**Project Report**

**Rushikesh Khankar**

**Comcast Telecom Consumer Complaints**

**Objective:**

**Comcast Telecom Consumer Complaints**

Provide the trend chart for the number of complaints at monthly and daily granularity levels. Provide a table with the frequency of complaint types. Complaint types which are maximum i.e., around internet, network issues, or across any other domains. 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. Provide state wise status of complaints in a stacked bar chart. Check the state has the maximum complaints. State has the highest percentage of unresolved complaints. Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

**Data set names:**

Comcast Telecom Complaints data

**Source code:**

#- Import data into R environment.

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

#- 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.

#- 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.

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

# Which state has the maximum complaints

#Which state has the highest percentage of unresolved complaints

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

library(dplyr)

library(ggplot2)

library(stringi)

library(lubridate)

library(data.table)

library(plyr)

library(ggpubr)

#Import data

Comcast\_data <- read.csv("~/Desktop/R/Comcast Telecom Complaints data.csv")

Comcast\_data

summary(Comcast\_data)

#Extracting monthly and daily count

Comcast\_data$Date <- dmy(Comcast\_data$Date)

Comcast\_data$datemonth <- months(as.Date(Comcast\_data$Date))

#Comcast\_data$datemonth

Months\_counts <- table(Comcast\_data$datemonth)

Months\_counts

month\_counts\_df <- as.data.frame(Months\_counts)

month\_counts\_df

names(month\_counts\_df)[1]<- "Months"

names(month\_counts\_df)[2]<- "Count"

month\_counts\_df

Comcast\_data$date\_of\_each\_month <- day(as.Date(Comcast\_data$Date))

Comcast\_data$date\_of\_each\_month

date\_counts <- table(Comcast\_data$date\_of\_each\_month)

date\_counts\_df <- as.data.frame(date\_counts)

date\_counts\_df

names(date\_counts\_df)[1]<- "date\_of\_each\_month"

names(date\_counts\_df)[2]<- "Count"

date\_counts\_df

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

#Trend chart for monthly count

ggplot(data = month\_counts\_df, aes(x = Months, y = Count, label = Count, group = 1)) +

geom\_line(color = "green") +

geom\_point(size = 0.8) + geom\_text()+

labs(title = "Monthly Ticket Count", x = "Months", y = "No. of Tickets")

#Trend Chart for daily count

ggplot(data = date\_counts\_df, aes(x= date\_of\_each\_month, y=Count, label = Count, group = 1)) +

geom\_line(color = "blue") +

geom\_text()+labs(title = "Daily Ticket Count", x="Dates of Each Month", y = "No. of Tickets")

#Complaint Type

Comcast\_data$Customer.Complaint

complaint\_counts <- table(Comcast\_data$Customer.Complaint)

complaint\_counts

network\_tickets <- data.frame(contains(Comcast\_data$Customer.Complaint, match = 'network', ignore.case = T))

internet\_tickets <- data.frame(contains(Comcast\_data$Customer.Complaint, match = 'internet', ignore.case = T))

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

datacap\_tickets <- data.frame(contains(Comcast\_data$Customer.Complaint, match = 'data cap', ignore.case = T))

customerservice\_tickets <- data.frame(contains(Comcast\_data$Customer.Complaint, match = 'customer service', ignore.case = T))

nrow(network\_tickets)

nrow(internet\_tickets)

nrow(billing\_tickets)

nrow(datacap\_tickets)

nrow(customerservice\_tickets)

#Complaint Type Thas Has Maximum Tickets - Internet issues has maximum tickets

if(nrow(network\_tickets) > nrow(internet\_tickets)){

print("Network Issues")

}else if(nrow(internet\_tickets) > nrow(billing\_tickets)){

print("Internet Issues")

}else if(nrow(billing\_tickets) > nrow(datacap\_tickets)){

print("Billing Issues")

}else if(nrow(datacap\_tickets) > nrow(customerservice\_tickets)){

print("Data cap Issues")

}else{

print("Customer Service Issues")

}

#Internet issues has maximum tickets

#Open And Pending Statuses Are Cons As Open

Open\_complaints <- (Comcast\_data$Status == "Open")

Comcast\_data$Complaint\_Status[Open\_complaints] <- "Open"

#Solved And Closed Statuses are considered as "Closed"

Comcast\_data$Status <- gsub('Solved','Closed', Comcast\_data$Status)

Closed\_complaints <- (Comcast\_data$Status == "Closed")

Comcast\_data$Complaint\_Status[Closed\_complaints] <- "Closed"

#Stacked bar chart for open and closed complaints

Comcast\_data <- group\_by(Comcast\_data, State, Complaint\_Status)

chart\_data <- dplyr::summarise(Comcast\_data, count = n())

ggplot(as.data.frame(chart\_data), mapping = aes(State,count)) +

geom\_col(aes(fill = Complaint\_Status), width = 0.95) +

theme(axis.text.x = element\_text(angle = 90),

axis.title.y = element\_text(size = 15),

axis.title.x = element\_text(size = 15),

title = element\_text(size = 16, color = "#0073C2FF"),

plot.title = element\_text(hjust = 0.5))+

labs(title = "Ticket Status Stacked Bar Chart",

x = "States", y = "No. of Tickets",

fill = "Status")

#State that has the maximum complaints

chart\_data%>%filter(Complaint\_Status=="Open") -> Open\_complaints

#max(Open\_complaints$count)

Open\_complaints[Open\_complaints$count == max(Open\_complaints$count), c(1,3)]

#Complaints which were recieved through the internet and customer care calls.

Resolved\_data <- group\_by(Comcast\_data, Complaint\_Status)

Total\_resolved <- dplyr::summarise(Resolved\_data, percentage = (n()/nrow(Resolved\_data)))

#Total\_resolved

Resolved\_data1 <- group\_by(Comcast\_data, Received.Via, Complaint\_Status)

Category\_resolved <- dplyr::summarise(Resolved\_data1, percentage = (n()/nrow(Resolved\_data)))

#Category\_resolved

#Pie Chart For Category Wise Ticket Status

par(mfrow = c(1,2))

total <- ggplot(data = Total\_resolved,

aes(x= "", y = percentage, fill = Complaint\_Status)) +

geom\_bar(stat = "identity", width = 1) +

coord\_polar("y", start = 0) +

geom\_text(aes(label = paste0(round(percentage\*100),"%")),

position = position\_stack(vjust = 0.5))+

labs(title = "Pie Chart based on Ticket Status", x = NULL, y = NULL, fill = NULL ) +

theme\_classic()+theme(axis.line = element\_blank(),

axis.text = element\_blank(),

axis.ticks = element\_blank())

category <- ggplot(data = Category\_resolved,

aes(x = "", y = percentage, fill = Complaint\_Status)) +

geom\_bar(stat = "identity", width = 1) +

coord\_polar("y", start = 0)+

geom\_text(aes(label = paste0(Received.Via, "", round(percentage\*100), "%")),

position = position\_stack(vjust = 0.5)) +

labs(title = "Pie Chart for Category wise Ticket Status", x = NULL, y =NULL, fill = NULL) +

theme\_classic()+theme(axis.line = element\_blank(),

axis.text = element\_blank(),

axis.ticks = element\_blank())

ggarrange(total, category, nrow = 1, ncol = 2)

Chart, line chart

Description automatically generated

Chart, line chart

Description automatically generated

Chart, bar chart

Description automatically generated

Chart, pie chart

Description automatically generated

**Insights:**

June month has maximum no of complaints.

More complaints are based on Internet Issues.

2nd half of each month has more number of complaints.

Georgia State has maximum complaints.

Georgia State has maximum Open/unsolved complaints.

Total resolved complaints are 77% in which 38% are received from the internet and 39% are from the customer care calls.

There are 23% complaints are still unresolved and in which 12% are received from the internet and 11% are from the customer care calls.