

# Comcast Telecom Customer Complaints

Nishit Kothari (I.D. No. 205453)

16/12/2020

```
#Assessment project by Nishit Kothari (I.D. No. 205453)
## Import data into R environment.
```

```
rm(list = ls()) #clear variables
ComcastData <- read.csv("Comcast Telecom Complaints data.csv", stringsAsFactors=FALSE)
View(ComcastData)
summary(ComcastData)
```

```
## Ticket.. Customer.Complaint Date Time
## Length:2224 Length:2224 Length:2224 Length:2224
## Class :character Class :character Class :character Class :character
## Mode :character Mode :character Mode :character Mode :character
##
##
##
## Received.Via City State Zip.code
## Length:2224 Length:2224 Length:2224 Min. : 1075
## Class :character Class :character Class :character 1st Qu.:30057
## Mode :character Mode :character Mode :character Median :37211
## Mean :47994
## 3rd Qu.:77059
## Max. :99223
##
## Status Filing.on.Behalf.of.Someone
## Length:2224 Length:2224
## Class :character Class :character
## Mode :character Mode :character
##
##
##
```

```
str(ComcastData)
```

```
## '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" ...
```

```
sum(is.na(ComcastData)) # check for missing values
```

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

```
names(ComcastData)=c("Ticket","CustomerComplaint","Date","Time","ReceivedVia","City","State","Zipcode",
```

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

```
library(lubridate)
```

```
##
```

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

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

```
##
```

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

```
ComcastData$DateLubridate <- parse_date_time(ComcastData$Date, orders = "%d!-%m!-%y!")
```

```
str(ComcastData)
```

```
## 'data.frame': 2224 obs. of 11 variables:
```

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

```
## $ CustomerComplaint: chr "Comcast Cable Internet Speeds" "Payment disappear - service got disconne
```

```
## $ 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" ...
```

```
## $ ReceivedVia : chr "Customer Care Call" "Internet" "Internet" "Internet" ...
```

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

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

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

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

```
## $ Is_4_Someone : chr "No" "No" "Yes" "Yes" ...
```

```
## $ DateLubridate : POSIXct, format: "2015-04-22" "2015-08-04" ...
```

```
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
```

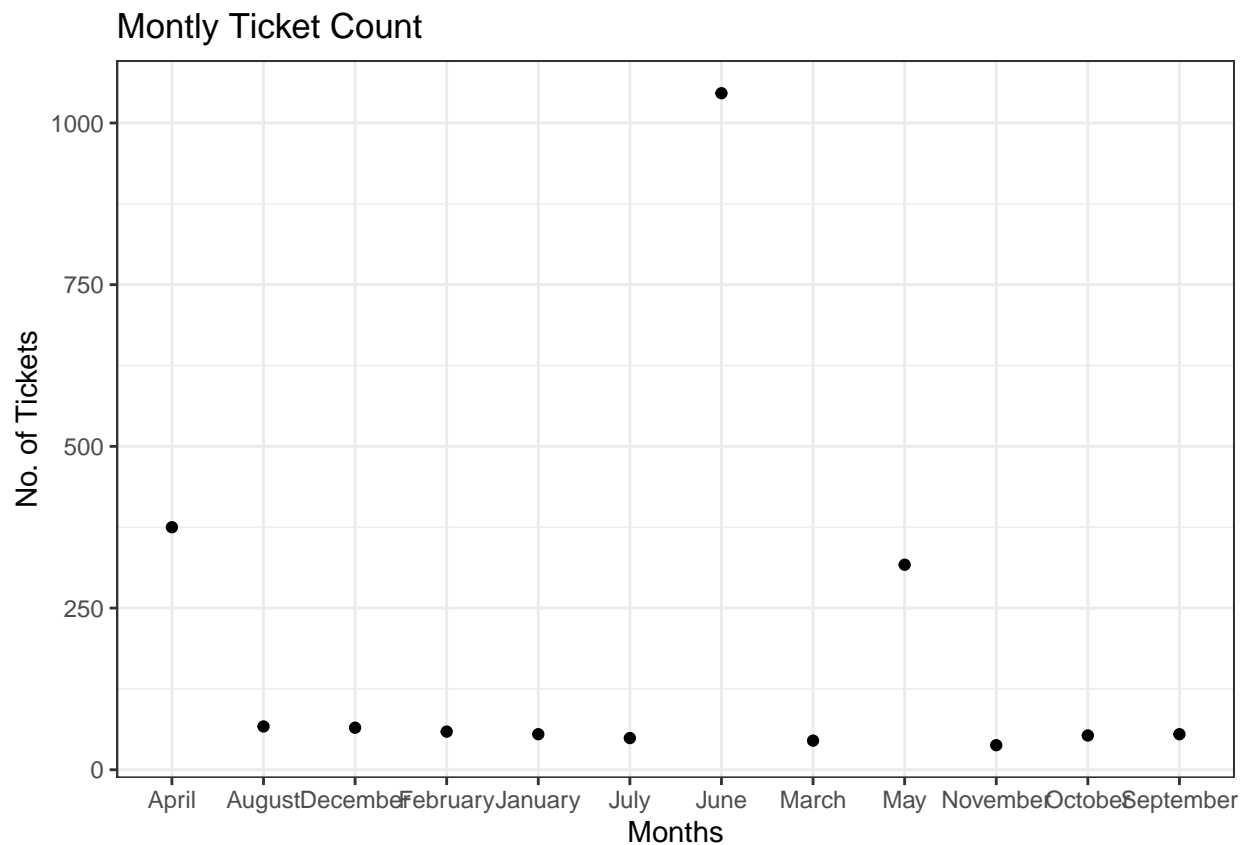
```
library(ggplot2)
## Number of complaints at a monthly level

ComcastData$Month<-months(x = ComcastData$DateLubridate)

monthly_level <- ComcastData %>% group_by(Month) %>% summarize(NumOfComplaints=n()) %>% arrange(desc(NumOfComplaints))

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

monthly <- ggplot(data = monthly_level,aes(x = Month, y = NumOfComplaints))+
  geom_point()+
  labs(title = "Montly Ticket Count",x= "Months",y ="No. of Tickets")+
  theme_bw()
plot(monthly)
```



```
###As we can see that in the month of April,May the tickets are increases but in
### the month of June it increases drastically, so there might be some reason
### for which they received high amount of tickets.

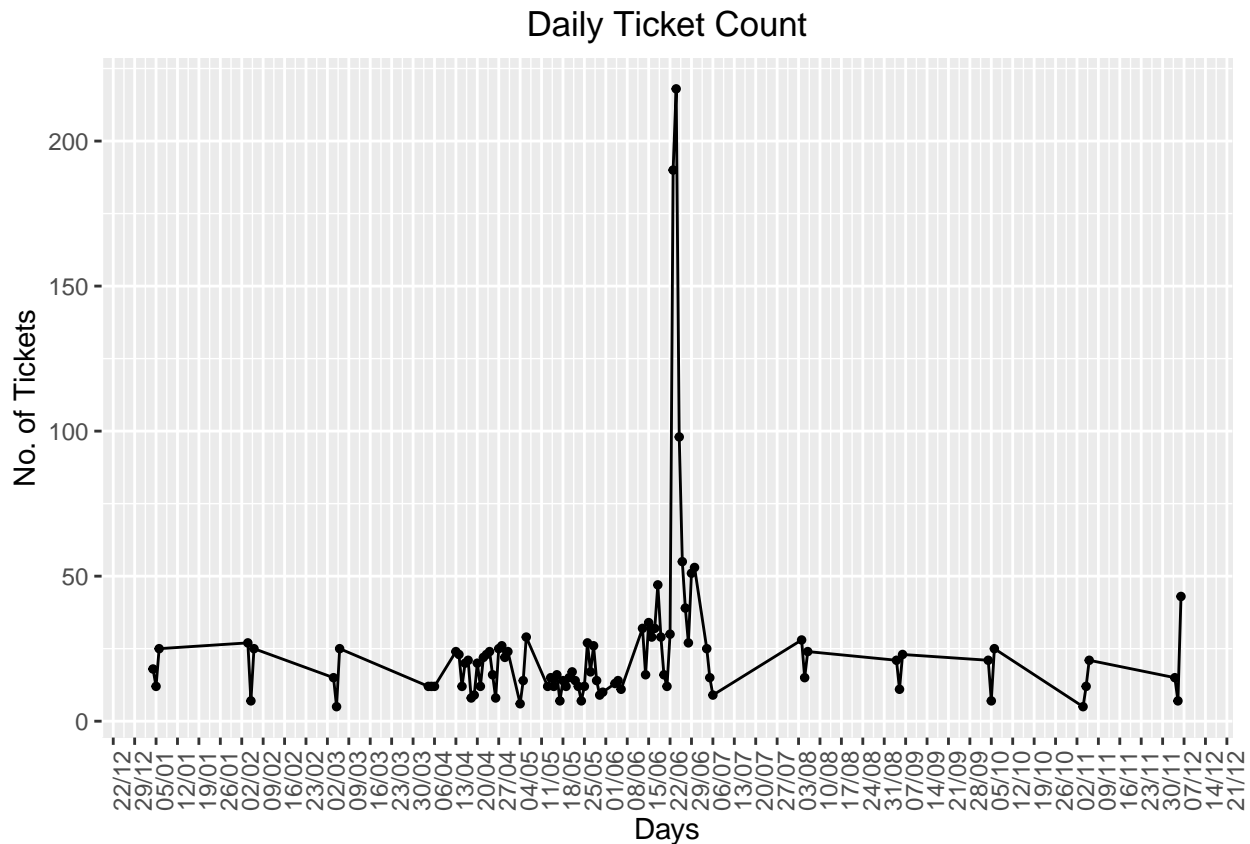
## Number of complaints at a daily level
daily_level <- ComcastData %>% group_by(DateLubridate) %>% summarize(NumOfComplaints=n()) %>% arrange(desc(NumOfComplaints))

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

```

daily <- ggplot(data = daily_level,aes(x = DateLubridate, y = 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 = 90),
        plot.title = element_text(hjust = 0.5))
plot(daily)

```



###And with the help of above daily chart of tickets we can observe that in  
 ###second half of June month we received more tickets with respect to normal days

### ## 3. Complaint type processing

#Check for common complaint words in data

```

network_tickets<- contains(ComcastData$CustomerComplaint,match = 'network',ignore.case = T)
internet_tickets<- contains(ComcastData$CustomerComplaint,match = 'internet',ignore.case = T)
billing_tickets<- contains(ComcastData$CustomerComplaint,match = 'bill',ignore.case = T)
email_tickets<- contains(ComcastData$CustomerComplaint,match = 'email',ignore.case = T)
charges_ticket<- contains(ComcastData$CustomerComplaint,match = 'charge',ignore.case = T)

```

#Label each complaint found as "Internet", "Network", "Billing", etc

```

ComcastData$ComplaintType[internet_tickets]<- "Internet"
ComcastData$ComplaintType[network_tickets]<- "Network"

```

```
ComcastData$ComplaintType[billing_tickets]<- "Billing"
ComcastData$ComplaintType[email_tickets]<- "Email"
ComcastData$ComplaintType[charges_ticket]<- "Charges"

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

table(ComcastData$ComplaintType)
```

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

*###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.*

*## 4. Creating new Variable ComplaintStatus with values Open and Closed.*

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

*## 5. Creating Stacked barchart for complaints based on State and Status.*

```
chart_data <- ComcastData %>% group_by(State, ComplaintStatus) %>% summarise(NumOfComplaints = n())
```

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

```
chart_data
```

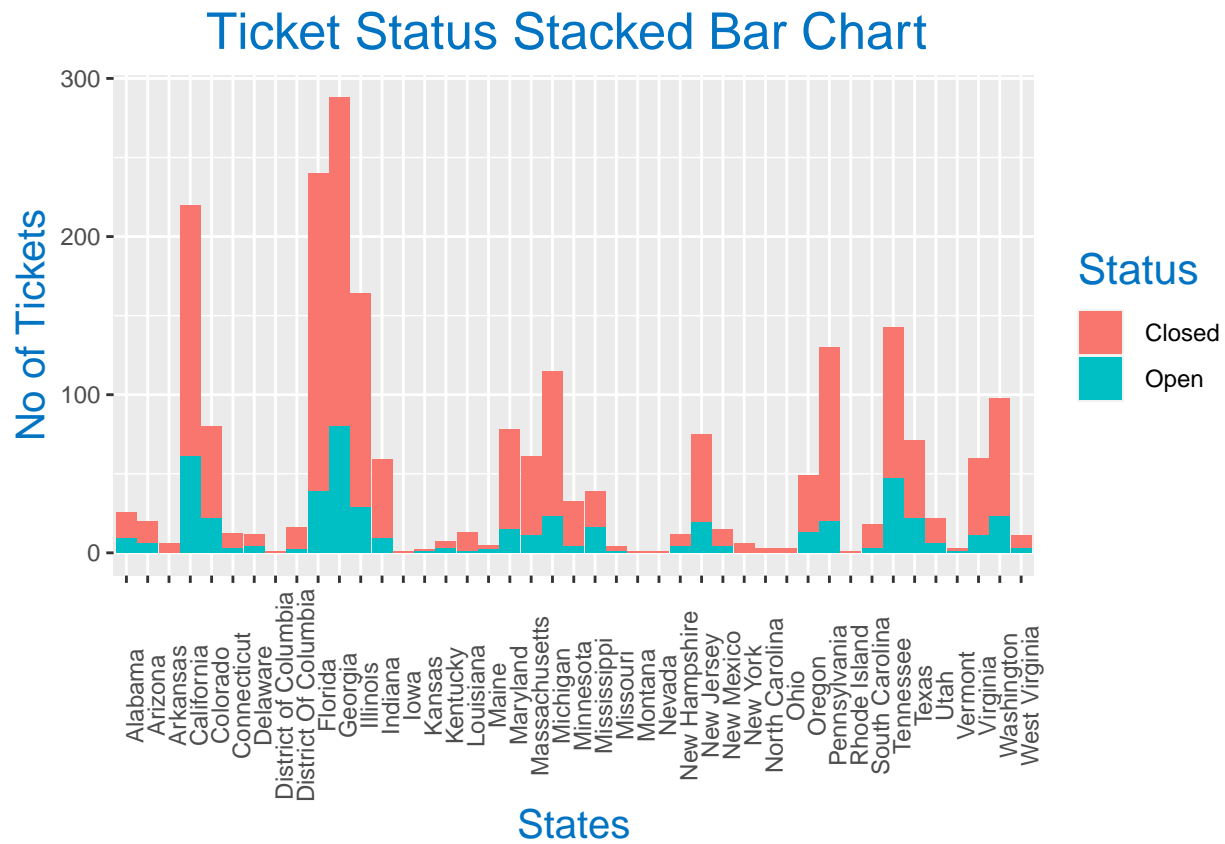
```
## # A tibble: 77 x 3
## # Groups:   State [43]
##   State      ComplaintStatus NumOfComplaints
##   <chr>      <chr>             <int>
## 1 Alabama    Closed                17
## 2 Alabama    Open                  9
## 3 Arizona    Closed                14
## 4 Arizona    Open                  6
## 5 Arkansas    Closed                6
## 6 California Closed               159
## 7 California Open                 61
## 8 Colorado    Closed                58
## 9 Colorado    Open                 22
## 10 Connecticut Closed                9
## # ... with 67 more rows
```

```
ss <- ggplot(as.data.frame(chart_data) ,mapping = aes(State,NumOfComplaints))+
  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")
plot(ss)

```



```

### 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 having highest number of unresolved complaints
open_complaints <- chart_data %>% filter(ComplaintStatus == "Open")
open_complaints[open_complaints$NumOfComplaints == max(open_complaints$NumOfComplaints),c(1,3)]

## # A tibble: 1 x 2
## # Groups:   State [1]
##   State   NumOfComplaints
##   <chr>         <int>
## 1 Georgia             80

```

```
###As we can observe that State Georgia has maximum number of unresolved tickets
###and these ticket count is 80
```

```
# 6. Calculating Resolution Percentage based on Total and Catagory.
```

```
resolved_data <- group_by(ComcastData,ComplaintStatus)
total_resolved<- summarise(resolved_data ,percentage =(n()/nrow(resolved_data)))
```

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

```
total_resolved
```

```
## # A tibble: 2 x 2
##   ComplaintStatus percentage
##   <chr>           <dbl>
## 1 Closed          0.768
## 2 Open            0.232
```

```
resolved_data <- group_by(ComcastData,ReceivedVia,ComplaintStatus)
Category_resolved<- summarise(resolved_data ,percentage =(n()/nrow(resolved_data)))
```

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

```
Category_resolved
```

```
## # A tibble: 4 x 3
## # Groups:   ReceivedVia [2]
##   ReceivedVia      ComplaintStatus percentage
##   <chr>           <chr>           <dbl>
## 1 Customer Care Call Closed          0.388
## 2 Customer Care Call Open           0.115
## 3 Internet        Closed           0.379
## 4 Internet        Open            0.118
```

```
###Insights:As per the above analysis we observe that in the 2nd half of the
###June month Comcast received high amount of complaints in which most of the
###complaints are related to internet service issue and the highest amount of
###complaints are received from the state Georgia. The highest unresolved
###complaints are related from the state Georgia and the total amount of
###resolved complaints are 77% in which 38% are received the internet and
###39% are from the customer care calls.
```