

```
#Comcast Telecom Consumer Complaints.
```

```
getwd()
```

```
# Import data into R environment.
```

```
comcast_data=read.csv("Comcast Telecom Complaints data.csv")
```

```
View(comcast_data)
```

```
# view top 6 rows
```

```
head(comcast_data)
```

```
summary(comcast_data)
```

```
str(comcast_data)
```

```
#import the libraries
```

```
library(stringi)
```

```
library(lubridate)
```

```
library(dplyr)
```

```
library(ggplot2)
```

```
head(comcast_data)
```

```
names(comcast_data)<- stri_replace_all(regex = "\\.",replacement =  
"",str =names(comcast_data))
```

```
names
```

```
head(comcast_data)
```

```
na_vector <- is.na(comcast_data)
```

```
na_vector
```

```
length(na_vector[na_vector==T])
```

```
#Provide the trend chart for the number of complaints at monthly and  
daily granularity levels.
```

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

```
head(comcast_data)
```

```
monthly_count=arrange(summarise(group_by(comcast_data,month=as.integer(m
```

```
daily_count=summarise(group_by(comcast_data,Date),Count=n())
```

```
monthly_count
```

```
daily_count
```

```
ggplot(data = monthly_count,aes(month,Count,label =
```

```
Count))+geom_line()+geom_text()+scale_x_continuous(breaks =
```

```
monthly_count$month)+labs(title = "Monthly Ticket Count",x= "Months",y  
="No. of Tickets")
```

```
ggplot(data =
```

```
daily_count,aes(as.POSIXct(Date),Count))+geom_line()+theme(axis.text.x
```

```
= element_text(angle = 75))+scale_x_datetime(breaks = "1
weeks",date_labels = "%d/%m")+labs(title = "Daily Ticket Count",x=
"Days",y ="No. of Tickets")
```

#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.

```
network_tickets=contains(comcast_data$CustomerComplaint,match='network',
= T)
```

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

```
bill_tickets=contains(comcast_data$CustomerComplaint,match='bill',ignore
= T)
```

```
email_tickets=contains(comcast_data$CustomerComplaint,match="email",igno
= T)
```

```
charge_tickets=contains(comcast_data$CustomerComplaint,match='charge',ig
= T)
```

```
comcast_data$ComplaintType[internet_tickets]='Internet'
```

```
comcast_data$ComplaintType[bill_tickets]='Billing'
```

```
comcast_data$ComplaintType[email_tickets]='Email'
```

```
comcast_data$ComplaintType[charge_tickets]='Charges'
```

```
comcast_data$ComplaintType[network_tickets]='Network'
```

```
comcast_data$ComplaintType[-
c(network_tickets,internet_tickets,bill_tickets,email_tickets,charge_tic
```

```
View(comcast_data)
```

```
table(comcast_data$ComplaintType)
```

#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.

```
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'
```

```
#Provide state wise status of complaints in a stacked bar chart. Use  
the categorized variable from Q3.
```

```
chart_data=summarize(group_by(comcast_data,State,ComplaintStatus),Count=
```

```
chart_data
```

```
chart_data=as.data.frame(chart_data)
```

```
chart_data
```

```
ggplot(chart_data ,mapping = aes(State,Count))+geom_col(aes(fill =  
ComplaintStatus),width = 0.95)+theme(axis.text.x = element_text(angle =  
90))+labs(title = "Ticket Status Stacked Bar Chart ",x = "States",y =  
"No of Tickets",fill= "Status")
```

```
#Which state has the maximum complaints
```

```
max(chart_data$Count)
```

```
arrange(select(chart_data,State,Count),desc(Count))
```

```
#Which state has the highest percentage of unresolved complaints
```

```
arrange(filter(chart_data,ComplaintStatus=="Open"),desc(Count))
```

```
#Provide the percentage of complaints resolved till date,
```

```
#which were received through theInternet and customer care calls.
```

```
resolved=summarise(filter(comcast_data,ComplaintStatus=='Closed'),count=
```

```
resolved
```

```
resolved_internet=summarise(filter(comcast_data,ComplaintStatus=='Closed
```

```
resolved_internet
```

```
resolved_CustomerCare=summarise(filter(comcast_data,ComplaintStatus=='Cl  
Care Call'),count=n())
```

```
resolved_CustomerCare
```

```
percentage_internet=(resolved_internet/resolved)*100
```

```
percentage_internet
```

```
percentage_CustomerCare=(resolved_CustomerCare/resolved)*100
```

```
percentage_CustomerCare
```

```

table_df=table(comcast_data$ReceivedVia,comcast_data$ComplaintStatus)
table_df

bar=ggplot(comcast_data,aes(ComplaintStatus,fill=ReceivedVia))+geom_bar(
bar

resolved_df=select(filter(comcast_data,ComplaintStatus=='Closed'),ComplaintStatus,ReceivedVia)

pie<- ggplot(resolved_df, aes(x="", y=
ComplaintStatus,fill=ReceivedVia)) +geom_bar(width = 1, stat =
"identity") +coord_polar("y")
pie

```