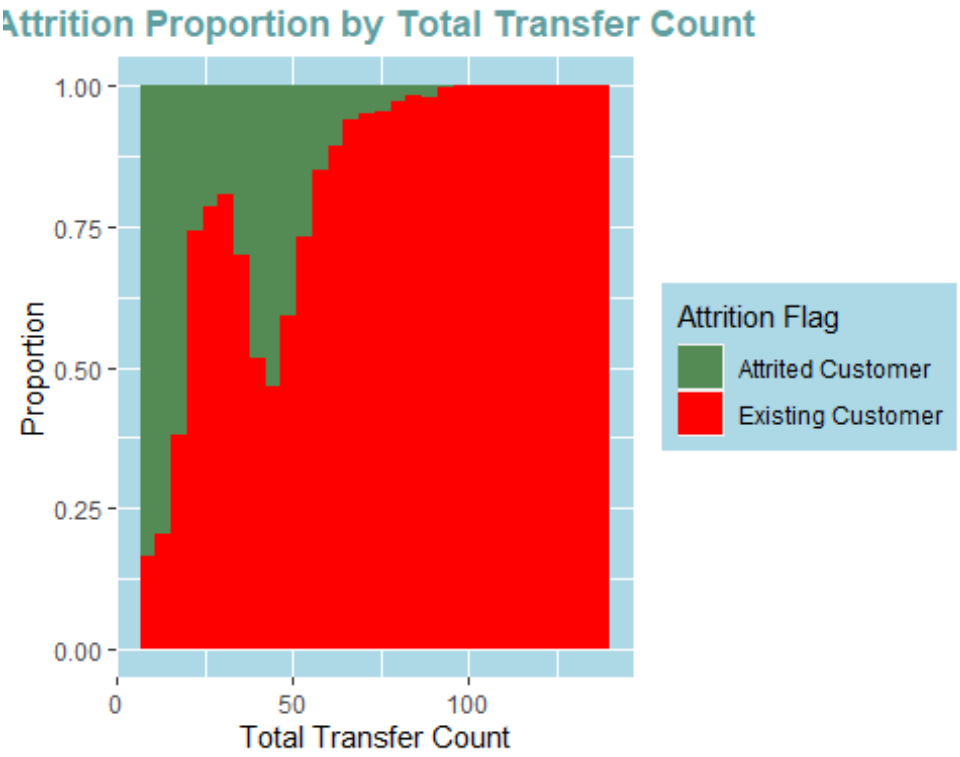
- The above graphs do not seem so regular. One thing we observe is that more than half of the customers lie below 6000. It means more than half of the bank's customers have their total transfer amount less than 6000.
- Also we observe that that the customers with the highest Total Transfer Amount approx. 13000 and above have 0 attrition rate which is indeed a good thing.
- Near 0, there is 100% attrition rate which is the area of concern for the bank. After slightly up an down again the attrition rate is very high near 2500 Total Transfer Amount and again near 6000. After that we see that the attrition rate continuously increases until Total Transfer Amount reaches nearly 1000 where we again have 100% attrition rate.
- Overall, we see that the area of concern is Total Transfer AMount near 0 and from 5000 to 10000 Total Transfer Amount.

Attrition_Flag - Total_Trans_Count

```
summary(BankChurners$Total_Trans_Ct)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   10.00   45.00   67.00   64.86   81.00  139.00

ggplot(data = BankChurners,aes(x=Total_Trans_Ct,fill=Attrition_Flag))+geom_hi
stogram(position = "stack",bins = 30)+labs(title = "Attrition Count by Total
Transfer Count",x="Total Transfer Count",fill="Attrition Flag")+theme(panel.b
ackground = element_rect(fill = "lightblue"))+theme(plot.title = element_text
(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(values=c(
"palegreen4","red"))+theme(legend.background = element_rect(fill = "lightblue
"))
```




```
ggplot(data = BankChurners,aes(x=Total_Trans_Ct,fill=Attrition_Flag))+geom_hi
stogram(position = "fill",bins = 30)+labs(title = "Attrition Proportion by To
tal Transfer Count",x="Total Transfer Count",fill="Attrition Flag")+theme(pan
el.background = element_rect(fill = "lightblue"))+theme(plot.title = element_
text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(value
s=c("palegreen4","red"))+theme(legend.background = element_rect(fill = "light
blue"))+scale_y_continuous("Proportion")
```



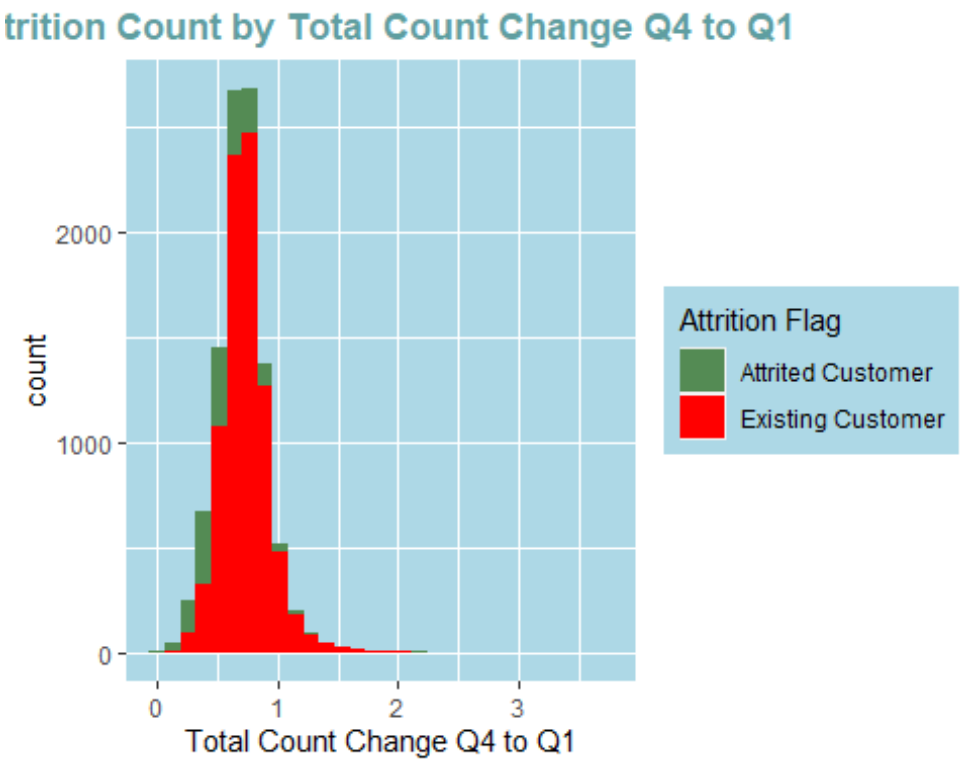
- From the above graphs we get that a major chunk of the customer has Total Transfer count from 50 to 100.
- Also we get that we have higher rate of attrition when the Total Transfer Count is les than 50.
- From the proportionate graph above we observe that as the Total Transfer Count increases, the attrition rate decreases.
- But one unusual pattern we observe that from 25 Total Transfer Count to 50 Total Transfer Count, again the attrition rate increases.
- Overall, the area of concern for the bank should be from 0 Total Transfer Count to 50 Total Transfer Count.

Attrition_Flag - Total_Count_Change_Q4_Q1

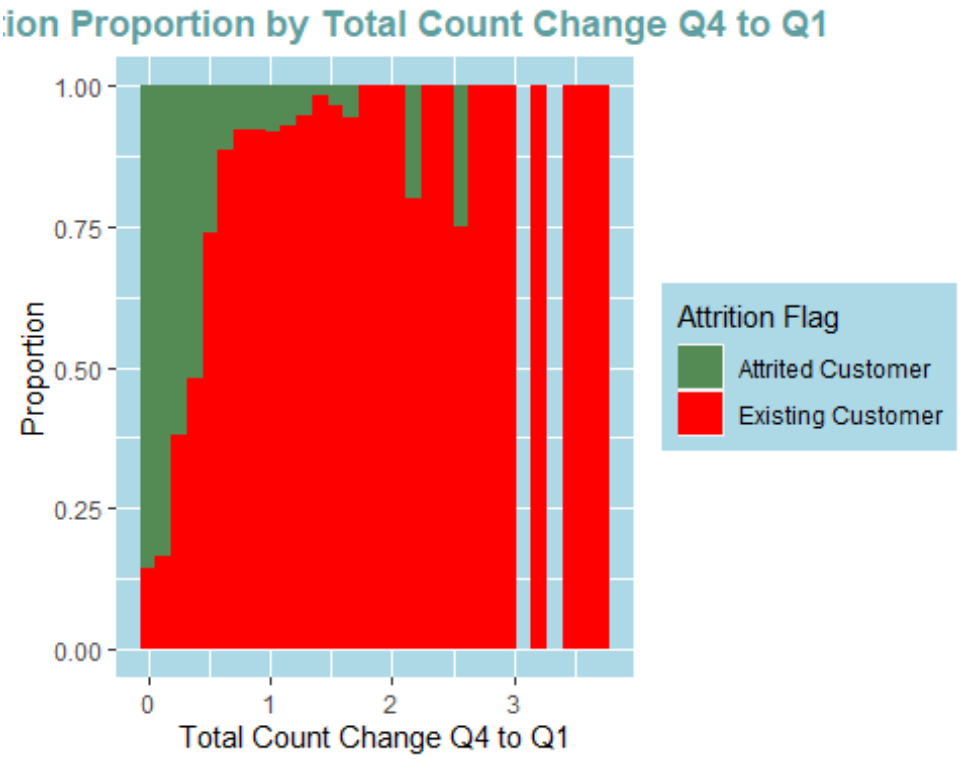
```
summary(BankChurners$Total_Ct_Chng_Q4_Q1)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.5820  0.7020  0.7122  0.8180  3.7140
```

```
ggplot(data = BankChurners,aes(x=Total_Ct_Chng_Q4_Q1,fill=Attrition_Flag))+geom_histogram(position = "stack",bins = 30)+labs(title = "Attrition Count by Total Count Change Q4 to Q1",x="Total Count Change Q4 to Q1",fill="Attrition Flag")+theme(panel.background = element_rect(fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.background = element_rect(fill = "lightblue"))
```




```
ggplot(data = BankChurners,aes(x=Total_Ct_Chng_Q4_Q1,fill=Attrition_Flag))+geom_histogram(position = "fill",bins = 30)+labs(title = "Attrition Proportion by Total Count Change Q4 to Q1",x="Total Count Change Q4 to Q1",fill="Attrition Flag")+theme(panel.background = element_rect(fill = "lightblue"))+theme(plot.title = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_manual(values=c("palegreen4","red"))+theme(legend.background = element_rect(fill = "lightblue"))+scale_y_continuous("Proportion")
```

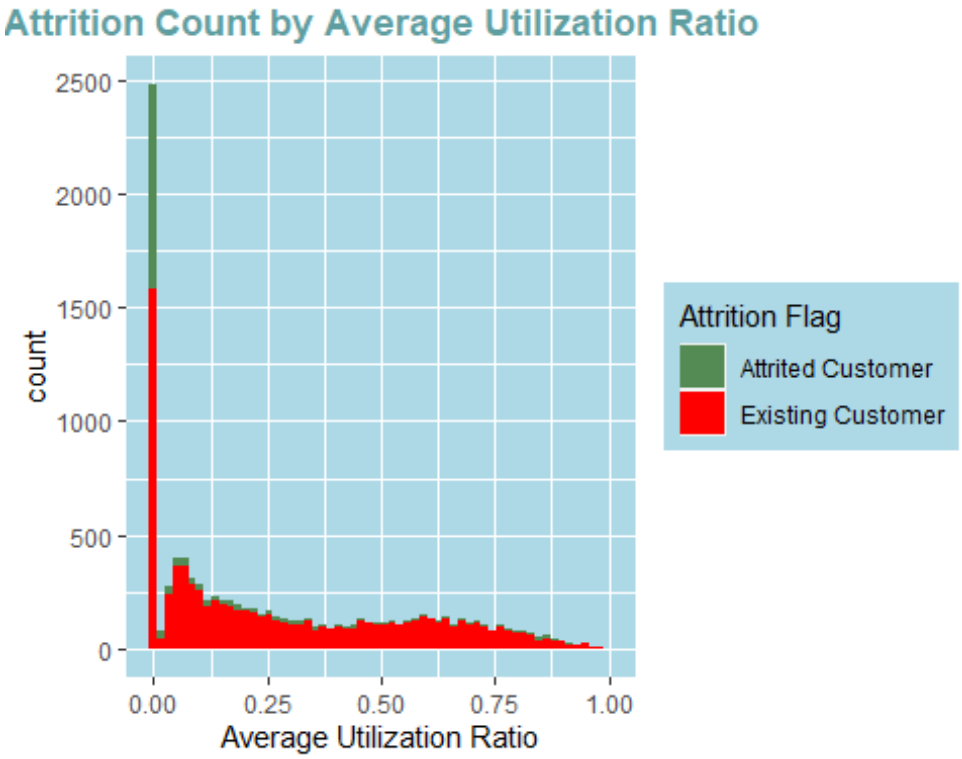


- From the above graph we get that approx half of the bank's customers lie between 0.5 to 1.
- From the proportionate graph above we see that attrition rate decreases as T0tal Count Change decreases. We also have nearly 75% attrition rate near 0 indicating that the customers who have decreased their Total Transaction Count in Q4 compared to Q1 have been more probable to leave the credit card facility.
- The customers with more than 2.5 Total Count Change Q4_Q1 have 0 attrition rate.

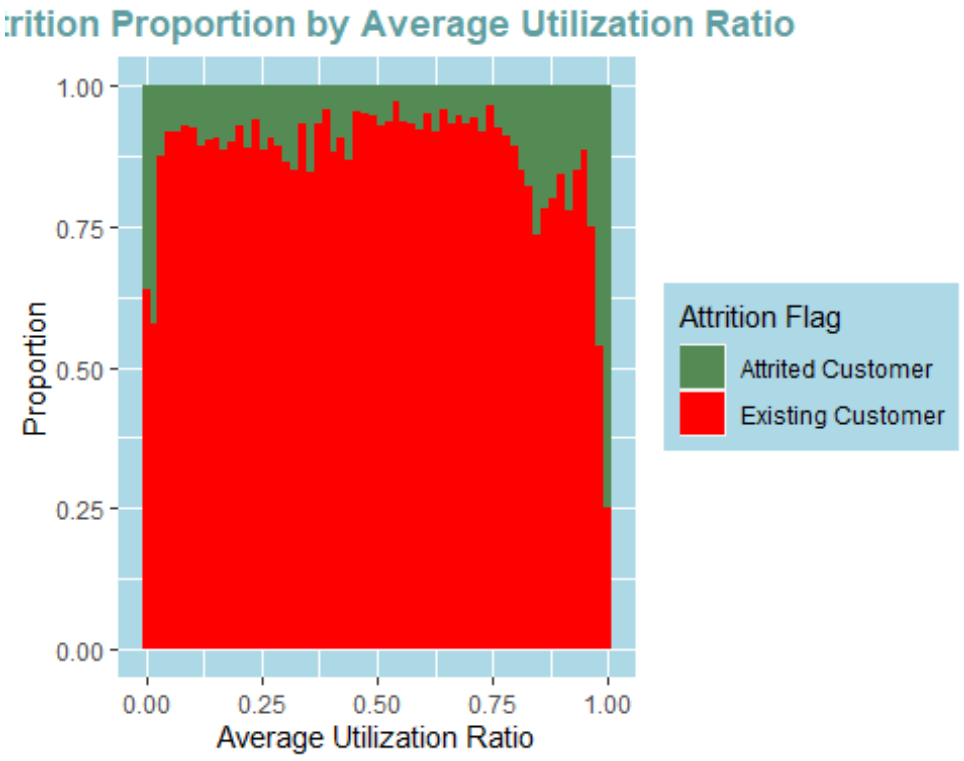
Attrition_Flag - Average_Utilization_Ratio

```
summary(BankChurners$Avg_Utilization_Ratio)

##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0000  0.0230  0.1760  0.2749  0.5030  0.9990

ggplot(data = BankChurners,aes(x=Avg_Utilization_Ratio,fill=Attrition_Flag))+
geom_histogram(position = "stack",bins = 60)+labs(title = "Attrition Count by
Average Utilization Ratio",x="Average Utilization Ratio",fill="Attrition Flag
```

```
")+theme(panel.background = element_rect(fill = "lightblue"))+theme(plot.titl
e = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_fill_
manual(values=c("palegreen4","red"))+theme(legend.background = element_rect(f
ill = "lightblue"))
```




```
ggplot(data = BankChurners,aes(x=Avg_Utilization_Ratio,fill=Attrition_Flag))+
geom_histogram(position = "fill",bins = 60)+labs(title = "Attrition Proportio
n by Average Utilization Ratio",x="Average Utilization Ratio",fill="Attrition
Flag")+theme(panel.background = element_rect(fill = "lightblue"))+theme(plot.
title = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+scale_f
ill_manual(values=c("palegreen4","red"))+theme(legend.background = element_re
ct(fill = "lightblue"))+scale_y_continuous("Proportion")
```

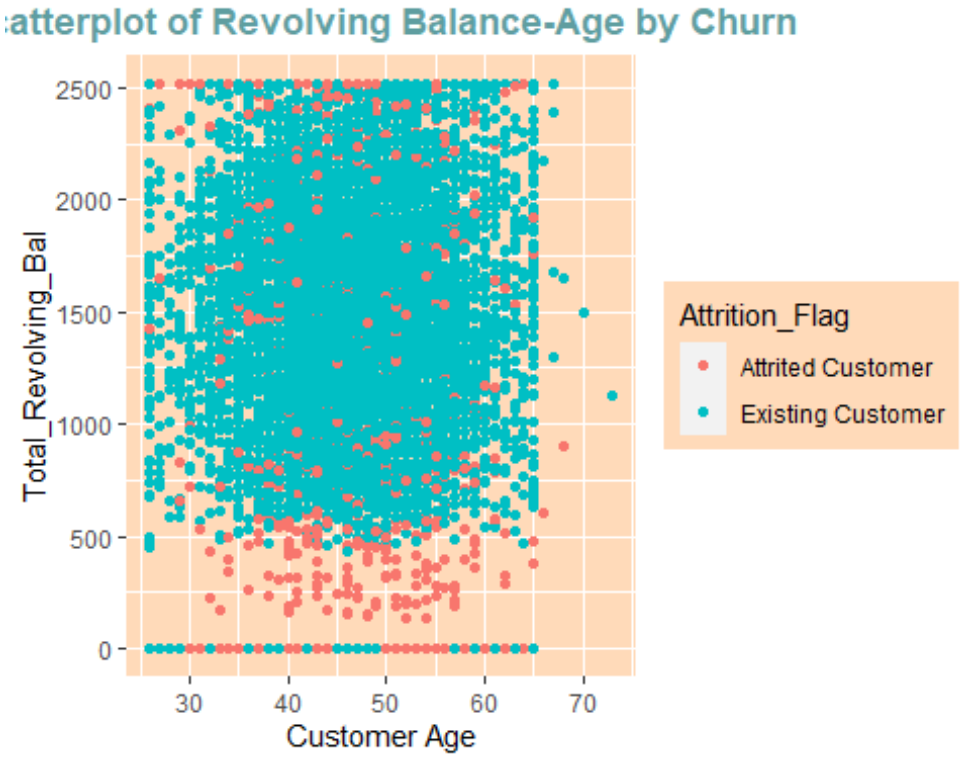


- From the above graphs we observe that there are 2500 customers who have their average utilization ratio 0 or near to 0.
- After that we see that as the average utilization ratio of card increases, their count decreases from 400 to nearly 0.
- From the proportionate graph we observe that the bank has maximm attrition rate of approx. 25% near 0, which also has maximum number of customers. This should clearly be the concern of the bank to deal with the customers with 0 average utilization ratio with great attention.
- Next we see that the attrition rate remains almost constant and even decreases as the average utilization ratio increases except for the average utilization ratio nearly 1 where the customers have the maximum churn rate of approximately 75%. The bank should handle this group very carefully.

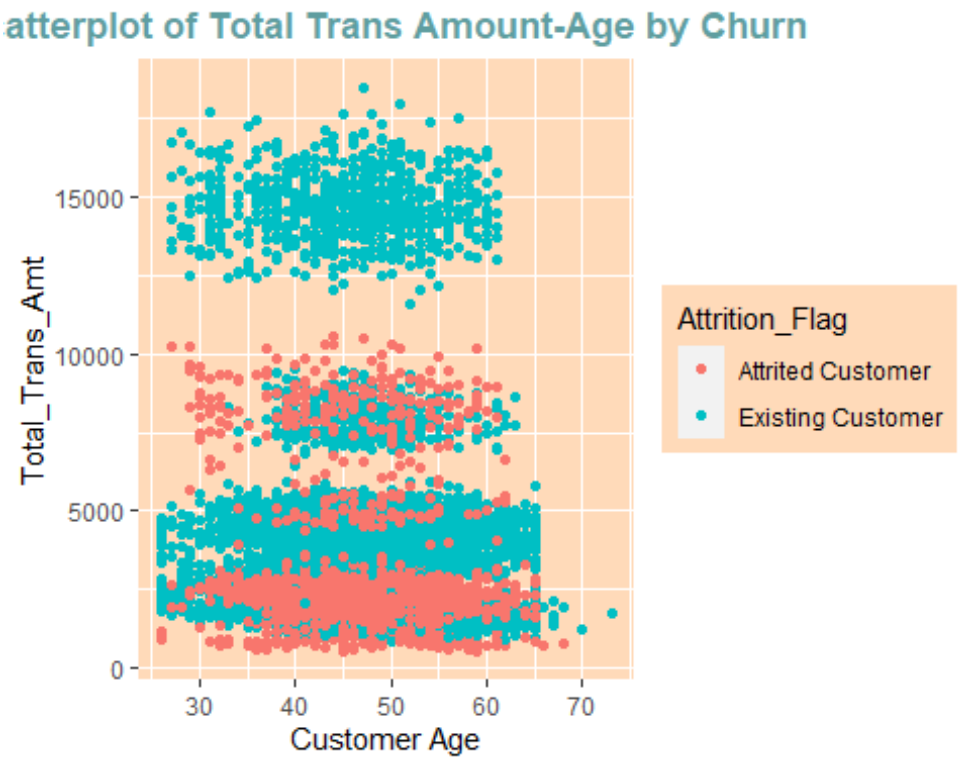
D. EXPLORING MULTIVARIATE RELATIONSHIPS

Visualizing various continuous variables with Customer Age colored by Attrition Flag

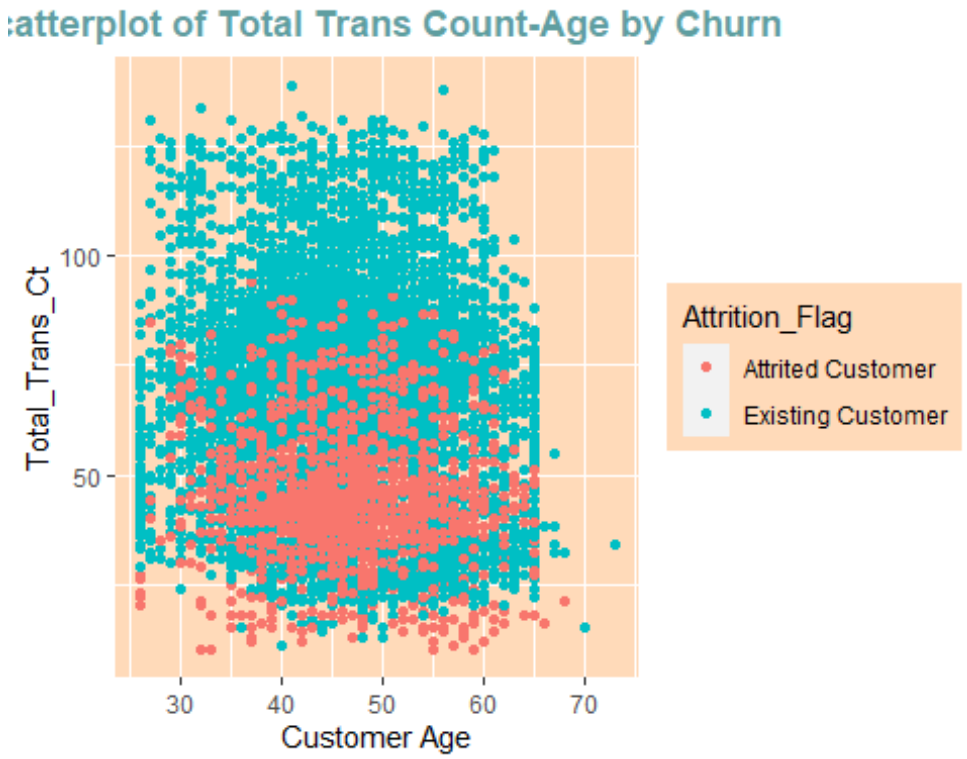
```
ggplot(data = BankChurners,aes(y=Total_Revolving_Bal,x=Customer_Age,col=Attri
tion_Flag))+geom_point()+labs(title = "Scatterplot of Revolving Balance-Age b
y Churn",x="Customer Age",fill="Attrition Flag")+theme(panel.background = ele
ment_rect(fill = "peachpuff"))+theme(plot.title = element_text(hjust = 0.5,fa
ce = "bold",colour = "cadetblue"))+theme(legend.background = element_rect(fil
l = "peachpuff"))
```



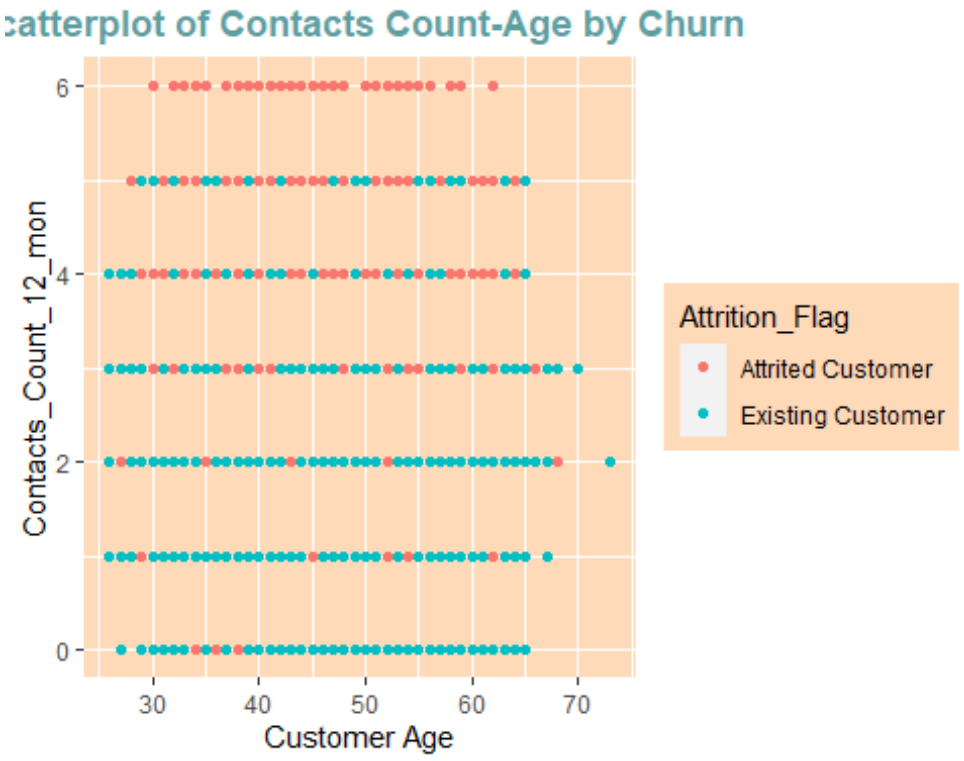

```
ggplot(data = BankChurners,aes(y=Total_Trans_Amt,x=Customer_Age,col=Attrition
_Flag))+geom_point()+labs(title = "Scatterplot of Total Trans Amount-Age by C
hurn",x="Customer Age",fill="Attrition Flag")+theme(panel.background = elemen
t_rect(fill = "peachpuff"))+theme(plot.title = element_text(hjust = 0.5,face
= "bold",colour = "cadetblue"))+theme(legend.background = element_rect(fill =
"peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Total_Trans_Ct,x=Customer_Age,col=Attrition_
Flag))+geom_point()+labs(title = "Scatterplot of Total Trans Count-Age by Chu
rn",x="Customer Age",fill="Attrition Flag")+theme(panel.background = element_
rect(fill = "peachpuff"))+theme(plot.title = element_text(hjust = 0.5,face =
"bold",colour = "cadetblue"))+theme(legend.background = element_rect(fill = "
peachpuff"))
```



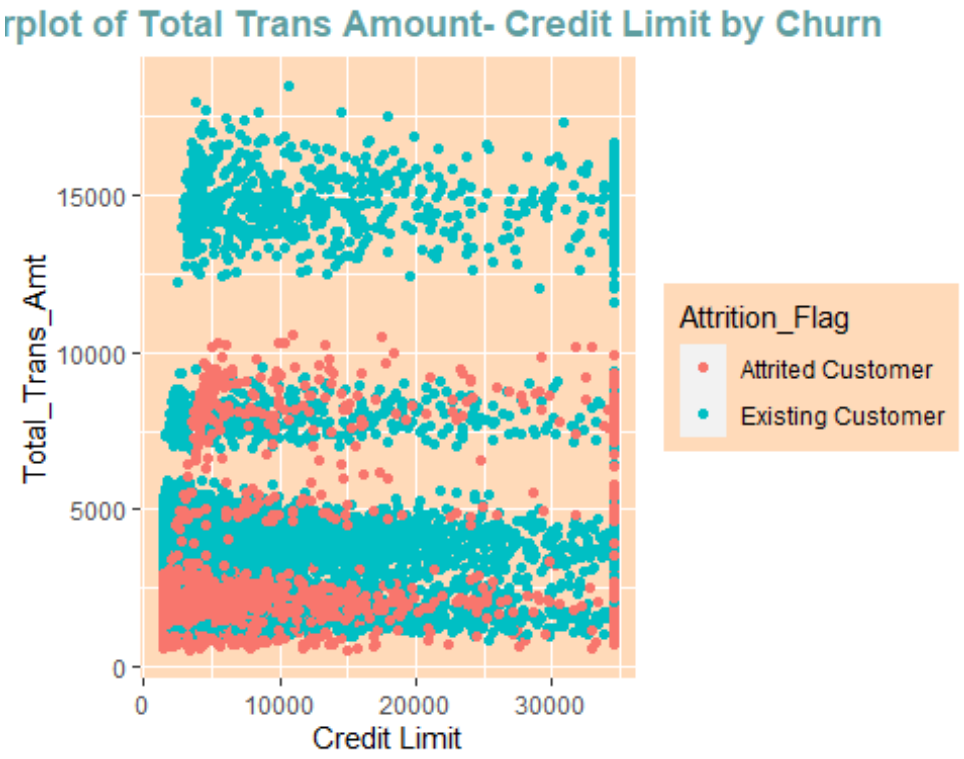

```
ggplot(data = BankChurners,aes(y=Contacts_Count_12_mon,x=Customer_Age,col=Att
rition_Flag))+geom_point()+labs(title = "Scatterplot of Contacts Count-Age by
Churn",x="Customer Age",fill="Attrition Flag")+theme(panel.background = eleme
nt_rect(fill = "peachpuff"))+theme(plot.title = element_text(hjust = 0.5,face
= "bold",colour = "cadetblue"))+theme(legend.background = element_rect(fill =
"peachpuff"))
```



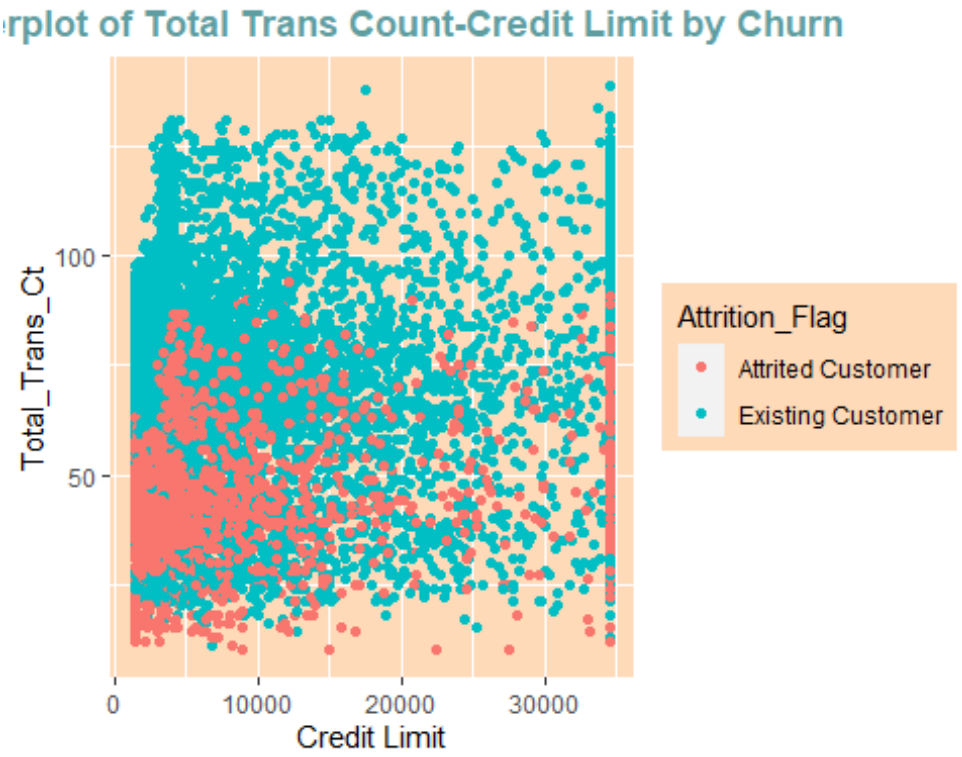
- From the first graph we observe that customers of all age are more attrited customers who have total revolving balance less than 500.
- From the second graph we see that customers of all age have more attrition rate when there total transfer amount is less than 3000. Also as the Total Transfer Amount from 750 to 1000, customers have higher attrition rate. The customers with more than 12500 Total Transfer Amount has no attrition at all.
- From the third figure above we observe that the area between Total Transfer Count of 25-75 and between 30-60 age group, the customers tend to churn the credit card service more which is the area of concern.
- From the fourth graph above, we observe that as the number of contact counts reaches 4 and beyond, the attrition rate increases in the customers of all age group.

Visualizing various continuous variables with Credit Limit colored by Attrition Flag

```
ggplot(data = BankChurners,aes(y=Total_Trans_Amt,x=Credit_Limit,col=Attrition
_Flag))+geom_point()+labs(title = "Scatterplot of Total Trans Amount- Credit
Limit by Churn",x="Credit Limit",fill="Attrition Flag")+theme(panel.backgroun
d = element_rect(fill = "peachpuff"))+theme(plot.title = element_text(hjust =
0.5,face = "bold",colour = "cadetblue"))+theme(legend.background = element_re
ct(fill = "peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Total_Trans_Ct,x=Credit_Limit,col=Attrition_
Flag))+geom_point()+labs(title = "Scatterplot of Total Trans Count-Credit Lim
it by Churn",x="Credit Limit",fill="Attrition Flag")+theme(panel.background =
element_rect(fill = "peachpuff"))+theme(plot.title = element_text(hjust = 0.5
,face = "bold",colour = "cadetblue"))+theme(legend.background = element_rect(
fill = "peachpuff"))
```

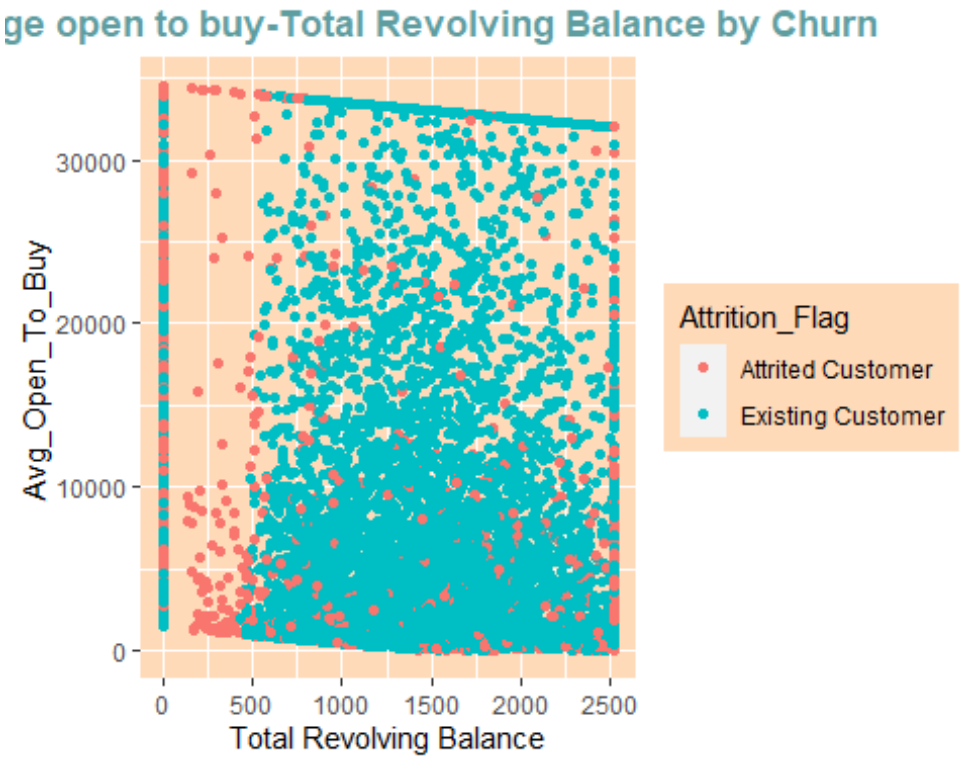


- From the first figure above, we see that the customers with less than 10000 credit limit and Total Transfer Amount less than 3500 have the highest attrition rate followed by the customers with nearly 5000 credit limit and Total Transfer Amount from 7500-10000. The customers with more than 12500 Total Transfer Amount have no attrition as we have seen earlier.
- From the second figure above we see that the customers with less than 10000 credit limit and less than 60 Total Transaction Count have high attrition rate. Also, we see that the customers with nearly 35000 credit limit and less than 80 Total transfer count have almost 90% attrition rate which is the area of concern for the bank.

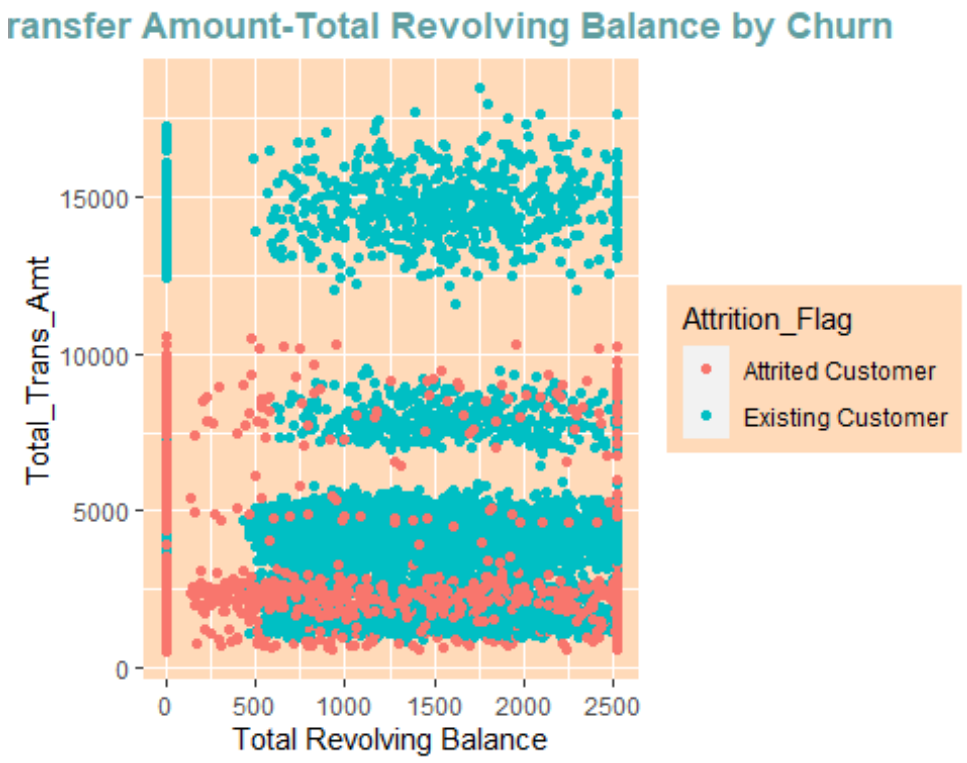
Visualizing various continuous variables with Total Revolving Balance colored by Attrition Flag

```
ggplot(data = BankChurners,aes(y=Avg_Open_To_Buy,x=Total_Revolving_Bal,col=At
trition_Flag))+geom_point()+labs(title = "Average open to buy-Total Revolving
Balance by Churn",x="Total Revolving Balance",fill="Attrition Flag")+theme(pa
nel.background = element_rect(fill = "peachpuff"))+theme(plot.title = element
```

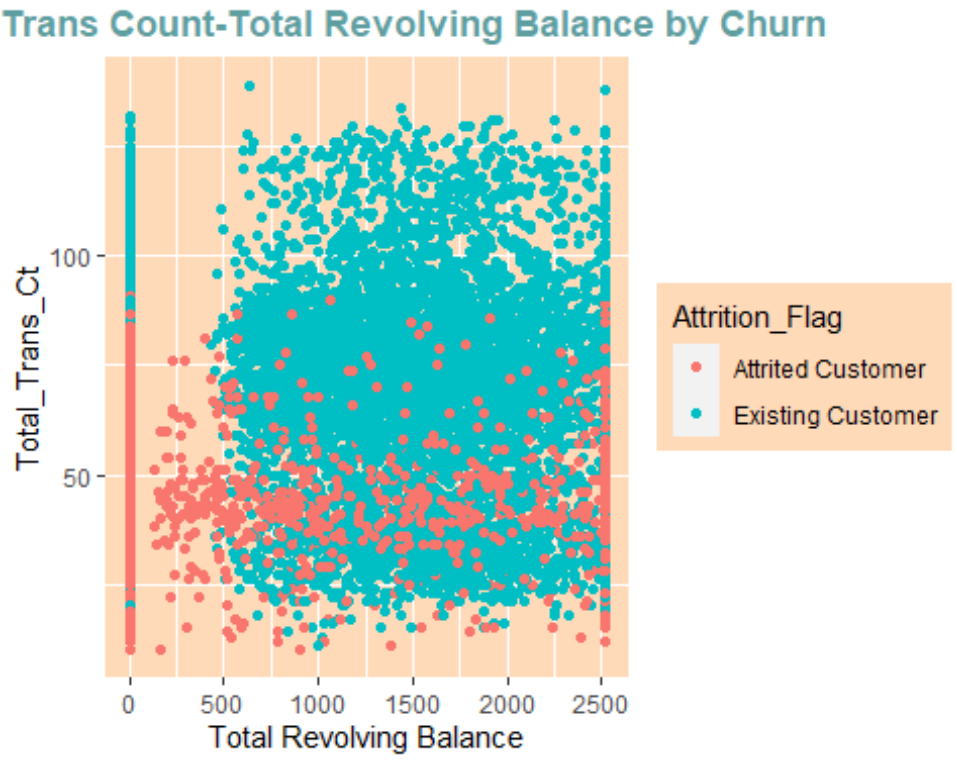
```
_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backgroun
d = element_rect(fill = "peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Total_Trans_Amt,x=Total_Revolving_Bal,col=At
trition_Flag))+geom_point()+labs(title = "Total Transfer Amount-Total Revolvi
ng Balance by Churn",x="Total Revolving Balance",fill="Attrition Flag")+theme
(panel.background = element_rect(fill = "peachpuff"))+theme(plot.title = elem
ent_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backgr
ound = element_rect(fill = "peachpuff"))
```



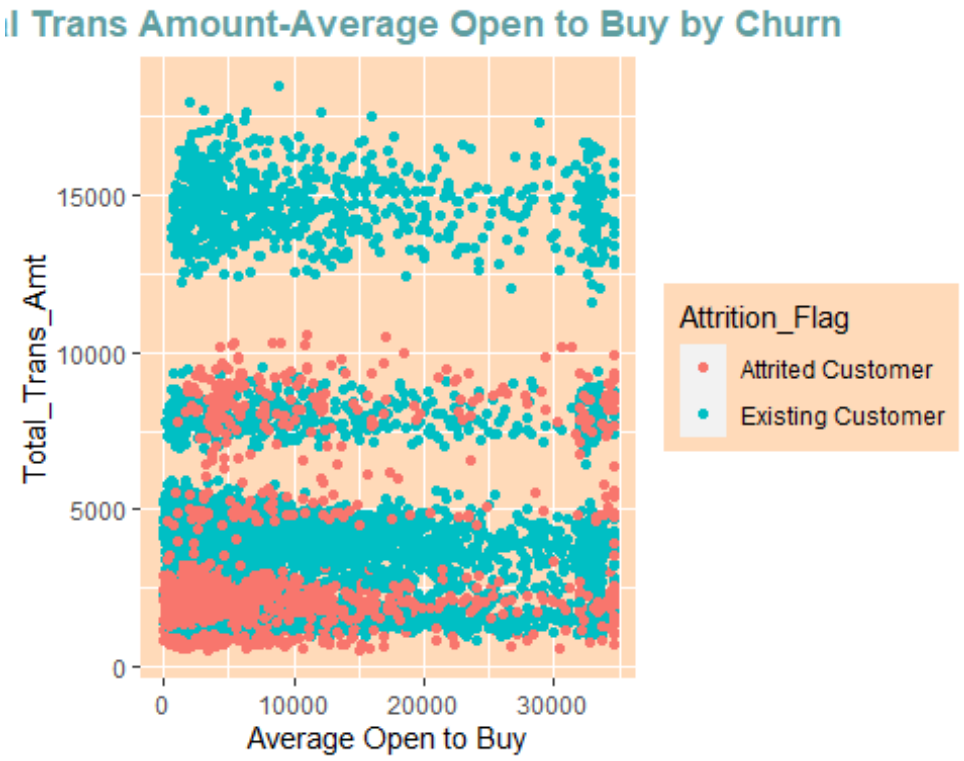

```
ggplot(data = BankChurners,aes(y=Total_Trans_Ct,x=Total_Revolving_Bal,col=Att
rition_Flag))+geom_point()+labs(title = "Total Trans Count-Total Revolving Ba
lance by Churn",x="Total Revolving Balance",fill="Attrition Flag")+theme(pane
l.background = element_rect(fill = "peachpuff"))+theme(plot.title = element_t
ext(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.background
= element_rect(fill = "peachpuff"))
```



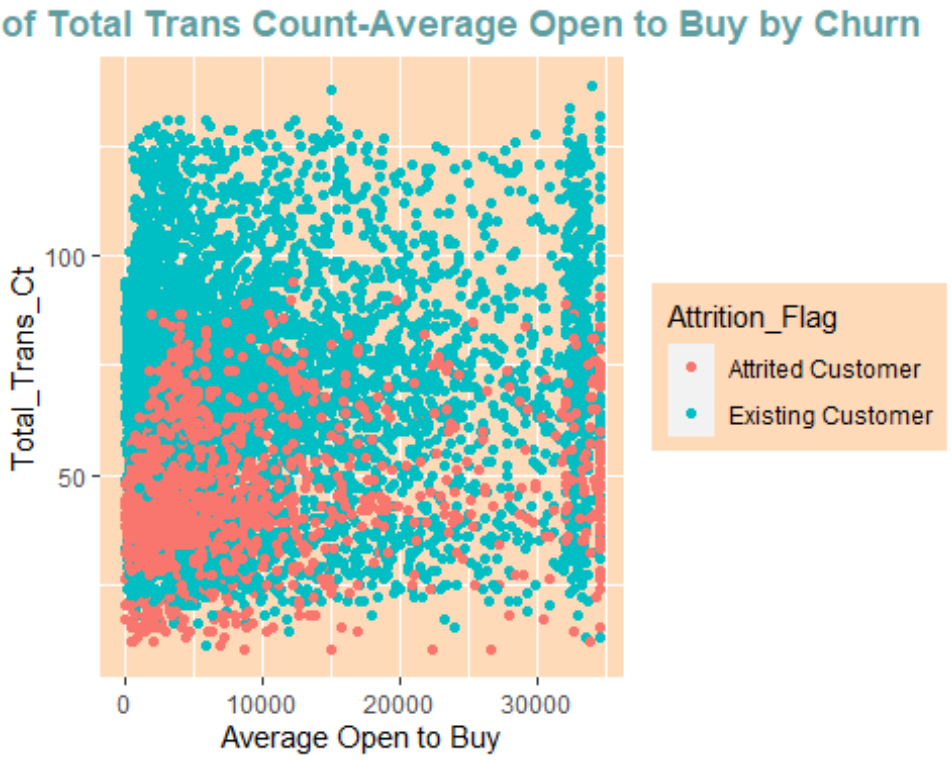
- From the first graph above we see that the customers with 0-5000 Total Revolving Balance have the high attrition rate.
- From the second graph above we see that the customers with 0 Revolving Balance have the high attrition rate as we have seen earlier. Also we see that the customers with less than 3000 Total Transfer Amount have high attrition rate. Also the customers Total Revolving Balance above 2500 have high attrition Rate. The customers with Total Transfer Amount of more than 12500 have no attrition at all.
- From the third graph above we see that the customers with 0 and more than 2500 of Total Revolving Balance and Total Transaction Count of less than 100 have higher attrition rate. Also, the customers with Total Transaction Count between 30-50 have higher rate of attrition.

Visualizing various continuous variables with Average Open To Buy colored by Attrition Flag

```
ggplot(data = BankChurners,aes(y=Total_Trans_Amt,x=Avg_Open_To_Buy,col=Attrit
ion_Flag))+geom_point()+labs(title = "Total Trans Amount-Average Open to Buy
by Churn",x="Average Open to Buy",fill="Attrition Flag")+theme(panel.backgrou
nd = element_rect(fill = "peachpuff"))+theme(plot.title = element_text(hjust
= 0.5,face = "bold",colour = "cadetblue"))+theme(legend.background = element_
rect(fill = "peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Total_Trans_Ct,x=Avg_Open_To_Buy,col=Attriti
on_Flag))+geom_point()+labs(title = "Scatterplot of Total Trans Count-Average
Open to Buy by Churn",x="Average Open to Buy",fill="Attrition Flag")+theme(pa
nel.background = element_rect(fill = "peachpuff"))+theme(plot.title = element
_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backgroun
d = element_rect(fill = "peachpuff"))
```

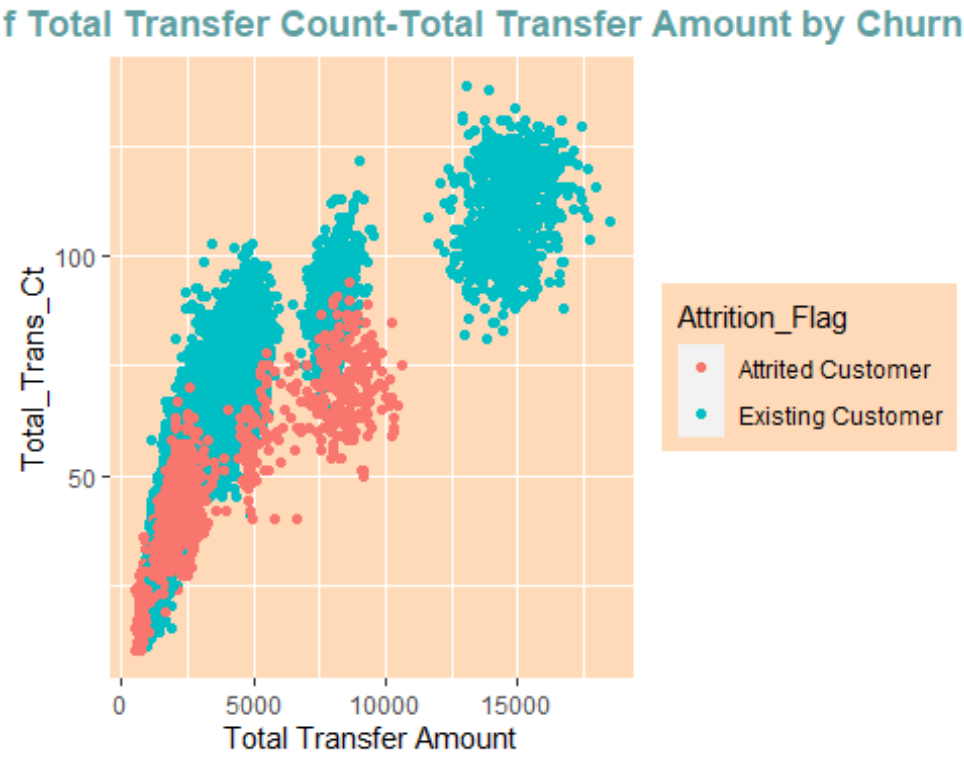


- From the graph above we see that the area of the customers with less than 20000 Average Open to Buy and Total Transfer Amount of less than 3000 have higher attrition rate followed by the area of the batch of the customers with Average Open to buy of less than 7000 and total transfer amount between 7500-1000. The customers with the total transfer amount of more than 12500 have no attrition at all as we have seen earlier also.
- From the second graph above we see that the area between the average open to buy of less than 100000 and the Total transaction count of less than 50 contains attrited customers. Alsoo the customers with average open to buy of 34000 and total transaction count of less than 75 have the high attrition.

Visualizing Total Transfer Count - Total Transfer Amount by Attrition Flag

```
ggplot(data = BankChurners,aes(y=Total_Trans_Ct,x=Total_Trans_Amt,col=Attriti
on_Flag))+geom_point()+labs(title = "Scatterplot of Total Transfer Count-Tota
l Transfer Amount by Churn",x="Total Transfer Amount",fill="Attrition Flag")+
theme(panel.background = element_rect(fill = "peachpuff"))+theme(plot.title =
```

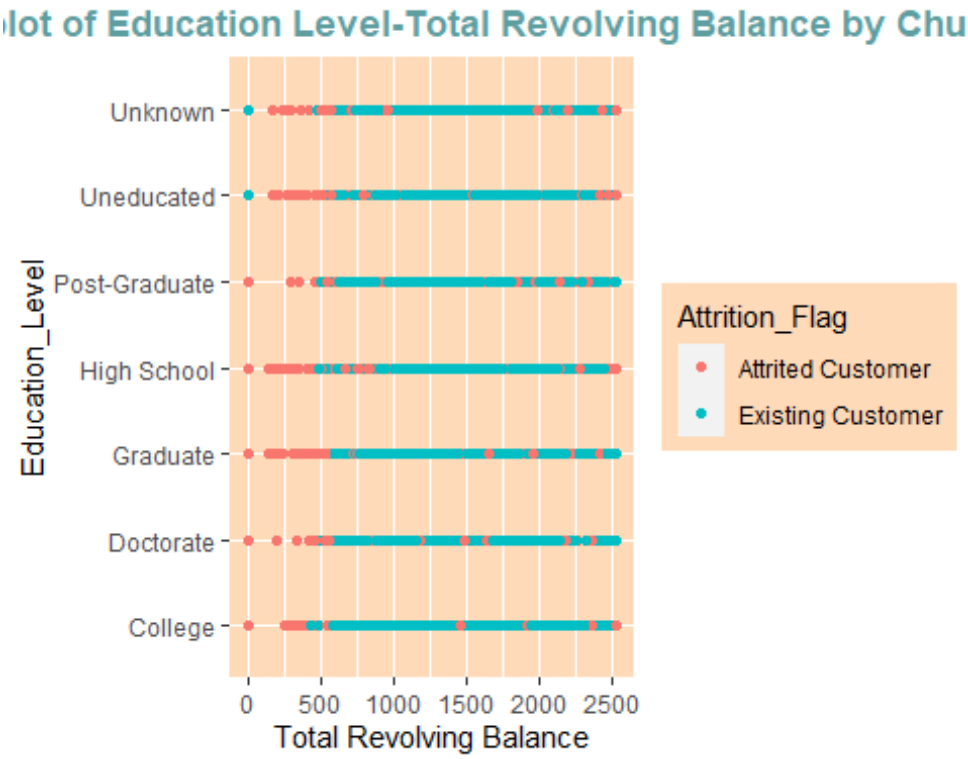
```
element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.ba
ckground = element_rect(fill = "peachpuff"))
```



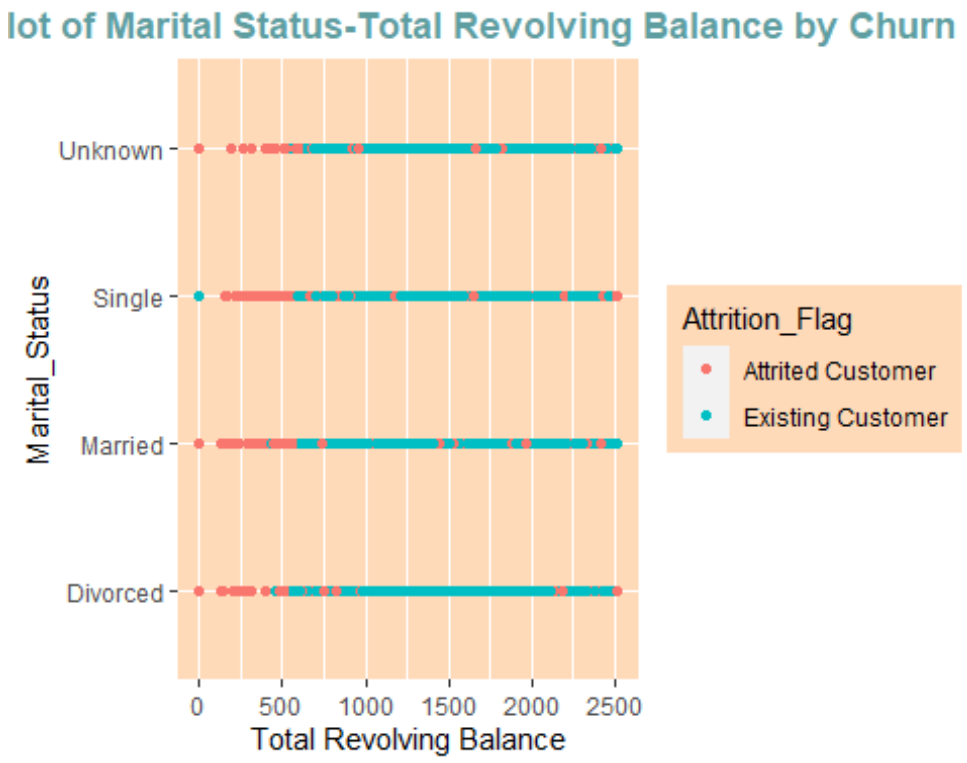
- From the above graph we see that there are 3 chunks. In the first one we see that the customers with less than 4000 Total Transfer Amount and Total Trannsfer COunt of less than 70 have the high attrition rate. Also we see high attrition near the Total Transfer Amount of 5000 and Total Transaction Count between 35-75. In the second chunk we see that the customers with Total Transfer Amount of 7500-10000 and Total Transaction Count of 50-80 have high attrition rate. While in the third chunk attrition rate is nil.

Visualizing various categorical variables with Total Revolving Balance colored by Attrition Flag

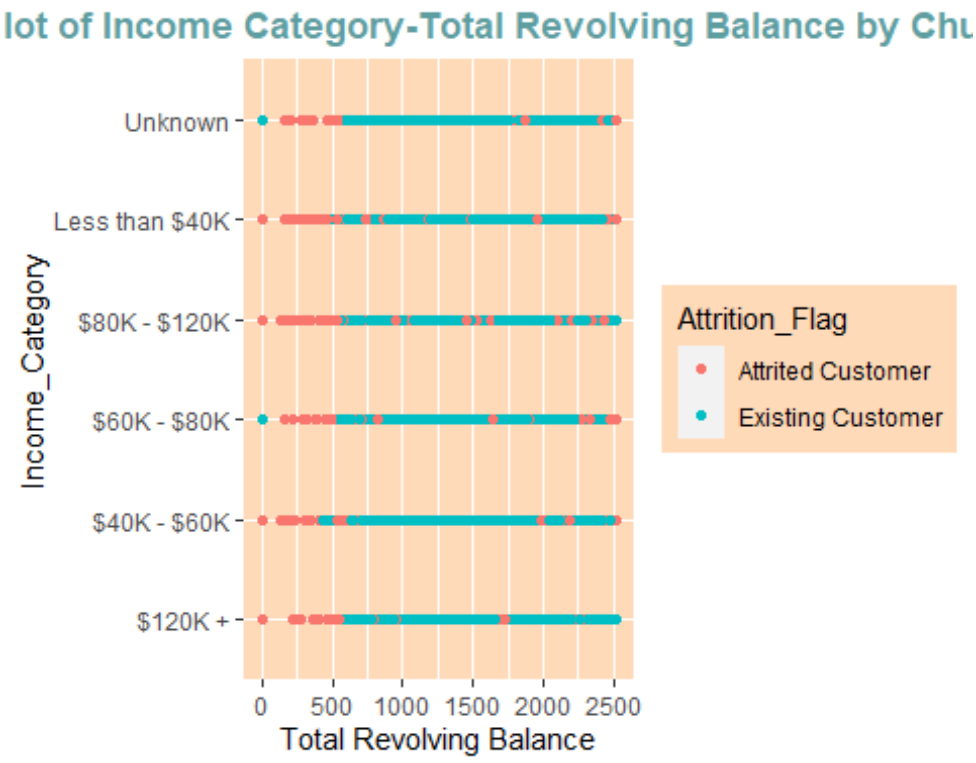
```
ggplot(data = BankChurners,aes(y=Education_Level,x=Total_Revolving_Bal,col=At
trition_Flag))+geom_point()+labs(title = "Scatterplot of Education Level-Tota
l Revolving Balance by Churn",x="Total Revolving Balance",fill="Attrition Fla
g")+theme(panel.background = element_rect(fill = "peachpuff"))+theme(plot.tit
le = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(lege
nd.background = element_rect(fill = "peachpuff"))
```



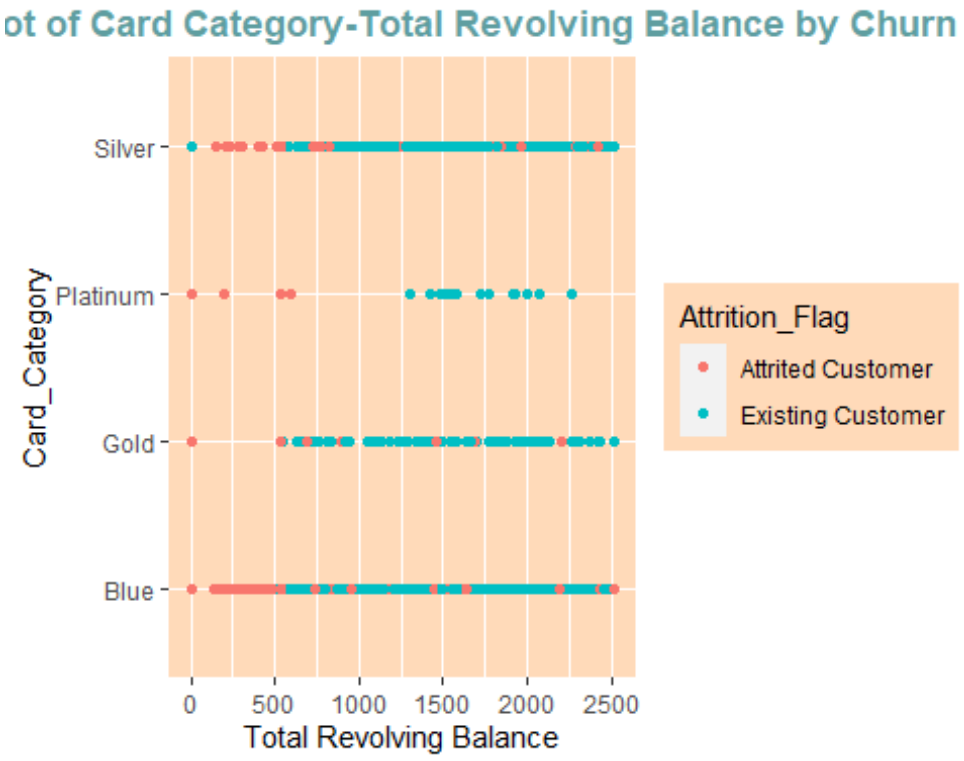

```
ggplot(data = BankChurners,aes(y=Marital_Status,x=Total_Revolving_Bal,col=Att
rition_Flag))+geom_point()+labs(title = "Scatterplot of Marital Status-Total
Revolving Balance by Churn",x="Total Revolving Balance",fill="Attrition Flag"
)+theme(panel.background = element_rect(fill = "peachpuff"))+theme(plot.title
= element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.
background = element_rect(fill = "peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Income_Category,x=Total_Revolving_Bal,col=At
trition_Flag))+geom_point()+labs(title = "Scatterplot of Income Category-Tota
l Revolving Balance by Churn",x="Total Revolving Balance",fill="Attrition Fla
g")+theme(panel.background = element_rect(fill = "peachpuff"))+theme(plot.tit
le = element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(lege
nd.background = element_rect(fill = "peachpuff"))
```



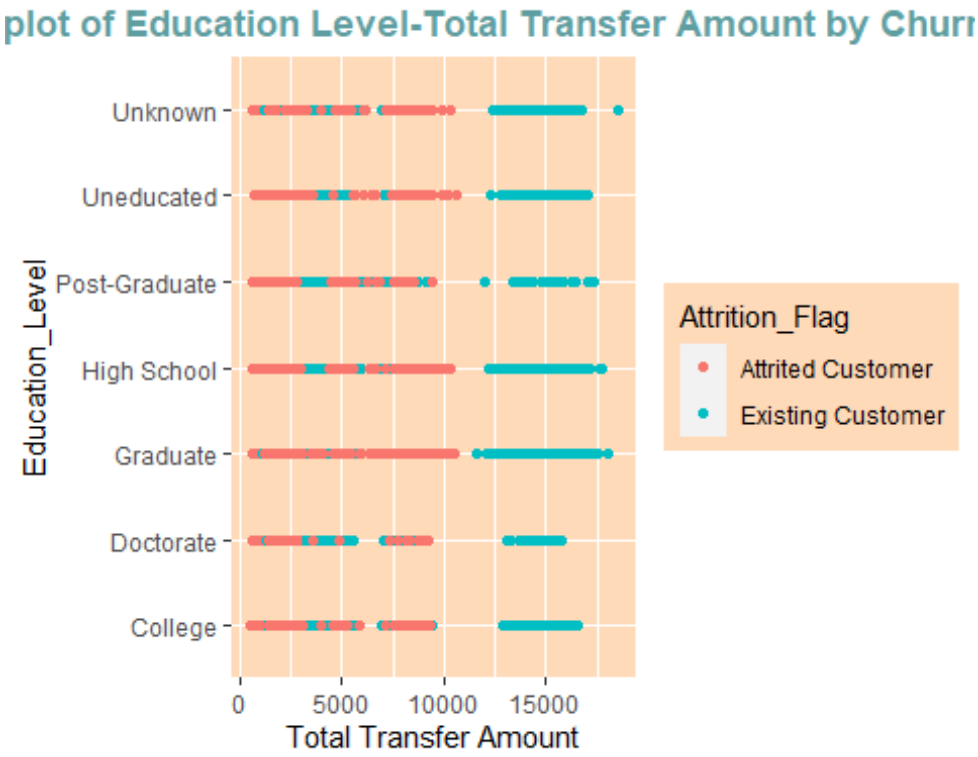

```
ggplot(data = BankChurners,aes(y=Card_Category,x=Total_Revolving_Bal,col=Attr
ition_Flag))+geom_point()+labs(title = "Scatterplot of Card Category-Total Re
volving Balance by Churn",x="Total Revolving Balance",fill="Attrition Flag")+
theme(panel.background = element_rect(fill = "peachpuff"))+theme(plot.title =
element_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.ba
ckground = element_rect(fill = "peachpuff"))
```



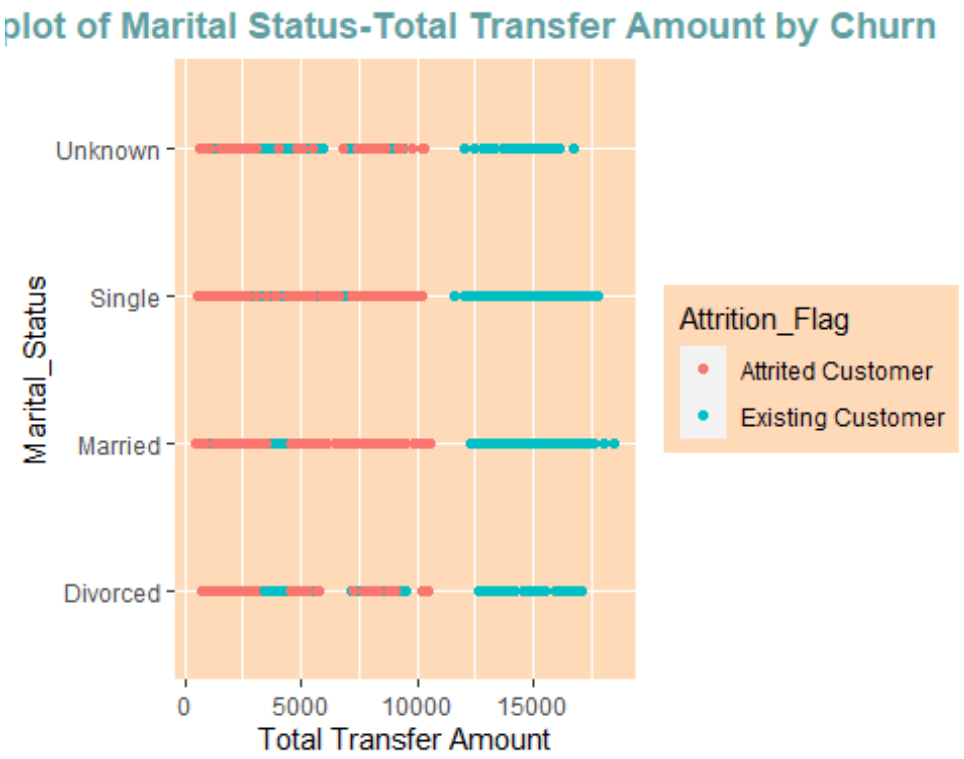
- As is clear from the earlier observation, that the total revolving balance of less than 500 shows a higher attrition rate, in all the above four figures we observe that the total revolving Balance of less than 500 has a higher rate of attrition irrespective of any of the mentioned category.

Visualizing various categorical variables with Total Transfer Amount colored by Attrition Flag

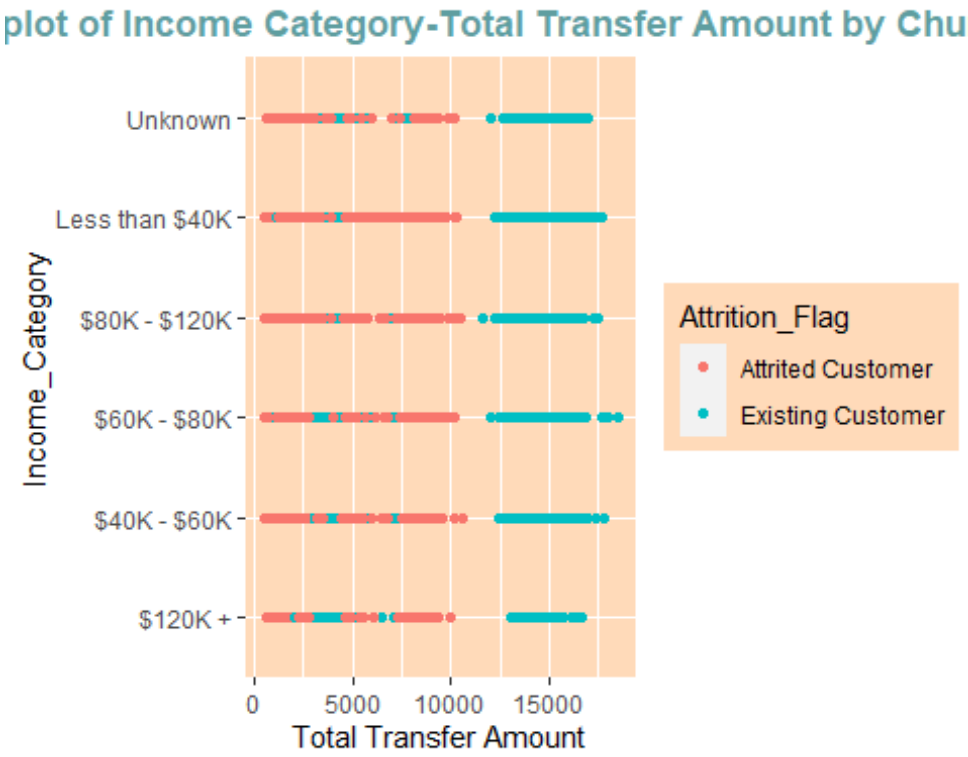
```
ggplot(data = BankChurners,aes(y=Education_Level,x=Total_Trans_Amt,col=Attrit
ion_Flag))+geom_point()+labs(title = "Scatterplot of Education Level-Total Tr
ansfer Amount by Churn",x="Total Transfer Amount",fill="Attrition Flag")+them
e(panel.background = element_rect(fill = "peachpuff"))+theme(plot.title = ele
ment_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backg
round = element_rect(fill = "peachpuff"))
```



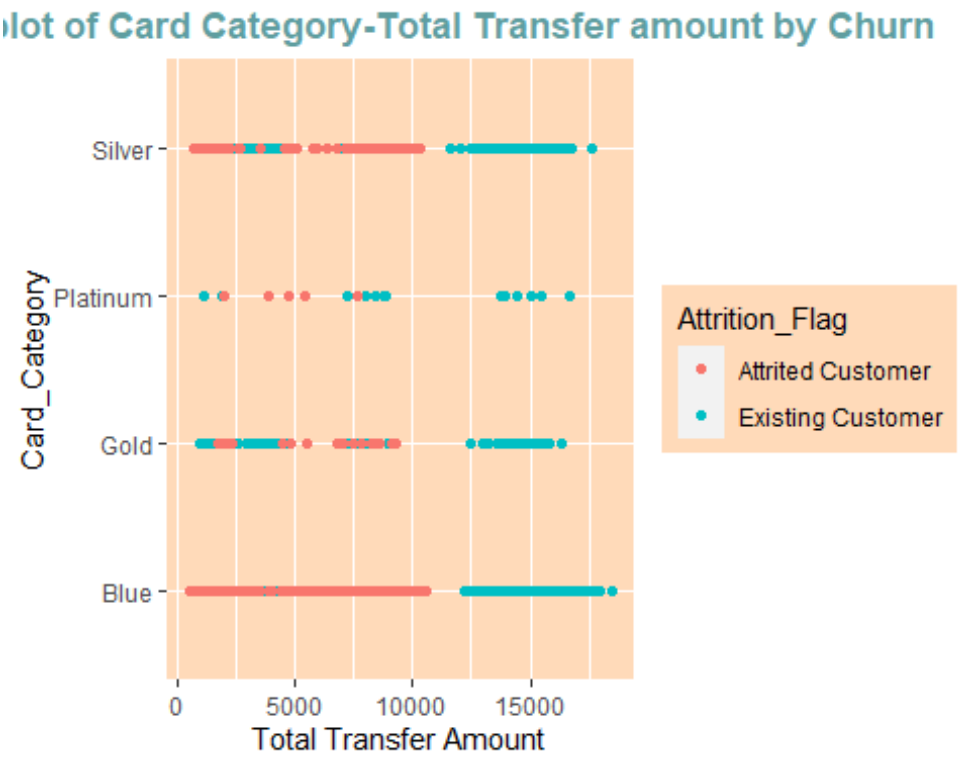

```
ggplot(data = BankChurners,aes(y=Marital_Status,x=Total_Trans_Amt,col=Attriti
on_Flag))+geom_point()+labs(title = "Scatterplot of Marital Status-Total Tran
sfer Amount by Churn",x="Total Transfer Amount",fill="Attrition Flag")+theme(
panel.background = element_rect(fill = "peachpuff"))+theme(plot.title = eleme
nt_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backgro
und = element_rect(fill = "peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Income_Category,x=Total_Trans_Amt,col=Attrit
ion_Flag))+geom_point()+labs(title = "Scatterplot of Income Category-Total Tr
ansfer Amount by Churn",x="Total Transfer Amount",fill="Attrition Flag")+them
e(panel.background = element_rect(fill = "peachpuff"))+theme(plot.title = ele
ment_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backg
round = element_rect(fill = "peachpuff"))
```



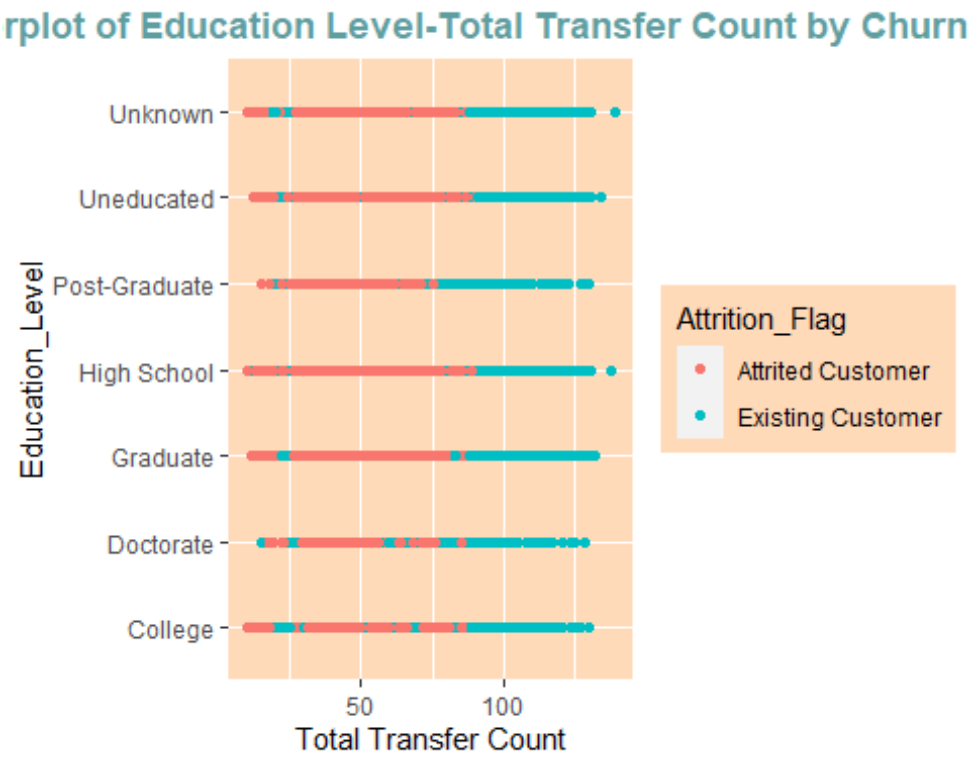

```
ggplot(data = BankChurners,aes(y=Card_Category,x=Total_Trans_Amt,col=Attritio
n_Flag))+geom_point()+labs(title = "Scatterplot of Card Category-Total Transf
er amount by Churn",x="Total Transfer Amount",fill="Attrition Flag")+theme(pa
nel.background = element_rect(fill = "peachpuff"))+theme(plot.title = element
_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backgroun
d = element_rect(fill = "peachpuff"))
```



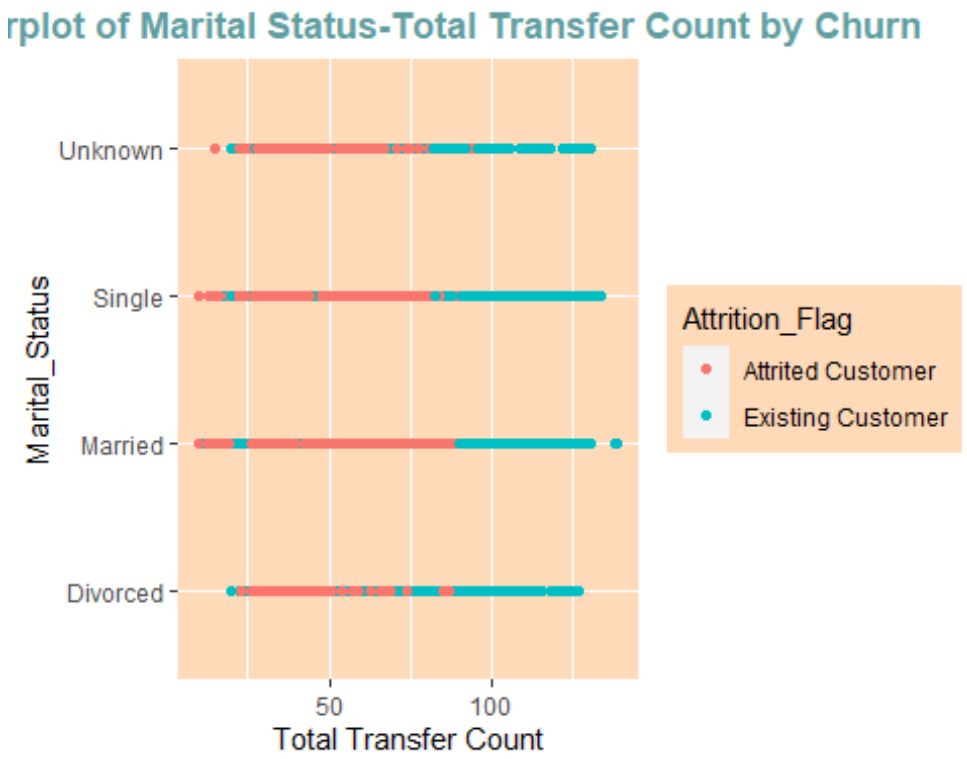
- From all the four figures above we the the similar results. The total transfer Amount from 0 to 1000 has a higher attrition rate irrespective of any of the mentioned categorical variables.

Visualizing various categorical variables with Total Transfer Count colored by Attrition Flag

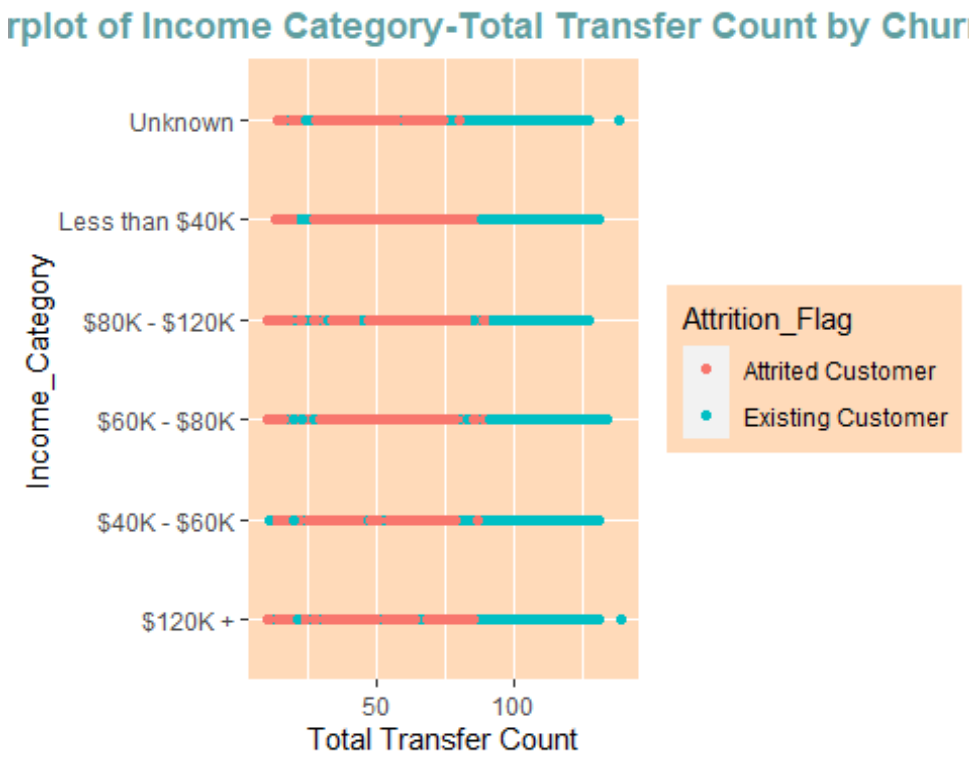
```
ggplot(data = BankChurners,aes(y=Education_Level,x=Total_Trans_Ct,col=Attriti
on_Flag))+geom_point()+labs(title = "Scatterplot of Education Level-Total Tra
nsfer Count by Churn",x="Total Transfer Count",fill="Attrition Flag")+theme(p
anel.background = element_rect(fill = "peachpuff"))+theme(plot.title = elemen
t_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backgrou
nd = element_rect(fill = "peachpuff"))
```



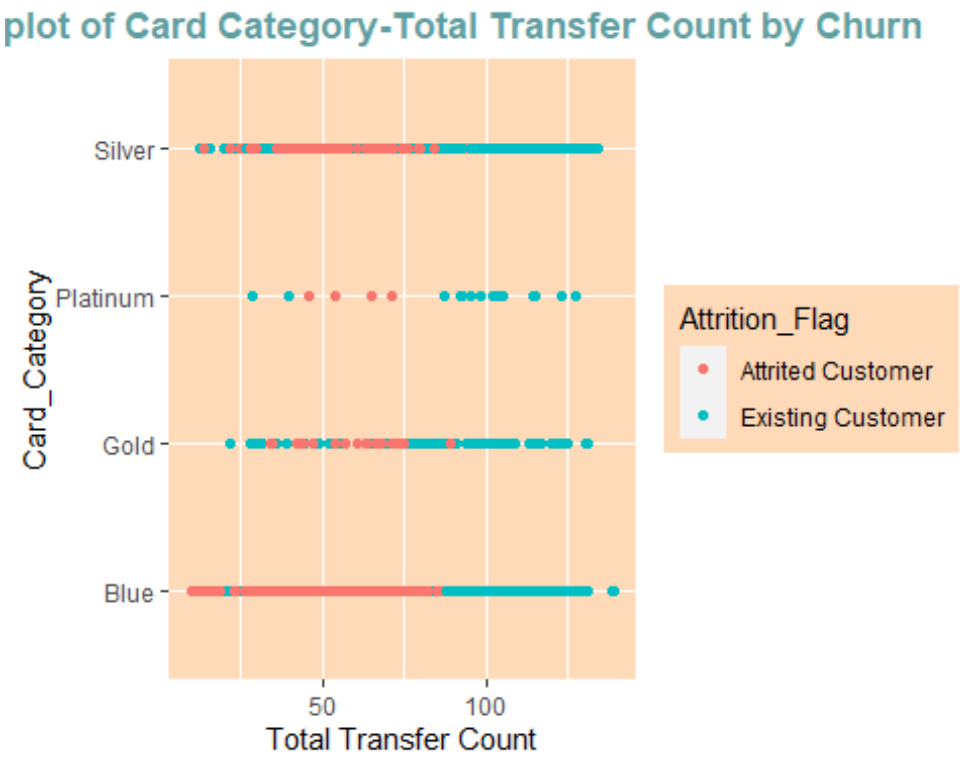

```
ggplot(data = BankChurners,aes(y=Marital_Status,x=Total_Trans_Ct,col=Attritio
n_Flag))+geom_point()+labs(title = "Scatterplot of Marital Status-Total Trans
fer Count by Churn",x="Total Transfer Count",fill="Attrition Flag")+theme(pan
el.background = element_rect(fill = "peachpuff"))+theme(plot.title = element_
text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.background
= element_rect(fill = "peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Income_Category,x=Total_Trans_Ct,col=Attriti
on_Flag))+geom_point()+labs(title = "Scatterplot of Income Category-Total Tra
nsfer Count by Churn",x="Total Transfer Count",fill="Attrition Flag")+theme(p
anel.background = element_rect(fill = "peachpuff"))+theme(plot.title = elemen
t_text(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.backgrou
nd = element_rect(fill = "peachpuff"))
```




```
ggplot(data = BankChurners,aes(y=Card_Category,x=Total_Trans_Ct,col=Attrition
_Flag))+geom_point()+labs(title = "Scatterplot of Card Category-Total Transfe
r Count by Churn",x="Total Transfer Count",fill="Attrition Flag")+theme(panel
.background = element_rect(fill = "peachpuff"))+theme(plot.title = element_te
xt(hjust = 0.5,face = "bold",colour = "cadetblue"))+theme(legend.background =
element_rect(fill = "peachpuff"))
```



- From the above four graphs we observe that the Bank has high attrition rates of upto 75 Total Transaction Count irrespective of any mentioned category.

5. SUMMARY OF FINDINGS

A. Findings

- The "Attrition Rate of the customers leaving the Credit Card Service facility is 16.07%.

Categorical Variables Findings

- After Visualizing the categorical variables independently, we get that the customers in few areas need special attention.
- The customers with the highest "Income Category" are more likely to leave the Credit Card Service facility.
- The customers with platinum and gold cards are very less and at the same time have the highest "Attrition Rates".
- The customers with the highest education qualification are more likely to leave the service.

Numeric Variables Findings

- After that, visualizing the numerical variables independently we observe that some areas of the customers need special attention.
- The customers with the age group of 40-50 require more attention as it contains maximum proportion of the customers along with the higher "Attrition Rate".
- We also observe that the most regular customers who have 0 month inactive in the last 12 months churn at the rate of 50%.
- The Attrition Rate" of the customers increases with the increase in the contacts counts in the last 12 months. At contacts count 6, the attrition rate is 100%.
- The customers with the "Total Revolving Balance" of less than 500 have 100% attrition rate. The "Attrition Rate" decreases with increase in the "Total Revolving Balance".
- We also see that the customers whose "Total Amount Change Q4-Q1" is near 0 i.e. huge decrease in Total Balance, The "Attrition Rate" is 100%.
- The customers with the "Total Transfer Amount" of 0 have maximum attrition of about 95%. Also, as the "Total Transfer Amount" increases from 5,000 to 10,000 the "Attrition Rate" reaches to 100%.

- The customers with “Total Transfer Count” and “Total Count Change Q4-Q1” have high “Attrition Rate” of 80% and starts decreasing as the “Total Transfer Count” and “Total Count Change Q4-Q1” increases.

Multivariate Relationship Findings

- The customers with lower “Total Revolving Balance”, “Total Transfer Amount” and “Total Transfer Count” have high “Attrition Rates” and decreases with increase in these variables.

B. LIMITATIONS

- Due to page limit few of multivariate visualizations are not done as they failed to show the clear relationship with the target variable.
- Visualizing more than three variables at once on a two-dimensional plane is quite difficult and result becomes so vague. Other statistical methods might be used to do the detect the patterns in the given variables.
- Using the Visualization, we have estimated result, but it may not be the accurate result. While the data is accurate in predicting the situations, the visualization of the same just gives the estimation.

C. REFERENCES

Books

Data Mining and Predictive Analysis by Daniel T. Larose and Chantal T. Larose.

Data Mining and Business Analytics with R by Johannes Ledolter.

Websites

https://www.indianbank.in/#!

https://blog.csgsolutions.com/data-visualization-banking

https://www.assk.in/blog/the-growing-demand-of-data-analytics-in-banking-sector/

https://www.researchgate.net/publication/321804138

YouTube link

(11570) ggplot2 Tutorial | ggplot2 In R Tutorial | Data Visualization In R | R Training | Edureka - YouTube