Call_Center_Final-MiriamDalfin.R

mdalf

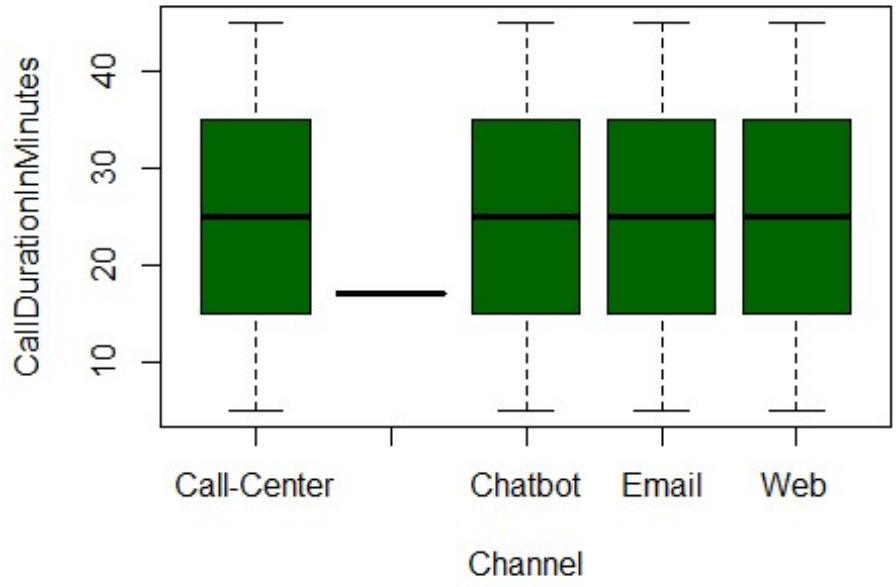
2024-12-22

```r
# What is the correlation between sentiment and reason?
# Which channels have the longest and shortest call durations?
# Which reasons are the most common?

# I changed the csv file to not have spaces in the column names

call_data <- read.csv('Call_center.csv')

# Looking at the data
head(call_data)
```

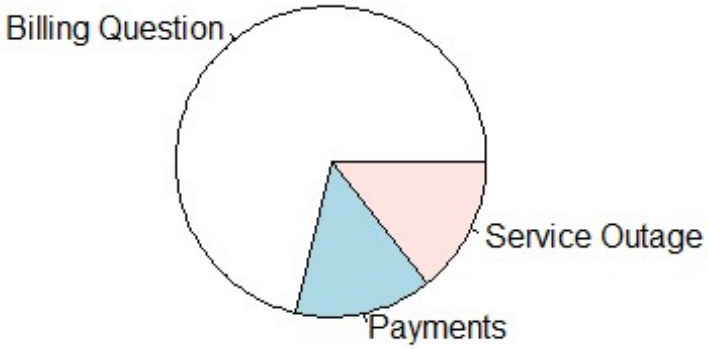
```
##                         Id   CallTimestamp CallCentresCity     Channel
## 1 DKK-57076809-w-055481-fU 10/29/2020 0:00     Los Angeles  CallCenter
## 2 QGK-72219678-w-102139-KY  10/5/2020 0:00       Baltimore     Chatbot
## 3 GYJ-30025932-A-023015-LD  10/4/2020 0:00     Los Angeles Call-Center
## 4 ZJI-96807559-i-620008-m7 10/17/2020 0:00     Los Angeles     Chatbot
## 5 DDU-69451719-O-176482-Fm 10/17/2020 0:00     Los Angeles Call-Center
## 6 JVI-79728660-U-224285-4a 10/28/2020 0:00       Baltimore Call-Center
##             City       CustomerName           Reason ResponseTime
Sentiment
## 1        Detroit   Analise Gairdner Billing Question   Within SLA
Neutral
## 2    Spartanburg   Crichton Kidsley   Service Outage   Within SLA Very
Positive
## 3    Gainesville  Averill Brundrett Billing Question    Above SLA
Negative
## 4       Portland    Noreen Lafflina Billing Question   Within SLA Very
Negative
## 5     Fort Wayne Toma Van der Beken         Payments   Within SLA Very
Positive
## 6 Salt Lake City       Kaylyn Emlen Billing Question   Within SLA
Neutral
##            State CallDurationInMinutes CsatScore
## 1       Michigan                    17         7
## 2 South Carolina                    23        NA
## 3        Florida                    45        NA
## 4         Oregon                    12         1
## 5        Indiana                    23        NA
## 6           Utah                    25         5
```

```r
# Changing sentiment to ordinal values
sentiment_numbers <- c("Very Positive" = 5, "Positive" = 4, "Neutral" = 3,
"Negative" = 2, "Very Negative" = 1)
```

```r
call_data$Sentiment <-
factor(as.numeric(sentiment_numbers[call_data$Sentiment]), ordered = TRUE)

# What is the impact of Reason, ResponseTime, Channel, and CallDuration is on
Sentiment
library(MASS)

## Warning: package 'MASS' was built under R version 4.4.2

call_model <- polr(Sentiment ~ Reason + ResponseTime + CallDurationInMinutes
+ Channel, data = call_data, method = "logistic")

summary(call_model)

##
## Re-fitting to get Hessian

## Call:
## polr(formula = Sentiment ~ Reason + ResponseTime + CallDurationInMinutes +
##     Channel, data = call_data, method = "logistic")
##
## Coefficients:
##                              Value Std. Error   t value
## ReasonPayments         -0.0190833  0.0350415 -0.544593
## ReasonService Outage    0.0167019  0.0295968  0.564314
## ResponseTimeBelow SLA   0.0073454  0.0341485  0.215103
## ResponseTimeWithin SLA -0.0025815  0.0304712 -0.084720
## CallDurationInMinutes  -0.0012338  0.0008345 -1.478401
## ChannelCallCenter       0.6770130  1.4804327  0.457308
## ChannelChatbot          0.0001535  0.0317777  0.004831
## ChannelEmail           -0.0153556  0.0318459 -0.482186
## ChannelWeb              0.0204482  0.0323474  0.632143
##
## Intercepts:
##     Value    Std. Error t value
## 1|2  -1.5272   0.0420   -36.3683
## 2|3   0.0448   0.0410     1.0921
## 3|4   1.2621   0.0417    30.2981
## 4|5   2.2097   0.0436    50.6649
##
## Residual Deviance: 99359.05
## AIC: 99385.05

# It looks like none of the above columns have a significant effect on
sentiment because the t-value is less than 2.

#Does the Channel change the length of Call Duration?
boxplot(CallDurationInMinutes ~ Channel, data = call_data, col = 'darkgreen')
```




```r
# All of the channels have an average of about 25 minutes, so it looks like
the channel doesn't matter.

# Which reasons are the most common?
reasons <- table(call_data$Reason)
pie(reasons)
```




```
# Billing questions are the most common reason, and payments and service
outage seem about equal
```