

project.r

lenovo

2021-04-28

```
#importing 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(lubridate)
```

```
##
```

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

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      date, intersect, setdiff, union
```

```
library(ggplot2)
```

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

```
View(Com_data)
```

```
str(Com_data)
```

```
## 'data.frame': 2224 obs. of 10 variables:
```

```
## $ Ticket.. : chr "250635" "223441" "242732" "277946" ...
```

```
## $ Customer.Complaint : chr "Comcast Cable Internet Speeds" "Payment disappear - service go
```

```
## $ Date : chr "22-04-2015" "4/8/2015" "18-04-2015" "5/7/2015" ...
```

```
## $ Time : chr "3:53:50 PM" "10:22:56 AM" "9:55:47 AM" "11:59:35 AM" ...
```

```
## $ Received.Via : chr "Customer Care Call" "Internet" "Internet" "Internet" ...
```

```
## $ City : chr "Abingdon" "Acworth" "Acworth" "Acworth" ...
```

```
## $ State : chr "Maryland" "Georgia" "Georgia" "Georgia" ...
```

```
## $ Zip.code : int 21009 30102 30101 30101 30101 30101 30101 30101 49221 94502 94501 ...
```

```
## $ Status : chr "Closed" "Closed" "Closed" "Open" ...
```

```
## $ Filing.on.Behalf.of.Someone: chr "No" "No" "Yes" "Yes" ...
```

```
#checking if NA present
```

```
na_value <- is.na(Com_data)
```

```
length(na_value[na_value==TRUE])
```

```
## [1] 0
```

```
head(Com_data$Date)
```

```
## [1] "22-04-2015" "4/8/2015" "18-04-2015" "5/7/2015" "26-05-2015"  
## [6] "6/12/2015"
```

```
Com_data$Date_New<- dmy(Com_data$Date)  
head(Com_data$Date_New)
```

```
## [1] "2015-04-22" "2015-08-04" "2015-04-18" "2015-07-05" "2015-05-26"  
## [6] "2015-12-06"
```

```
#monthly and daily complaint count
```

```
Com_data$month = month(Com_data$Date_New)  
monthly_count<- summarise(group_by(Com_data,month),Count = n())  
daily_count<- summarise(group_by(Com_data,Date_New),Count =n())  
monthly_count<-arrange(monthly_count,month)  
daily_count
```

```
## # A tibble: 91 x 2  
##   Date_New Count  
##   <date>   <int>  
## 1 2015-01-04    18  
## 2 2015-01-05    12  
## 3 2015-01-06    25  
## 4 2015-02-04    27  
## 5 2015-02-05     7  
## 6 2015-02-06    25  
## 7 2015-03-04    15  
## 8 2015-03-05     5  
## 9 2015-03-06    25  
## 10 2015-04-04    12  
## # ... with 81 more rows
```

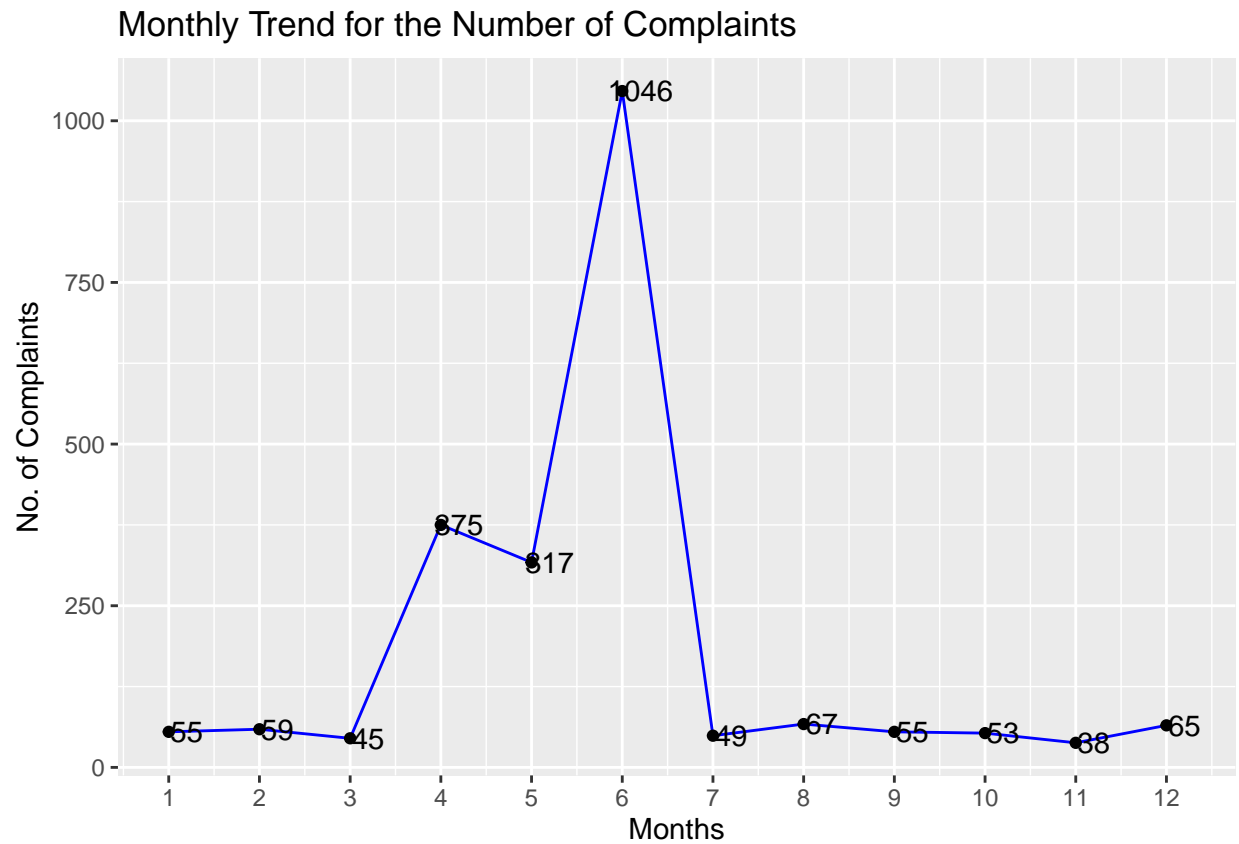
```
monthly_count
```

```
## # A tibble: 12 x 2  
##   month Count  
##   <dbl> <int>  
## 1     1     55  
## 2     2     59  
## 3     3     45  
## 4     4    375  
## 5     5    317  
## 6     6   1046  
## 7     7     49  
## 8     8     67  
## 9     9     55  
## 10    10     53  
## 11    11     38  
## 12    12     65
```

```
#count distribution - monthly plot.
```

```
ggplot(data = monthly_count, aes(x= month, y= Count))+  
  geom_line(color="blue")+  
  geom_point()+  
  geom_text(label= monthly_count$Count, nudge_x = 0.2, nudge_y = 0.3, check_overlap = T)+  
  scale_x_continuous(breaks = monthly_count$month)+
```

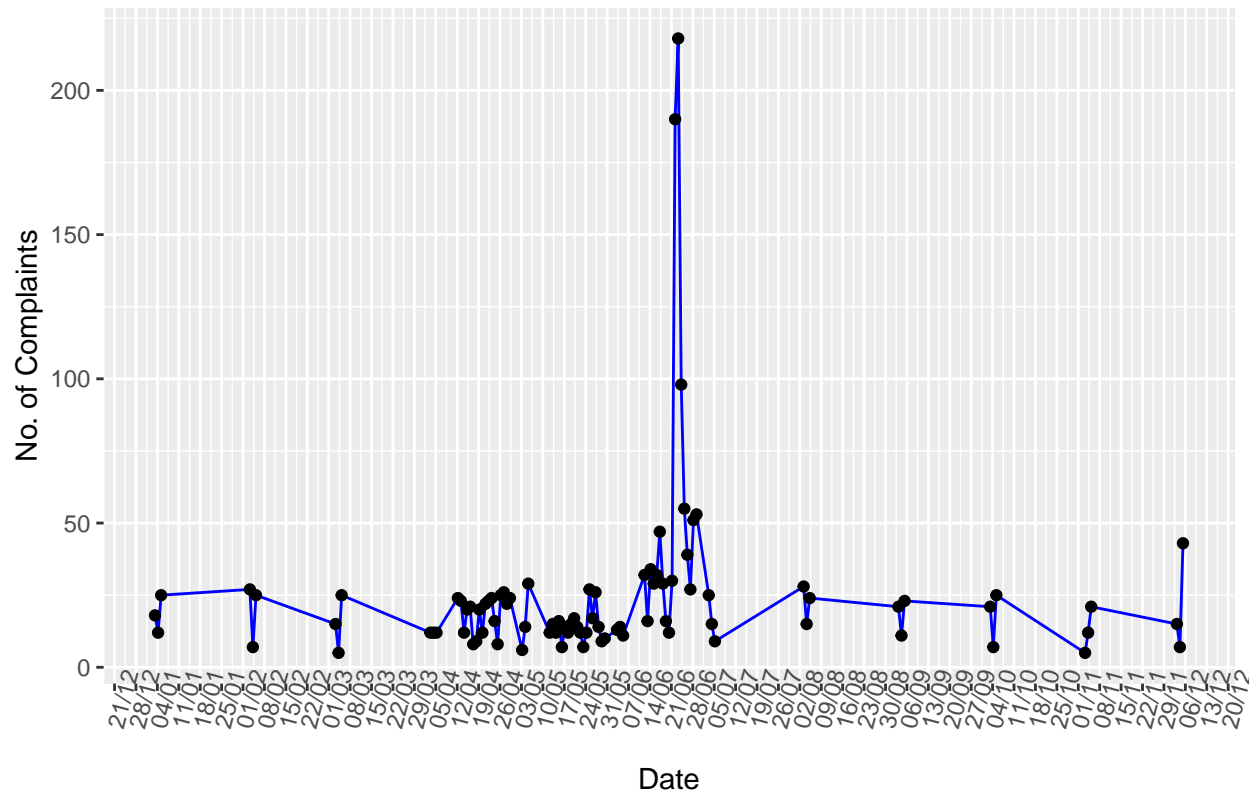
```
labs(title = "Monthly Trend for the Number of Complaints", x= "Months", y ="No. of Complaints")
```



#from above plot it is determined that June has the highest number of complaints

```
#count distribution - daily plot
ggplot(data = daily_count, aes(x= as.POSIXct(Date_New), y= Count))+
  geom_line(color="blue")+
  geom_point()+
  scale_x_datetime(breaks = "1 weeks", date_labels = "%d/%m")+
  theme(axis.text.x = element_text(angle = 75))+
  labs(title = "Daily Trend for the Number of Complaints", x = "Date", y ="No. of Complaints")
```

Daily Trend for the Number of Complaints



#from above graph it is determined that June 21 has the highest number of complaints with 14 and 28 as

Complaint Type Processing

```
network_tickets<- contains(Com_data$Customer.Complaint,match = 'network',ignore.case = T)
internet_tickets<- contains(Com_data$Customer.Complaint,match = 'internet',ignore.case = T)
billing_tickets<- contains(Com_data$Customer.Complaint,match = 'bill',ignore.case = T)
email_tickets<- contains(Com_data$Customer.Complaint,match = 'email',ignore.case = T)
charges_ticket<- contains(Com_data$Customer.Complaint,match = 'charge',ignore.case = T)
Com_data$ComplaintType[internet_tickets]<- "Internet"
Com_data$ComplaintType[network_tickets]<- "Network"
Com_data$ComplaintType[billing_tickets]<- "Billing"
Com_data$ComplaintType[email_tickets]<- "Email"
Com_data$ComplaintType[charges_ticket]<- "Charges"
Com_data$ComplaintType[-c(internet_tickets,network_tickets,
                           billing_tickets,charges_ticket,email_tickets)]<- "Others"
table(Com_data$ComplaintType)
```

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

#make a new categorical variable for Complaint Status.

```
open_complaints<-(Com_data$Status == 'Open' | Com_data$Status == 'Pending')
closed_complaints<-(Com_data$Status == 'Closed' | Com_data$Status == 'Solved')
Com_data$ComplaintStatus[open_complaints]<- 'Open'
Com_data$ComplaintStatus[closed_complaints]<- 'Closed'
stack<-table(Com_data$ComplaintStatus,Com_data$State)
```

```
stack
```

```
##
##      Alabama Arizona Arkansas California Colorado Connecticut Delaware
## Closed      17      14       6      159      58          9       8
## Open       9       6       0       61      22          3       4
##
##      District of Columbia District Of Columbia Florida Georgia Illinois
## Closed           1           14      201      208      135
## Open            0           2       39      80      29
##
##      Indiana Iowa Kansas Kentucky Louisiana Maine Maryland Massachusetts
## Closed      50      1      1       4       12      3       63       50
## Open       9      0      1       3       1      2       15       11
##
##      Michigan Minnesota Mississippi Missouri Montana Nevada New Hampshire
## Closed      92      29       23       3       1       1       8
## Open       23      4       16       1       0       0       4
##
##      New Jersey New Mexico New York North Carolina Ohio Oregon Pennsylvania
## Closed      56      11       6          3      3      36      110
## Open       19      4       0          0      0      13       20
##
##      Rhode Island South Carolina Tennessee Texas Utah Vermont Virginia
## Closed           1           15      96      49      16       2       49
## Open            0           3      47      22      6       1       11
##
##      Washington West Virginia
## Closed      75           8
## Open       23           3
```

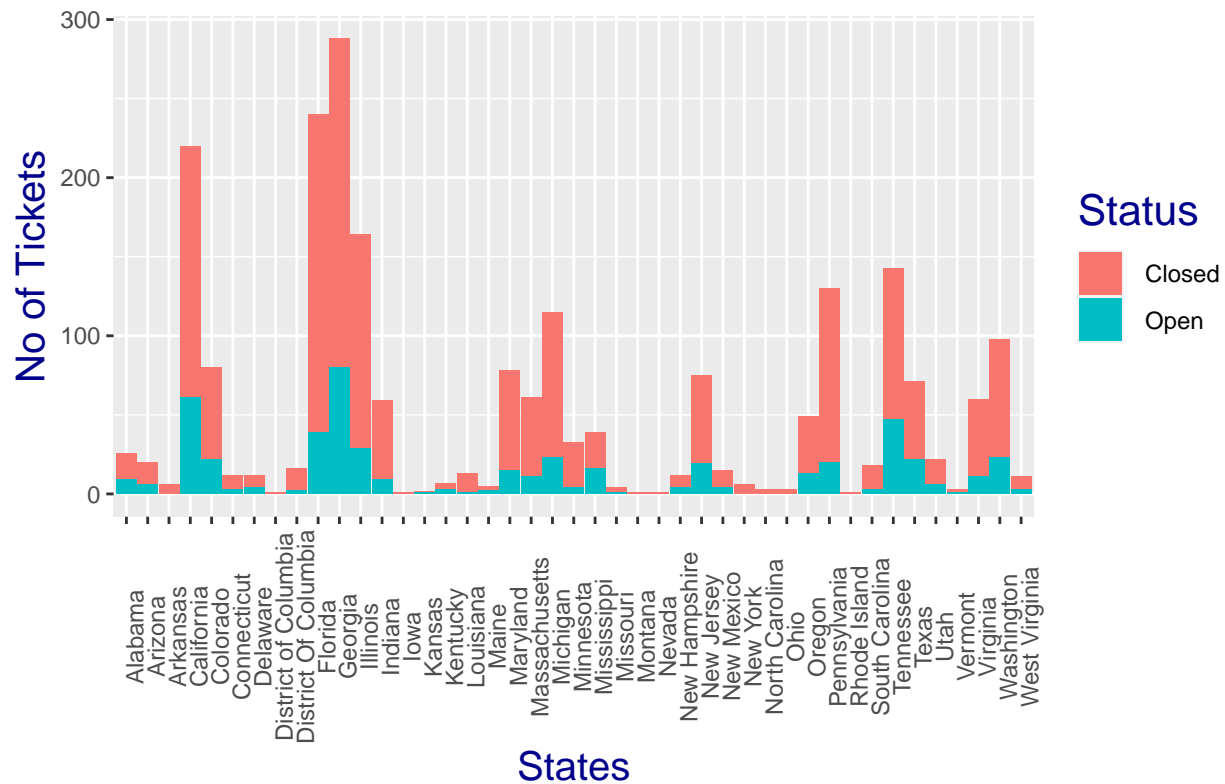
```
Com_data<- group_by(Com_data,State,ComplaintStatus)
chart_data<- summarise(Com_data,Count = n())
```

```
## `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 = "darkblue"),
        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



#which state has maximum unresolved complaints

```
Com_data %>% filter(ComplaintStatus=='Open') %>% group_by(State) %>% summarize(NumOfComplaints=n())
```

```
## # A tibble: 34 x 2
##   State                               NumOfComplaints
##   <chr>                               <int>
## 1 Alabama                               9
## 2 Arizona                               6
## 3 California                           61
## 4 Colorado                             22
## 5 Connecticut                           3
## 6 Delaware                              4
## 7 District Of Columbia                 2
## 8 Florida                              39
## 9 Georgia                              80
## 10 Illinois                             29
## # ... with 24 more rows
```

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

```
## # A tibble: 1 x 2
## # Groups:   State [1]
##   State Count
##   <chr>   <int>
```

```
## 1 Georgia      80
```

```
#Georgia has the highest number of open complaints
```

```
#the percentage of resolved complaints.
```

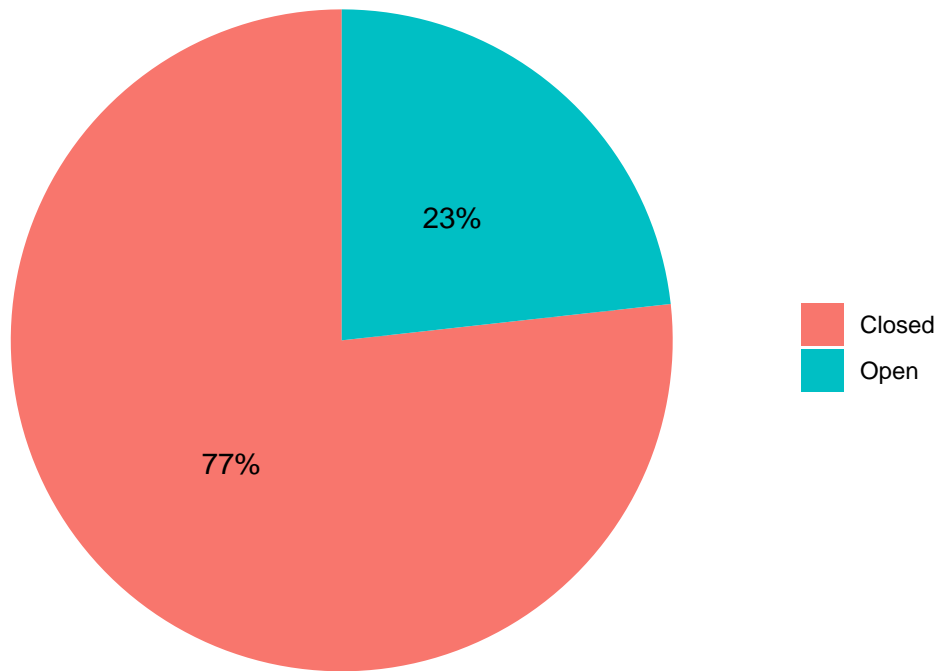
```
resolved_data <- group_by(Com_data,ComplaintStatus)
total_resolved<- summarise(resolved_data ,percentage =(n()/nrow(resolved_data)))
resolved_data <- group_by(Com_data,Received.Via,ComplaintStatus)
Category_resolved<- summarise(resolved_data ,percentage =(n()/nrow(resolved_data)))
```

```
## `summarise()` has grouped output by 'Received.Via'. You can override using the `.groups` argument.
```

```
par(mfrow = c(1,2))
total<-ggplot(total_resolved,
              aes(x= "",y =percentage,fill = ComplaintStatus))+
  geom_bar(stat = "identity",width = 1)+
  coord_polar("y",start = 0)+
  geom_text(aes(label = paste0(round(percentage*100),"%")),
            position = position_stack(vjust = 0.5))+
  labs(x = NULL,y = NULL,fill = NULL)+
  theme_classic()+theme(axis.line = element_blank(),
                        axis.text = element_blank(),
                        axis.ticks = element_blank())

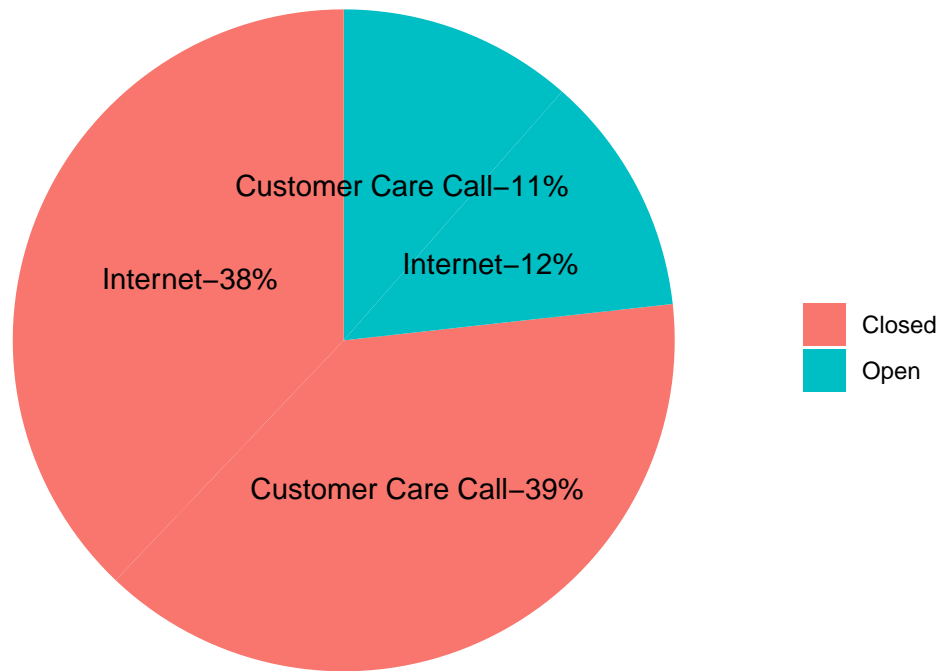
# Pie Chart for Category wise Ticket Status
category<-ggplot(Category_resolved,
                 aes(x= "",y =percentage,fill = ComplaintStatus))+
  geom_bar(stat = "identity",width = 1)+
  coord_polar("y",start = 0)+
  geom_text(aes(label = paste0(Received.Via,"-",round(percentage*100),"%")),
            position = position_stack(vjust = 0.5))+
  labs(x = NULL,y = NULL,fill = NULL)+
  theme_classic()+theme(axis.line = element_blank(),
                        axis.text = element_blank(),
                        axis.ticks = element_blank())

total
```



from the chart it can be determined that 77% of complaints are closed and 23% are open.

category



#out of closed complaints 39% are taken from customer care calls, 38% are taken from Internet
#out of open complaints 11% are taken from customer care calls, 12% are taken from Internet