Assignment\_1

Abhinav

2024-02-05

# Dataset taken from

<https://www.kaggle.com/datasets>

library(readxl)

## Warning: package 'readxl' was built under R version 4.3.2

df1 <- read.csv("D:/SUBJECTS/FML/Assignment-1/BankChurners.csv")  
View(df1)

#printing the descriptive Statistics

summary(df1)

## CLIENTNUM Attrition\_Flag Customer\_Age Gender   
## Min. :708508758 Length:30 Min. :32.00 Length:30   
## 1st Qu.:711297108 Class :character 1st Qu.:41.25 Class :character   
## Median :736896858 Mode :character Median :47.00 Mode :character   
## Mean :754245686 Mean :47.87   
## 3rd Qu.:805726177 3rd Qu.:53.25   
## Max. :818906208 Max. :65.00   
## Dependent\_count Education\_Level Marital\_Status Income\_Category   
## Min. :0.000 Length:30 Length:30 Length:30   
## 1st Qu.:1.250 Class :character Class :character Class :character   
## Median :3.000 Mode :character Mode :character Mode :character   
## Mean :2.567   
## 3rd Qu.:3.750   
## Max. :5.000   
## Card\_Category Months\_on\_book Total\_Relationship\_Count  
## Length:30 Min. :21.00 Min. :2.000   
## Class :character 1st Qu.:34.00 1st Qu.:3.250   
## Mode :character Median :36.00 Median :5.000   
## Mean :38.73 Mean :4.467   
## 3rd Qu.:43.50 3rd Qu.:5.750   
## Max. :56.00 Max. :6.000   
## Months\_Inactive\_12\_mon Contacts\_Count\_12\_mon Credit\_Limit   
## Min. :0.00 Min. :0.0 Min. : 1438   
## 1st Qu.:1.00 1st Qu.:1.0 1st Qu.: 4293   
## Median :2.00 Median :2.0 Median : 8821   
## Mean :2.00 Mean :1.8 Mean :10921   
## 3rd Qu.:2.75 3rd Qu.:3.0 3rd Qu.:13324   
## Max. :6.00 Max. :3.0 Max. :34516   
## Total\_Revolving\_Bal Avg\_Open\_To\_Buy Total\_Amt\_Chng\_Q4\_Q1 Total\_Trans\_Amt  
## Min. : 0.0 Min. : 676 Min. :0.5250 Min. : 692   
## 1st Qu.: 704.2 1st Qu.: 3419 1st Qu.:0.9097 1st Qu.:1094   
## Median :1326.5 Median : 7450 Median :1.3555 Median :1204   
## Mean :1213.8 Mean : 9707 Mean :1.4466 Mean :1262   
## 3rd Qu.:1668.2 3rd Qu.:12237 3rd Qu.:1.7078 3rd Qu.:1350   
## Max. :2517.0 Max. :32252 Max. :3.3970 Max. :1904   
## Total\_Trans\_Ct Total\_Ct\_Chng\_Q4\_Q1 Avg\_Utilization\_Ratio  
## Min. :16.00 Min. :0.3530 Min. :0.00000   
## 1st Qu.:21.00 1st Qu.:0.8363 1st Qu.:0.06225   
## Median :26.50 Median :1.0715 Median :0.10200   
## Mean :26.97 Mean :1.3623 Mean :0.17667   
## 3rd Qu.:31.75 3rd Qu.:1.6250 3rd Qu.:0.21650   
## Max. :42.00 Max. :3.7140 Max. :0.78800   
## Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_1  
## Min. :0.0000077   
## 1st Qu.:0.0000556   
## Median :0.0000948   
## Mean :0.0333130   
## 3rd Qu.:0.0001829   
## Max. :0.9961600   
## Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_2  
## Min. :0.003836   
## 1st Qu.:0.999815   
## Median :0.999905   
## Mean :0.966687   
## 3rd Qu.:0.999940   
## Max. :0.999990

#Printing the Quantitative Varaibles

mean(df1$Total\_Relationship\_Count  
)

## [1] 4.466667

median(df1$Contacts\_Count\_12\_mon  
)

## [1] 2

summary(df1$Avg\_Open\_To\_Buy  
)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 676 3419 7450 9707 12237 32252

# Printing the Qualitative Variables

summary(df1$Gender)

## Length Class Mode   
## 30 character character

summary(df1$Card\_Category)

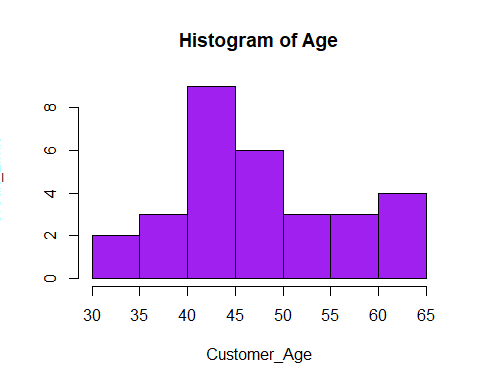
## Length Class Mode   
## 30 character character

# Transformation of INPUT Months on book and printing it’s value

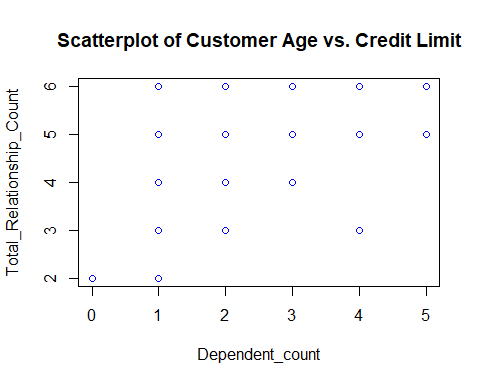
# Assuming df1 is your data frame  
df1$Months\_on\_book <- sqrt(df1$Months\_on\_book)  
  
# Print the first few rows to see the changes  
head(df1)

## CLIENTNUM Attrition\_Flag Customer\_Age Gender Dependent\_count  
## 1 768805383 Existing Customer 45 M 3  
## 2 818770008 Existing Customer 49 F 5  
## 3 713982108 Existing Customer 51 M 3  
## 4 769911858 Existing Customer 40 F 4  
## 5 709106358 Existing Customer 40 M 3  
## 6 713061558 Existing Customer 44 M 2  
## Education\_Level Marital\_Status Income\_Category Card\_Category Months\_on\_book  
## 1 High School Married $60K - $80K Blue 6.244998  
## 2 Graduate Single Less than $40K Blue 6.633250  
## 3 Graduate Married $80K - $120K Blue 6.000000  
## 4 High School Unknown Less than $40K Blue 5.830952  
## 5 Uneducated Married $60K - $80K Blue 4.582576  
## 6 Graduate Married $40K - $60K Blue 6.000000  
## Total\_Relationship\_Count Months\_Inactive\_12\_mon Contacts\_Count\_12\_mon  
## 1 5 1 3  
## 2 6 1 2  
## 3 4 1 0  
## 4 3 4 1  
## 5 5 1 0  
## 6 3 1 2  
## Credit\_Limit Total\_Revolving\_Bal Avg\_Open\_To\_Buy Total\_Amt\_Chng\_Q4\_Q1  
## 1 12691 777 11914 1.335  
## 2 8256 864 7392 1.541  
## 3 3418 0 3418 2.594  
## 4 3313 2517 796 1.405  
## 5 4716 0 4716 2.175  
## 6 4010 1247 2763 1.376  
## Total\_Trans\_Amt Total\_Trans\_Ct Total\_Ct\_Chng\_Q4\_Q1 Avg\_Utilization\_Ratio  
## 1 1144 42 1.625 0.061  
## 2 1291 33 3.714 0.105  
## 3 1887 20 2.333 0.000  
## 4 1171 20 2.333 0.760  
## 5 816 28 2.500 0.000  
## 6 1088 24 0.846 0.311  
## Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_1  
## 1 0.00009340  
## 2 0.00005690  
## 3 0.00002110  
## 4 0.00013366  
## 5 0.00002170  
## 6 0.00005510  
## Naive\_Bayes\_Classifier\_Attrition\_Flag\_Card\_Category\_Contacts\_Count\_12\_mon\_Dependent\_count\_Education\_Level\_Months\_Inactive\_12\_mon\_2  
## 1 0.99991  
## 2 0.99994  
## 3 0.99998  
## 4 0.99987  
## 5 0.99998  
## 6 0.99994

hist(df1$Customer\_Age,xlab ="Customer\_Age",ylab ="Credit\_Limit  
",main ="Histogram of Age",col ="purple")



# Assuming df1 is your data frame  
plot(df1$Dependent\_count, df1$Total\_Relationship\_Count,main = "Scatterplot of Customer Age vs. Credit Limit",xlab = "Dependent\_count",ylab = "Total\_Relationship\_Count",col = "blue") # Optional: Set the color of the points



# Assuming df1 is your data frame  
barplot(table(df1$Gender),  
 main = "Barplot of Attrition\_Flag",  
 xlab = "Gender",  
 ylab = "Frequency",  
 col = "lightblue") # Optional: Set the color of the bars

