

CUSTOMER CHURN DATA ANALYSIS

Wajiha Shaukat

#Importing Data

```
data <- read.csv("customer churn data.csv")
head(data)
```

```
##   customerID gender SeniorCitizen Partner Dependents tenure PhoneService
## 1  CUST0000  Male                0      No          Yes    23           No
## 2  CUST0001 Female                0      Yes         No    43           No
## 3  CUST0002  Male                1      No          No    51           Yes
## 4  CUST0003  Male                1      No          No    72           Yes
## 5  CUST0004  Male                1      No          No    25           Yes
## 6  CUST0005 Female                0      Yes         No    35           Yes
##   MultipleLines InternetService OnlineSecurity OnlineBackup
## 1 No phone service No No internet service No internet service
## 2 No phone service DSL Yes No No
## 3 No No DSL No Yes
## 4 Yes DSL Yes No
## 5 Yes DSL No No
## 6 No DSL No No
##   DeviceProtection TechSupport StreamingTV
## 1 No internet service No internet service No internet service
## 2 Yes No Yes
## 3 Yes Yes No
## 4 Yes No No
## 5 No Yes No
## 6 No Yes Yes
##   StreamingMovies Contract PaperlessBilling PaymentMethod
## 1 No internet service Month-to-month Yes Bank transfer
## 2 No Month-to-month No Mailed check
## 3 No One year No Electronic check
## 4 No Month-to-month No Credit card
## 5 Yes Month-to-month No Electronic check
## 6 Yes One year No Credit card
##   MonthlyCharges TotalCharges Churn
## 1 49.85 1146.55 No
## 2 100.70 4330.10 Yes
## 3 97.33 4963.83 Yes
## 4 101.38 7299.36 No
## 5 52.22 1305.50 Yes
## 6 116.96 4093.60 No
```

#Checking Data

```
str(data)
```

```
## 'data.frame': 5880 obs. of 21 variables:
## $ customerID : chr "CUST0000" "CUST0001" "CUST0002" "CUST0003" ...
```

```
## $ gender      : chr "Male" "Female" "Male" "Male" ...
## $ SeniorCitizen : int 0 0 1 1 1 0 0 0 0 ...
## $ Partner      : chr "No" "Yes" "No" "No" ...
## $ Dependents   : chr "Yes" "No" "No" "No" ...
## $ tenure       : int 23 43 51 72 25 35 17 18 27 15 ...
## $ PhoneService : chr "No" "No" "Yes" "Yes" ...
## $ MultipleLines : chr "No phone service" "No phone service" "No" "Yes" ...
## $ InternetService : chr "No" "DSL" "DSL" "DSL" ...
## $ OnlineSecurity : chr "No internet service" "Yes" "No" "Yes" ...
## $ OnlineBackup : chr "No internet service" "No" "Yes" "No" ...
## $ DeviceProtection: chr "No internet service" "Yes" "Yes" "Yes" ...
## $ TechSupport : chr "No internet service" "No" "Yes" "No" ...
## $ StreamingTV : chr "No internet service" "Yes" "No" "No" ...
## $ StreamingMovies : chr "No internet service" "No" "No" "No" ...
## $ Contract      : chr "Month-to-month" "Month-to-month" "One year" "Month-to-month" ...
## $ PaperlessBilling: chr "Yes" "No" "No" "No" ...
## $ PaymentMethod : chr "Bank transfer" "Mailed check" "Electronic check" "Credit card" ...
## $ MonthlyCharges : num 49.9 100.7 97.3 101.4 52.2 ...
## $ TotalCharges : num 1147 4330 4964 7299 1306 ...
## $ Churn         : chr "No" "Yes" "Yes" "No" ...
```

#DATA CLEANING -> removing column 1 i.e., customerID

#DATA CLEANING

```
dat1 <- subset(data, select = -customerID)
head(dat1)
```

```
##   gender SeniorCitizen Partner Dependents tenure PhoneService MultipleLines
## 1  Male              0      No         Yes    23         No No phone service
## 2 Female              0      Yes        No    43         No No phone service
## 3  Male              1      No         No    51         Yes             No
## 4  Male              1      No         No    72         Yes             Yes
## 5  Male              1      No         No    25         Yes             Yes
## 6 Female              0      Yes        No    35         Yes             No
##   InternetService OnlineSecurity OnlineBackup DeviceProtection
## 1         No No internet service No internet service No internet service
## 2         DSL              Yes              No              Yes
## 3         DSL              No              Yes              Yes
## 4         DSL              Yes              No              Yes
## 5         DSL              No              No              No
## 6         DSL              No              No              No
##   TechSupport StreamingTV StreamingMovies Contract
## 1 No internet service No internet service No internet service Month-to-month
## 2         No              Yes              No Month-to-month
## 3         Yes              No              No       One year
## 4         No              No              No Month-to-month
## 5         Yes              No              Yes Month-to-month
## 6         Yes              Yes              Yes       One year
##   PaperlessBilling PaymentMethod MonthlyCharges TotalCharges Churn
## 1         Yes Bank transfer         49.85      1146.55     No
## 2         No  Mailed check        100.70      4330.10    Yes
## 3         No Electronic check        97.33      4963.83    Yes
## 4         No  Credit card        101.38      7299.36     No
## 5         No Electronic check        52.22      1305.50    Yes
## 6         No  Credit card        116.96      4093.60     No
```

#CHECKING THE NULL VALUES

```
colSums(is.na(dat1) | dat1 == "")
```

```
##      gender      SeniorCitizen      Partner      Dependents
##      0          0          0          0
##      tenure      PhoneService      MultipleLines      InternetService
##      0          0          0          0
##      OnlineSecurity      OnlineBackup      DeviceProtection      TechSupport
##      0          0          0          0
##      StreamingTV      StreamingMovies      Contract      PaperlessBilling
##      0          0          0          0
##      PaymentMethod      MonthlyCharges      TotalCharges      Churn
##      0          0          0          0
```

#Checking Data Structure

```
summary(data)
```

```
##      customerID      gender      SeniorCitizen      Partner
##      Length:5880      Length:5880      Min. :0.0000      Length:5880
##      Class :character      Class :character      1st Qu.:0.0000      Class :character
##      Mode :character      Mode :character      Median :1.0000      Mode :character
##
##
##      Mean :0.5007
##      3rd Qu.:1.0000
##      Max. :1.0000
##      Dependents      tenure      PhoneService      MultipleLines
##      Length:5880      Min. : 1.00      Length:5880      Length:5880
##      Class :character      1st Qu.:18.00      Class :character      Class :character
##      Mode :character      Median :37.00      Mode :character      Mode :character
##
##      Mean :36.55
##      3rd Qu.:55.00
##      Max. :72.00
##      InternetService      OnlineSecurity      OnlineBackup      DeviceProtection
##      Length:5880      Length:5880      Length:5880      Length:5880
##      Class :character      Class :character      Class :character      Class :character
##      Mode :character      Mode :character      Mode :character      Mode :character
##
##
##      TechSupport      StreamingTV      StreamingMovies      Contract
##      Length:5880      Length:5880      Length:5880      Length:5880
##      Class :character      Class :character      Class :character      Class :character
##      Mode :character      Mode :character      Mode :character      Mode :character
##
##
##
##      PaperlessBilling      PaymentMethod      MonthlyCharges      TotalCharges
##      Length:5880      Length:5880      Min. : 20.00      Min. : 20.03
##      Class :character      Class :character      1st Qu.: 45.72      1st Qu.:1020.22
##      Mode :character      Mode :character      Median : 70.16      Median :2136.45
##
##      Mean : 70.16      Mean :2566.81
##      3rd Qu.: 95.46      3rd Qu.:3767.66
##      Max. :119.99      Max. :8589.60
##      Churn
##      Length:5880
```

```
## Class :character
## Mode :character
##
##
##

#mutate () modifies the column
library(dplyr)

##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##   filter, lag

## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

recoding <- data %>%
  mutate(Partner = if_else(Partner == 0, "No", "Yes"))
head(recoding)
```

| ## | customerID | gender | SeniorCitizen | Partner | Dependents | tenure | PhoneService |
|------|------------|--------|---------------|---------|------------|--------|--------------|
| ## 1 | CUST0000 | Male | 0 | Yes | Yes | 23 | No |
| ## 2 | CUST0001 | Female | 0 | Yes | No | 43 | No |
| ## 3 | CUST0002 | Male | 1 | Yes | No | 51 | Yes |
| ## 4 | CUST0003 | Male | 1 | Yes | No | 72 | Yes |
| ## 5 | CUST0004 | Male | 1 | Yes | No | 25 | Yes |
| ## 6 | CUST0005 | Female | 0 | Yes | No | 35 | Yes |

| ## | MultipleLines | InternetService | OnlineSecurity | OnlineBackup |
|------|------------------|-----------------|---------------------|---------------------|
| ## 1 | No phone service | No | No internet service | No internet service |
| ## 2 | No phone service | DSL | Yes | No |
| ## 3 | No | DSL | No | Yes |
| ## 4 | Yes | DSL | Yes | No |
| ## 5 | Yes | DSL | No | No |
| ## 6 | No | DSL | No | No |

| ## | DeviceProtection | TechSupport | StreamingTV |
|------|---------------------|---------------------|---------------------|
| ## 1 | No internet service | No internet service | No internet service |
| ## 2 | Yes | No | Yes |
| ## 3 | Yes | Yes | No |
| ## 4 | Yes | No | No |
| ## 5 | No | Yes | No |
| ## 6 | No | Yes | Yes |

| ## | StreamingMovies | Contract | PaperlessBilling | PaymentMethod |
|------|---------------------|----------------|------------------|------------------|
| ## 1 | No internet service | Month-to-month | Yes | Bank transfer |
| ## 2 | No | Month-to-month | No | Mailed check |
| ## 3 | No | One year | No | Electronic check |
| ## 4 | No | Month-to-month | No | Credit card |
| ## 5 | Yes | Month-to-month | No | Electronic check |
| ## 6 | Yes | One year | No | Credit card |

| ## | MonthlyCharges | TotalCharges | Churn |
|------|----------------|--------------|-------|
| ## 1 | 49.85 | 1146.55 | No |
| ## 2 | 100.70 | 4330.10 | Yes |
| ## 3 | 97.33 | 4963.83 | Yes |

```
## 4      101.38      7299.36    No
## 5       52.22      1305.50   Yes
## 6      116.96      4093.60    No
```

```
library(ggplot2)
library(gridExtra)
```

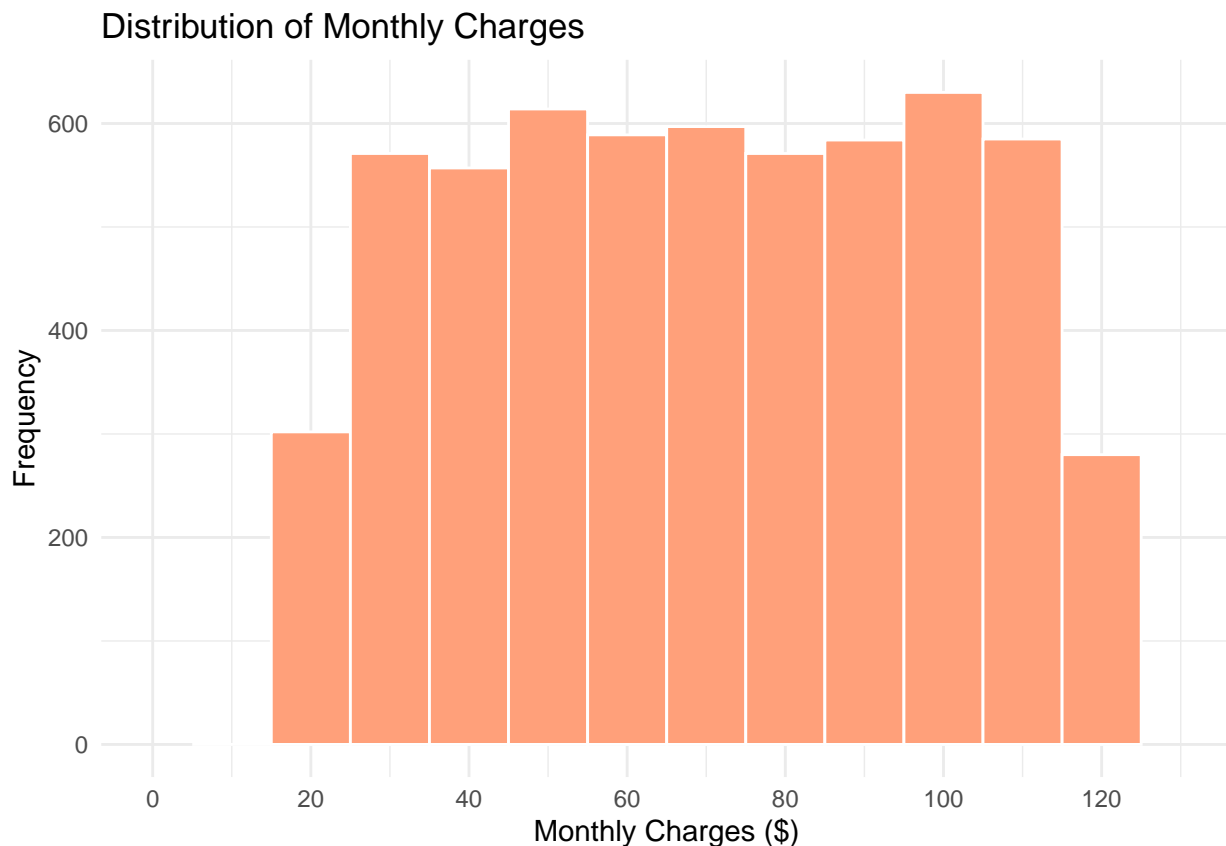
```
##
## Attaching package: 'gridExtra'

## The following object is masked from 'package:dplyr':
##
##   combine
```

```
# Create the histogram
p1 <- ggplot(data = recoding, aes(x = MonthlyCharges)) +
  geom_histogram(binwidth = 10, color = "white", fill = "#FFA07A") +
  labs(title = "Distribution of Monthly Charges",
       x = "Monthly Charges ($)",
       y = "Frequency") +
  scale_x_continuous(limits = c(0, 130), breaks = seq(0, 150, 20)) +
  theme_minimal()

# Display the plot
grid.arrange(p1)
```

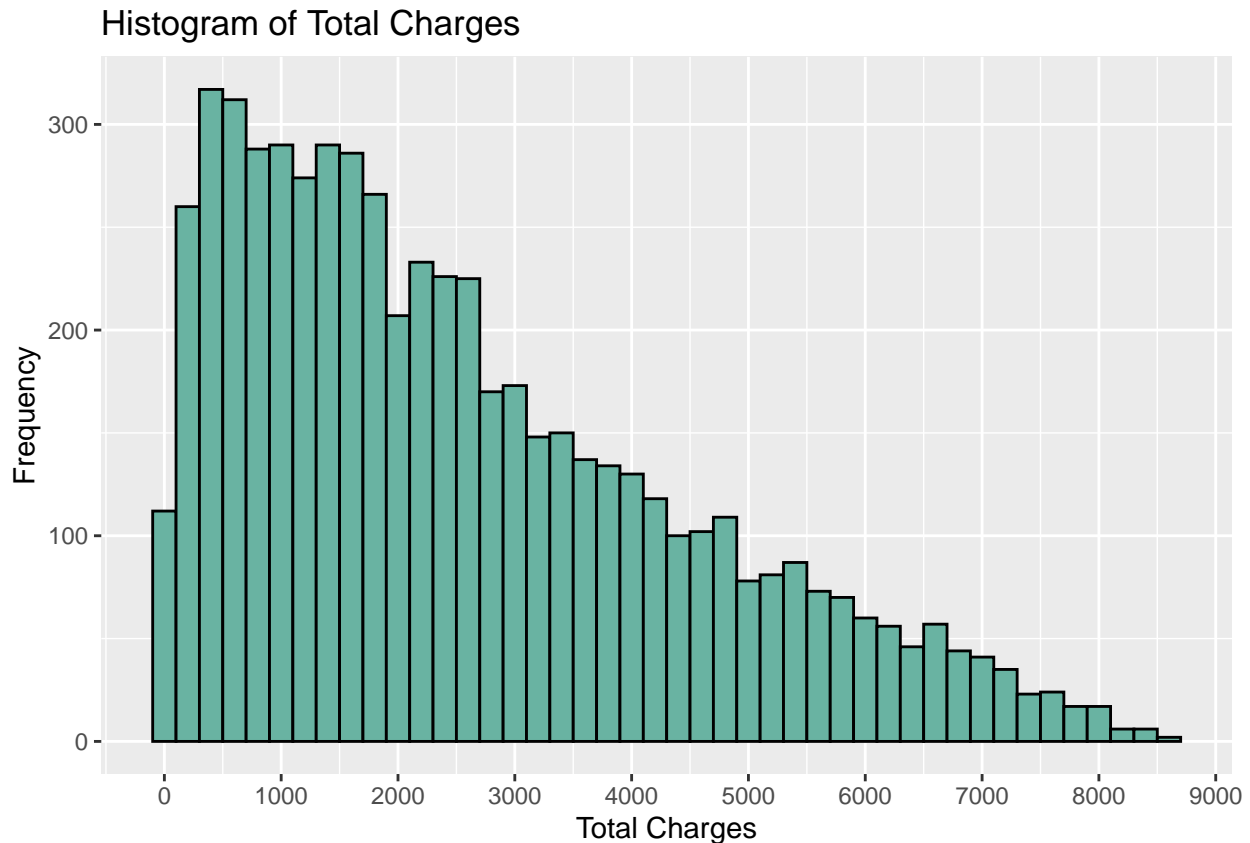
```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



```
library(ggplot2)
library(gridExtra)

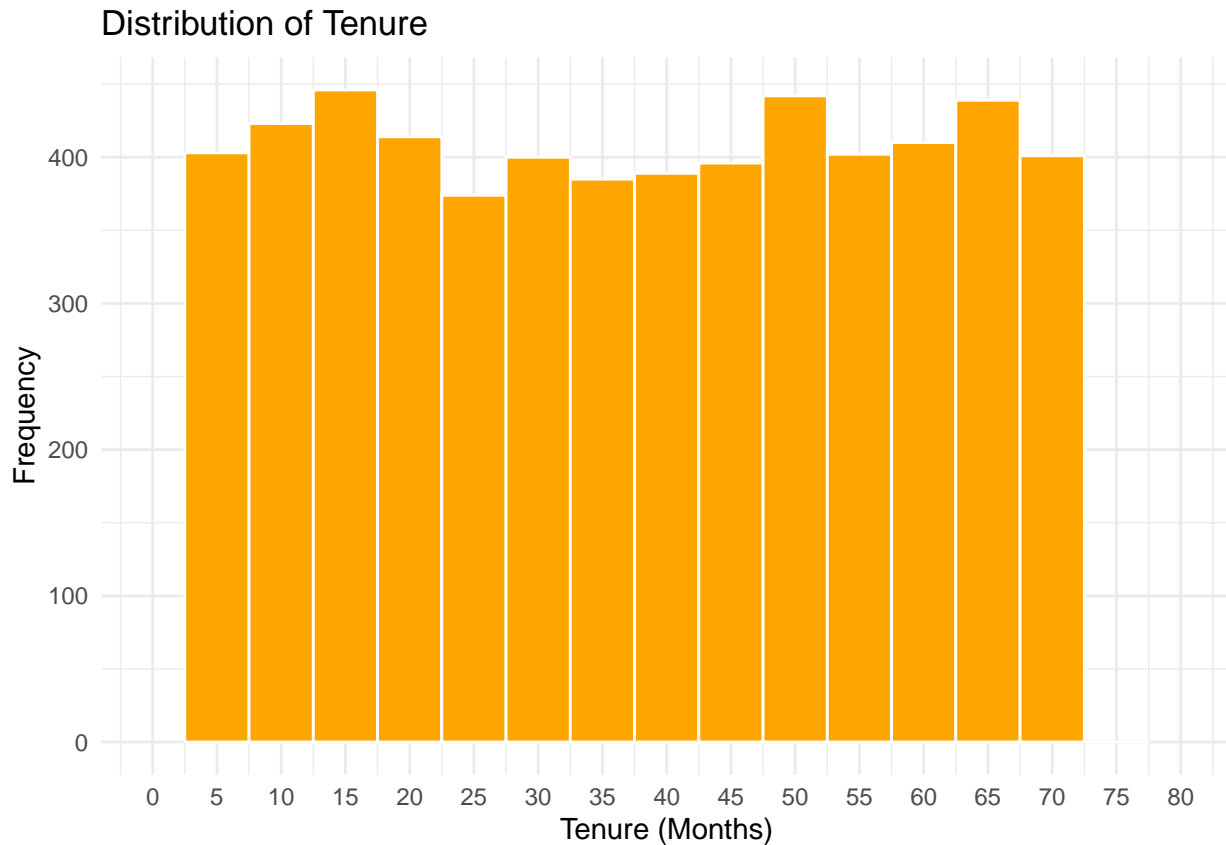
p2 <- ggplot(recoding, aes(x = TotalCharges)) +
  geom_histogram(binwidth = 200, fill = "#69b3a2", color = "black") +
  scale_x_continuous(breaks = seq(0, 9000, by = 1000)) +
  labs(x = "Total Charges", y = "Frequency",
       title = "Histogram of Total Charges")

grid.arrange(p2)
```



```
p3 <- ggplot(recoding, aes(x = tenure)) +
  geom_histogram(binwidth = 5, color = "white", fill = "#FFA500") +
  labs(title = "Distribution of Tenure",
       x = "Tenure (Months)",
       y = "Frequency") +
  scale_x_continuous(limits = c(0, 80), breaks = seq(0, 80, 5)) +
  theme_minimal()
grid.arrange(p3)
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
## (`geom_bar()`).
```



```
p4 <- ggplot(data = recoding, aes(x = SeniorCitizen)) +
  geom_histogram(binwidth = 1, color = "white", fill = "#FFA07A") +
  geom_text(aes(y = ..count.. -400, label = paste0(round(prop.table(..count..), 4)*100, "%")),
    stat = "count")+labs(title = "Distribution Senior Citizen",
    x = "numAdminTickets",
    y = "Frequency") +
  scale_x_continuous(limits = c(-1, 6), breaks = seq(0, 5, 1)) +
  theme_minimal()
grid.arrange(p4)
```

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.
```

```
## i Please use `after_stat(count)` instead.
```

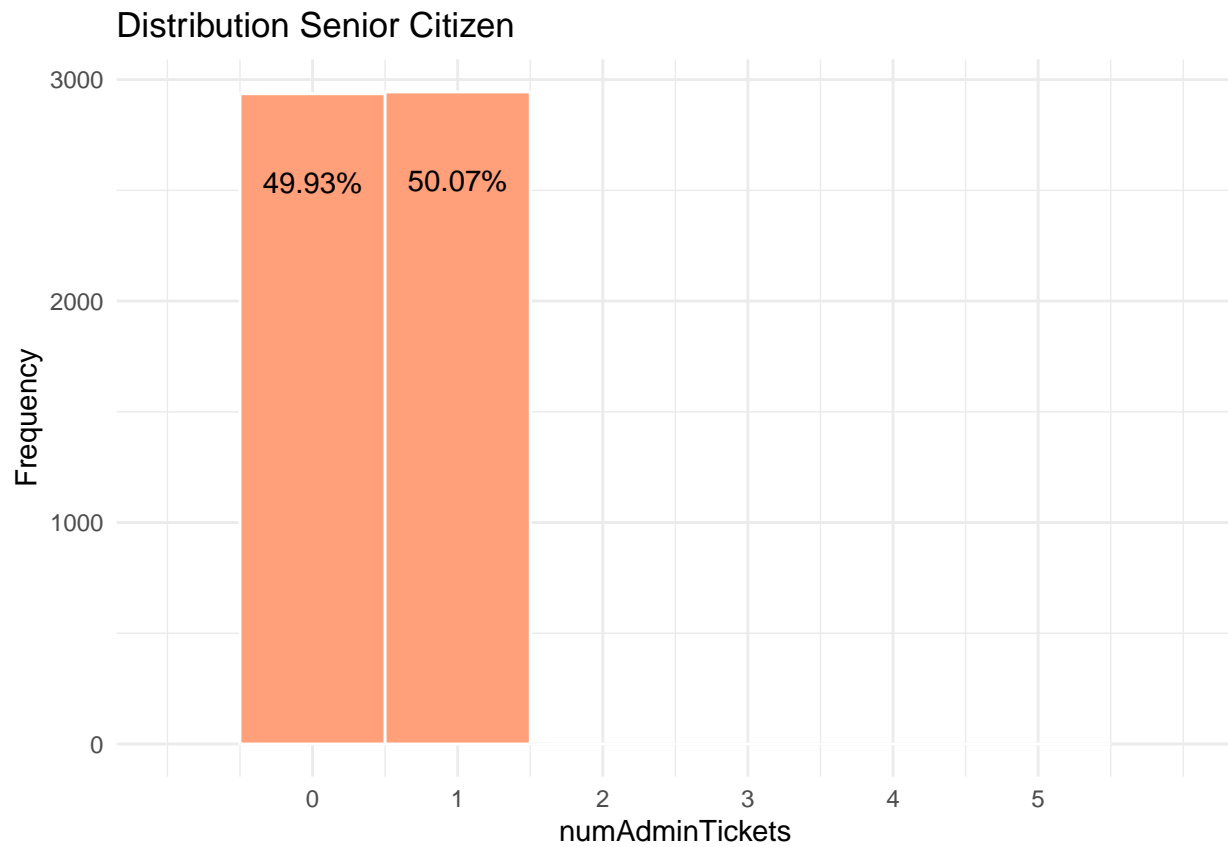
```
## This warning is displayed once every 8 hours.
```

```
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
```

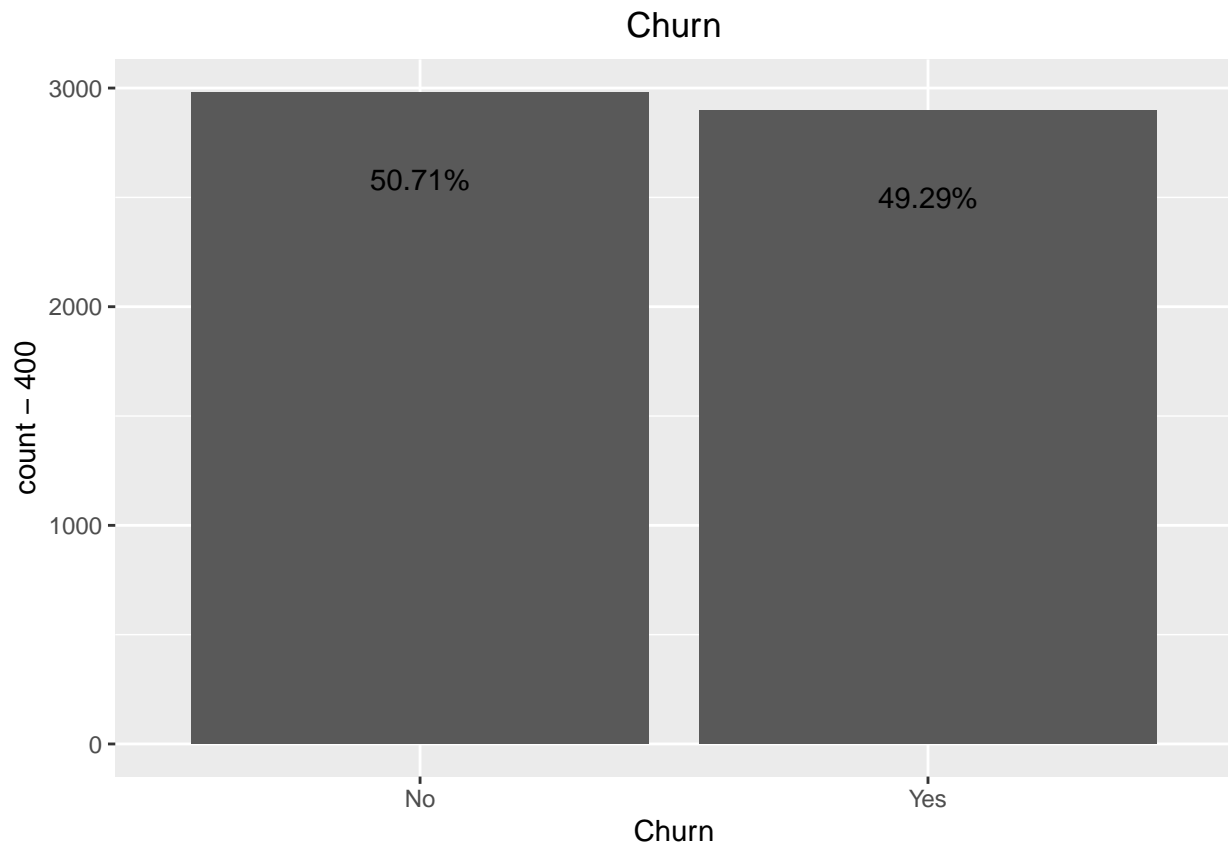
```
## generated.
```

```
## Warning: Removed 2 rows containing missing values or values outside the scale range
```

```
## (`geom_bar()`).
```



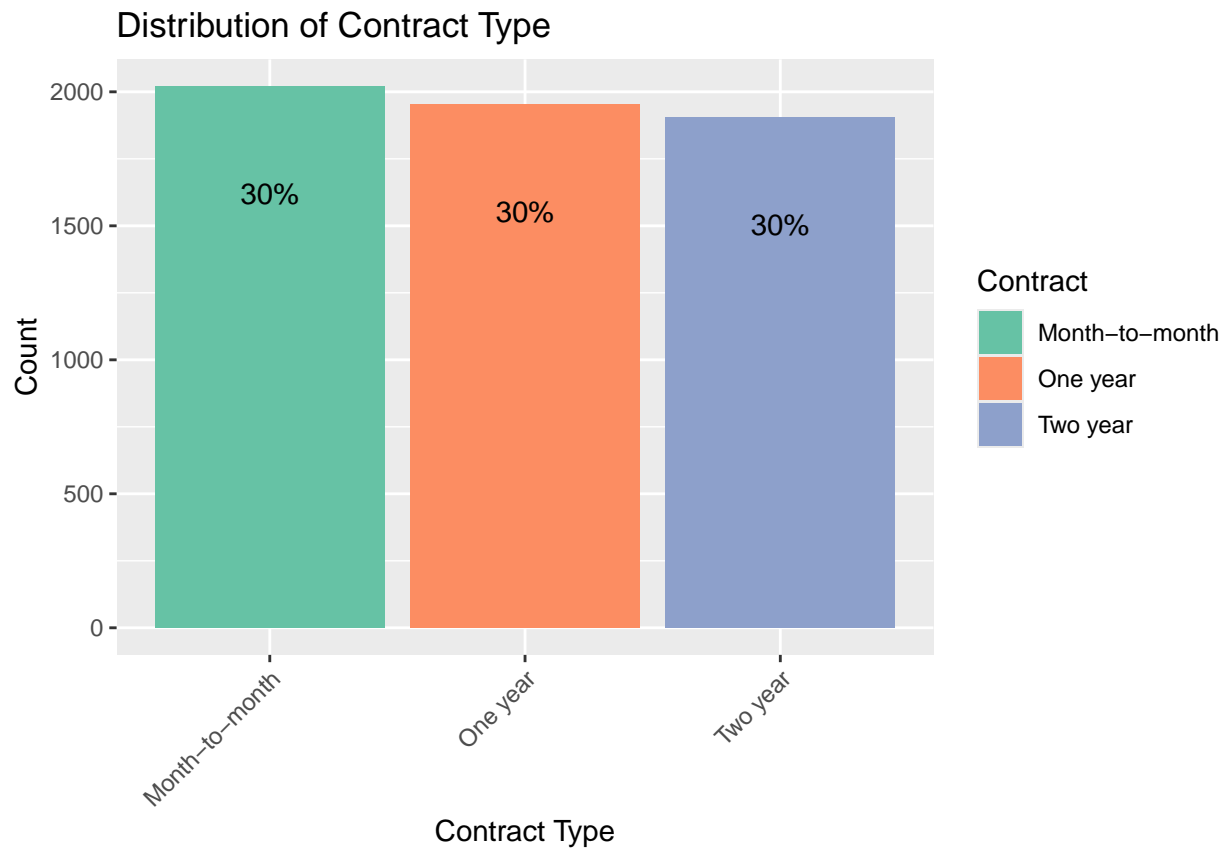
```
p5 <- ggplot(recoding, aes(x = Churn)) +  
  geom_bar() + geom_text(aes(y = ..count.. -400,  
                             label = paste0(round(prop.table(..count..),4)*100, '%'),  
                             stat = 'count')) + ggtitle("Churn") +  
  theme(plot.title = element_text(hjust = .5))  
p5
```

```
p6 <- ggplot(recoding, aes(x = Contract, fill = Contract)) +
  geom_bar() + geom_text(aes(y = ..count.. -400, label = paste0(round(prop.table(..count..),
                                                                    1)*100, "%")), stat = "count")+labs(
  x = "Contract Type",
  y = "Count") +
  scale_fill_brewer(palette = "Set2") +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))

# Bar graph of gender

colors <- c("#0072B2", "#E69F00")
grid.arrange(p6)
```



```
p7 <- ggplot(recoding, aes(x = PaymentMethod, fill = PaymentMethod)) +  
  geom_bar() +  
  labs(title = "Distribution of Payment Method",  
        x = "Payment Method",  
        y = "Count") +  
  scale_fill_brewer(palette = "Set2") +  
  theme(axis.text.x = element_text(angle = 45, hjust = 1))  
  
# Arrange plots in a grid  
grid.arrange(p7)
```

