

R-Programming Project

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.

Tasks to be 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 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.

```
#Importing necessary packages  
library(dplyr)
```

```
##  
## Attaching package: 'dplyr'  
  
## The following objects are masked from 'package:stats':  
##  
## filter, lag  
  
## The following objects are masked from 'package:base': ##  
## intersect, setdiff, setequal, union
```

```
library(ggplot2)
library(lubridate)
```

```
##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base' : ##
##   date, intersect, setdiff, union
```

#Importing Comcast Dataset

```
comcast_data<- read.csv("Comcast Telecom Complaints data.csv",header = TRUE)
```

#Manipulating Field Names

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

```
## [1] "Ticket"           "CustomerComplaint"
## [3] "Date"             "Time"
## [5] "ReceivedVia"      "City"
## [7] "State"            "Zipcode"
## [9] "Status"           "FilingonBehalfofSomeone"
```

```
View(comcast_data)
```

From the dataset, we can see that the format of Date column is not same throughout, so we need to make it same for analysis.

#Processing Date

```
comcast_data$Date<- dmy(comcast_data$Date)
View(comcast_data)
```

- Now we need to get the complaints on a daily level basis and plot a trend chart for it.

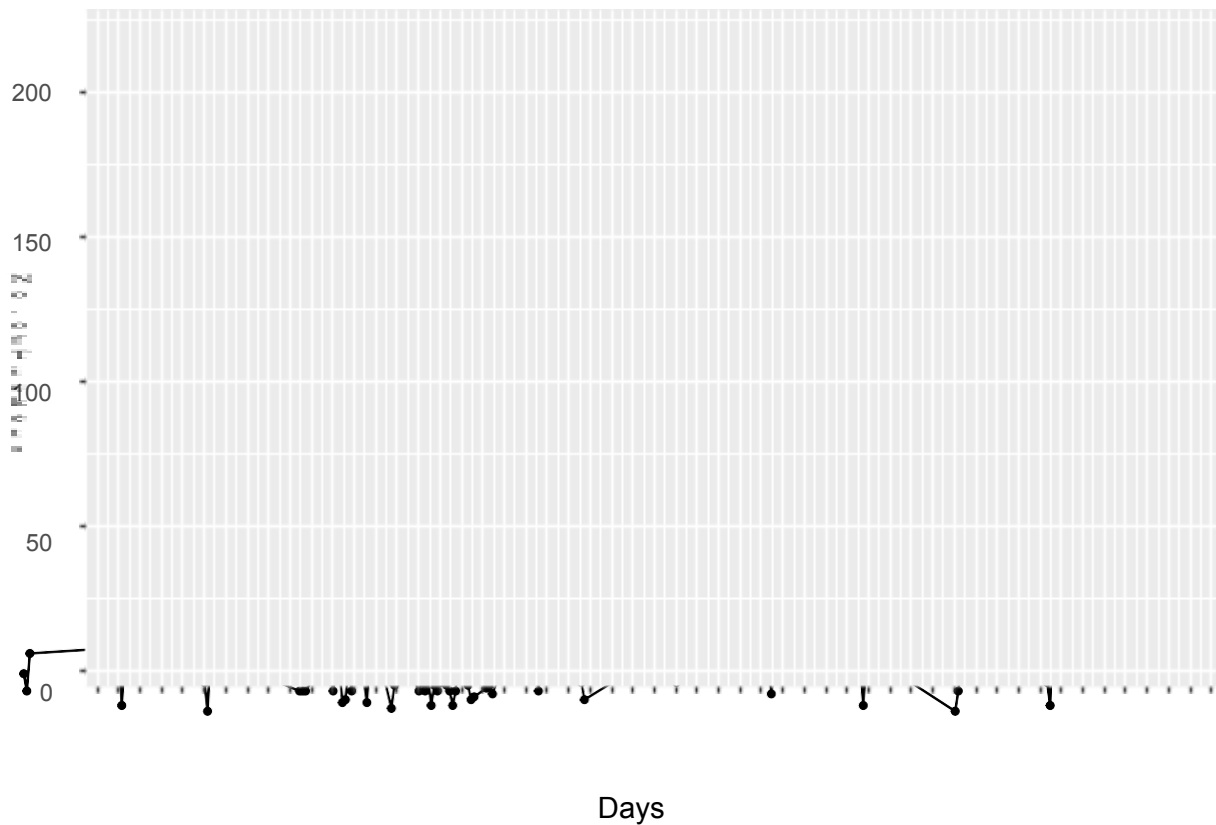
```
ans<-comcast_data %>% group_by(Date) %>% summarize(NumOfComplaints=n())
```

```
## 'summarise()' 'ungrouping output (override with '.groups' argument)
```

#Plotting 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))
```

Daily Ticket Count



- Now we need to get the complaints on a monthly level basis and plot a trend chart for it.

#Making month field

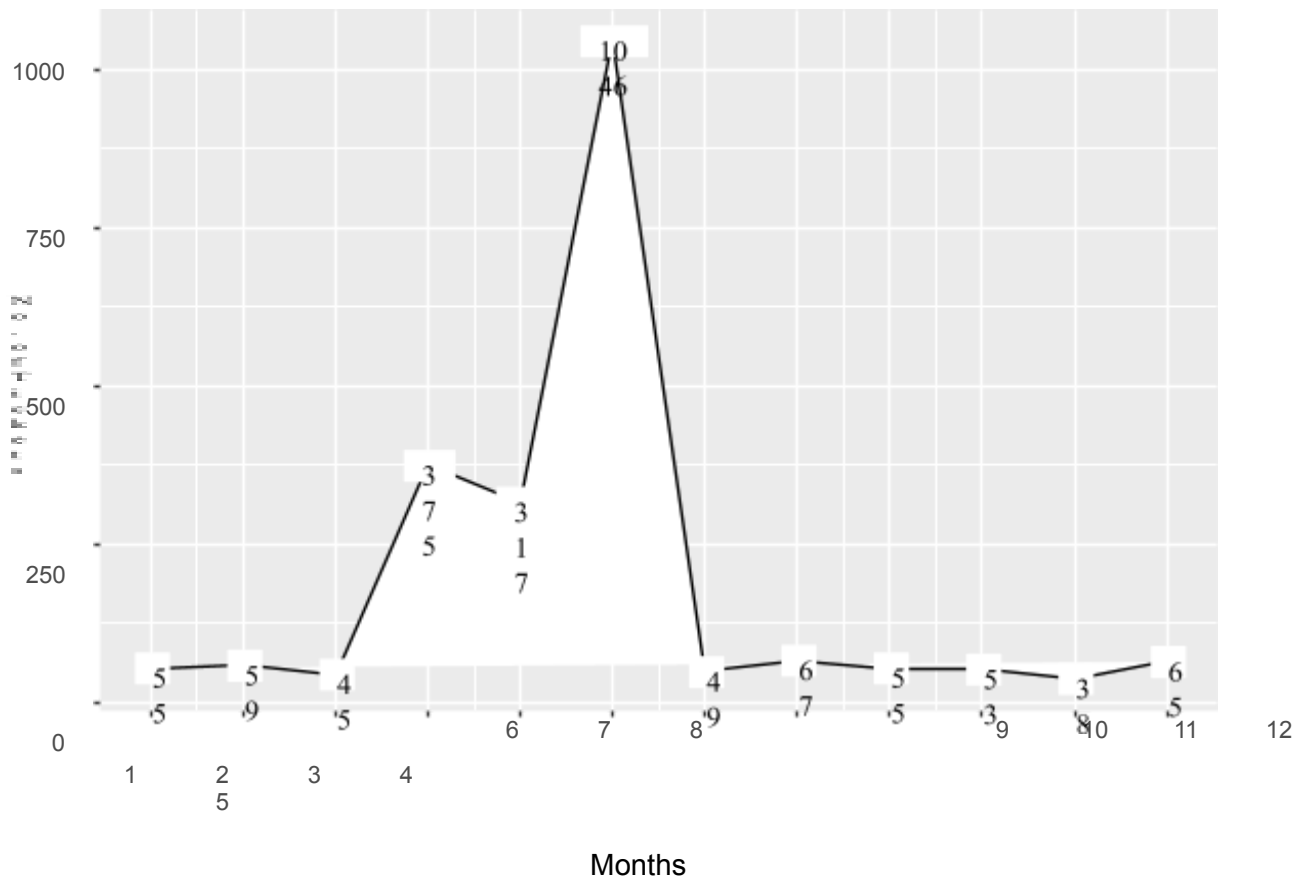
```
comcast_data$Month<-months(comcast_data$Date)
ans1<-comcast_data %>% group_by(Month =as.integer(month(Date))) %>% summarize(NumOfComplaints=n()) %>%
```

```
## 'summarise()' 'ungrouping output (override with '.groups' argument)
```

#Plotting for monthly granularity level

```
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))
```

Monthly Ticket Count



INSIGHTS:- From the above trend chart, we can clearly see that complaints for the month of June are maximum i.e.1046.

- Now we need to make a frequency table basis the complaint types.

```
# Complaint Type Processing
network_tickets<- contains(comcast_data$CustomerComplaint,match = ' network' ,ignore.case =
internet_tickets<- contains(comcast_data$CustomerComplaint,match = ' internet' ,ignore.case
billing_tickets<- contains(comcast_data$CustomerComplaint,match = ' bill' ,ignore.case = T)
email_tickets<- contains(comcast_data$CustomerComplaint,match = ' email' ,ignore.case = T)
charges_ticket<- contains(comcast_data$CustomerComplaint,match = ' charge' ,ignore.case = T)

comcast_data$ComplaintType[internet_tickets]<- "Internet"
comcast_data$ComplaintType[network_tickets]<- "Network"
comcast_data$ComplaintType[billing_tickets]<- "Billing"
comcast_data$ComplaintType[email_tickets]<- "Email"
comcast_data$ComplaintType[charges_ticket]<- "Charges"

comcast_data$ComplaintType[-c(internet_tickets,network_tickets,
billing_tickets,charges_ticket,email_tickets)]<- "Others"
table(comcast_data$ComplaintType)
```

```
##
## Billing ChargesEmail Internet Network      Others ##
363      139          16 472          1      1233
```

INSIGHTS:- From the above table we can see that the Internet type complaints are maximum.

- Now we need to make a new categorical variable for Complaint Status.

```
open_complaints<-(comcast_data$Status == ' Open' | comcast_data$Status == ' Pending' )
closed_complaints<-(comcast_data$Status == ' Closed' | comcast_data$Status == ' Solved' )
comcast_data$ComplaintStatus[open_complaints]<-' Open'
comcast_data$ComplaintStatus[closed_complaints]<-' Closed'
```

- Now we need to plot state wise status of complaints in a stacked bar chart.

```
stack<-table(comcast_data$ComplaintStatus,comcast_data$State) stack
```

```
##
```

District of Columbia District Of Columbia Florida Georgia Illinois

Kansas Kentucky Louisiana Maine Maryland Massa

South Carolina Tennessee Texas U

```
## Closed 75 8
## Open 23 3
```

```
comcast_data<- group_by(comcast_data, State, ComplaintStatus)
chart_data<- summarise(comcast_data, Count = n())
```

```
## 'summarise()' 'regrouping output by ' State' (override with '.groups' argument)
```

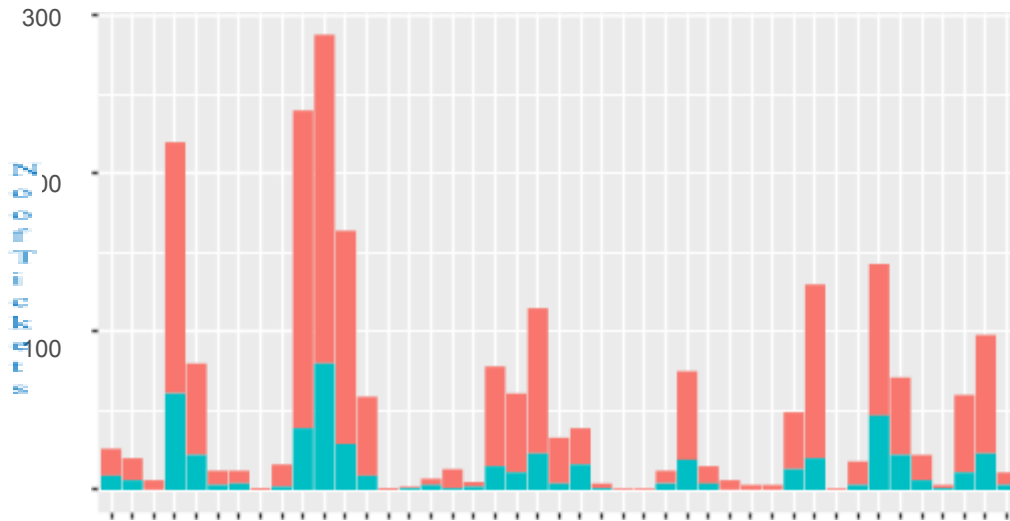
```
#Plotting on stacked bar chart
ggplot(as.data.frame(chart_data), mapping = aes(State, Count))+
  geom_col(aes(fill = ComplaintStatus), width = 0.95)+ theme(axis.text.x =
  element_text(angle = 90),
  axis.title.y = element_text(size = 15), axis.title.x
  = element_text(size = 15),
```

```

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")

```

Ticket Status Stacked Bar Chart



Status

```

## Alabama A
## Closed 17
## Open 9
##
##
## Closed
## Open
##
## Indiana I
## Closed 50
## Open 9
##
## Michigan M
## Closed 92
## Open 23
##
## New Jersey N
## Closed
## Open
##
## Rhode Is R
## Closed
## Open
##
## Washington W

```

C
I
o
s
e
d
O

p
e

n

0

Alabama Arizona Arkansas California Colorado Connecticut Delaware
District of Columbia District Of Columbia Florida
Georgia Illinois Indiana Iowa Kansas Kentucky Louisiana Maine Maryland
Massachusetts Michigan Minnesota Mississippi Missouri Montana Nevada
New Hampshire New Jersey New Mexico New York
North Carolina Ohio
Oregon Pennsylvania Rhode Island South Carolina Tennessee Texas
Utah Vermont Virginia Washington West Virginia

States

INSIGHTS:- From the above chart, we can clearly see that Georgia has maximum complaints.

- Now we need to see which state has maximum unresolved complaints

```
comcast_data %>% filter(ComplaintStatus==' Open' ) %>% group_by(State) %>% summarize(NumOfComplaints=n())  
## 'summarise()' ungrouping output (override with '.groups' argument)  
## # A tibble: 34 x 2  
##   State      NumOfComplaints  
##   <chr>          <int>
```



```
## 8 Colorado                22
## 9 Texas                   22
## 10 Pennsylvania           20
## # ... with 24 more rows
```

INSIGHTS:- From the table generated above we can see that Georgia has maximum unresolved complaints i.e. 80.

- Now we want to see the percentage of resolved complaints.

```
tot<-comcast_data %>% group_by(ComplaintStatus) %>% summarize(NumOfComplaints=n())
```

```
## 'summarise()' 'ungrouping output (override with '.groups' argument)
```

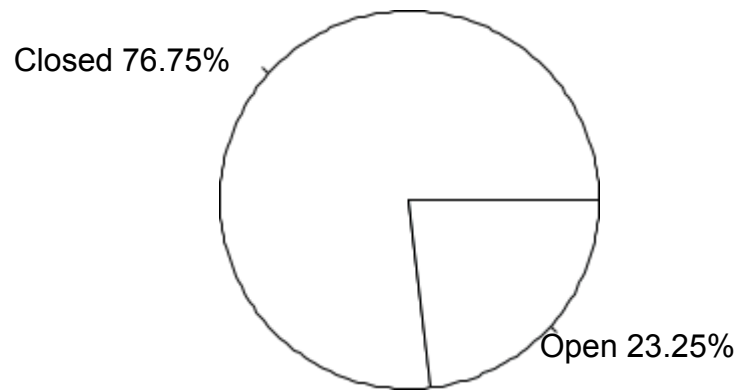
```
tot
```

```
## # A tibble: 2 x 2
##   ComplaintStatus NumOfComplaints
##   <chr>           <int>
## 1 Closed         1707
## 2 Open           517
```

```
slices<-tot$NumOfComplaints
pct<-round((slices/sum(slices)*100),2)
lbls<-paste(tot$ComplaintStatus, " ", pct, "%", sep=" ")
```

```
#Plotting pie chart
```

```
pie(slices, labels=lbls)
```



INSIGHTS:- From the above pie chart we can clearly see that there are total 76.75% Complaints resolved.

```
int<-comcast_data %>% filter(ReceivedVia==' Internet' ,ComplaintStatus==' Closed' ) %>% group_by(ReceivedVia
```

```
## 'summarise()' 'regrouping output by ' ReceivedVia' (override with '.groups' argument)
```

```
ccc<-comcast_data %>% filter(ReceivedVia==' Customer Care Call' ,ComplaintStatus==' Closed' ) %>% group_by
```

```
## 'summarise()' 'regrouping output by ' ReceivedVia' (override with '.groups' argument)
```

```
#Percentage of resolved internet Complaints
```

```
intpct<-round(int$NumOfComplaints/sum(tot$NumOfComplaints)*100,2) intpct
```

```
## [1] 37.9
```

```
#Percentage of resolved Customer Care Call Complaints
```

```
cccpct<-round(ccc$NumOfComplaints/sum(tot$NumOfComplaints)*100,2) cccpct
```

```
## [1] 38.85
```

INSIGHTS:- From the above output we can see that of the 76.75% resolved Complaints, 37.9% complaints are Internet type while 38.85% are Customer Care Call type.