DESCRIPTIVE ANALYITICS FOR COMCAST TELECOME

KRISHNAVENI RAJAN
201902960

SYMBIOSIS CENTRE FOR DISTANCE LEARNING (SCDL)
JULY 2019 – JUNE 2021

Table of Content

INTRODUCTION

The analysis request by the Comcast Telecom is to provide the insights of the complaints based on various factor with the data provided for the year of 2015. This is requires the "**Descriptive Analytics**" to be done to analyse the data from the historic data and provide the insights about what has happened in the past. This information of Descriptive Analytics could be used the company to come-up with the corrective actions, improvements required and identify the gap in the existing process. 'R' Program is used for this project.

Objective

- Analyse the Test data provided by the Comcast Telecom
- Perform Descriptive analytics
- Provide details insight to the Comcast Telecom

Scope and Background

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.

Comcast Telecom Requirement

Using the dataset, help to pin down what is wrong with Comcast's customer service.

1. Provide the trend chart for the number of complaints at monthly and daily granularity levels.
2. Provide a table with the frequency of complaint types.
3. Which complaint types are maximum i.e., around internet, network issues, or across any other domains.

4. 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?"
5. Which state has the highest percentage of unresolved complaints
6. Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

ANALYSIS OF WORK DONE AND DESIGN

Phases of Project Work

The project work has been done using the following phases

1. Understanding the Problem statement
2. Review the Dataset to understand the data provided
3. Identify the columns in the Dataset that requires Data Wrangling
4. Break the analysis required for the solution into smaller chunks
    a. Come up with the step-by-step activity to be done (without starting the coding)
    b. Check if the solution required can be generated using the Excel so this can be used for cross verification post R coding
5. Write the R code
    a. Execute the code and get results
    b. Ensure there is no errors
    c. Cross validate the results with the #4.b
    d. Confirm the solution for all the analysis is done. If not done repeat #5.
6. Complete the project and submit for Grading.

Data Dictionary

Data dictionary of the Comcast Telecom Complaints data.csv is as follows

- Ticket #: Ticket number assigned to each complaint
- Customer Complaint: Description of complaint
- Date: Date of complaint
- Time: Time of complaint
- Received Via: Mode of communication of the complaint

- City: Customer city
- State: Customer state
- Zipcode: Customer zip
- Status: Status of complaint
- Filing on behalf of someone : Confirm if the ticket is filled by the customer directly or being filed on on-behalf of others

Data Wrangling Performed

Date Format

Date column provided in the data set had different formats which required formatting all the column to single format. This was done with multiple steps

a. Replace the / with – in the Date column (dd/mm/yyyy → dd-mm-yyyy)
b. Format the data with dd-mm-yyyy format into yyyy-mm-dd format

Complaint Type

To get the "Complaint type", data cleaning was required to come up with the new categories based on the Keywords present in the description provided for "Customer Complaint" column.

Grouping & New Columns added

Following columns added to the data set to break the complexity of data and easy data access

Date Based Columns

a. Identify and add the column 'Monthly' in the format of MMM as new column based on the dates in the Date column
b. Identify and add the column 'Quarters' in the format of Q1, Q2, Q3 and Q4 based on the dates in the Date column
    1. Q1 – January, February, March
    2. Q2 – April, May, June
    3. Q3 – July, August, September
    4. Q4 – October, November, December

Ticket # Based Changes

Count of total complaints using the sum of Ticket # based on the following groupings

a. Date (Daily range)
b. Month (Monthly Trend)
c. Quarter
d. City (State wise)
e. By the complaints received source (Received Via)
f. By the complaint type based on the description provided for Customer Complaint (Grouping the complaints)

Status Based Changes

Values in the Status column was worked on to retain only Open and Closed status

a. Status with value 'Pending' was replaced with 'Open'
b. Status with value 'Solved' was replaced with 'Closed'

Source code

Source code of the project is as follows

```
#Libraries
library(tidyverse)
library(forecast)
library(ggplot2)
library(lubridate)
library(dplyr)
library(data.table)
library(scales)
library(base)
library(zoo)

#Load data
comcast <- read.csv("/Users/renu/Learnings/Data Science With R/Project/Projects for
Submission/Comcasr Telecom Consumer Complaints/Comcast Telecom Complaints data.csv")
View(comcast)
str(comcast)
```

```
names(comcast) [1] <- "TickNo" #Rename the Coulmn
View(comcast)
# Converting all the dates to same format
comcast$Date <- gsub("/", "-", comcast$Date) #Replace the / with - in the Date column
data1 <- data.frame(initialDiagnose = comcast$Date) #Assign the Date column data into
data1
data1 <- as.Date(comcast$Date, "%d-%m-%y") #Convert the data in the data1 to specific date
format
comcast$Date <- data1 #Update the comcast's Date column with data1
year(comcast$Date) <- 2015
summary(comcast)
View(comcast)
# Month Wise Data
comcast <- comcast %>% mutate(Months_Complaints = month(Date))
View(comcast)
comcast_monthly <- comcast %>% group_by(Months_Complaints) %>%
summarise(Total_Complaints = n())
comcast_monthly$Months_Complaints <- month.abb[comcast_monthly$Months_Complaints]
View(comcast_monthly)

comcast_monthly$Months_Complaints <-factor(comcast_monthly$Months_Complaints,
                    levels = c("Jan", "Feb", "Mar", "Apr", "May", "Jun",
                              "Jul", "Aug", "Sep", "Oct", "Nov", "Dec"))

#Plot for Monthly Trend
ggplot(comcast_monthly, aes(Months_Complaints,Total_Complaints)) +
  geom_point() +
  geom_bar(stat = "identity") +
  xlab("Months") +
  ylab("Total Complaints Raised")

# Daily Wise Data

comcast_daily <- comcast %>% group_by(Date) %>% summarise(Total_Complaints = n())
View(comcast_daily)
#Plot of data for daily trend
ggplot(comcast_daily, aes(Date,Total_Complaints)) +
  geom_point() +
  geom_line() +
  xlab("Date") +
  ylab("Total Complaints Raised") +
  scale_x_date(date_breaks = "1 month", date_minor_breaks = "1 week",
```

```
            date_labels = "%b-%y")

## Create Table with the Compliant Types & Frequency
#Load data
comcast_cleaned <- read.csv("/Users/renu/Learnings/Data Science With R/Project/Projects for
Submission/Comcasr Telecom Consumer Complaints/Comcast Telecom Complaints data
Cleaned.csv")
unique(comcast_cleaned$Customer.Complaint)

MaxTickets_Domain <- comcast_cleaned %>% group_by(Customer.Complaint) %>%
summarise(Ticket_Count =n()) %>% arrange(desc(Ticket_Count))
View(MaxTickets_Domain)
top_n(MaxTickets_Domain,5)


ggplot(MaxTickets_Domain, aes(Ticket_Count,Customer.Complaint)) +
  geom_point() +
  geom_line() +
  xlab("Total Complaints Raised") +
  ylab("Domain")


# State with Maximum Complaints
statewise <- comcast %>% group_by(City) %>% summarise(Total_Complaints = n()) %>%
arrange(desc(Total_Complaints))
View(statewise)
top_n(statewise,1)

statewise1 <- statewise
statewise1$Total_Complaints=cut(statewise1$Total_Complaints,c(0,5,10,15,20,25,30,35,40,45,
50,55,60,65,70))
grid::current.viewport()
ggplot(statewise1, aes(City,Total_Complaints, fill=Total_Complaints)) +
  geom_point() +
  geom_bar(stat = "identity", position="stack") +
  xlab("State") +
  ylab("Total_Complaints")

#State with Maximum Unresolved tickets for Q3
#Adding Quarters to the data
comcast_status_open$month <- factor(format(comcast_status_open$Date, format = "%b"),
levels = month.abb)
```

```
comcast_status_open$quarter <- character(length = NROW(comcast_status_open))
comcast_status_open$quarter[comcast_status_open$month %in% month.abb[c(1,2,3)]] <-
"Q1"
comcast_status_open$quarter[comcast_status_open$month %in% month.abb[c(4,5,6)]] <-
"Q2"
comcast_status_open$quarter[comcast_status_open$month %in% month.abb[c(7,8,9)]] <-
"Q3"
comcast_status_open$quarter[comcast_status_open$month %in% month.abb[c(10,11,12)]] <-
"Q4"
comcast_status_open$quarter <- factor(comcast_status_open$quarter, levels =
c("Q1","Q2","Q3","Q4"))

#Filter Data for Open and Q3
comcast_status1 <- filter(comcast_status_open, Status == "Open")
comcast_status2 <- filter(comcast_status1, quarter == "Q3")
MaxUnresolved_State_Q3 <- comcast_status2 %>% group_by(City) %>%
summarise(OpenStatus = n()) %>% arrange(desc(OpenStatus))
top_n(MaxUnresolved_State_Q3,1)

# Status update to Open and Close
unique(comcast$Status)
cstatus <- comcast$Status
cstatus <- gsub("Pending", "Open", comcast$Status) #Replace Status with value Pending to
Open
comcast_status <- comcast
comcast_status$Status <- cstatus
cstatus1 <- comcast_status$Status
cstatus1 <- gsub("Solved", "Closed", comcast_status$Status) #Replace Status with value Solved
to Closed
comcast_status$Status <- cstatus1
unique(comcast_status$Status)

#State with Maximum Unresolved tickets for given year
comcast_status_open <- filter(comcast_status, Status == "Open")
MaxUnresolved_State <- comcast_status_open %>% group_by(City) %>%
summarise(OpenStatus = n()) %>% arrange(desc(OpenStatus))
MaxUnresolved_State
top_n(MaxUnresolved_State,1)

# % of complaints received by Internet, Customer Calls and are Resolved Successfully
unique(comcast_status$Received.Via)
unique(comcast_status$Status)
```

```
comcast_status_closed <- filter(comcast_status, Status == "Closed")
comcast_status_closed <- comcast_status_closed %>% group_by(Received.Via) %>%
summarise(Resolved_Complaints = n())

Total_Complaints <- comcast %>% group_by(Received.Via) %>%
summarise(Total_Complaints = n())
Total_Complaints

comcast_total_resolved <- merge(Total_Complaints,comcast_status_closed)
comcast_total_resolved

comcast_resolved_percentage <- comcast_total_resolved %>%
mutate("ResolvedComplaints%"= (Resolved_Complaints/sum(Total_Complaints)*100))
roundof1 <- round(comcast_resolved_percentage$ResolvedComplaints, digits = 2)
comcast_resolved_percentage$`ResolvedComplaints%` <- roundof1
comcast_resolved_percentage
```

Snapshots of solution using r-code

1. Daily and Monthly Trends of Complaints

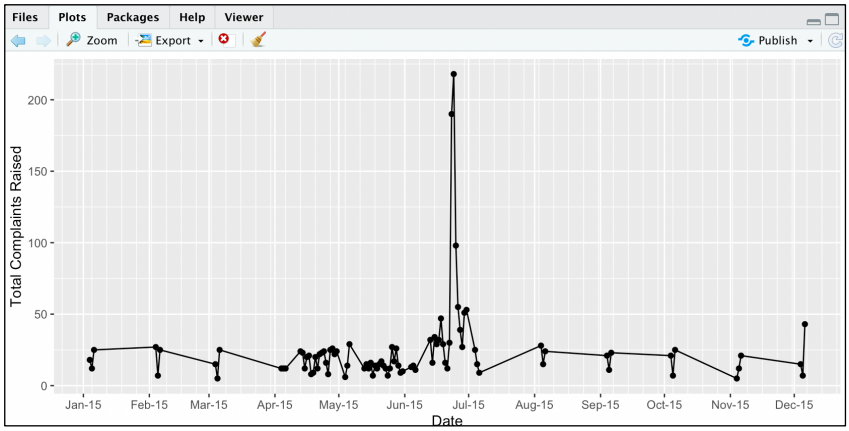


Figure 1 - Daily Ticket Trend

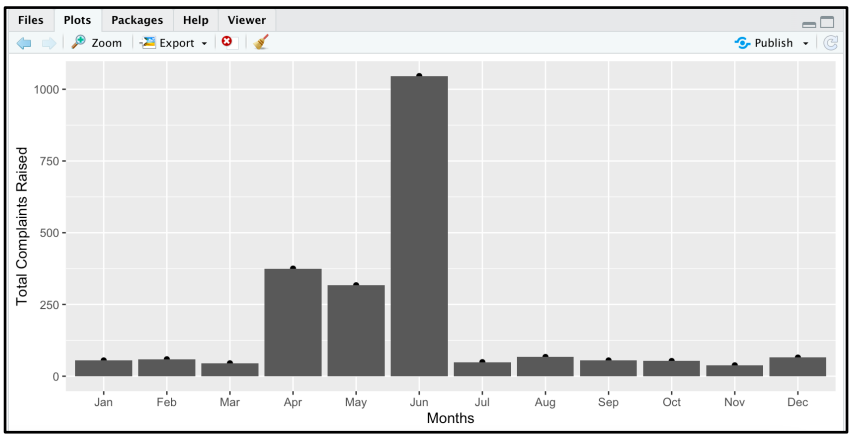


Figure 2 - Monthly Ticket Trend

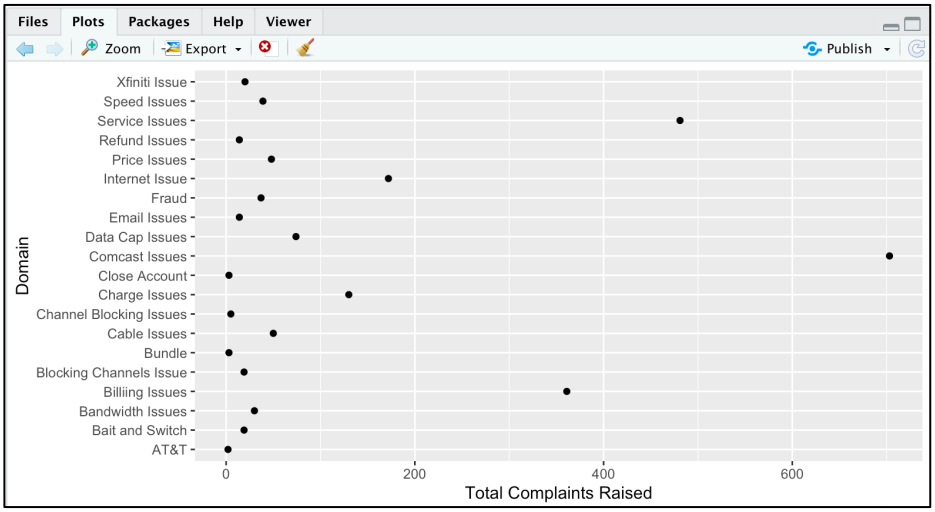
2. Table to represent the frequency of complaint types.

Go to file/function

data_ori | comcast | Comcast Telecome C

Filter

	Customer.Complaint	Ticket_Count
1	Comcast Issues	703
2	Service Issues	481
3	Billiing Issues	361
4	Internet Issue	172
5	Charge Issues	130
6	Data Cap Issues	74
7	Cable Issues	50
8	Price Issues	48
9	Speed Issues	39
10	Fraud	37
11	Bandwidth Issues	30
12	Xfiniti Issue	20
13	Blocking Channels Issue	19
14	Bait and Switch	19
15	Refund Issues	14
16	Email Issues	14
17	Channel Blocking Issues	5
18	Close Account	3
19	Bundle	3
20	AT&T	2



3. Top 5 maximum complaint types based domain

```
View(MaxTickets_Domain)
top_n(MaxTickets_Domain,5)

74:1  (Top Level)
Console  Terminal  Jobs
~/
> top_n(MaxTickets_Domain,5)
Selecting by Ticket_Count
# A tibble: 5 x 2
  Customer.Complaint Ticket_Count
  <chr>                     <int>
1 Comcast Issues              703
2 Service Issues              481
3 Billiing Issues             361
4 Internet Issue              172
5 Charge Issues               130
```

4. Maximum complaints based on 'State' for Q3

```
> MaxUnresolved_State_Q3 <- comcast_status2 %>% group_by(City) %>% summarise(OpenStatus = n()) %>% ar
range(desc(OpenStatus))
> top_n(MaxUnresolved_State_Q3,1)
Selecting by OpenStatus
# A tibble: 1 x 2
  City  OpenStatus
  <chr>      <int>
1 Miami          2
>
```

Figure 3 State with Maximum Open Tickets in Q3

5. Which state has the highest percentage of unresolved complaints

```
#State with Maximum Unresolved tickets for given year
comcast_status_open <- filter(comcast_status, Status == "Open")
MaxUnresolved_State <- comcast_status_open %>% group_by(City) %>% summarise(OpenStatus = n()
MaxUnresolved_State
top_n(MaxUnresolved_State,1)

# % of complaints received by Internet, Customer Calls and are Resolved Successfully

130:29  (Top Level)  R Script
Console  Terminal  Jobs
~/
# A tibble: 337 x 2
   City       OpenStatus
   <chr>           <int>
 1 Atlanta            20
 2 Knoxville          15
 3 Houston            13
 4 Miami               7
 5 Nashville           7
 6 Denver              6
 7 Tucson              6
 8 Baltimore           5
 9 Boca Raton          5
10 Savannah            5
#  with 327 more rows
```

6. *Percentage of complaints resolved based on source*

```
144 comcast_resolved_percentage <- comcast_total_resolved %>% mutate("ResolvedComplaints%"= (Resc
145 roundof1 <- round(comcast_resolved_percentage$ResolvedComplaints, digits = 2)
146 comcast_resolved_percentage$`ResolvedComplaints%` <- roundof1
147 comcast_resolved_percentage
148
149
148:1 (Top Level) R Script
Console Terminal Jobs
~/
> comcast_resolved_percentage <- comcast_total_resolved %>% mutate("ResolvedComplaints%"= (Resolved_C
omplaints/sum(Total_Complaints)*100))
> roundof1 <- round(comcast_resolved_percentage$ResolvedComplaints, digits = 2)
> comcast_resolved_percentage$`ResolvedComplaints%` <- roundof1
> comcast_resolved_percentage
        Received.Via Total_Complaints Resolved_Complaints ResolvedComplaints%
1 Customer Care Call             1119                 864               38.85
2           Internet             1105                 843               37.90
```

LEARNING EXPERIENCE ON BUSINESS

The project has helped in understanding about how the customer problem must be understood and interfered. The data set provided gave an clarity on how the real-time data collected by Secondary resources could be. The data in the columns did not have proper format and had multiple entries that required very deep data cleansing.

The project also helped in exploring the R program, different libraries available. I have learnt how to search of the required information from the provided data as well as in the internet and books to come up with the solution. I have learnt about the following functions and libraries.

Functions and Libraries Learnt

Different functions and libraries were used in the R code while coming up with the program to deliver the required results

a. Summarise from dplyr library was mainly used to come up with the solution to get the Total ticket counts, Maximum Counts etc.
b. Built-In libraries used are
    a. library(tidyverse)
    b. library(forecast)
    c. library(ggplot2)
    d. library(lubridate)
    e. library(dplyr)

f. library(data.table)
g. library(scales)
h. library(base)
i. library(zoo)

c. Built-in functions used include the following (not all are listed, few of the used are listed)
    a. Unique
    b. Filter
    c. As.date
    d. Gsub
    e. Gplot
    f. Top_n
    g. Round
    h. Group_by
    i. Month.abb
    j. Factor

CONCLUSION

With the data analysis done on the dataset the 'Descriptive Analytics' inference is as follows

1. Total tickets raised in the year 2015 was 2224. In the year of 2015, 63 tickets were raised by the Atlanta had the maximum ticket reported in total and total of 20 tickets were in "Open" Status.
2. June-2015 had the maximum number of complaints raised which was 1046.
3. 2015-06-24 was the day which had the maximum number of ticket raised and the count was 218.
4. The Top 5 category of the complaint type and their frequency was as follows
    a. Comcast Issues = 703
    b. Service Issues = 481
    c. Billing Issues = 361
    d. Internet Issue = 172
    e. Charge Issues = 130

5. In the Q3 of 2015, Miami had the maximum number tickets in “Open” Status which counted to 2.
6. Out of the total 2224 tickets raised, “Resolved %” of tickets was as follows
    a. Customer Care was 38.85%
    b. Internet was 37.90%

REFERENCES

- https://www.r-project.org/other-docs.html
- https://cran.r-project.org/manuals.html
- https://www.codecademy.com/catalog/language/r
- https://www.tutorialspoint.com/r/index.htm