Call-Center-Analysis.R

wilhe

2023-12-10

call\_center <- read.csv("C:/Users/wilhe/Downloads/Call\_Center.csv")  
  
#load all necessary libraries  
library(ggplot2)  
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

library(gridExtra)

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

##   
## Attaching package: 'gridExtra'

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

library(pivottabler)

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

library(tidyr)

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

#looking at the data  
head(call\_center)

## Id Call.Timestamp Call.Centres.City Channel  
## 1 DKK-57076809-w-055481-fU 10/29/20 0:00 Los Angeles Call-Center  
## 2 QGK-72219678-w-102139-KY 10/5/20 0:00 Baltimore Chatbot  
## 3 GYJ-30025932-A-023015-LD 10/4/20 0:00 Los Angeles Call-Center  
## 4 ZJI-96807559-i-620008-m7 10/17/20 0:00 Los Angeles Chatbot  
## 5 DDU-69451719-O-176482-Fm 10/17/20 0:00 Los Angeles Call-Center  
## 6 JVI-79728660-U-224285-4a 10/28/20 0:00 Baltimore Call-Center  
## City Customer.Name Reason Response.Time  
## 1 Detroit Analise Gairdner Billing Question Within SLA  
## 2 Spartanburg Crichton Kidsley Service Outage Within SLA  
## 3 Gainesville Averill Brundrett Billing Question Above SLA  
## 4 Portland Noreen Lafflina Billing Question Within SLA  
## 5 Fort Wayne Toma Van der Beken Payments Within SLA  
## 6 Salt Lake City Kaylyn Emlen Billing Question Within SLA  
## Sentiment State Call.Duration.In.Minutes Csat.Score  
## 1 Neutral Michigan 17 7  
## 2 Very Positive South Carolina 23 NA  
## 3 Negative Florida 45 NA  
## 4 Very Negative Oregon 12 1  
## 5 Very Positive Indiana 23 NA  
## 6 Neutral Utah 25 5

summary(call\_center)

## Id Call.Timestamp Call.Centres.City Channel   
## Length:32941 Length:32941 Length:32941 Length:32941   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## City Customer.Name Reason Response.Time   
## Length:32941 Length:32941 Length:32941 Length:32941   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Sentiment State Call.Duration.In.Minutes  
## Length:32941 Length:32941 Min. : 5.00   
## Class :character Class :character 1st Qu.:15.00   
## Mode :character Mode :character Median :25.00   
## Mean :25.02   
## 3rd Qu.:35.00   
## Max. :45.00   
##   
## Csat.Score   
## Min. : 1.000   
## 1st Qu.: 4.000   
## Median : 5.000   
## Mean : 5.548   
## 3rd Qu.: 7.000   
## Max. :10.000   
## NA's :20670

#changed call timestamp class to date. it was character  
call\_center$Call.Timestamp <- as.Date(call\_center$Call.Timestamp, format = "%m/%d/%y")  
head(call\_center)

## Id Call.Timestamp Call.Centres.City Channel  
## 1 DKK-57076809-w-055481-fU 2020-10-29 Los Angeles Call-Center  
## 2 QGK-72219678-w-102139-KY 2020-10-05 Baltimore Chatbot  
## 3 GYJ-30025932-A-023015-LD 2020-10-04 Los Angeles Call-Center  
## 4 ZJI-96807559-i-620008-m7 2020-10-17 Los Angeles Chatbot  
## 5 DDU-69451719-O-176482-Fm 2020-10-17 Los Angeles Call-Center  
## 6 JVI-79728660-U-224285-4a 2020-10-28 Baltimore Call-Center  
## City Customer.Name Reason Response.Time  
## 1 Detroit Analise Gairdner Billing Question Within SLA  
## 2 Spartanburg Crichton Kidsley Service Outage Within SLA  
## 3 Gainesville Averill Brundrett Billing Question Above SLA  
## 4 Portland Noreen Lafflina Billing Question Within SLA  
## 5 Fort Wayne Toma Van der Beken Payments Within SLA  
## 6 Salt Lake City Kaylyn Emlen Billing Question Within SLA  
## Sentiment State Call.Duration.In.Minutes Csat.Score  
## 1 Neutral Michigan 17 7  
## 2 Very Positive South Carolina 23 NA  
## 3 Negative Florida 45 NA  
## 4 Very Negative Oregon 12 1  
## 5 Very Positive Indiana 23 NA  
## 6 Neutral Utah 25 5

summary(call\_center)

## Id Call.Timestamp Call.Centres.City Channel   
## Length:32941 Min. :2020-10-01 Length:32941 Length:32941   
## Class :character 1st Qu.:2020-10-08 Class :character Class :character   
## Mode :character Median :2020-10-16 Mode :character Mode :character   
## Mean :2020-10-15   
## 3rd Qu.:2020-10-23   
## Max. :2020-10-31   
##   
## City Customer.Name Reason Response.Time   
## Length:32941 Length:32941 Length:32941 Length:32941   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
##   
## Sentiment State Call.Duration.In.Minutes  
## Length:32941 Length:32941 Min. : 5.00   
## Class :character Class :character 1st Qu.:15.00   
## Mode :character Mode :character Median :25.00   
## Mean :25.02   
## 3rd Qu.:35.00   
## Max. :45.00   
##   
## Csat.Score   
## Min. : 1.000   
## 1st Qu.: 4.000   
## Median : 5.000   
## Mean : 5.548   
## 3rd Qu.: 7.000   
## Max. :10.000   
## NA's :20670

class(call\_center$Call.Timestamp)

## [1] "Date"

#looking at the possible values for these variables  
unique(call\_center$Channel)

## [1] "Call-Center" "Chatbot" "Email" "Web"

unique(call\_center$Reason)

## [1] "Billing Question" "Service Outage" "Payments"

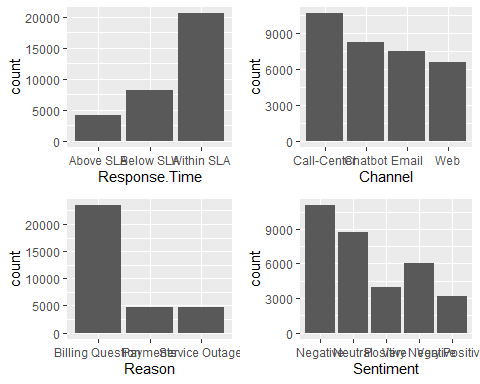
unique(call\_center$Response.Time)

## [1] "Within SLA" "Above SLA" "Below SLA"

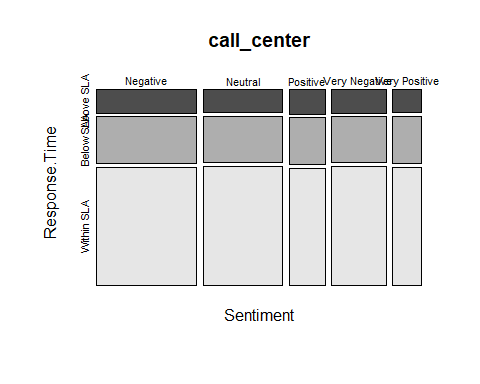
unique(call\_center$Sentiment)

## [1] "Neutral" "Very Positive" "Negative" "Very Negative"  
## [5] "Positive"

#plot each of the above to visualize the distributions  
response\_time\_plot <- ggplot(call\_center) +  
 geom\_bar(aes(x=Response.Time))   
  
channel\_plot <- ggplot(call\_center) +  
 geom\_bar(aes(x=Channel))   
   
  
reason\_plot <- ggplot(call\_center) +  
 geom\_bar(aes(x=Reason))  
  
sentiment\_plot <- ggplot(call\_center) +  
 geom\_bar(aes(x=Sentiment))   
  
#arrange plots to display together  
grid.arrange(response\_time\_plot, channel\_plot, reason\_plot, sentiment\_plot, nrow=2)



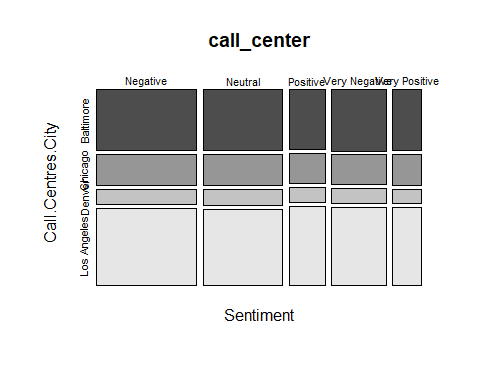
####Seeing the relationship of Sentiment with other variables  
  
#plot of relationship of sentiment and response time  
mosaicplot(Sentiment~Response.Time, data=call\_center, color=TRUE)



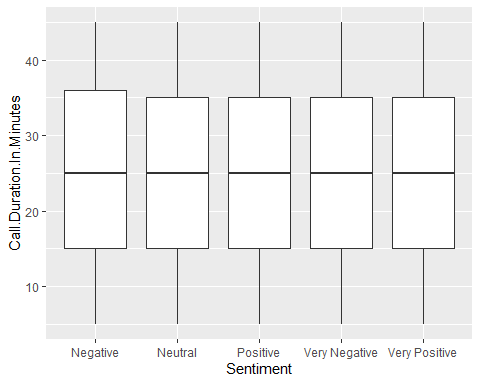
#pivot table of the above information to see it in numbers  
call\_center |>   
 count(Sentiment, Response.Time) |>   
 pivot\_wider(names\_from = Response.Time, values\_from = n, values\_fill = 0)

## # A tibble: 5 × 4  
## Sentiment `Above SLA` `Below SLA` `Within SLA`  
## <chr> <int> <int> <int>  
## 1 Negative 1406 2745 6912  
## 2 Neutral 1076 2169 5509  
## 3 Positive 520 979 2429  
## 4 Very Negative 766 1472 3788  
## 5 Very Positive 400 783 1987

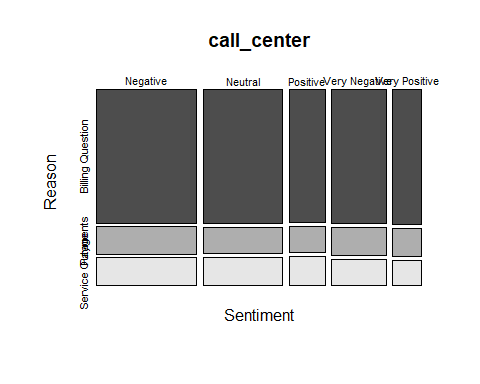
#plot of relationship of sentiment and each specific call center  
mosaicplot(Sentiment~Call.Centres.City, data=call\_center, color=TRUE)



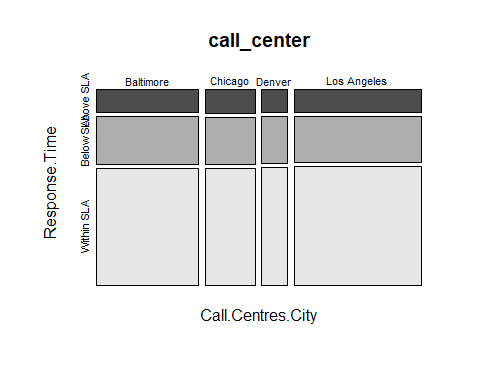
#plot of relationship of sentiment and the call duration  
ggplot(call\_center, aes(x=Sentiment, y= Call.Duration.In.Minutes)) +  
 geom\_boxplot()



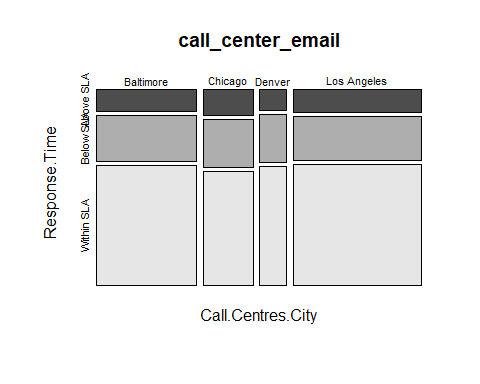
#plot of relationship of sentiment and the call reason  
mosaicplot(Sentiment~Reason, data=call\_center, color=TRUE)



######Is there one call center that is significantly better or worse than the others  
  
#Does one center have more people waiting for longer  
mosaicplot(Call.Centres.City~Response.Time, data=call\_center, color=TRUE)



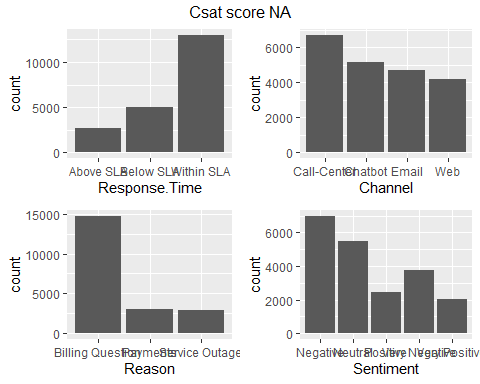
#looking at response time specifically for email  
call\_center\_email <- call\_center |>   
 filter(Channel == "Email")  
mosaicplot(Call.Centres.City~Response.Time, data=call\_center\_email, color=TRUE)



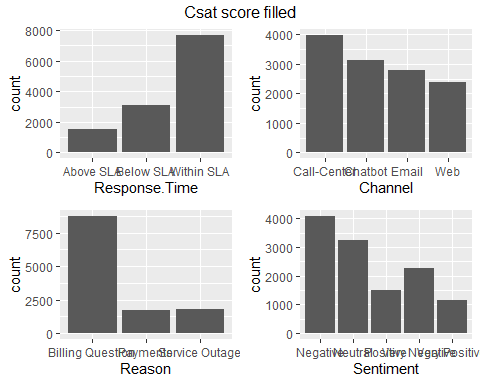
##### Why are there so many NAs in the customer satisfaction score column. Is there a specific reason  
  
#Separate the data based on the the Csat score column is filled in or not  
csat\_NA <- call\_center |>   
 filter(is.na(Csat.Score))  
  
csat\_fill <- call\_center |>   
 filter(!is.na(Csat.Score))  
  
#histogram of csat scores  
ggplot(csat\_fill, aes(x=Csat.Score)) +  
 geom\_histogram(bins=20)



#plot each of the above to visualize the distributions  
response\_time\_NA <- ggplot(csat\_NA) +  
 geom\_bar(aes(x=Response.Time))   
  
channel\_NA <- ggplot(csat\_NA) +  
 geom\_bar(aes(x=Channel))   
   
   
reason\_NA <- ggplot(csat\_NA) +  
 geom\_bar(aes(x=Reason))  
  
sentiment\_NA <- ggplot(csat\_NA) +  
 geom\_bar(aes(x=Sentiment))   
  
#arrange plots to display together  
grid.arrange(response\_time\_NA, channel\_NA, reason\_NA, sentiment\_NA, nrow=2, top="Csat score NA")



#plot each of the above to visualize the distributions  
response\_time\_fill <- ggplot(csat\_fill) +  
 geom\_bar(aes(x=Response.Time))   
  
channel\_fill <- ggplot(csat\_fill) +  
 geom\_bar(aes(x=Channel))   
  
  
reason\_fill <- ggplot(csat\_fill) +  
 geom\_bar(aes(x=Reason))  
  
sentiment\_fill <- ggplot(csat\_fill) +  
 geom\_bar(aes(x=Sentiment))   
  
#arrange plots to display together  
grid.arrange(response\_time\_fill, channel\_fill, reason\_fill, sentiment\_fill, nrow=2, top="Csat score filled")



#finding the mean of the two tables  
mean(csat\_fill$Call.Duration.In.Minutes)

## [1] 25.17928

mean(csat\_NA$Call.Duration.In.Minutes)

## [1] 24.92729

#Seeing if specific states filled in the score more than others  
sort(table(csat\_fill$State), decreasing = TRUE)[1:5]

##   
## Texas California Florida New York Virginia   
## 1402 1273 1089 714 434

sort(table(csat\_NA$State), decreasing = TRUE)[1:5]

##   
## California Texas Florida New York Virginia   
## 2358 2170 1745 1072 730

##Seeing if specific dates that people filled in the score more than others  
sort(table(csat\_fill$Call.Timestamp), decreasing = TRUE)[1:5]

##   
## 2020-10-23 2020-10-21 2020-10-28 2020-10-30 2020-10-20   
## 439 438 435 434 431

sort(table(csat\_NA$Call.Timestamp), decreasing = TRUE)[1:5]

##   
## 2020-10-22 2020-10-06 2020-10-21 2020-10-13 2020-10-18   
## 765 735 732 711 709