

Comcast Telecom Consumer Complaints:

Project 2:

The source code:

```
# Import data into R environment.
setwd("D:/Downloads/GDrive/SEU Files/Training files/IT499/R/project")
getwd()
file <- read.csv("Comcast Telecom Complaints data.csv", stringsAsFactors = TRUE)
library(dplyr)
library(stringi)
library(lubridate)
library(ggplot2)

str(file)
summary(file)
names(file)<- stri_replace_all(regex = "\\.",replacement = "",str =names(file))
head(file)

# - Provide the trend chart for the number of complaints at monthly and daily granula
rity levels.
file$Date<- dmy(file$Date)
head(file)
by_day <- summarise(group_by(file,Date),Count=n())
by_month <- arrange(summarise(group_by(file,month=as.integer(month(Date))),Count=n())
,month)

ggplot(data = by_month,aes(month,Count,label = Count))+geom_line(color="red")+geom_te
xt()+scale_x_continuous(breaks = by_month$month)+labs(title = "Tickets Count by Month
",x= "Month",y = "Count")
ggplot(data = by_day,aes(as.POSIXct(Date),Count))+geom_line(color="blue")+theme(axis.
text.x = element_text(angle = 90))+scale_x_datetime(breaks = "1 weeks",date_labels =
"%d/%m")+labs(title = "Tickets Count by Day",x= "Day",y = "Count")

# - Provide a table with the frequency of complaint types.
technical <- contains(file$CustomerComplaint,match=("speed"),ignore.case = T)
technical2 <- contains(file$CustomerComplaint,match="internet",ignore.case = T)
technical3 <- contains(file$CustomerComplaint,match="network",ignore.case = T)
financial <- contains(file$CustomerComplaint,match=("Pay"),ignore.case = T)
financial2 <- contains(file$CustomerComplaint,match=("bill"),ignore.case = T)
financial3 <- contains(file$CustomerComplaint,match=("price"),ignore.case = T)
financial4 <- contains(file$CustomerComplaint,match=("charge"),ignore.case = T)
services <- contains(file$CustomerComplaint,match="service",ignore.case = T)

file$Complaint_type[technical]<-"technical"
file$Complaint_type[technical2]<-"technical"
file$Complaint_type[technical3]<-"technical"
```

```

file$Complaint_type[financial]<-"financial"
file$Complaint_type[financial2]<-"financial"
file$Complaint_type[financial3]<-"financial"
file$Complaint_type[financial4]<-"financial"
file$Complaint_type[services]<-"services"
file$Complaint_type[-c(technical, financial, services)]<-"others"

table(file$Complaint_type)

# Which complaint types are maximum i.e., around internet, network issues, or across
any other domains.
# *the table shows that the maximum complaints are of type "others"
# *and the minimum are of type "financial"

# - Create a new categorical variable with value as Open and Closed. Open & Pending i
s to be categorized as Open and Closed & Solved is to be categorized as Closed.

file <- cbind(file,new_status= (ifelse(file$Status=="Open","Open",ifelse(file$Status=
="Pending","Open","Closed"))))

# - Provide state wise status of complaints in a stacked bar chart. Use the categoriz
ed variable from Q3. Provide insights on:
by_state <- summarize(group_by(file,State,new_status), Count=n())
ggplot(by_state ,aes(State,Count))+geom_col(aes(fill = new_status),width = 0.75)+them
e(axis.text.x = element_text(angle = 90))+labs(title = "Tickets Count by State",x = "
State",y = "Count",fill= "Status")

# Which state has the maximum complaints
by_state <- as.data.frame(by_state)
filter(by_state, Count ==max(Count))

# Which state has the highest percentage of unresolved complaints
filter(filter(by_state,new_status=="Open"),Count ==max(Count))

# - Provide the percentage of complaints resolved till date, which were received thro
ugh the Internet and customer care calls.
resolved_tickets <- summarise(filter(file, new_status=="Closed"),count=n())
resolved_tickets
via_Internet <- summarise(filter(file,new_status=="Closed", ReceivedVia=="Internet"),
count=n())
via_Calls <- summarise(filter(file,new_status=="Closed", ReceivedVia == "Customer Care
Call"),count=n())

internet_tickets_percentage <- (via_Internet/resolved_tickets)*100
internet_tickets_percentage
call_tickets_percentage <- (via_Calls/resolved_tickets)*100
call_tickets_percentage

```

The output as it appears in the output window:

```
# <Import data into R environment.

<setwd("D:/Downloads/GDrive/SEU Files/Training files/IT499/R/project")

<getwd()

" [\\]D:/Downloads/GDrive/SEU Files/Training files/IT499/R/project"

<file <- read.csv("Comcast Telecom Complaints data.csv", stringsAsFactors = TRUE)

<library(dplyr)

<library(stringi)

<library(lubridate)

<library(ggplot2)

<

<str(file)

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

 $ Ticket..      : Factor w/ 2224 levels "211255","211472",...: 371 124 307 611 849 1214 1763
1590 967 2110...

 $ Customer.Complaint : Factor w/ 1841 levels "(Comcast is not my complaint!) Cyber Tele-
marketing is my complaint!",...: 329 1519 1660 520 668 1353 1715 733 468 717...

 $ Date          : Factor w/ 91 levels "13-04-2015","13-05-2015",...: 28 66 16 77 41 83 36 33 80
48...

 $ Time          : Factor w/ 2190 levels "1:00:18 AM","1:00:32 PM",...: 1198 291 2165 652 89
2189 252 1666 594 1648...

 $ Received.Via   : Factor w/ 2 levels "Customer Care Call",...: 1 2 2 2 2 2 1 2 1 1...

 $ City          : Factor w/ 928 levels "Abingdon","Acworth",...: 1 2 2 2 2 2 2 3 4 4...

 $ State         : Factor w/ 43 levels "Alabama","Arizona",...: 19 11 11 11 11 11 11 21 4 4...

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

 $ Status        : Factor w/ 4 levels "Closed","Open",...: 1 1 1 2 4 4 3 4 1 2...

 $ Filing.on.Behalf.of.Someone: Factor w/ 2 levels "No","Yes": 1 1 2 2 1 1 1 1 1 2...

<summary(file)

  Ticket..      Customer.Complaint      Date      Time      Received.Via      City
State      Zip.code

  1 : 211255 Comcast      : 83  24-06-2015: 218  1:07:49 PM : 2  Customer Care Call:1119
Atlanta   : 63  Georgia   : 288  Min.   : 1075

  1 : 211472 Comcast Internet : 18  23-06-2015: 190  1:12:13 PM : 2  Internet      :1105
Chicago   : 47  Florida   : 240  1st Qu.:30057
```

1 : 211478 Comcast Data Cap : 17 25-06-2015: 98 1:26:23 PM : 2
Knoxville : 36 California : 220 Median : 37211

1 : 211677 comcast : 13 26-06-2015: 55 1:29:58 PM : 2 Houston :
33 Illinois : 164 Mean : 47994

1 : 211770 Comcast Billing : 11 30-06-2015: 53 1:30:10 PM : 2
Jacksonville: 31 Tennessee : 143 3rd Qu.: 77059

1 : 211784 Comcast Data Caps: 11 29-06-2015: 51 10:02:10 AM: 2
Philadelphia: 25 Pennsylvania: 130 Max. : 99223

(Other): 2218 (Other) : 2071 (Other) : 1559 (Other) : 2212
(Other) : 1989 (Other) : 1039

Status Filing.on.Behalf.of.Someone

Closed : 734 No : 2021

Open : 363 Yes: 203

Pending: 154

Solved : 973

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

```
<head(file)
```

Ticket	CustomerComplaint	Date	Time	ReceivedVia
City	State	Zipcode	Status	
1	200630 Comcast Cable Internet Speeds	22-04-2015	3:53:50 PM	Customer Care Call
Abingdon	Maryland	21009	Closed	
2	223441 Payment disappear - service got disconnected	4/8/2015	10:22:56 AM	
Internet	Acworth	Georgia	30102	Closed
3	242732 Speed and Service	18-04-2015	9:55:47 AM	Internet
Acworth	Georgia	30101	Closed	
4	277946 Comcast Imposed a New Usage Cap of 300GB that punishes streaming.	5/7/2015	11:59:35 AM	
Internet	Acworth	Georgia	30101	Open
5	307170 Comcast not working and no service to boot	26-05-2015	1:25:26 PM	
Internet	Acworth	Georgia	30101	Solved
6	338019 ISP Charging for arbitrary data limits with overage fees	6/12/2015	9:59:40 PM	
Internet	Acworth	Georgia	30101	Solved

FilingonBehalfofSomeone

1 No
2 No
3 Yes

4 Yes
5 No
6 No

<

<

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

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

```
<head(file)
```

Ticket	CustomerComplaint	Date	Time	ReceivedVia
1	250630 Comcast Cable Internet Speeds	2015-04-22	3:53:50 PM	Customer Care Call
2	223441 Payment disappear - service got disconnected	2015-08-04	10:22:56 AM	Internet
3	242732 Speed and Service	2015-04-18	9:55:47 AM	Internet
4	277946 Comcast Imposed a New Usage Cap of 300GB that punishes streaming.	2015-07-05	11:59:35 AM	Internet
5	307170 Comcast not working and no service to boot	2015-05-26	1:25:26 PM	Internet
6	338019 ISP Charging for arbitrary data limits with overage fees	2015-12-06	9:59:40 PM	Internet

FilingonBehalfofSomeone

1 No
2 No
3 Yes
4 Yes
5 No
6 No

```
<by_day <- summarise(group_by(file,Date),Count=n())
```

```
<by_month <-
```

```
arrange(summarise(group_by(file,month=as.integer(month(Date))),Count=n()),month)
```

<

```
<ggplot(data = by_month,aes(month,Count,label =
Count))+geom_line(color="red")+geom_text()+scale_x_continuous(breaks =
by_month$month)+labs(title = "Tickets Count by Month",x= "Month",y ="Count")
```

```
<ggplot(data = by_day,aes(as.POSIXct(Date),Count))+geom_line(color="blue")+theme(axis.text.x =
element_text(angle = 90))+scale_x_datetime(breaks = "1 weeks",date_labels = "%d/%m")+labs(title
= "Tickets Count by Day",x= "Day",y ="Count")
```

```
<
```

```
<
```

```
- # <Provide a table with the frequency of complaint types.
```

```
<technical <- contains(file$CustomerComplaint,match="speed"),ignore.case = T)
```

```
<technical2 <- contains(file$CustomerComplaint,match="internet",ignore.case = T)
```

```
<technical3 <- contains(file$CustomerComplaint,match="network",ignore.case = T)
```

```
<financial <- contains(file$CustomerComplaint,match="Pay"),ignore.case = T)
```

```
<financial2 <- contains(file$CustomerComplaint,match="bill"),ignore.case = T)
```

```
<financial3 <- contains(file$CustomerComplaint,match="price"),ignore.case = T)
```

```
<financial4 <- contains(file$CustomerComplaint,match="charge"),ignore.case = T)
```

```
<services <- contains(file$CustomerComplaint,match="service",ignore.case = T)
```

```
<
```

```
<file$Complaint_type[technical]<-"technical"
```

```
<file$Complaint_type[technical2]<-"technical"
```

```
<file$Complaint_type[technical3]<-"technical"
```

```
<file$Complaint_type[financial]<-"financial"
```

```
<file$Complaint_type[financial2]<-"financial"
```

```
<file$Complaint_type[financial3]<-"financial"
```

```
<file$Complaint_type[financial4]<-"financial"
```

```
<file$Complaint_type[services]<-"services"
```

```
<file$Complaint_type[-c(technical, financial, services)]<-"others"
```

```
<
```

```
<table(file$Complaint_type)
```

```
financial    others    services technical
```

```
144        489        1036        50
```

```
<
```

```
<
```

```
# <Which complaint types are maximum i.e., around internet, network issues, or across any other
domains.
```

```
* # <the table shows that the maximum complaints are of type "others"
```

```
* # <and the minimum are of type "financial"
```

```
<
```

```
<
```

- # <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.

```
<
```

```
<file <- cbind(file,new_status=
(ifelse(file$Status=="Open","Open",ifelse(file$Status=="Pending","Open","Closed"))))
```

```
<
```

```
<
```

- # <Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:

```
<by_state <- summarize(group_by(file,State,new_status), Count=n())
```

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

```
<ggplot(by_state ,aes(State,Count))+geom_col(aes(fill = new_status),width =
0.75)+theme(axis.text.x = element_text(angle = 90))+labs(title = "Tickets Count by State",x =
"State",y = "Count",fill= "Status")
```

```
<
```

```
<
```

<Which state has the maximum complaints

```
<by_state <- as.data.frame(by_state)
```

```
<filter(by_state, Count ==max(Count))
```

```
State new_status Count
```

```
1 Georgia    Closed  208
```

```
<
```

<Which state has the highest percentage of unresolved complaints

```
<filter(filter(by_state,new_status=="Open"),Count ==max(Count))
```

```
State new_status Count
```

```
1 Georgia    Open    80
```

```
<
```

- # <Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

```
<resolved_tickets <- summarise(filter(file, new_status=="Closed"),count=n())
```

```
<resolved_tickets
```

```
count
```

```
1 100.0
```

```
<via_Internet <- summarise(filter(file,new_status=="Closed", ReceivedVia=="Internet"),count=n())
```

```
<via_Calls <- summarise(filter(file,new_status=="Closed", ReceivedVia == "Customer Care Call"),count=n())
```

```
<
```

```
<internet_tickets_percentage <- (via_Internet/resolved_tickets)*100
```

```
<internet_tickets_percentage
```

```
count
```

```
149,38489
```

```
<call_tickets_percentage <- (via_Calls/resolved_tickets)*100
```

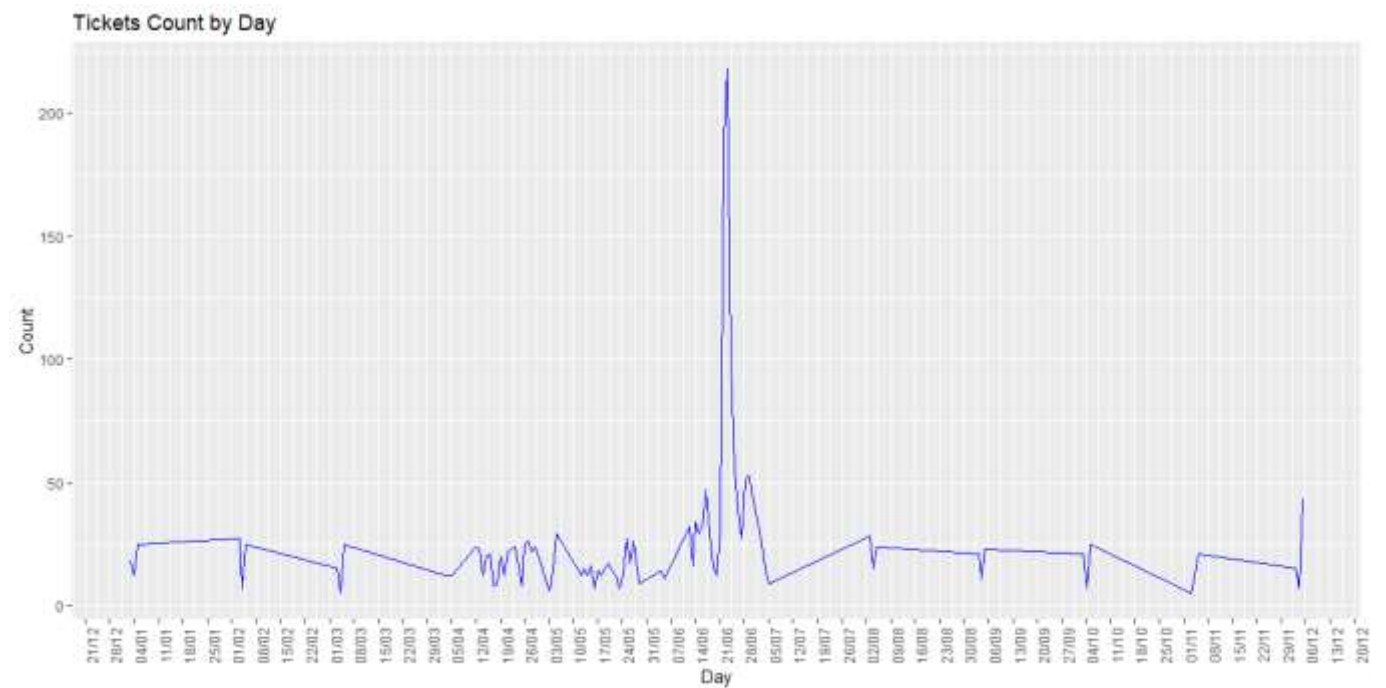
```
<call_tickets_percentage
```

```
count
```

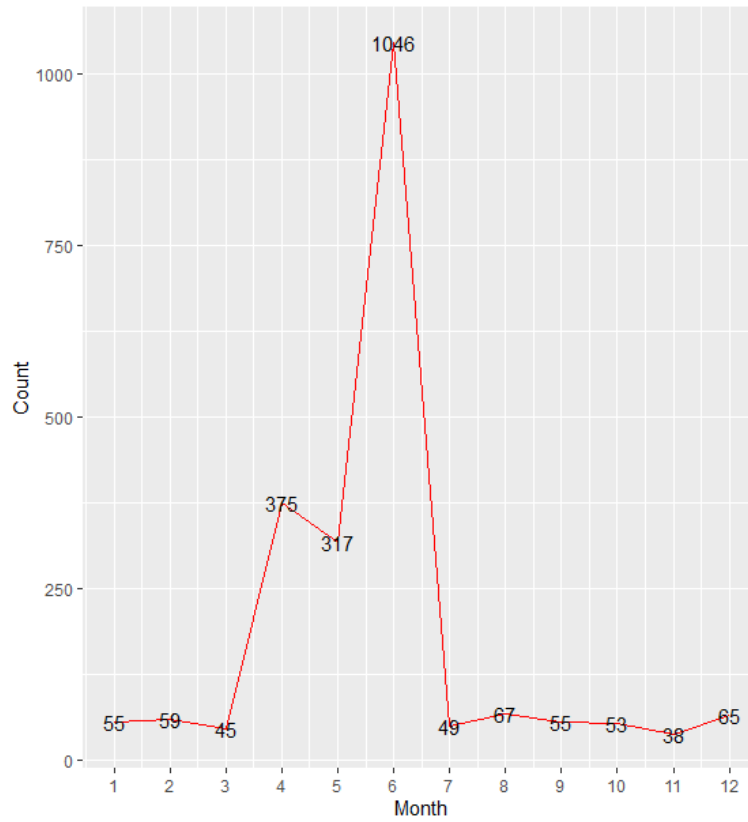
```
100,61011
```

```
>
```

The graphs appears as the following images:



Tickets Count by Month



Tickets Count by State

