Project : Comcast Telecom in Data Science with R

Business Scenario :

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.

Analysis Task

- 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 the Internet and customer care calls.

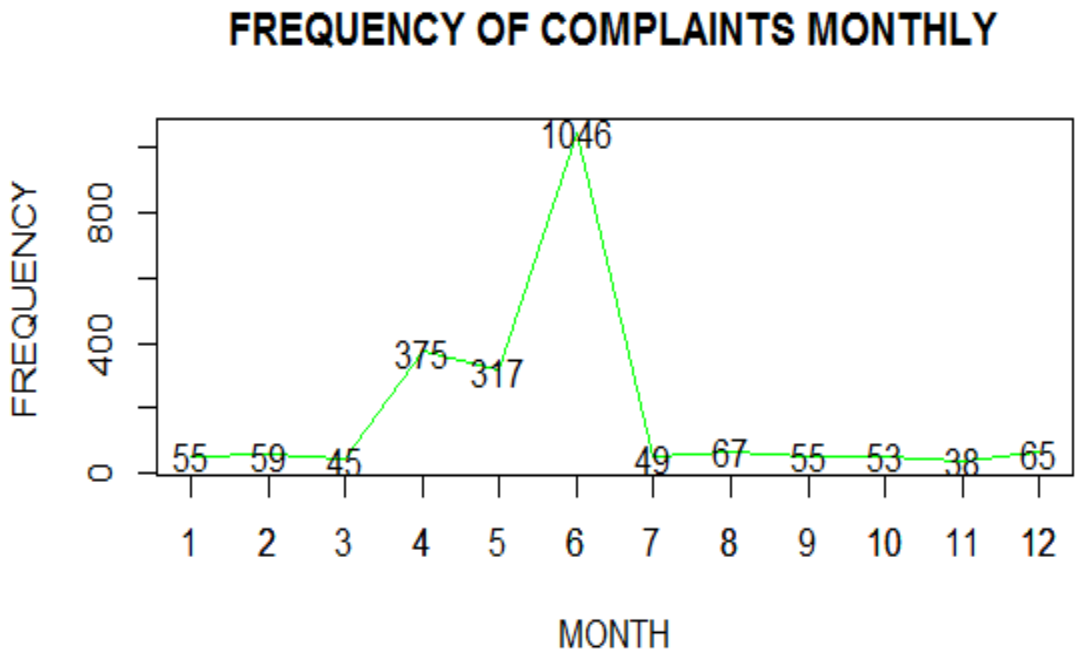
CODE :

```
library(dplyr)

library(lubridate)

library(ggplot2)


#IMPORTING DATA

data = read.csv(choose.files())

View(data)


#CHECK MISSING VALUE

anyNA(data)


#DATE CLEANING

data$Date = gsub("-","/",data$Date)


#CHANGING FORMAT TO DATE

data$Date = as.Date(data$Date,format = "%d/%m/%Y")


#FREQUENCY OF COMPLAINTS ON MONTHLY AND DAILY BASIS
```

```
month_freq = summarise(group_by(data,month = month(Date)),freq = n())

daily_freq = summarise(group_by(data,Date),freq =n())

month_freq = arrange(month_freq,month)
```

#MONTHLY PLOT

```
plot(month_freq$month, month_freq$freq ,xlab = "MONTH",ylab = "FREQUENCY",
    main = "FREQUENCY OF COMPLAINTS MONTHLY",type = "l",col="green")

axis(1,at=1:length(month_freq$month))

text(month_freq$month,month_freq$freq,labels = month_freq$freq)
```

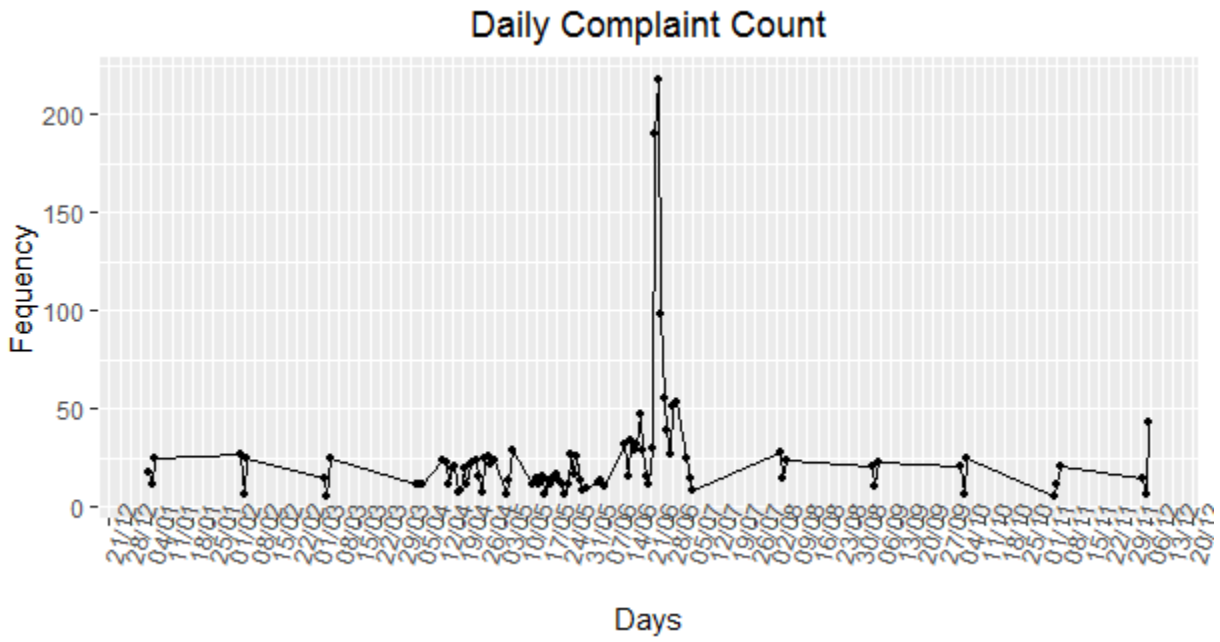


##From the data it can be observed that June had the most complains.

#DAILY PLOT

```
ggplot(data = daily_freq,aes(as.POSIXct(Date),freq))+
```

```
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))
```



##above chart is daily complaint received by Comcast. Which is displayed here as a break=1 week . And it can be observed that week of 6th June had the most complaints.

#CONVERTING DATA TO LOWER CASE

```
data$Customer.Complaint=tolower(data$Customer.Complaint)
```

```
#DIFFERENTIATING INTO COMPLAINT TYPES
service_complains = contains(data$Customer.Complaint,match = '',
                             ignore.case = T)
internet_complains = contains(data$Customer.Complaint,match = 'internet',
                              ignore.case = T)
charges_complains = contains(data$Customer.Complaint,match = 'charges',
                             ignore.case = T)
data$complain_type[service_complains] = "NETWORK"
data$complain_type[internet_complains] = "INTERNET"
data$complain_type[charges_complains] = "CHARGES"
data$complain_type[-c(service_complains,internet_complains,
                      charges_complains)] = "OTHERS"

#FREQUENCY OF COMPLAINT TYPES
table(data$complain_type)
```

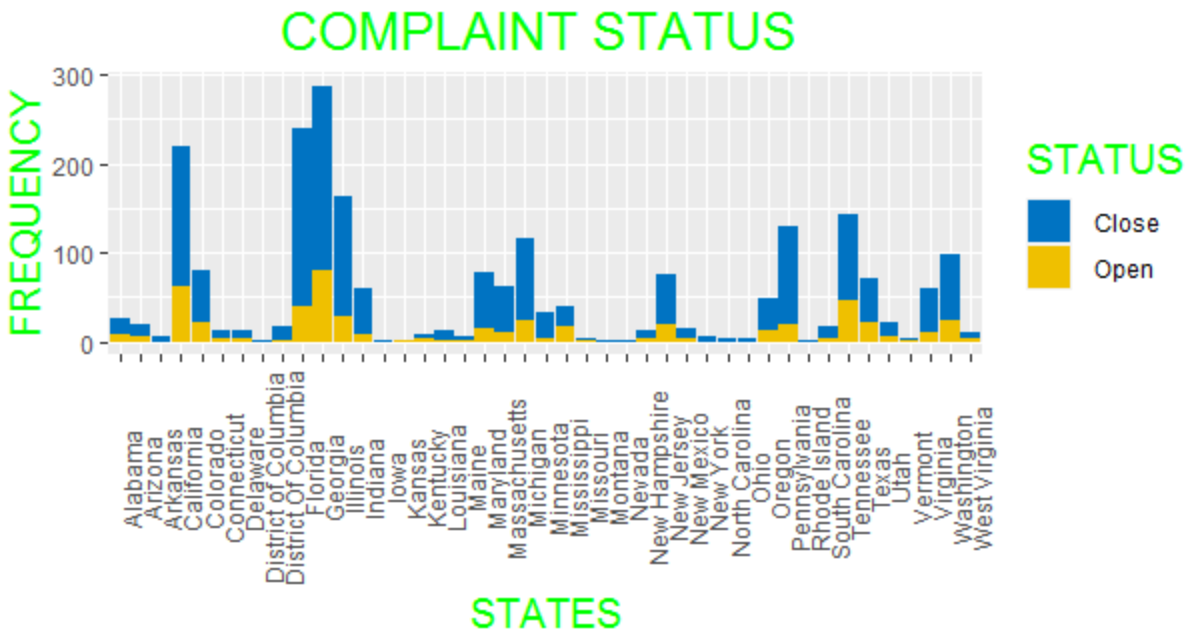
##OUTPUT

CHARGES	INTERNET	NETWORK	OTHERS
77	524	348	1275

##Internet related problems are more observed if others is excepted.

```
#CLASSIFYING DATA INTO OPEN AND CLOSE
open = (data$Status == "Open" | data$Status == "Pending")
close = (data$Status == "Closed" | data$Status == "Solved")
```

```
#NEW VARIABLE AS COMPLAINT STATUS
data$complaint_status[open] = "Open"
data$complaint_status[close] = "Close"

#STACKED BAR CHART
state_status=group_by(data,State,complaint_status)
state_count = summarise(state_status,Count =n())

ggplot(as.data.frame(state_count) ,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,colour = "green"),
        plot.title = element_text(hjust =  0.5))+
  labs(title = "COMPLAINT STATUS ",
       x = "STATES",y = "FREQUENCY",
       fill= "STATUS")+scale_fill_manual(values = c("#0073C2FF", "#EFC000FF"))
```



it can be observed that from Georgia Comcast had the most Complaints. And most of the unresolved complaints are also from there.

#COMPLAINTS RECEIVED VIA

```
rec_via = factor(data$Received.Via)

levels(rec_via)
```

[1] "Customer Care Call" "Internet"

After calling level function it can be observed that there are only two levels which are asked in tasks.

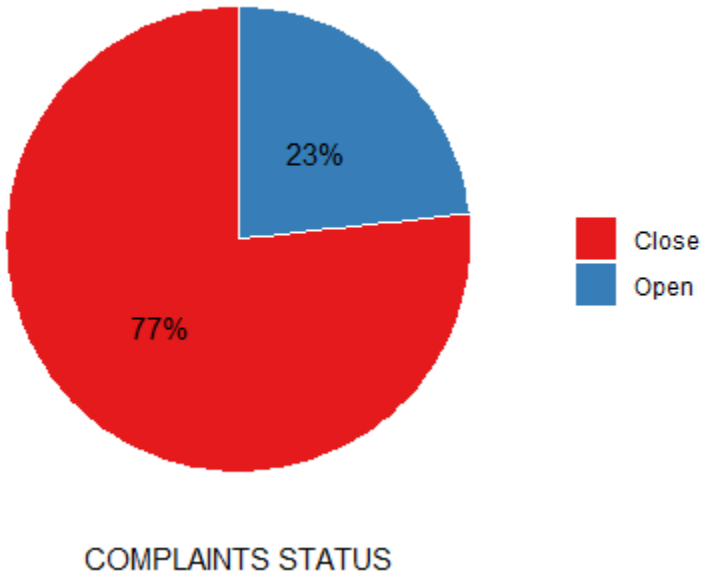
#COMPALINTS RESOLVED

```
status_data= group_by(data,complaint_status)

total_resolved<- summarise(status_data ,
```

```
                percentage =(n()/nrow(status_data)))
```

#CHART WISE STATUS

```
ggplot(total_resolved,
          aes(x= "",y =percentage,fill = complaint_status))+
 geom_bar(stat = "identity",width = 1,color = "white")+
 coord_polar("y",start = 0)+
 geom_text(aes(label = paste0(round(percentage*100),"%")),
          position = position_stack(vjust = 0.5))+
 labs(x = "",y = "COMPLAINTS STATUS",fill = "")+
 theme_classic()+theme(axis.line = element_blank(),
                axis.text = element_blank(),
                axis.ticks = element_blank())+
 scale_fill_brewer(palette = "Set1")
```



##Complaints resolved = 77%

ANALYSIS:

As per the above analysis we observe that in the June month Comcast received high amount of complaints in which most of the complaints are related to Internet service issue and the highest amount of complaints are received from the state Georgia. The highest unresolved complaints are related from the state Georgia and the total amount of resolved complaints are 77%.