Q1.R

Hannan

Sun Dec 09 03:51:46 2018

# Loading necessary libraries and dataset  
# install.packages("readr")  
   
 library(readr)  
 Telco\_Customer\_Churn <- read\_csv("Telco-Customer-Churn.csv")

## Parsed with column specification:  
## cols(  
## customerID = col\_character(),  
## gender = col\_character(),  
## SeniorCitizen = col\_double(),  
## Partner = col\_character(),  
## Dependents = col\_character(),  
## tenure = col\_double(),  
## PhoneService = col\_character(),  
## MultipleLines = col\_character(),  
## InternetService = col\_character(),  
## StreamingTV = col\_character(),  
## StreamingMovies = col\_character(),  
## Contract = col\_character(),  
## PaperlessBilling = col\_character(),  
## PaymentMethod = col\_character(),  
## MonthlyCharges = col\_double(),  
## TotalCharges = col\_double(),  
## Churn = col\_character()  
## )

# 1.1. What is the proportion of males/females in the dataset?  
 gend <- Telco\_Customer\_Churn$gender  
 gend\_agreg <- aggregate(data.frame(count = gend), list(value = gend), length)  
 gend\_ratio <- gend\_agreg$count[2]/gend\_agreg$count[1]  
  
 cat('The ratio of male/femal in data is: ',gend\_agreg$count[2],'/' ,gend\_agreg$count[1],' or ', round(gend\_ratio, 4),'\n')

**## The ratio of male/femal in data is: 3555 / 3488 or 1.0192**

# 1.2. What is the proportion of senior citizens in the dataset?   
 Sen\_Cit <- Telco\_Customer\_Churn$SeniorCitizen   
 SC\_agreg <- aggregate(x = data.frame(Sen\_Cit), list(value = Sen\_Cit),length)  
 Sen\_Cit\_ratio <- round(SC\_agreg$Sen\_Cit[2]/sum(SC\_agreg$Sen\_Cit),4)  
   
 cat('The Proportion of Senior Citizens in data is: ',Sen\_Cit\_ratio,' or ',Sen\_Cit\_ratio\*100 ,'%\n')

**## The Proportion of Senior Citizens in data is: 0.1621 or 16.21%**

# 1.3. What is the most common type of contract in the dataset?   
 Contract <- Telco\_Customer\_Churn$Contract  
 Cont\_agreg <- aggregate(x = data.frame(Contract), list(value=Contract),length)  
 max\_Cont\_field <- Cont\_agreg$value[which.max(Cont\_agreg$Contract)]  
   
 cat('The most common type of contract in the dataset is: ',max\_Cont\_field,'\n')

**## The most common type of contract in the dataset is: Month-to-month**

# 1.4. What is the most common type of internet service in the dataset?   
 Int\_Serv <- Telco\_Customer\_Churn$InternetService  
 IS\_agreg <- aggregate(x = data.frame(Int\_Serv), list(value=Int\_Serv),length)  
 max\_IS\_field <- IS\_agreg$value[which.max(IS\_agreg$Int\_Serv)]  
   
 cat('The most common type of internet service in the dataset is: ',max\_IS\_field,'\n')

**## The most common type of internet service in the dataset is: Fiber optic**

# 1.5. What is the least common payment method in the dataset?   
 Pay\_Method <- Telco\_Customer\_Churn$PaymentMethod  
 PM\_agreg <- aggregate(x = data.frame(Pay\_Method), list(value=Pay\_Method),length)  
 min\_PM\_field <- PM\_agreg$value[which.min(PM\_agreg$Pay\_Method)]  
   
 cat('The least common payment method in the dataset is: ',min\_PM\_field,'\n')

**## The least common payment method in the dataset is: Credit card (automatic)**

# 1.6. What is the average tenure of male and female customers?  
 Ten\_Cal <- Telco\_Customer\_Churn[, c(2,6)]  
 males\_ten = subset(Ten\_Cal, subset = (Ten\_Cal[,1] == "Male"))  
 female\_ten = subset(Ten\_Cal, subset = (Ten\_Cal[,1] == "Female"))  
 male\_ave = round(sum(males\_ten$tenure)/length(males\_ten$tenure),4)  
 female\_ave = round(sum(female\_ten$tenure)/length(female\_ten$tenure),4)  
  
 cat('The average tenures of males and females are ',male\_ave,' & ',female\_ave,' respectively.\n')

**## The average tenures of males and females are 32.4954 & 32.2446 respectively.**

# 1.7. Is Streaming TV favored over the Streaming Movies service?  
 TV\_count = length(subset(Telco\_Customer\_Churn$StreamingTV,subset = (Telco\_Customer\_Churn$StreamingTV == 'Yes')))  
 Mov\_count = length(subset(Telco\_Customer\_Churn$StreamingMovies,subset = (Telco\_Customer\_Churn$StreamingMovies == 'Yes')))  
   
 cat('Is Streaming TV favored over the Streaming Movies service? \t', ifelse(TV\_count > Mov\_count,"YES!","NO!"),'\n')

**## Is Streaming TV favored over the Streaming Movies service? NO!**

# 1.8. How many customers churned out of the total customers?   
 churn\_count = length(subset(Telco\_Customer\_Churn$Churn,subset = (Telco\_Customer\_Churn$Churn == 'Yes')))  
   
 cat(churn\_count ,'customers churned out of the total customers!,\n')

**## 1869 customers churned out of the total customers!**

# 1.9. What is the mean “monthly charges” amongst customers with month-to-month contracts?   
 Cont\_MonCharge = Telco\_Customer\_Churn[,c('Contract','MonthlyCharges')]  
 MonChages\_for\_montomon = subset(Cont\_MonCharge,Cont\_MonCharge$Contract == "Month-to-month")  
   
 cat('The mean “monthly charges” amongst customers with month-to-month contracts is:',round(mean(MonChages\_for\_montomon$MonthlyCharges),4),'\n')

**## The mean “monthly charges” amongst customers with month-to-month contracts is: 66.3985**

# 1.10. What is the gender wise average of “total charges” amongst customers?   
 Gend\_Charges = Telco\_Customer\_Churn[,c('gender','TotalCharges')]  
 male\_charges = subset(Gend\_Charges, Gend\_Charges$gender == "Male")  
 female\_charges = subset(Gend\_Charges, Gend\_Charges$gender == "Female")  
 male\_charges\_ave = round(sum(male\_charges$TotalCharges,na.rm = TRUE)/length(male\_charges$TotalCharges),4)   
 female\_charges\_ave = round(sum(female\_charges$TotalCharges,na.rm = TRUE)/length(female\_charges$TotalCharges),4)   
   
 gend\_wise\_ave = data.frame(c("Male","Female"),c(male\_charges\_ave,female\_charges\_ave))  
 colnames(gend\_wise\_ave) <- c("Gender","Total Charges Average")  
 show(gend\_wise\_ave)

**## Gender Total Charges Average  
## 1 Male 2279.554  
## 2 Female 2279.918**

View(gend\_wise\_ave)