

```

> #Importing necessary packages
> library(dplyr)
> library(ggplot2)
> library(lubridate)
>
> #Importing Comcast Dataset
> comcast_data<- read.csv("Comcast Telecom Complaints data.csv",header = TRUE)
> View(comcast_data)
>
>
> #Manipulating Field Names
> names(comcast_data)<-gsub(pattern = '\\\\.',replacement = "",x=names(comcast_data))
> names(comcast_data)
[1] "Ticket"           "CustomerComplaint"      "Date"
[4] "Time"             "ReceivedVia"            "City"
[7] "State"            "Zipcode"                "Status"
[10] "FilingonBehalfofSomeone"
> View(comcast_data)
>
>
> #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())
> #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))
> #Making month field
> comcast_data$Month<-months(comcast_data$Date)
> ans1<-comcast_data %>% group_by(Month =as.integer(month(Date))) %>%
summarize(NumOfComplaints=n()) %>% arrange(desc(NumOfComplaints))
>
> #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))
> #INSIGHTS:- From the above trend chart, we can clearly see that complaints for

```

the month of June is 1046 which is maximum.

```
>
>
> #Now we need to make a frequency table basis the complaint types.
>
>
> # Complaint Type Processing
> network_tickets<- contains(comcast_data$CustomerComplaint,match =
'network',ignore.case = T)
> internet_tickets<- contains(comcast_data$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)

Billing  Charges    Email 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
```

Alabama Arizona Arkansas California Colorado Connecticut Delaware District
of Columbia

Closed	17	14	6	159	58	9	8
1							
Open	9	6	0	61	22	3	4
0							

		District Of Columbia	Florida	Georgia	Illinois	Indiana	Iowa	Kansas	Kentucky
Louisiana	Maine								
Closed		14	201	208	135	50	1	1	4
12	3								
Open		2	39	80	29	9	0	1	3
1	2								

	Maryland	Massachusetts	Michigan	Minnesota	Mississippi	Missouri	Montana
Nevada	New Hampshire						
Closed	63	50	92	29	23	3	1
1	8						
Open	15	11	23	4	16	1	0
0	4						

	New Jersey	New Mexico	New York	North Carolina	Ohio	Oregon	Pennsylvania
Rhode Island							
Closed	56	11	6		3	3	36
1							110
Open	19	4	0		0	0	13
0							20

	South Carolina	Tennessee	Texas	Utah	Vermont	Virginia	Washington	West
Virginia								
Closed	15	96	49	16	2	49		75
8								
Open	3	47	22	6	1	11		23
3								

```
> comcast_data<- group_by(comcast_data,State,ComplaintStatus)
> chart_data<- summarise(comcast_data,Count = n())
`summarise()` has grouped output by 'State'. You can override using the `.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")
> INSIGHTS:- From the above chart, we can clearly see that Georgia has maximum
complaints.
Error: unexpected symbol in "INSIGHTS:- From the"
```

```

>
> #Now we need to see which state has maximum unresolved complaints
>
> comcast_data %>% filter(ComplaintStatus=='Open') %>% group_by(State) %>%
summarize(NumOfComplaints=n()) %>% arrange(desc(NumOfComplaints))
# A tibble: 34 x 2
  State      NumOfComplaints
  <chr>          <int>
1 Georgia             80
2 California          61
3 Tennessee          47
4 Florida            39
5 Illinois           29
6 Michigan           23
7 Washington         23
8 Colorado           22
9 Texas              22
10 Pennsylvania      20
# ... with 24 more rows
>
> #Now we want to see the percentage of resolved complaints.
> tot<-comcast_data %>% group_by(ComplaintStatus) %>%
summarize(NumOfComplaints=n())
> 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)
> #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())
`summarise()` has grouped output by 'ReceivedVia'. You can override using the
`.groups` argument.
> ccc<-comcast_data %>% filter(ReceivedVia=='Customer Care
Call',ComplaintStatus=='Closed') %>% group_by(ReceivedVia,ComplaintStatus) %>%
summarize(NumOfComplaints=n())
`summarise()` has grouped output by 'ReceivedVia'. You can override using the
`.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.
>
```