COMCAST TELECOM CONSUMER COMPLAINTS ANALYSIS

Sushrut Kalyan Patnaik

Data Dictionary

- **Ticket #:** Ticket number assigned to each complaint
- **Customer Complaint:** Description of complaint
- **Date:** Date of complaint
- **Time:** Time of complaint
- **Received Via:** Mode of communication of the complaint
- **City:** Customer city
- **State:** Customer state
- **Zip code:** Customer zip
- **Status:** Status of complaint
- **Filing on behalf of someone**

ANALYSIS TASK

1. Import data into R environment.
2. Provide the trend chart for the number of complaints at monthly and daily granularity levels.
3. Provide a table with the frequency of complaint types.
4. Which complaint types are maximum i.e., around internet, network issues, or across any other domains.
   a. 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.
   b. Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:
      i. Which state has the maximum complaints
      ii. Which state has the highest percentage of unresolved complaints
5. Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

1. Import data into R environment:

```
# Use the required library/packages for your code.

library(dplyr)     # Data frame manipulation
library(ggplot2)   # Data visualization package
library(stringi)   # String/text/natural language processing
library(tidyverse) # Collection of required Data Science R packages
library(lubridate) # Deal with Date-Time data


# You can read you source file from anywhere from you PC. I prefer to use below.
comcast_data = read.csv (choose. files ()) #File: Comcast Telecom Complaints
data.csv

# View sample or complete data and Structure of the data set
head(comcast_data) # First 5 record from data set
tail(comcast_data) # Last 5 records from data set
View(comcast_data) # Complete data set like Tabular view
str(comcast_data) # Structure of the data set

# Check if there is any missing data point
any(is.na(comcast_data)) # False is NA/blank is available and vice versa

# Convert the DATE to one format
comcast_data$Date = dmy(comcast_data$Date)
```

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

```
# Filter for observations daily
comcast_daily = dplyr::summarise(group_by(comcast_data,Date), Count = n())
View(comcast_daily)

# Filter for observations monthly
comcast_monthly = dplyr::summarise(group_by(comcast_data, Month =
as.integer(month(Date))), Count = n())
comcast_monthly.Total = arrange(comcast_monthly, Month)
View(comcast_monthly)

# Renaming the months to factor
month.name = c("Jan", "Feb", "Mar",
        "Apr", "May", "Jun",
        "Jul", "Aug", "Sep",
        "Oct", "Nov", "Dec")

comcast_monthly$Month = month.name[comcast_monthly$Month]
comcast_monthly$Month <- as.character(comcast_monthly$Month)
```
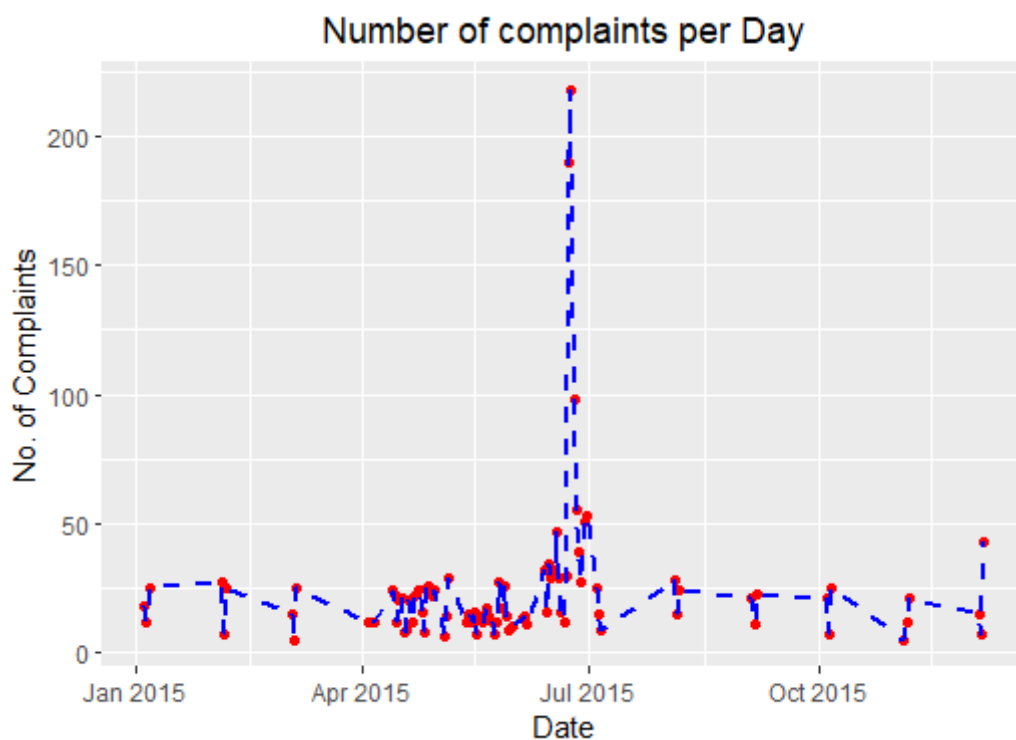
```
comcast_monthly$Month <- factor(comcast_monthly$Month , levels =
comcast_monthly$Month )
# Display the monthly complaints
View(comcast_monthly)

# Plotting the number of complaints per day

ggplot_daily = ggplot(comcast_daily, aes(x = comcast_daily$Date, y =
comcast_daily$Count)) +
  geom_point(col = "red", size = 1.5) +
  geom_line(col = 'blue', linetype = "dashed", size = .75) +
  xlab("Date")+
  ylab("No. of Complaints") +
  ggtitle("Number of complaints per Day") +
  theme(plot.title = element_text(hjust = 0.5))
```
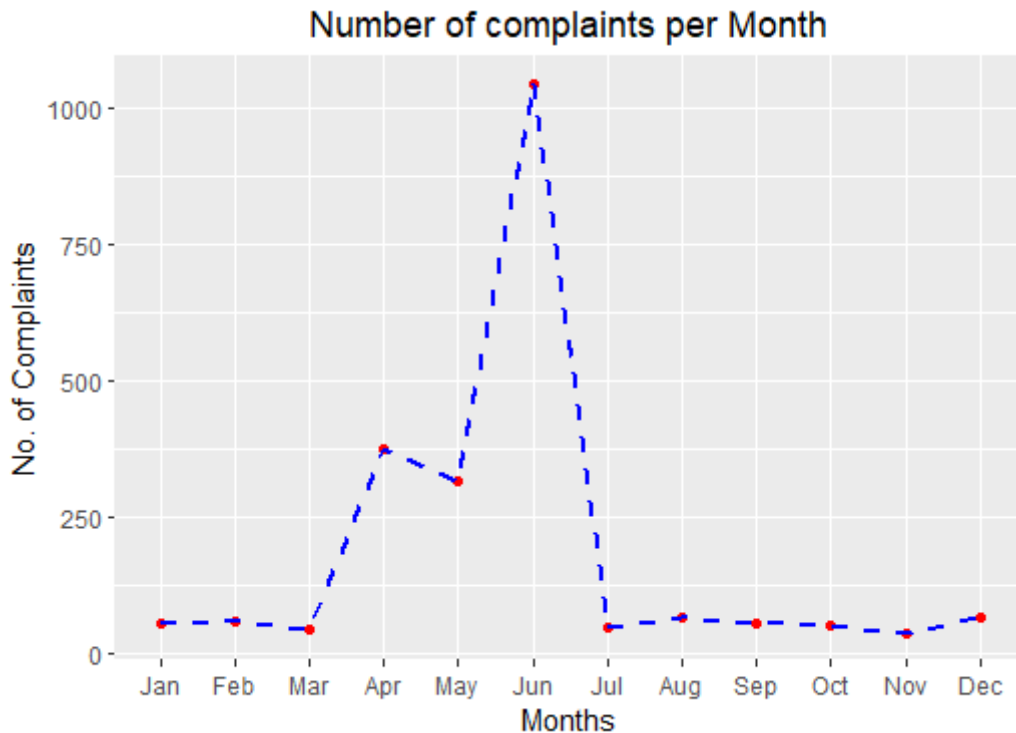


```
# Plotting the number of complaints per Month

ggplot_monthly = ggplot(comcast_monthly, aes(x = comcast_monthly$Month, y =
comcast_monthly$Count, group = 1)) +
  geom_point(col = "red", size = 1.5) +
  geom_line(col = 'blue', linetype = "dashed", size = .75) +
  xlab("Months")+
  ylab("No. of Complaints") +
  ggtitle("Number of complaints per Month") +
  theme(plot.title = element_text(hjust = 0.5))
```

Number of complaints per Month

3. Provide a table with the frequency of complaint types:

names(comcast_data)<-gsub(pattern = '\\.',replacement =
'',x=names(comcast_data))

names(comcast_data)

network_complaint = contains(comcast_data$CustomerComplaint, match =
'network', ignore.case = T)

internet_complaint = contains(comcast_data$CustomerComplaint, match =
'internet', ignore.case = T)

bill_complaint = contains(comcast_data$CustomerComplaint, match = 'bill',
ignore.case = T)

email_complaint = contains(comcast_data$CustomerComplaint, match = 'email',
ignore.case = T)

charge_complaint = contains(comcast_data$CustomerComplaint, match = 'charge',
ignore.case = T)

speed_complaint = contains(comcast_data$CustomerComplaint, match = 'speed',
ignore.case = T)

#datacap_complaint = contains(comcast_data$CustomerComplaint, match = 'data
cap', ignore.case = T)

```r
data_complaint = contains(comcast_data$CustomerComplaint, match = 'data',
ignore.case = T)


comcast_data$ComplaintType[network_complaint] = "Network"

comcast_data$ComplaintType[internet_complaint] = "Internet"

comcast_data$ComplaintType[bill_complaint] = "Bill"

comcast_data$ComplaintType[email_complaint] = "e-Mail"

comcast_data$ComplaintType[charge_complaint] = "Charge"

comcast_data$ComplaintType[speed_complaint] = "Speed"

#comcast_data$ComplaintType[datacap_complaint] = "Data Cap"

comcast_data$ComplaintType[data_complaint] = "Data"

comcast_data$ComplaintType[-
c(network_complaint,internet_complaint,bill_complaint,

          email_complaint,charge_complaint,speed_complaint,

          data_complaint)] = "Others"


complaint_frequency = table(comcast_data$ComplaintType)

complaint_frequency

View(complaint_frequency)


freq = c(348,122,219,15,355,2,973,190)

lab = c("Bill","Charge","Data","e-Mail","Internet", "Network", "Others", "Speed")


par(mfrow = c(1,2))


pie3D(complaint_frequency,
    labels = complaint_frequency,
    radius = 1,
    height = 0.1,
    border = "white",
```

```
        explode = 0.2,

        main = "Pie Chart of Frequency of complaint types")+

        theme(plot.title = element_text(hjust = 0.5))


pie3D(freq,

        labels = lab,

        radius = 1,

        height = 0.1,

        border = "white",

        explode = 0.2)
```
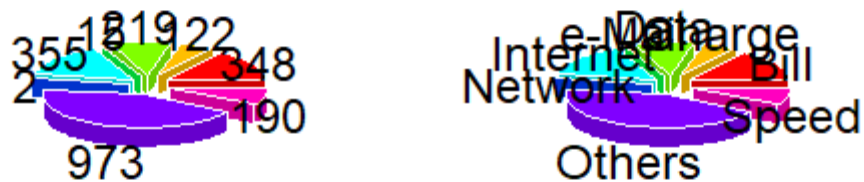
art of Frequency of complain

355 1819 22 348
2   190
    973

Internet e-Recharge Bill
Network   Data
            Speed
    Others

From the above table we can see that the Others and 2nd largest is Internet type complaints are maximum.

4. Which complaint types are maximum i.e., around internet, network issues, or across any other domains:

```
names(comcast_data)
# To achieve this goal we need to eliminate duplicate CustomerComplaint

ComplaintType_lower = comcast_data %>%
  mutate(tolower(CustomerComplaint))

ComplaintType = table(ComplaintType_lower$CustomerComplaint)
ComplaintType = data.frame(ComplaintType)
names(ComplaintType)
ComplaintType_filter = ComplaintType %>%
            rename(Complaint_Type = Var1, Frequency = Freq)
most_freq = ComplaintType_filter %>% arrange(desc(Frequency))

most_freq


wordcloud::wordcloud(words = most_freq$Complaint_Type, freq =
most_freq$Frequency, min.freq = 3,
            max.words =300, random.order = TRUE, colors=brewer.pal(8, "Dark2"),
rot.per=0.55)
```
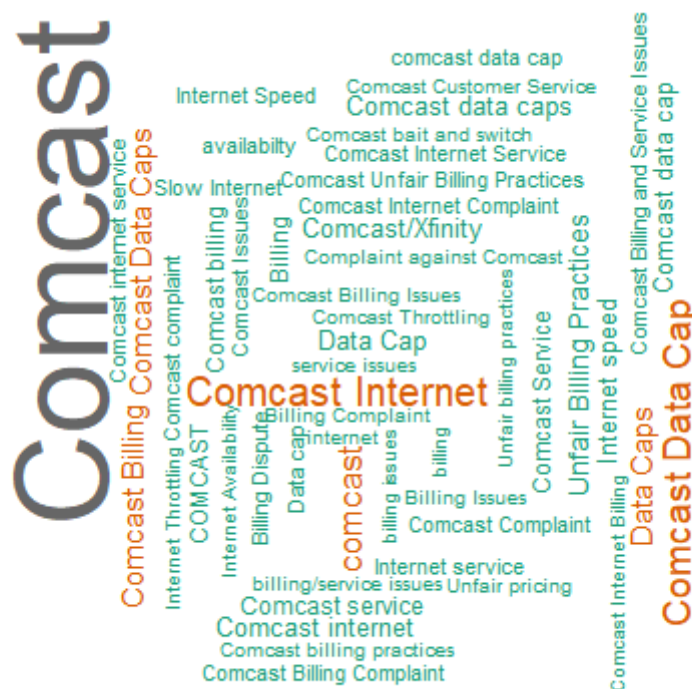

   a. Create a new categorical variable with value as Open and Closed. Open & Pending is to be categorized as Open and Closed & Solved is to be categorized as Closed:

```
names(comcast_data)
str(comcast_data)
unique(comcast_data$Status)

open <- ( comcast_data$Status   == "Open"| comcast_data$Status
=="Pending")

closed <- (comcast_data$Status   == "Closed"| comcast_data$Status
=="Solved")

comcast_data$ComplaintStatus[open]  <-"Open"
comcast_data$ComplaintStatus[closed]<- "Closed"

comcast_data_1 <- group_by(comcast_data,State,ComplaintStatus)
status_data<- dplyr::summarise(comcast_data_1,Count = n())

View(chart_data)
```

b. Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:
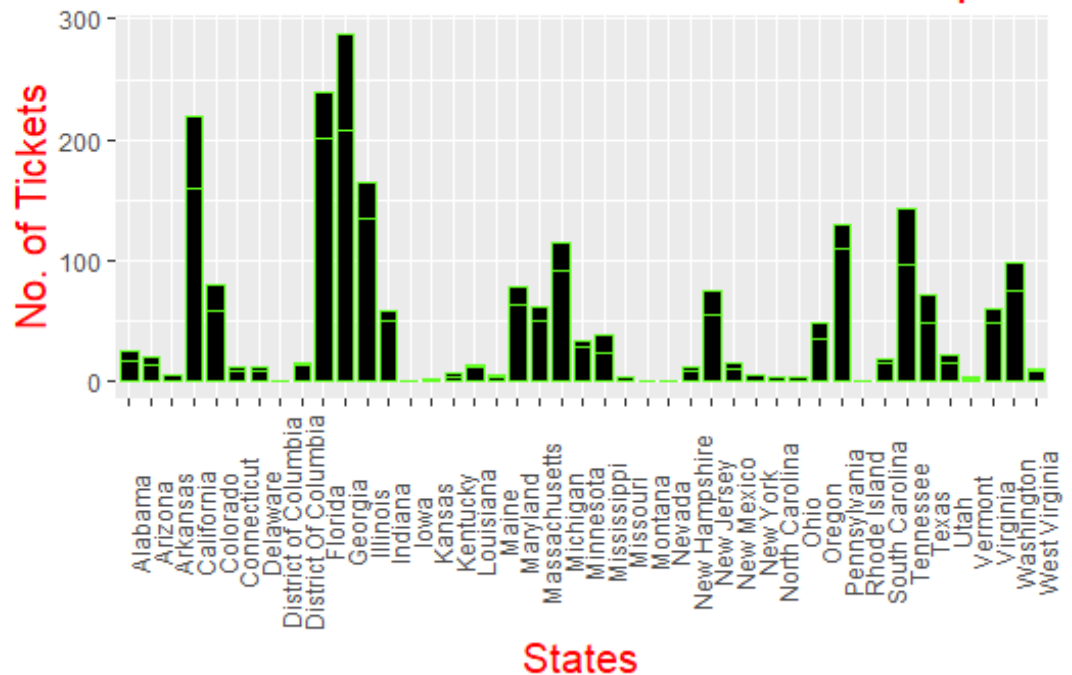
    i. Which state has the maximum complaints:

```
# plot the Ticket Status Distribution per States chart
p1 <- ggplot(status_data, aes(x =status_data$State, y =
status_data$Count)) +
  geom_bar(stat="identity", color= "#60ff20", fill= "black", width = .8)+
   #geom_text(aes(y = status_data$Count,
          #label=status_data$Count,
          #group=status_data$ComplaintStatus), vjust = 1, nudge_y =
40,
          #angle = 45, col = "blue", size = 4)+
   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 = "red"),
      plot.title = element_text(hjust =  0.5))+
   labs(title = "Total Number of Ticket Status Distribution per States",
      x = "States",y = "No. of Tickets", fill= "Status")
p1
```

ii. Which state has the highest percentage of unresolved complaints

```
State_by_Complian =  comcast_data %>% filter(ComplaintStatus ==
"Open") %>%
 group_by(State) %>% summarise(NumberOfComplian = n())

str(State_by_Complian)

ggplot(State_by_Complian, aes(x = State_by_Complian$State, y =
State_by_Complian$NumberOfComplian, group = 1))+
 geom_point(col = "red", size = 1.5) +
 geom_line(col = 'blue', linetype = "dashed", size = .75) +
 xlab("State")+
 ylab("No. of unresolved complaints") +
 ggtitle("Highest % of unresolved complaints by State") +
 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 = "red"),
     plot.title = element_text(hjust = 0.5))
```

Highest % of unresolved complaints by Stat
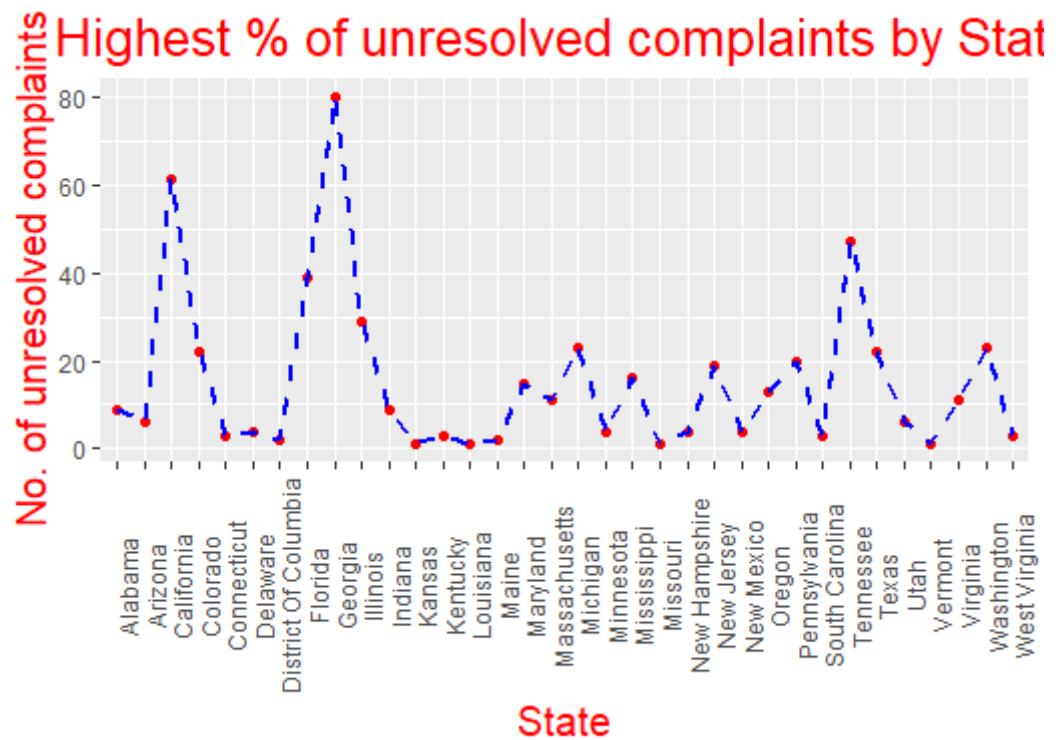
5. Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls:

cs_total = comcast_data %>%

       group_by(comcast_data$ComplaintStatus) %>%

       summarise(NumOfComplaintStatus = n())

cs_total

A = cs_total$NumOfComplaintStatus

Percentage = round((A/sum(A))*100,1)

lbl = paste(cs_total$`comcast_data$ComplaintStatus`," ", Percentage,"%", sep = " ")
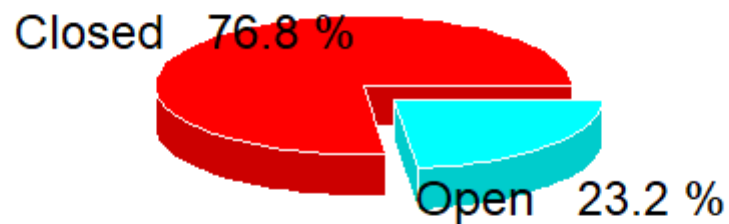

pie3D(A, labels = lbl,

   height = 0.1,

   border = "white",

   explode = 0.1,

   main = "Pie Chart of Complaints resolved v/s unresolved")+

```
theme(plot.title = element_text(hjust = 0.5))
```

Pie Chart of Complaints resolved v/s unresolved

Closed   76.8 %

Open   23.2 %

\# Pie chart we can clearly see that there is total **76.8%** Complaints **resolved**.

```
unique(comcast_data$ReceivedVia)

unique(comcast_data$ComplaintStatus)


Internet = comcast_data %>%

  filter(comcast_data$ReceivedVia == 'Internet', comcast_data$ComplaintStatus ==
'Closed') %>%

  summarise(NumOfComplaints = n())


Internet_Percent =
round((Internet$NumOfComplaints/sum(cs_total$NumOfComplaintStatus)*100),1)


CCC = comcast_data %>%
```

```
  filter(comcast_data$ReceivedVia == 'Customer Care Call',
comcast_data$ComplaintStatus == 'Closed') %>%

  summarise(NumOfComplaints = n())


CCC_Percent =
round((CCC$NumOfComplaints/sum(cs_total$NumOfComplaintStatus)*100),1)
```

From **76.75%** resolved Complaints, **37.9%** complaints are **Internet** type while **38.8%** are **Customer Care Call** type.