```r
library(stringi)
library(lubridate)
library(dplyr)
library(ggplot2)

comcast_data<- read.csv("E:/R Project/Comcast Telecom Complaints
data.csv",header = TRUE)

#Manipulating column names
names(comcast_data)<- stri_replace_all(regex =  "\\.",replacement = "",str
=names(comcast_data))
head(comcast_data)

na_vector <- is.na(comcast_data)
length(na_vector[na_vector==T])

#This shows that there is no missing values in dataset,so now data is tidy
and available to process further or do EDA based on requriment. •
Processing Date.
comcast_data$Date<- dmy(comcast_data$Date)


#Extracting Monthly and Daily Ticket Count
monthly_count<- summarise(group_by(comcast_data,Month
=as.integer(month(Date))),Count = n())
daily_count<- summarise(group_by(comcast_data,Date),Count =n())
monthly_count<-arrange(monthly_count,Month)


#Comparing Monthly and Daily Complaints
ggplot(data = monthly_count,aes(Month,Count,label = Count))+
  geom_line()+
  geom_point(size = 0.8)+
  geom_text()+
  scale_x_continuous(breaks = monthly_count$Month)+
  labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")+
  theme(plot.title = element_text(hjust = 0.5))


#As we can see that in the month of April,May the tickets are increses but
in the month of June it increases drastically, so there might be some
reason for which they received high amount of tickets.
ggplot(data = daily_count,aes(as.POSIXct(Date),Count))+
  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))


#And with the help of above daily chart of tickets we can observe that in
second half of June month we recived more tickets with respect to normal
days
# Complaint Type Processing
network_tickets<- contains(comcast_data$CustomerComplaint,match =
'network',ignore.case = T)
```

```r
internet_tickets<- contains(comcast_data$CustomerComplaint,match =
'internet',ignore.case = T)
billing_tickets<- contains(comcast_data$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)
```

#As we can observe that there are some complaints from different-different
categories and we combined them into one, i.e.- others. So most of the
complaints are related to Internet issue. • Creating new Variable
ComplaintStatus with values Open and Closed.

```r
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"
```

#Creating Stacked barchart for complaints based on State and Status

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

#Now it`s clearly shown that the highest number of complaints recorded from
the state Georgia and the second highest number of complaints recorded from
the state Florida. • Finding State which has Highest number of Unresolved
Tickets.

```r
chart_data%>%
  filter(ComplaintStatus == "Open")->
  open_complaints
open_complaints[open_complaints$Count == max(open_complaints$Count),c(1,3)]
```

#As we can observe that State Georgia has maximum number of unresolved
tickets and these ticket count is 80. • Calculating Resolution Percentage
based on Total and Catagory

```
resolved_data <- group_by(comcast_data,ComplaintStatus)
total_resloved<- summarise(resolved_data ,percentage =(n()/
nrow(resolved_data)))
resolved_data <- group_by(comcast_data,ReceivedVia,ComplaintStatus)
Category_resloved<- summarise(resolved_data ,percentage =(n()/
nrow(resolved_data)))

# Now we want to see the percentage of resolved complaints.
tot<-comcast_data %>% group_by(ComplaintStatus) %>%
summarize(NumOfComplaints=n())
tot
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,ComplaintStatus) %>% summarize(NumOfComplaints=n())
ccc<-comcast_data %>% filter(ReceivedVia=='Customer Care
Call',ComplaintStatus=='Closed') %>% group_by(ReceivedVia,ComplaintStatus)
%>% summarize(NumOfComplaints=n())

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

# 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.
```