Part-II.R

ae7le

2024-12-15

# Load and view data  
data = read.csv("C:/Users/ae7le/OneDrive/Documents/Sara Schenirer/Intro to Data Science/Final\_Project/Call\_Center.csv")  
summary(data)

## 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

head(data)

## 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

# Is the type of customer service channel significant to customer satisfaction?  
  
# View the different options of customer service channel  
unique(data$Channel)

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

# Drop the N/As in the Csat score column  
data1 = na.omit(data)  
summary(data1)

## Id Call.Timestamp Call.Centres.City Channel   
## Length:12271 Length:12271 Length:12271 Length:12271   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## City Customer.Name Reason Response.Time   
## Length:12271 Length:12271 Length:12271 Length:12271   
## Class :character Class :character Class :character Class :character   
## Mode :character Mode :character Mode :character Mode :character   
##   
##   
##   
## Sentiment State Call.Duration.In.Minutes  
## Length:12271 Length:12271 Min. : 5.00   
## Class :character Class :character 1st Qu.:15.00   
## Mode :character Mode :character Median :25.00   
## Mean :25.18   
## 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

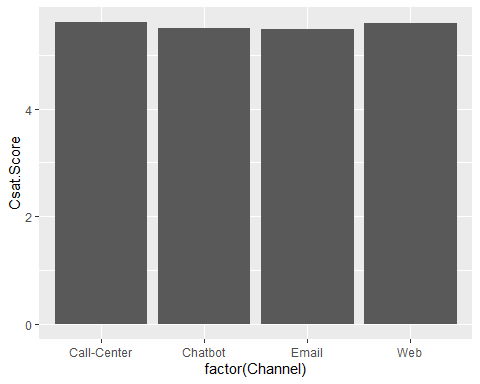
# Create a linear regression model for Csat score by channel  
model1 <- lm(Csat.Score ~ Channel, data = data1)  
summary(model1)

##   
## Call:  
## lm(formula = Csat.Score ~ Channel, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.6133 -1.6133 -0.4817 1.5183 4.5183   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.61331 0.03764 149.139 <2e-16 \*\*\*  
## ChannelChatbot -0.12084 0.05672 -2.130 0.0332 \*   
## ChannelEmail -0.13159 0.05857 -2.247 0.0247 \*   
## ChannelWeb -0.02158 0.06136 -0.352 0.7250   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.371 on 12267 degrees of freedom  
## Multiple R-squared: 0.0006288, Adjusted R-squared: 0.0003844   
## F-statistic: 2.573 on 3 and 12267 DF, p-value: 0.05226

# Show a bar plot of the mean Csat score by channel  
library(ggplot2)

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

ggplot(data1, aes(x = factor(Channel), y = Csat.Score)) +   
 stat\_summary(fun = "mean", geom = "bar")



# Use the chi-squared test to look for correlation between Sentiment and Channel (no need to drop rows)  
chisq.test(data$Sentiment, data$Channel)

##   
## Pearson's Chi-squared test  
##   
## data: data$Sentiment and data$Channel  
## X-squared = 12.91, df = 12, p-value = 0.3756

# Are some call centers better than others i.e. is the call center city significant in the two customer satisfaction scores?  
  
# Linear regression model for Csat score by Call Centre City  
model2 <- lm(Csat.Score ~ Call.Centres.City, data = data1)  
summary(model2)

##   
## Call:  
## lm(formula = Csat.Score ~ Call.Centres.City, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.6176 -1.5597 -0.4794 1.5206 4.5206   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.559700 0.036900 150.670 <2e-16 \*\*\*  
## Call.Centres.CityChicago -0.080261 0.064661 -1.241 0.215   
## Call.Centres.CityDenver 0.057864 0.081673 0.708 0.479   
## Call.Centres.CityLos Angeles -0.007725 0.049662 -0.156 0.876   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.371 on 12267 degrees of freedom  
## Multiple R-squared: 0.0002195, Adjusted R-squared: -2.501e-05   
## F-statistic: 0.8977 on 3 and 12267 DF, p-value: 0.4414

# Chi-squared test between Sentiment and Call Centre City  
chisq.test(data$Sentiment, data$Call.Centres.City)

##   
## Pearson's Chi-squared test  
##   
## data: data$Sentiment and data$Call.Centres.City  
## X-squared = 12.486, df = 12, p-value = 0.4075

# What is the most significant factor for customer satisfaction?  
  
# Create a linear model for CSat score with all likely factors for customer satisfaction  
model3 <- lm(Csat.Score ~ Call.Centres.City + Channel + Reason + Response.Time + Call.Duration.In.Minutes, data = data1)  
summary(model3)

##   
## Call:  
## lm(formula = Csat.Score ~ Call.Centres.City + Channel + Reason +   
## Response.Time + Call.Duration.In.Minutes, data = data1)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -4.7507 -1.6122 -0.3843 1.5466 4.6505   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.728367 0.093304 61.395 <2e-16 \*\*\*  
## Call.Centres.CityChicago -0.079451 0.064666 -1.229 0.219   
## Call.Centres.CityDenver 0.062628 0.081684 0.767 0.443   
## Call.Centres.CityLos Angeles -0.006536 0.049673 -0.132 0.895   
## ChannelChatbot -0.110118 0.068298 -1.612 0.107   
## ChannelEmail -0.119079 0.068900 -1.728 0.084 .   
## ChannelWeb -0.008634 0.070425 -0.123 0.902   
## ReasonPayments 0.038426 0.075803 0.507 0.612   
## ReasonService Outage 0.016655 0.064265 0.259 0.796   
## Response.TimeBelow SLA -0.056151 0.074308 -0.756 0.450   
## Response.TimeWithin SLA -0.089652 0.066629 -1.346 0.178   
## Call.Duration.In.Minutes -0.002015 0.001810 -1.113 0.266   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 2.371 on 12259 degrees of freedom  
## Multiple R-squared: 0.001143, Adjusted R-squared: 0.0002463   
## F-statistic: 1.275 on 11 and 12259 DF, p-value: 0.2319