Project 2 : Comcast Telecom Consumer Complaints

DESCRIPTION

Comcast is an American global telecommunication company. The firm has been providing terrible customer service. They continue to fall short despite repeated promises to improve. Only last month (October 2016) the authority fined them a $2.3 million, after receiving over 1000 consumer complaints.

The existing database will serve as a repository of public customer complaints filed against Comcast.
It will help to pin down what is wrong with Comcast's customer service.

#Title: Comcast Telecom Consumer Analysis
#Name : Divya Mukunde
#Date : 16/09/2021
#Output : pdf document
#Task performed
#- 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
 # 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.

#Load necessary packages

install.packages("dplyr")
install.packages("ggplot2")
install.packages("lubridate")
install.packages("readxl")
library(readxl)
library(dplyr)
library(ggplot2)
library(lubridate)

#Load necessary dataset

setwd(choose.dir())
getwd()
telecom_data <- read.csv("Comcast Telecom Complaints data.csv")
View(telecom_data)

```
#Converting field names
names(telecom_data) <- gsub(pattern = '\\.',replacement = "",x=names(telecom_data))
names(telecom_data)
View(telecom_data)

#Dataset shows the format of date column is not appropriate for analysis, so need to make it same.
telecom_data$Date <- dmy(telecom_data$Date)
View(telecom_data)

#Now to get the complaints on daily basis and plot a trend chart for it.
ans <- telecom_data %>% group_by(Date) %>% summarize(NumOfComplaints=n())

#Plot for daily granularity level
ggplot(data = ans,aes(as.POSIXct(Date),NumOfComplaints))+
  geom_line()+
  geom_point(size = 1)+
  scale_x_datetime(breaks = "1 weeks",date_labels = "%d/%m")+
  labs(title = "Daily Ticket Count",x= "Days",y ="No. of Tickets")+
  theme(axis.text.x = element_text(angle = 75),
      plot.title = element_text(hjust = 0.5))
```
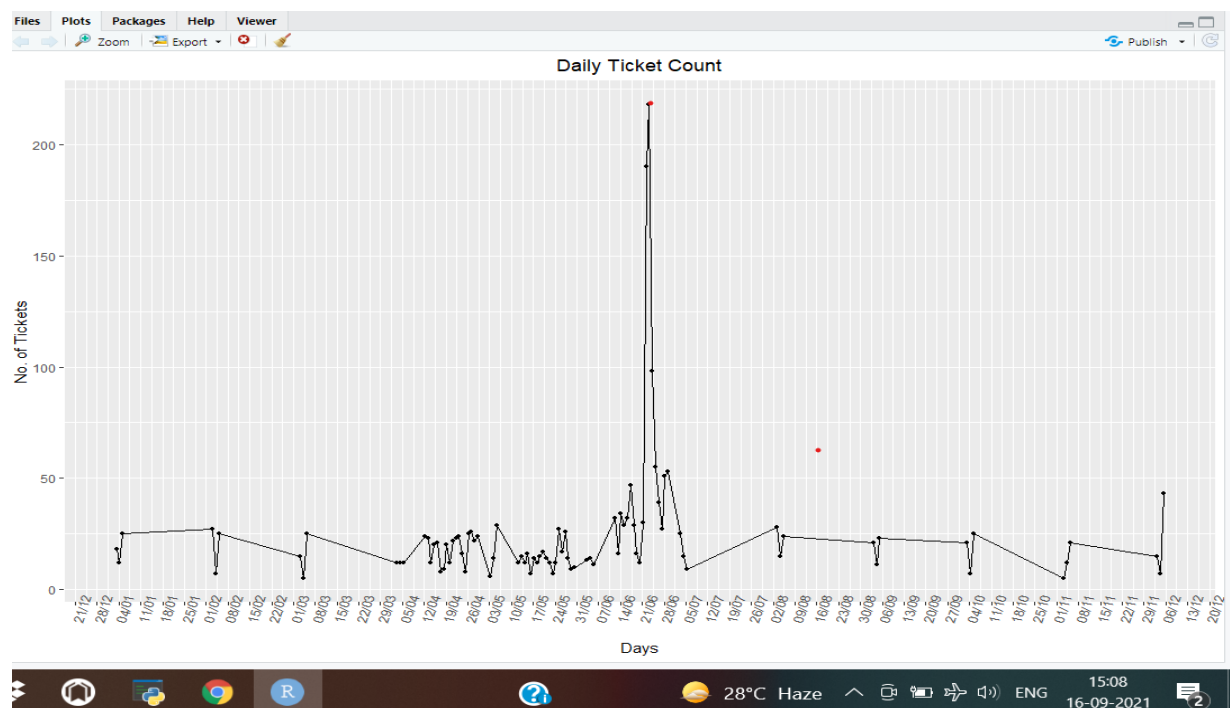


```
#Plot for monthly granularity level
telecom_data$Month<-months(telecom_data$Date)
ans1 <- telecom_data %>% group_by(Month =as.integer(month(Date))) %>%
summarize(NumOfComplaints=n()) %>% arrange(desc(NumOfComplaints))
ggplot(data = ans1,aes(Month,NumOfComplaints,label = NumOfComplaints))+
  geom_line()+
```

```
geom_point(size = 0.8)+
geom_text()+
scale_x_continuous(breaks = ans1$Month)+
labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")+
theme(plot.title = element_text(hjust = 0.5))
```
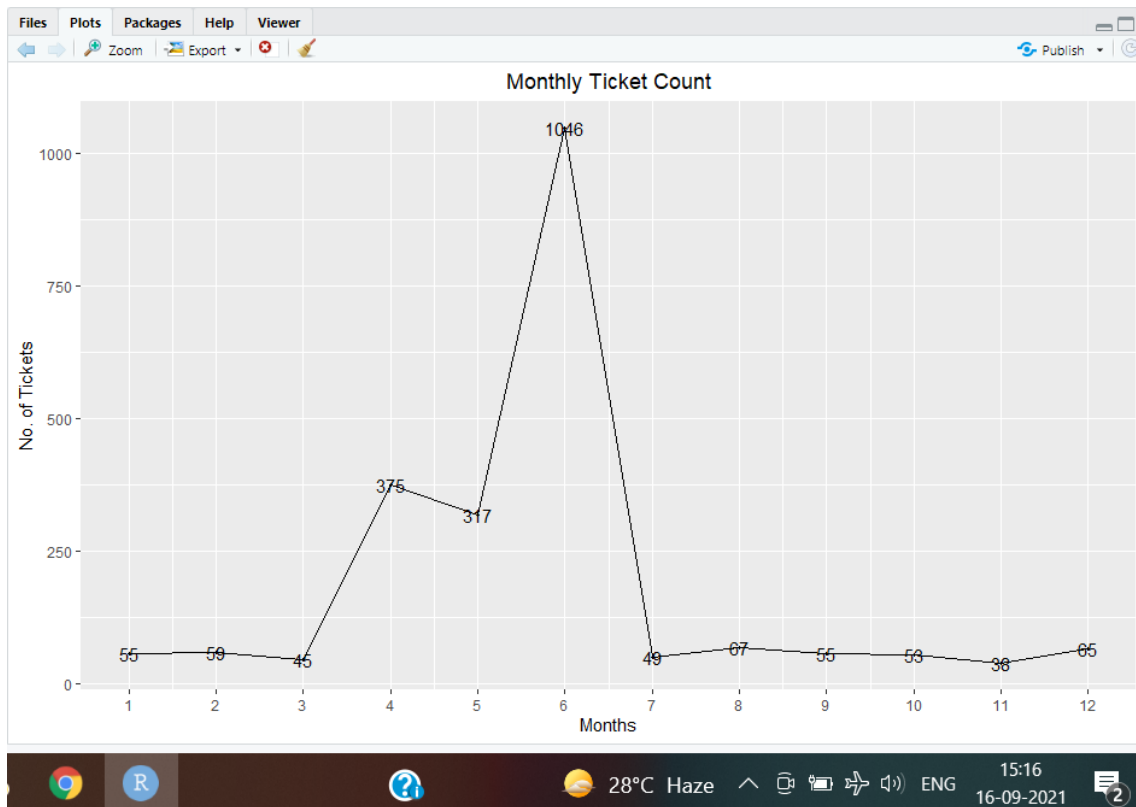

#INSIGHTS : From the trend chart we can conclude that the number of complaints are maximum in June month i.e.1046

#Table with the frequency of complaint types

```
# Complaint Type Processing
network_tickets <- contains(telecom_data$CustomerComplaint,match = 'network',ignore.case = T)
internet_tickets <- contains(telecom_data$CustomerComplaint,match = 'internet',ignore.case = T)
billing_tickets <- contains(telecom_data$CustomerComplaint,match = 'bill',ignore.case = T)
email_tickets <- contains(telecom_data$CustomerComplaint,match = 'email',ignore.case = T)
charges_ticket <- contains(telecom_data$CustomerComplaint,match = 'charge',ignore.case = T)
telecom_data$ComplaintType[internet_tickets]<- "Internet"
telecom_data$ComplaintType[network_tickets]<- "Network"
telecom_data$ComplaintType[billing_tickets]<- "Billing"
telecom_data$ComplaintType[email_tickets]<- "Email"
telecom_data$ComplaintType[charges_ticket]<- "Charges"
telecom_data$ComplaintType[-c(internet_tickets,network_tickets,
  billing_tickets,charges_ticket,email_tickets)]<- "Others"
table(telecom_data$ComplaintType)
```

#To make new categorical variable for Complaint Status

```
open_complaints<-(telecom_data$Status == 'Open' | telecom_data$Status == 'Pending')
closed_complaints<-(telecom_data$Status == 'Closed' | telecom_data$Status == 'Solved')
telecom_data$ComplaintStatus[open_complaints]<-'Open'
telecom_data$ComplaintStatus[closed_complaints]<-'Closed'
stack <- table(telecom_data$ComplaintStatus,telecom_data$State)
stack
telecom_data<- group_by(telecom_data,State,ComplaintStatus)
chart_data<- summarise(telecom_data,Count = n())
```

#Plotting on stacked bar chart
```
ggplot(as.data.frame(chart_data) ,mapping = aes(State,Count))+
  geom_col(aes(fill = ComplaintStatus),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,colour = "#0073C2FF"),
      plot.title = element_text(hjust =  0.5))+
```

labs(title = "Ticket Status Stacked Bar Chart ",
    x = "States",y = "No of Tickets",
    fill= "Status")
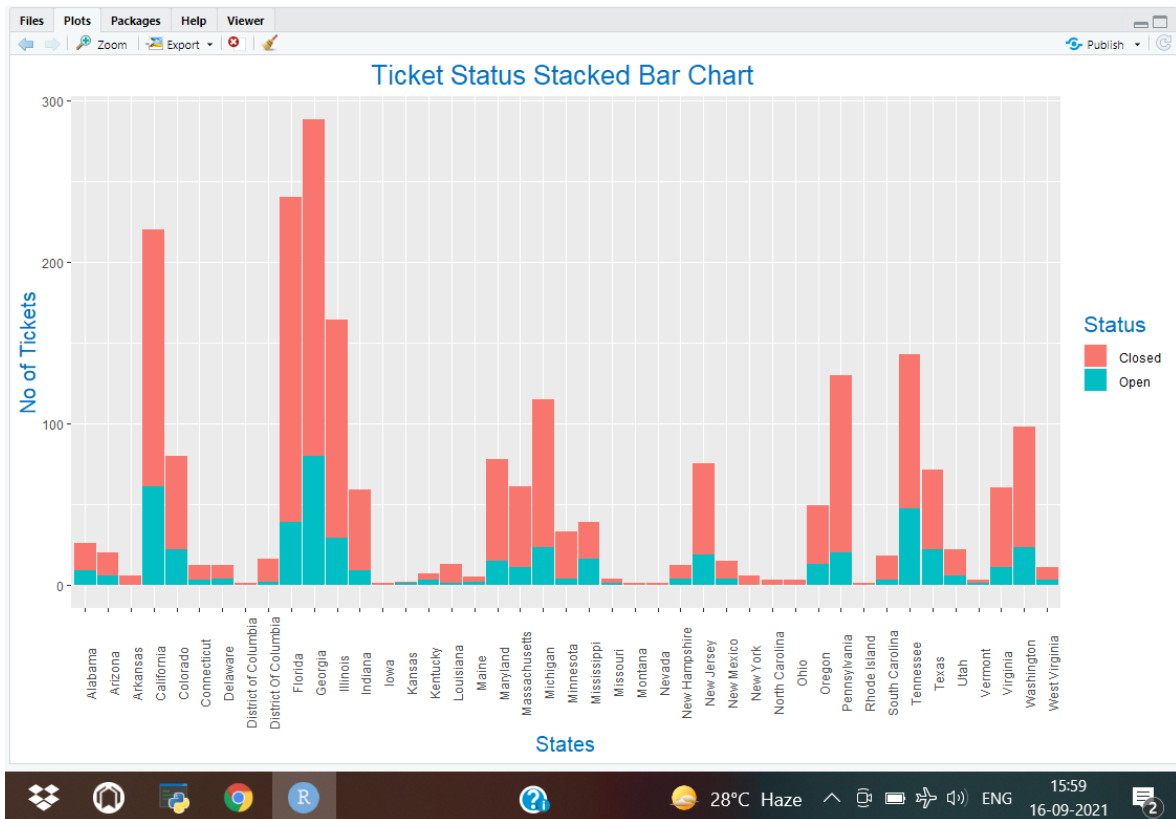
#INSIGHTS: From the Ticket Status Stacked Bar Chart, Georgia shows maximum number of complaints

#To find out the percentage of resolved complaints
total <- telecom_data %>% group_by(ComplaintStatus) %>% summarize(NumOfComplaints=n())
total
slices<-total$NumOfComplaints
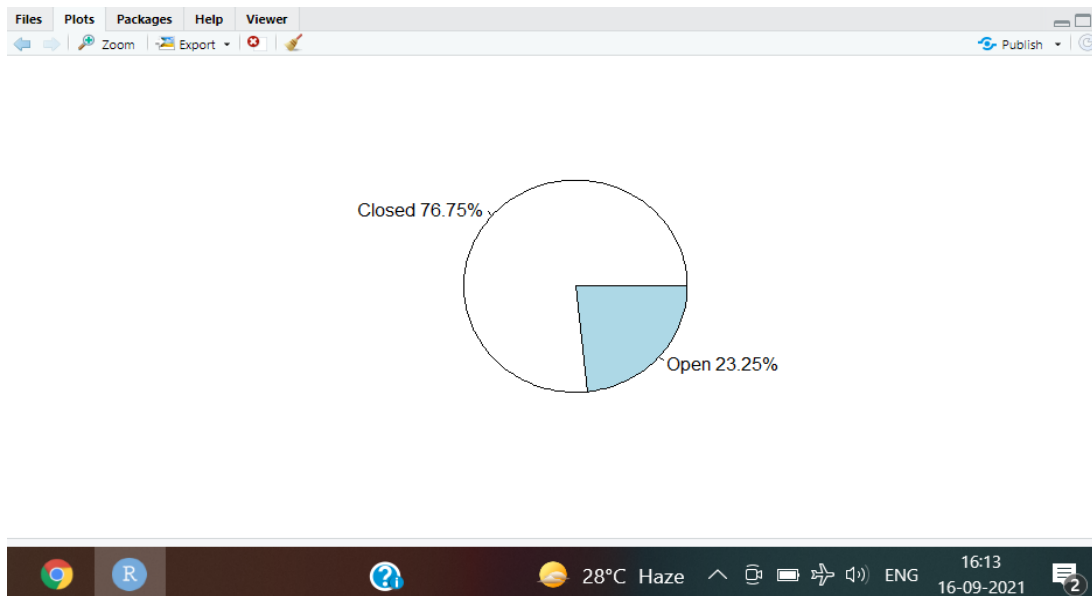*pct<-round((slices/sum(slices)*100),2)*
lbls<-paste(total$ComplaintStatus," ",pct,"%",sep="")

ComplaintStatus	NumOfComplaints
<chr>	<int>
1 Closed	1707
2 Open	517

#Plotting pie chart

pie(slices,labels=lbls)

int <- telecom_data %>% filter(ReceivedVia=='Internet',ComplaintStatus=='Closed') %>%
group_by(ReceivedVia,ComplaintStatus) %>% summarize(NumOfComplaints=n())
ccc <- telecom_data %>% filter(ReceivedVia=='Customer Care Call',ComplaintStatus=='Closed') %>%
group_by(ReceivedVia,ComplaintStatus) %>% summarize(NumOfComplaints=n())

#Percentage of resolved internet Complaints
*intpct<-round(int$NumOfComplaints/sum(total$NumOfComplaints)*100,2)*
intpct

#Percentage of resolved Customer Care Call Complaints
*cccpct<-round(ccc$NumOfComplaints/sum(total$NumOfComplaints)*100,2)*
cccpct

> intpct<-round(int$NumOfComplaints/sum(total$NumOfComplaints)*100,2)
> intpct
[1] 37.9
> cccpct<-round(ccc$NumOfComplaints/sum(total$NumOfComplaints)*100,2)
> cccpct
[1] 38.85