Comcast_telecom_complaint

Suprava Sahoo

26/06/2020

```
## Problem need to be solved
#Importing data into R environment.
#Provide the trend chart for the number of complaints at monthly and daily gr
anularity levels.
#Provide a table with the frequency of complaint types. -Which complaint type
s are maximum i.e., around internet, network issues, or across any other doma
ins.
#Create a new categorical variable with value as Open and Closed. Open & Pend
ing 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 cate
gorized variable from Q3. Provide insights on: -Which state has the maximum c
omplaints -Which state has the highest percentage of unresolved complaints
#Provide the percentage of complaints resolved till date, which were received
through the Internet and customer care calls.

# Including required Packages

library(stringi)
library(lubridate)

## Warning: package 'lubridate' was built under R version 4.0.2

##
## Attaching package: 'lubridate'

## The following objects are masked from 'package:base':
##
##     date, intersect, setdiff, union

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
```

```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.0.2
```

```r
#Loading Dataset:
comcast_data<- read.csv("d:/dataset/Comcast Telecom Complaints data.csv",head
er = TRUE)
#Manipulating column names
names(comcast_data)<- stri_replace_all(regex =  "\\.",replacement = "",str =n
ames(comcast_data))
head(comcast_data)
```

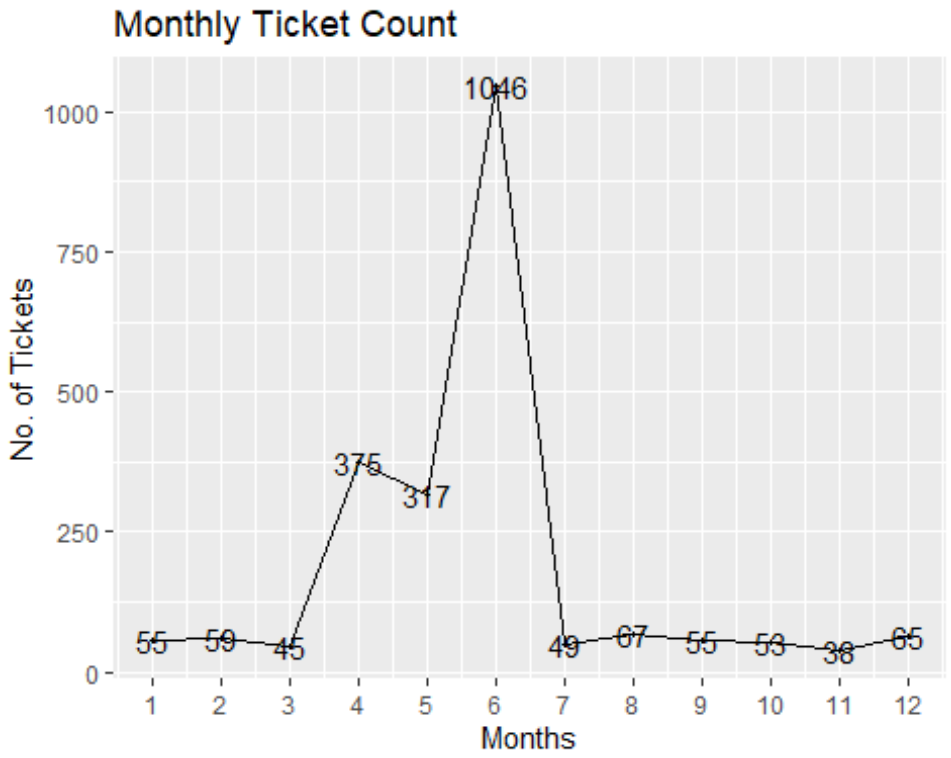
```
##   Ticket                                                CustomerComplaint
## 1 250635                                       Comcast Cable Internet Speeds
## 2 223441                         Payment disappear - service got disconnected
## 3 242732                                                Speed and Service
## 4 277946 Comcast Imposed a New Usage Cap of 300GB that punishes streaming.
## 5 307175                            Comcast not working and no service to boot
## 6 338519          ISP Charging for arbitrary data limits with overage fees
##         Date        Time        ReceivedVia     City    State Zipcode Stat
us
## 1 22-04-2015  3:53:50 PM Customer Care Call Abingdon Maryland   21009 Clos
ed
## 2   4/8/2015 10:22:56 AM           Internet  Acworth  Georgia   30102 Clos
ed
## 3 18-04-2015  9:55:47 AM           Internet  Acworth  Georgia   30101 Clos
ed
## 4   5/7/2015 11:59:35 AM           Internet  Acworth  Georgia   30101   Op
en
## 5 26-05-2015  1:25:26 PM           Internet  Acworth  Georgia   30101 Solv
ed
## 6  6/12/2015  9:59:40 PM           Internet  Acworth  Georgia   30101 Solv
ed
##   FilingonBehalfofSomeone
## 1                      No
## 2                      No
## 3                     Yes
## 4                     Yes
## 5                      No
## 6                      No
```

```r
#Finding NA`s in dataset:
na_vector <- is.na(comcast_data)
length(na_vector[na_vector==T])
```

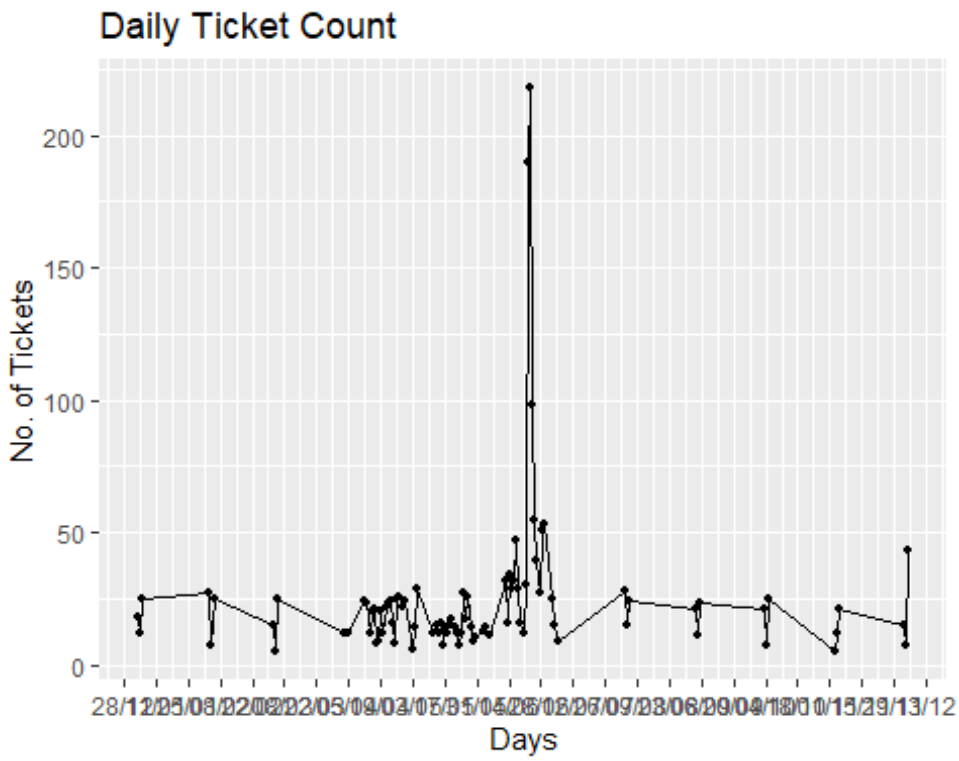
```
## [1] 0
```

```r
#Processing date:
comcast_data$Date<- dmy(comcast_data$Date)

#Extracting monthly and daily
```

```r
monthly_count<- summarise(group_by(comcast_data,Month =as.integer(month(Date)
)),Count = n())

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

daily_count<- summarise(group_by(comcast_data,Date),Count =n())

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

monthly_count<-arrange(monthly_count,Month)
#making count of monthly and daily complaints

#Comparing Monthly and daily Complaints
ggplot(data = monthly_count,aes(Month,Count,label = Count))+
  geom_line()+
  geom_point(size = 0.8)+
  geom_text()+
  scale_x_continuous(breaks = monthly_count$Month)+
  labs(title = "Monthly Ticket Count",x= "Months",y ="No. of Tickets")
```




```r
#As we can see there is a increases in tickets in the month of April,May and
this also increased in the month of June, so there might be some reseon due t
o that that they received high amount of tickets.
ggplot(data = daily_count,aes(as.POSIXct(Date),Count))+
  geom_line()+
  geom_point(size = 1)+
  scale_x_datetime(breaks = "2 weeks",date_labels = "%d/%m")+
  labs(title = "Daily Ticket Count",x= "Days",y ="No. of Tickets")
```


Daily Ticket Count
No. of Tickets
200
150
100
50
0
Days