Comcast Telecom Consumer Complaints

Comcast is an American global telecommunication company. The firm has been providing terrible customer service. They continue to fall short despite repeated promises to improve. Only last month (October 2016) the authority fined them a $2.3 million, after receiving over 1000 consumer complaints.

The existing database will serve as a repository of public customer complaints filed against Comcast. It will help to pin down what is wrong with Comcast’s customer service.

Tasks to be performed:-

* Import data into R environment.
* Provide the trend chart for the number of complaints at monthly and daily granularity levels.
* 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.
* 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.
* Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:
* Which state has the maximum complaints
* Which state has the highest percentage of unresolved complaints
* Provide the percentage of complaints resolved till date, which were received through theInternet and customer care calls.

*#Importing necessary 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**(ggplot2) **library**(lubridate)

##

## Attaching package: ’lubridate’

## The following objects are masked from ’package:base’: ##

## date, intersect, setdiff, union

*#Importing Comcast Dataset*

comcast\_data<- **read.csv**("Comcast Telecom Complaints data.csv",header = TRUE)

*#Manipulating Field Names*

**names**(comcast\_data)<-**gsub**(pattern = ’\\.’,replacement = "",x=**names**(comcast\_data))

**names**(comcast\_data)

## [1] "Ticket" "CustomerComplaint"

## [3] "Date" "Time"

## [5] "ReceivedVia" "City"

## [7] "State" "Zipcode"

## [9] "Status" "FilingonBehalfofSomeone"

**View**(comcast\_data)

From the dataset, we can see that the format of Date column is not same throughout, so we need to make it same for analysis.

*#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**())

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

*#Plotting for daily granularity level*

**ggplot**(data = ans,**aes**(**as.POSIXct**(Date),NumOfComplaints))**+ geom\_line**()**+**

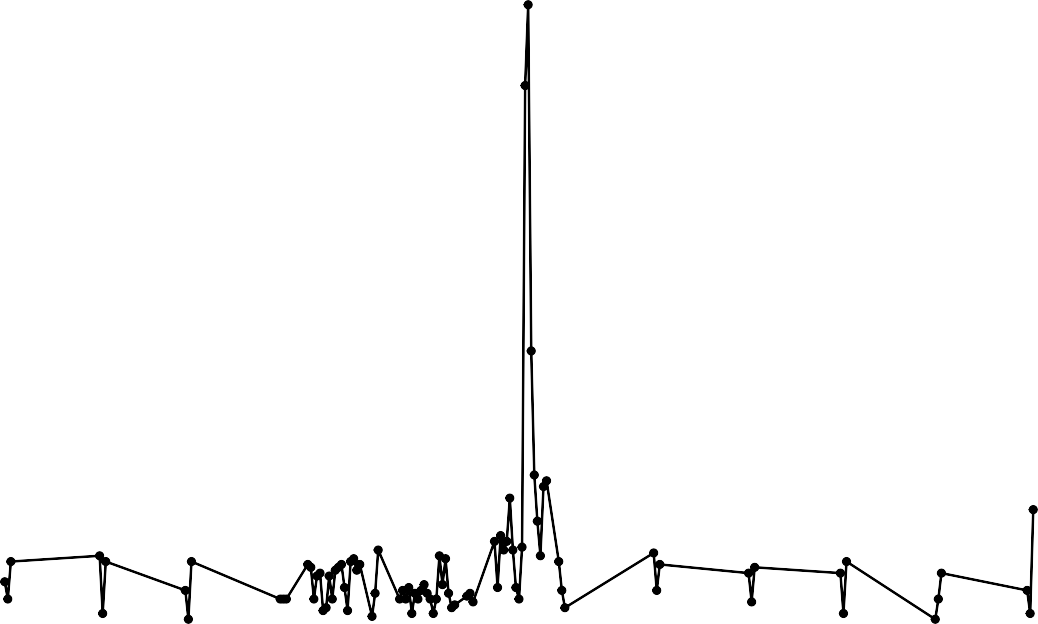
**geom\_point**(size = 1)**+**

**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))

## Daily Ticket Count

200



150

100

No. of Tickets

50

0

21/12

28/12

04/01

11/01

18/01

25/01

01/02

08/02

15/02

22/02

01/03

08/03

15/03

22/03

29/03

05/04

12/04

19/04

26/04

03/05

10/05

17/05

24/05

31/05

07/06

14/06

21/06

28/06

05/07

12/07

19/07

26/07

02/08

09/08

16/08

23/08

30/08

06/09

13/09

20/09

27/09

04/10

11/10

18/10

25/10

01/11

08/11

15/11

22/11

29/11

06/12

13/12

20/12

#### Days

* Now we need to get the complaints on a monthly level basis and plot a trend chart for it.

*#Making month field*

comcast\_data**$**Month<-**months**(comcast\_data**$**Date)

ans1<-comcast\_data **%>% group\_by**(Month =**as.integer**(**month**(Date))) **%>% summarize**(NumOfComplaints=**n**()) **%>%**

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

*#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))

## Monthly Ticket Count

1000

1046

375

317

55

59

45

49

67

55

53

38

65

750

500

No. of Tickets

250

0

1 2 3 4 5

6 7 8

9 10 11 12

#### Months

INSIGHTS:- From the above trend chart, we can clearly see that complaints for the month of June are maximum i.e.1046.

* Now we need to make a frequency table basis the complaint types.

T)

*# Complaint Type Processing*

network\_tickets<- **contains**(comcast\_data**$**CustomerComplaint,match = ’network’,ignore.case = internet\_tickets<- **contains**(comcast\_data**$**CustomerComplaint,match = ’internet’,ignore.case 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)

= T)

##

## 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 | | | | | | | | | | | | | | |
| ## | 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 | | | | | | | | | | | | | | |

comcast\_data<- **group\_by**(comcast\_data,State,ComplaintStatus) chart\_data<- **summarise**(comcast\_data,Count = **n**())

## ‘summarise()‘ regrouping output by ’State’ (override with ‘.groups‘ argument)

|  |  |  |  |
| --- | --- | --- | --- |
| ## | Closed | 75 | 8 |
| ## | Open | 23 | 3 |

*#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")

Ticket Status Stacked Bar Chart

300

200

No of Tickets

100

Status

Closed Open

0

# States

Alabama Arizona Arkansas California Colorado Connecticut Delaware

District of Columbia District Of Columbia Florida

Georgia Illinois Indiana Iowa Kansas Kentucky Louisiana Maine Maryland

Massachusetts Michigan Minnesota Mississippi Missouri Montana Nevada

New Hampshire New Jersey New Mexico New York

North Carolina Ohio

Oregon Pennsylvania Rhode Island South Carolina Tennessee Texas

Utah Vermont Virginia Washington West Virginia

INSIGHTS:- From the above chart, we can clearly see that Georgia has maximum complaints.

* Now we need to see which state has maximum unresolved complaints

comcast\_data **%>% filter**(ComplaintStatus**==**’Open’) **%>% group\_by**(State) **%>% summarize**(NumOfComplaints=**n**())

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

INSIGHTS:- From the table generated above we can see that Georgia has maximum unresolved complaints

i.e. 80.

* Now we want to see the percentage of resolved complaints.

tot<-comcast\_data **%>% group\_by**(ComplaintStatus) **%>% summarize**(NumOfComplaints=**n**())

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

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)

### Closed 76.75%

Open 23.25%

INSIGHTS:- 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**(ReceivedVi

## ‘summarise()‘ regrouping output by ’ReceivedVia’ (override with ‘.groups‘ argument)

ccc<-comcast\_data **%>% filter**(ReceivedVia**==**’Customer Care Call’,ComplaintStatus**==**’Closed’) **%>% group\_by**

## ‘summarise()‘ regrouping output by ’ReceivedVia’ (override with ‘.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.