**Programming Codes:**

#Reading Comcast Data and loading libraries

rm(list=ls())

library(dplyr)

library(ggplot2)

library(lubridate)

library(plyr)

library(zoo)

library(NLP)

library(tm)

library(RColorBrewer)

library(wordcloud)

library(gridExtra)

setwd("E:\\Simplilearn\\Data Science with R\\Project\\Comcast")

comcast<- read.csv("Comcast Telecom Complaints data.csv")

head(comcast)

str(comcast)

sum(is.na(comcast))

# Cleaning the date

comcast$Date<- dmy(comcast$Date)

#Provide the trend chart for the number of complaints at monthly and daily granularity levels.

#Daily Trend

a = aggregate(comcast$Date, by=list(comcast$Date), FUN=length)

str(a)

ggplot(a, aes(Group.1, x, group = 1)) +

geom\_line() +

xlab("Date") +

ylab("Number of Complaints")

#Monthly Trend

month<- as.yearmon(comcast$Date, format = "%d-%m-%Y")

b<-cbind(comcast,month)

c = aggregate(b$month, by=list(b$month), FUN=length)

str(c)

ggplot(c, aes(Group.1, x, group = 1)) +

geom\_line() +

xlab("Date") +

ylab("Number of Complaints")

#Provide a table with the frequency of complaint types. Which complaint types are maximum i.e., around internet, network issues, or across any other domains.

complaint\_list = data.frame(comcast$Customer.Complaint)

colnames(complaint\_list)=c("Complaint")

corpus= Corpus(VectorSource(complaint\_list$Complaint))

#Text Cleaning

corpus <- tm\_map(corpus,content\_transformer(tolower)) #Converting all text into lower case

corpus<- tm\_map(corpus,removeNumbers) #Remove Numbers

corpus = tm\_map(corpus,removeWords,stopwords(kind="en"))#Removing common stop words

corpus = tm\_map(corpus,removePunctuation)#Remove Punctuation

corpus = tm\_map(corpus,stripWhitespace)#Removing white spaces

corpus= tm\_map(corpus,removeWords,c("get","took","can","can","comcast"))#Remove additional words

#Create Term Document Matrix (TDM)

tdm = TermDocumentMatrix(corpus)

m=as.matrix(tdm)

v=sort(rowSums(m),decreasing = T)

#List with Frequency of Compliant Types

d=data.frame(word=names(v),freq=v)

#word cloud

set.seed(2)

wordcloud(d$word,d$freq,random.order=F,min.freq = 5, max.words=1000, rot.per=0.2, colors=brewer.pal(5, "Dark2"), scale=c(4,0.2))

title(main = "Complaint Types - Word Cloud",font.main=1,cex.main=1.5)

# Complaint Type Processing as seen from word cloud

internet\_tickets<- contains(comcast$Customer.Complaint,match = 'internet',ignore.case = T)

billing\_tickets<- contains(comcast$Customer.Complaint,match = 'bill',ignore.case = T)

data\_tickets<- contains(comcast$Customer.Complaint,match = 'data',ignore.case = T)

service\_tickets<- contains(comcast$Customer.Complaint,match = 'service',ignore.case = T)

speed\_tickets<- contains(comcast$Customer.Complaint,match = 'speed',ignore.case = T)

charges\_tickets<- contains(comcast$Customer.Complaint,match = 'charge',ignore.case = T)

comcast$ComplaintType[internet\_tickets]<- "Internet"

comcast$ComplaintType[billing\_tickets]<- "Billing"

comcast$ComplaintType[data\_tickets]<- "data"

comcast$ComplaintType[service\_tickets]<- "service"

comcast$ComplaintType[speed\_tickets]<- "speed"

comcast$ComplaintType[charges\_tickets]<- "charges"

comcast$ComplaintType[-c(internet\_tickets,

billing\_tickets,data\_tickets,service\_tickets,speed\_tickets,charges\_tickets)]<- "Others"

table(comcast$ComplaintType)

#Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed.

comcast$Status\_New<-revalue(comcast$Status, c(Pending = "Open", Solved = "Closed"))

head(comcast)

#Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:

ggplot(comcast, aes(y = State)) + geom\_bar(aes(fill = Status\_New))

#Which state has the maximum complaints

count\_complaints = aggregate(comcast$Status\_New, by=list(comcast$State), FUN=length)

df <-count\_complaints[order(-count\_complaints$x),]

head(df,5)

#Which state has the highest percentage of unresolved complaints

resolved\_data<- subset(comcast, Status\_New == "Closed")

unresolved\_data<- subset(comcast, Status\_New == "Open")

resolved = aggregate(resolved\_data$Status\_New, by=list(resolved\_data$State), FUN=length)

unresolved = aggregate(unresolved\_data$Status\_New, by=list(unresolved\_data$State), FUN=length)

per = merge(resolved, unresolved, by = "Group.1")

per$per = (per$x.y/(per$x.x+per$x.y)\*100)

high\_unresolved = aggregate(per$per, by=list(per$Group.1), FUN=max)

high\_unresolved[(high\_unresolved$x == max(high\_unresolved$x)),]

df1 <-high\_unresolved[order(-high\_unresolved$x),]

head(df1,5)

#Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

resolved1 = aggregate(resolved\_data$Status\_New, by=list(resolved\_data$Received.Via), FUN=length)

unresolved1 = aggregate(unresolved\_data$Status\_New, by=list(unresolved\_data$Received.Via), FUN=length)

per1 = merge(resolved1, unresolved1, by = "Group.1")

per1$per = (per1$x.x/(per1$x.x+per1$x.y)\*100)

per1

df2 <- table(comcast$Received.Via, comcast$Status\_New)

df2 <- cbind(df2, Total = rowSums(df2))

df2

# Pie Chart with Percentages

slices <- c(864, 255)

lbls <- c("Closed", "Open")

pct <- round(slices/sum(slices)\*100)

lbls <- paste(lbls, pct) # add percents to labels

lbls <- paste(lbls,"%",sep="") # ad % to labels

pie(slices,labels = lbls, col=rainbow(length(lbls)),

main="Pie Chart of Received Via Call")

# Pie Chart with Percentages

slices <- c(843, 262)

lbls <- c("Closed", "Open")

pct <- round(slices/sum(slices)\*100)

lbls <- paste(lbls, pct) # add percents to labels

lbls <- paste(lbls,"%",sep="") # ad % to labels

pie(slices,labels = lbls, col=rainbow(length(lbls)),

main="Pie Chart of Received Via Internet")

-------------------------------------------------------------------The End-------------------------------------------------------------------