

the Customer Churn Dataset in Module 3.2.

Marba iquen l

2024-11-16

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

```
summary(cars)
```

```
##      speed      dist
##  Min.   : 4.0    Min.   :  2.00
##  1st Qu.:12.0    1st Qu.: 26.00
##  Median :15.0    Median : 36.00
##  Mean   :15.4    Mean   : 42.98
##  3rd Qu.:19.0    3rd Qu.: 56.00
##  Max.   :25.0    Max.   :120.00
```

```
set.seed(123)
```

```
customerID = 1:200
monthlyUsage = round(rnorm(200, mean = 30, sd = 10), 1)
satisfactionScore = sample(1:10, 200, replace = TRUE)
subscriptionLength = sample(1:24, 200, replace = TRUE)
paymentStatus = sample(c(0, 1), 200, replace = TRUE, prob = c(0.2, 0.8))
churn = sample(c(0, 1), 200, replace = TRUE, prob = c(0.7, 0.3))
```

```
customerData <- data.frame(customerID, monthlyUsage, satisfactionScore,
subscriptionLength, paymentStatus, churn)
```

```
head(customerData)
```

```
##   customerID monthlyUsage satisfactionScore subscriptionLength paymentStatus
## 1          1          24.4                7                 23             1
## 2          2          27.7                3                  5             1
## 3          3          45.6                9                 17             1
## 4          4          30.7                7                 11             1
## 5          5          31.3                9                 23             1
## 6          6          47.2                4                  4             0
##   churn
## 1      0
```

```
## 2      0
## 3      0
## 4      0
## 5      0
## 6      0
```

```
str
```

```
## function (object, ...)
## UseMethod("str")
## <bytecode: 0x000001a1b079f4f8>
## <environment: namespace:utils>
```

```
nrow(customerData)
```

```
## [1] 200
```

```
summary(customerData)
```

```
##      customerID      monthlyUsage      satisfactionScore      subscriptionLength
## Min.   : 1.00    Min.   : 6.90    Min.   : 1.00    Min.   : 1.00
## 1st Qu.: 50.75    1st Qu.:23.70    1st Qu.: 3.00    1st Qu.: 6.00
## Median :100.50    Median :29.40    Median : 6.00    Median :10.50
## Mean   :100.50    Mean   :29.91    Mean   : 5.45    Mean   :11.72
## 3rd Qu.:150.25    3rd Qu.:35.65    3rd Qu.: 8.00    3rd Qu.:17.00
## Max.   :200.00    Max.   :62.40    Max.   :10.00    Max.   :24.00
## paymentStatus      churn
## Min.   :0.000    Min.   :0.00
## 1st Qu.:1.000    1st Qu.:0.00
## Median :1.000    Median :0.00
## Mean   :0.835    Mean   :0.27
## 3rd Qu.:1.000    3rd Qu.:1.00
## Max.   :1.000    Max.   :1.00
```

```
str(customerData)
```

```
## 'data.frame': 200 obs. of 6 variables:
## $ customerID : int 1 2 3 4 5 6 7 8 9 10 ...
## $ monthlyUsage : num 24.4 27.7 45.6 30.7 31.3 47.2 34.6 17.3 23.1 25.5 ...
## $ satisfactionScore : int 7 3 9 7 9 4 2 6 10 9 ...
## $ subscriptionLength: int 23 5 17 11 23 4 13 8 15 4 ...
## $ paymentStatus : num 1 1 1 1 1 0 1 0 1 1 ...
## $ churn : num 0 0 0 0 0 0 1 1 0 0 ...
```

```
nrow(customerData)
```

```
## [1] 200
```

Question number 1 What are the minimum, maximum, and range of subscription lengths in the dataset?

```
summary(customerData$subscriptionLength)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.00   6.00   10.50   11.72   17.00   24.00
```

```
min_s<- min(customerData$subscriptionLength)
max_s<- max(customerData$subscriptionLength)
range_s<- range(customerData$subscriptionLength)
print(min_s)
```

```
## [1] 1
```

```
print(max_s)
```

```
## [1] 24
```

```
print(range_s)
```

```
## [1] 1 24
```

Question number 2 What does this tell you about the customer base

*Answer:*In minimum subscription Some customers only subscribe for a very short period and in the maximum subscription Loyal customers remain subscribed for up to 24 months, demonstrating satisfaction and long-term engagement with the service

Question number 3 Are customers generally satisfied (based on the average satisfaction score)?

```
avgSatisfaction <- mean(customerData$satisfactionScore)
cat("Average Satisfaction Score:", avgSatisfaction, "\n")
```

```
## Average Satisfaction Score: 5.45
```

Question number 4 What percentage of customers have missed a payment versus those who haven't?

```
missed_payments <- mean(customerData$paymentStatus == 0) * 100
cat("Percentage of Customers with Missed Payments:", missed_payments, "%\n")
```

```
## Percentage of Customers with Missed Payments: 16.5 %
```

Question number 5 Do customers who churn tend to have lower satisfaction scores than those who stay?

```

paymentIssuesChrnd <-
mean(customerData$paymentStatus[customerData$Churn == 1] == 0)

paymentIssuesR <-
mean(customerData$paymentStatus[customerData$Churn == 0] == 0)

cat("Percentage of Payment Issues (Churned):", paymentIssuesChrnd * 100, "%\n")

```

```
## Percentage of Payment Issues (Churned): NaN %
```

```
cat("Percentage of Payment Issues (Retained):", paymentIssuesR * 100, "%\n")
```

```
## Percentage of Payment Issues (Retained): NaN %
```

Question number 6 Is monthly usage lower for customers who churn?

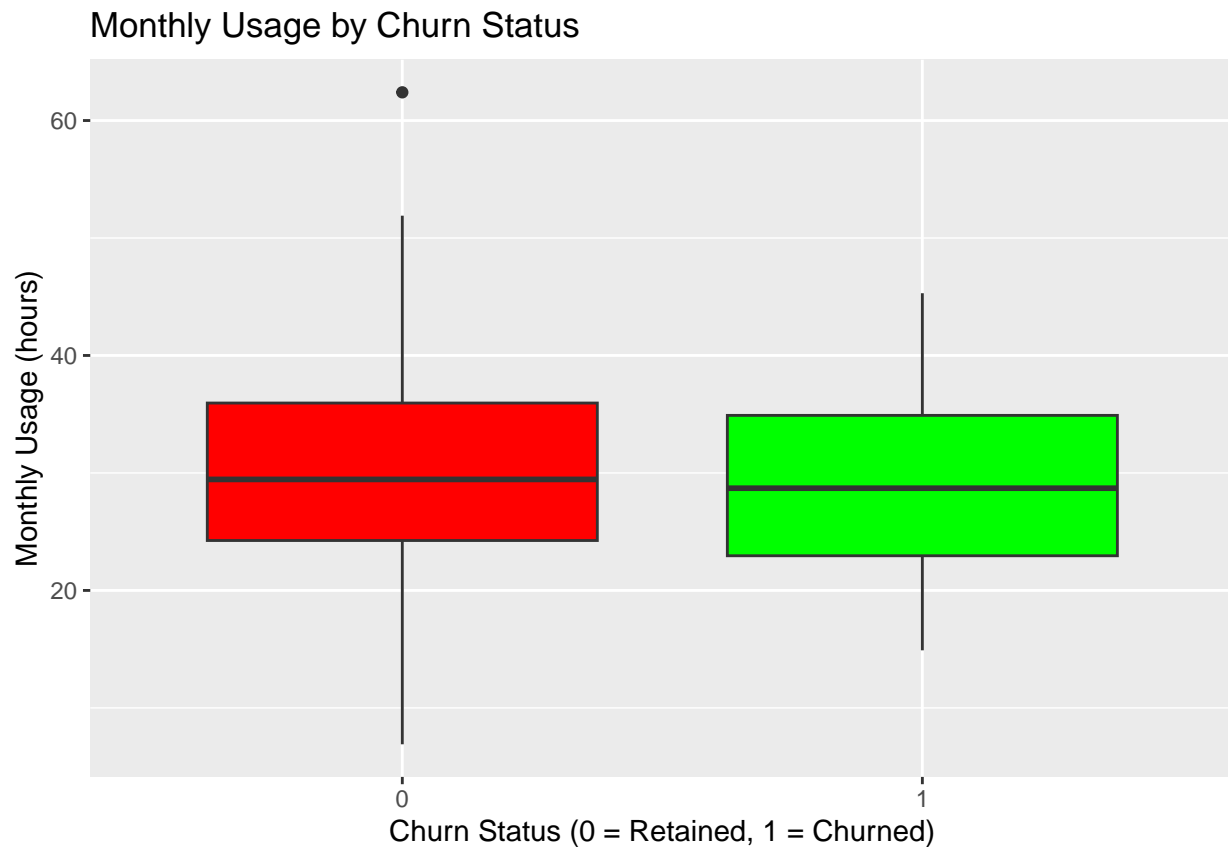
```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.4.2
```

```

ggplot(customerData, aes(x = as.factor(churn), y = monthlyUsage)) +
  geom_boxplot(fill = c("red", "green")) +
  labs(title = "Monthly Usage by Churn Status", x = "Churn Status (0 = Retained, 1 = Churned)", y = "Monthly Usage (hours)")

```



Question number 7

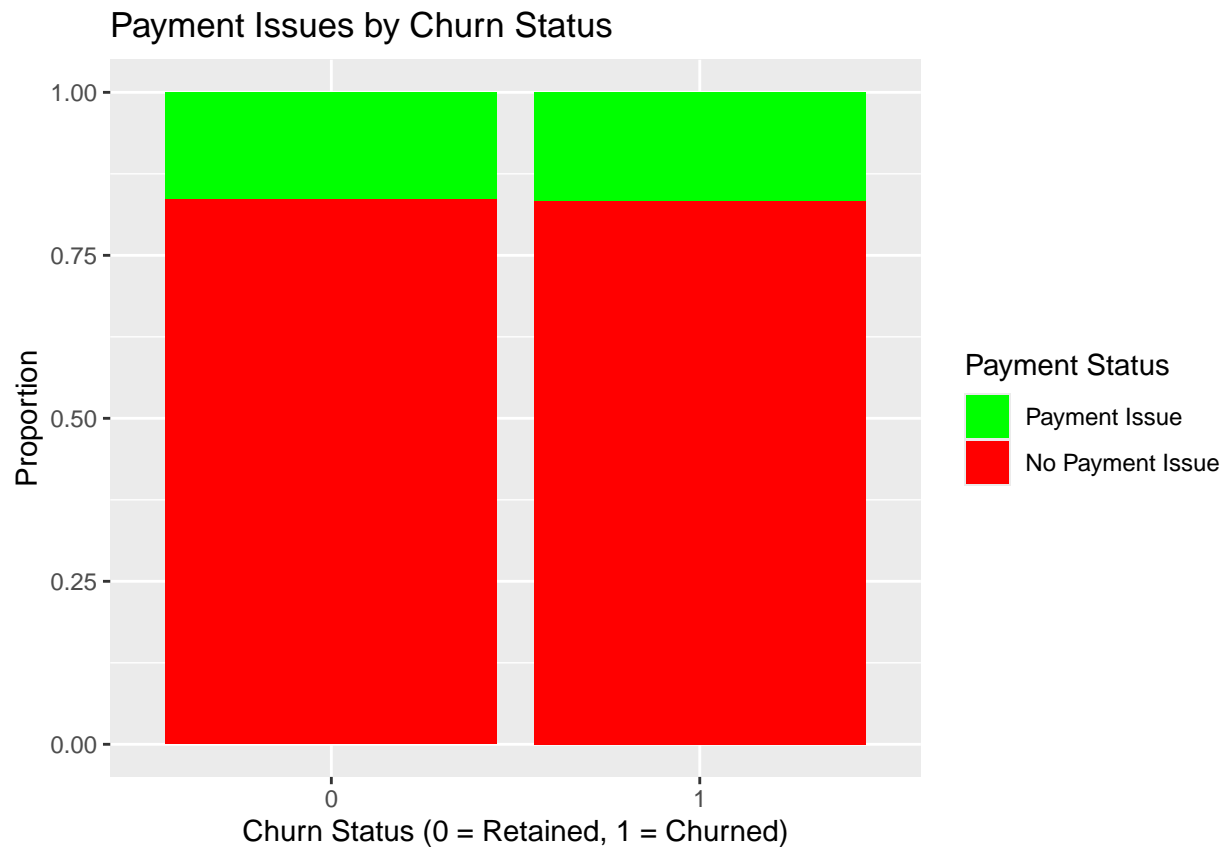
```
table(customerData$paymentStatus, customerData$churn)
```

```
##  
##      0  1  
## 0  24  9  
## 1 122 45
```

```
prop.table(table(customerData$paymentStatus, customerData$churn), margin = 2)
```

```
##  
##      0      1  
## 0 0.1643836 0.1666667  
## 1 0.8356164 0.8333333
```

```
ggplot(customerData, aes(x = as.factor(churn), fill = as.factor(paymentStatus))) +  
  geom_bar(position = "fill") +  
  scale_fill_manual(values = c("green", "red"),  
                    labels = c("Payment Issue", "No Payment Issue")) +  
  labs(title = "Payment Issues by Churn Status",  
       x = "Churn Status (0 = Retained, 1 = Churned)",  
       y = "Proportion",  
       fill = "Payment Status")
```



Question number 8 Are customers who leave the service more likely to have payment issues?

Answer If more churned customers miss payments, it suggests that reliable payments may help retain customers.

Question number 9 Does it seem that missed payments contribute to customer churn?

Answer Yes the chart shows that more churn customers have payment issues than those who stay, suggesting missed payments might lead to churn.

Question number 10 What does this distribution (based on Questions #7-9) tell you about payment reliability?

Answer distribution shows that reliable payments are important for keeping customers. Many churn customers had payment issues that improving payment consistency could lower churn rates.

Question number 11 Do customers with lower usage and satisfaction scores churn more often?

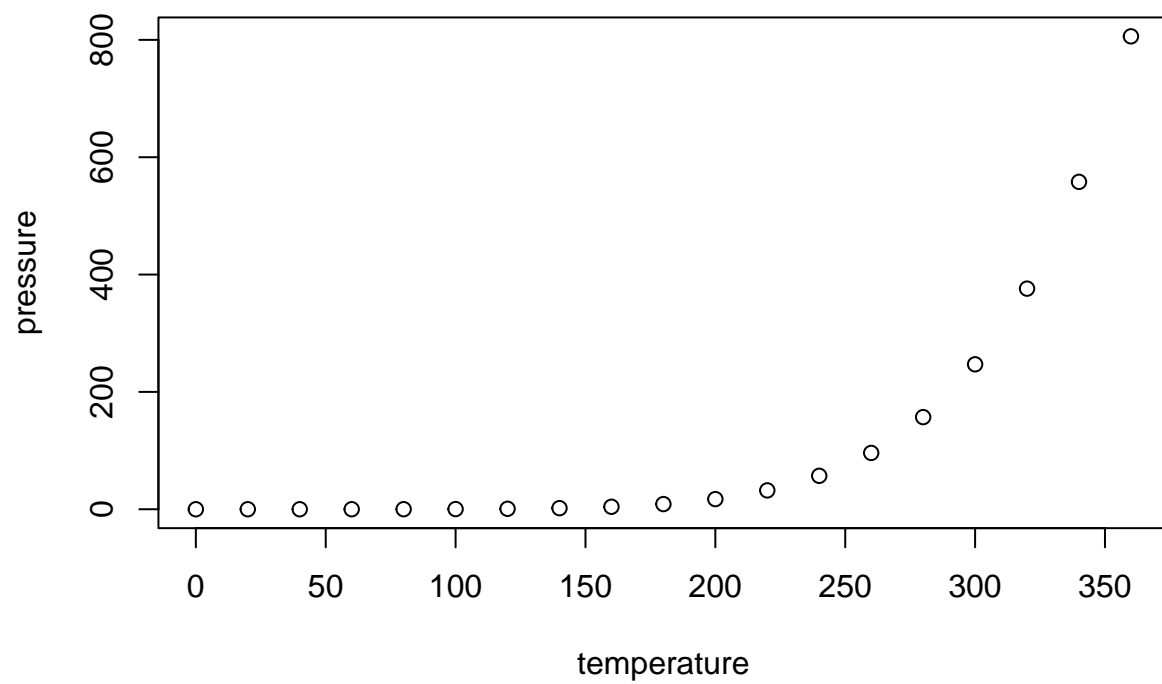
Answer Yes, lower usage and satisfaction to higher churn. Customers who use the service less and feel less satisfied are more likely to leave.

Conclusion

In the analysis shows that low satisfaction, less usage, and missed payments are linked to customer churn. To reduce churn, improving customer satisfaction, increasing engagement, and offering flexible payment options could help

Including Plots

You can also embed plots, for example:



Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.