DESCRIPTIVE ANALYITICS FOR COMCAST TELECOM

KRISHNAVENI RAJAN
201902960

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

Table of Content

INTRODUCTION

Data science is an interdisciplinary field that uses scientific methods, processes, algorithms and systems to extract knowledge and insights from structured and unstructured data, and apply knowledge and actionable insights from data across a broad range of application domains. Use case of Data Science is very vast and it has been helping out the world to get better day on day in various field like Health Care, Education, Entertainment, Retail & Banking.

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.

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

CASE PRESENTATION

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.

MANAGEMENT AND OUTCOME

The Case study 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

Comcast Telecom Complaints data

Ticket #	Customer Complaint	Date	Time	Received Via	City	State	Zip code	Status	Filing on Behalf of Someone
250635	Comcast Cable Internet Speeds	22-04-2015	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
223441	Payment disappear - service got disconnected	4/8/2015	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No
242732	Speed and Service	18-04-2015	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes
277946	Comcast Imposed a New Usage Cap of 300GB that punishes streaming.	5/7/2015	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes
307175	Comcast not working and no service to boot	26-05-2015	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No
338519	ISP Charging for arbitrary data limits with overage fees	6/12/2015	9:59:40 PM	Internet	Acworth	Georgia	30101	Solved	No
361148	Throttling service and unreasonable data caps	24-06-2015	10:13:55 AM	Customer Care Call	Acworth	Georgia	30101	Pending	No
359792	Comcast refuses to help troubleshoot and correct my service.	23-06-2015	6:56:14 PM	Internet	Adrian	Michigan	49221	Solved	No
318072	Comcast extended outages	6/1/2015	11:46:30 PM	Customer Care Call	Alameda	California	94502	Closed	No
371214	Comcast Raising Prices and Not Being Available To Ask Why	28-06-2015	6:46:31 PM	Customer Care Call	Alameda	California	94501	Open	Yes
255938	Billing after service was asked to be disconnected	24-04-2015	4:40:36 PM	Internet	Albuquerque	New Mexico	87106	Closed	No
276409	YAHOO FAILURE TO RESTORE EMAIL SEARCH FEATURE	5/6/2015	3:09:49 PM	Customer Care Call	Albuquerque	New Mexico	87109	Closed	No
339282	Comcast Violating Open Internet Rules by Blocking HBO Go Access on Sony Consoles	13-06-2015	4:03:18 PM	Internet	Albuquerque	New Mexico	87105	Open	Yes
360178	Internet speed	23-06-2015	9:23:23 PM	Internet	Albuquerque	New Mexico	87113	Solved	No
376268	Internet Disconnects Every Night	30-06-2015	10:30:02 PM	Customer Care Call	Albuquerque	New Mexico	87116	Solved	No
370137	Internet complaint	27-06-2015	3:25:03 PM	Customer Care Call	Albuquerqur	New Mexico	87102	Pending	No
363695	Internet Availability and Speed	24-06-2015	11:47:33 PM	Customer Care Call	Alexandria	Indiana	46001	Solved	No
238694	Comcast owes me $65 and claims I need to return equipment I never had	16-04-2015	10:04:57 AM	Internet	Alexandria	Virginia	22304	Closed	No
230876	Horrible Internet Service	4/11/2015	7:48:05 PM	Customer Care Call	Alexandria	Virginia	22305	Closed	No
318725	Failure to provide services that I am billed for.	6/2/2015	1:03:52 PM	Customer Care Call	Alexandria	Virginia	22314	Closed	No
327657	Internet out all the time but they have a monopoly	6/7/2015	8:55:43 PM	Customer Care Call	Alexandria	Virginia	22305	Solved	No
328742	horrible cable service and customer service	6/8/2015	3:18:58 PM	Internet	Alexandria	Virginia	22312	Solved	No
328165	Speed	6/8/2015	12:03:37 PM	Customer Care Call	Alexandria	Virginia	22304	Solved	No
370538	Comcast monopoly bundling practices	27-06-2015	9:04:34 PM	Internet	Alexandria	Virginia	22304	Open	No
370363	COMCAST!	27-06-2015	6:37:29 PM	Internet	Alexandria	Virginia	22305	Solved	No
270163	bait and switch	5/2/2015	3:55:24 PM	Internet	Algonquin	Illinois	60102	Open	Yes
355976	Comcast Customer Service	22-06-2015	2:10:39 PM	Internet	Aliquippa	Pennsylvania	15001	Open	No

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’

Snapshots of solution using r-code

1. *Daily and Monthly Trends of Complaints*

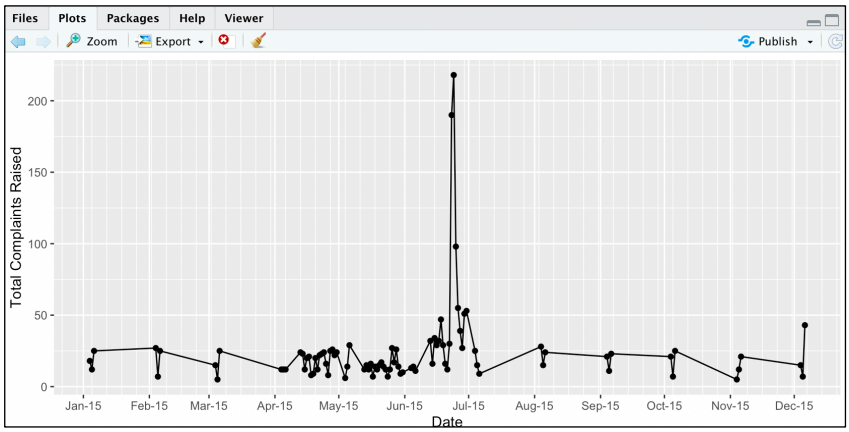


Figure 1 - Daily Ticket Trend

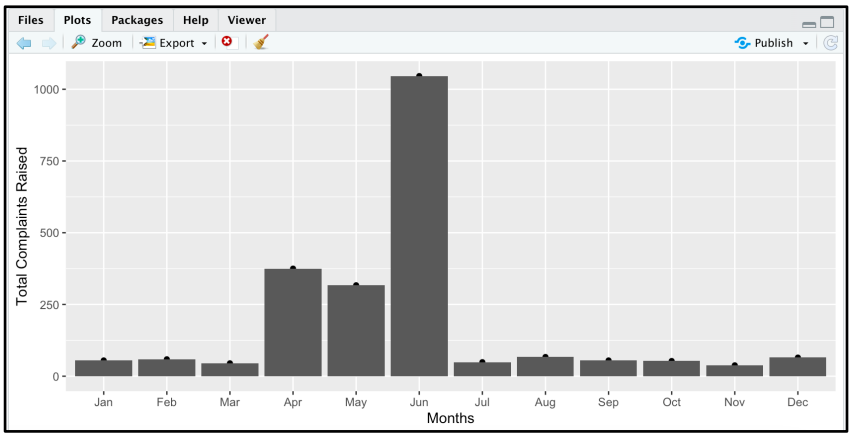
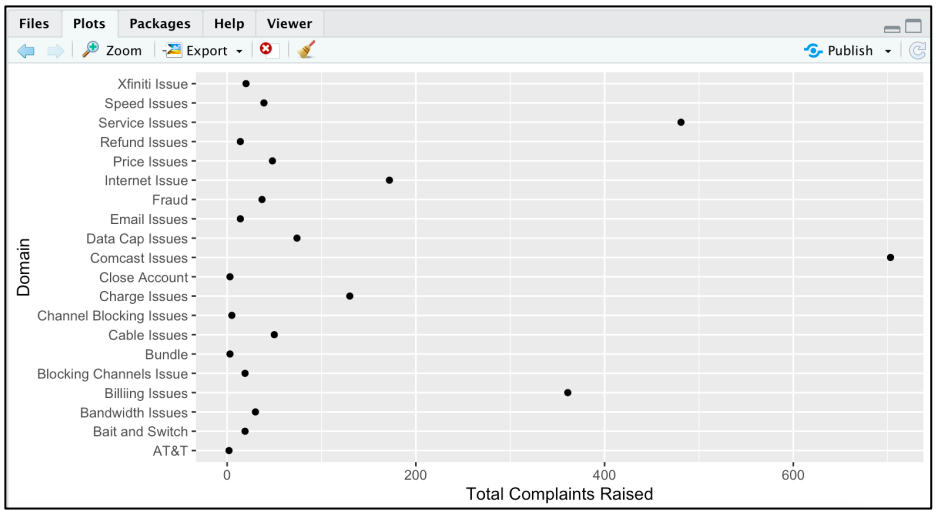
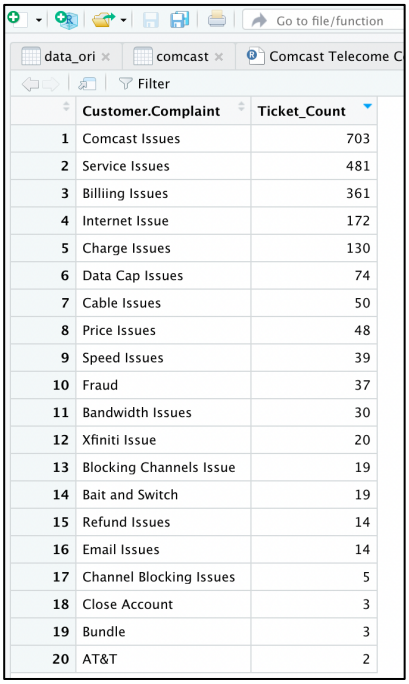


Figure 2 - Monthly Ticket Trend

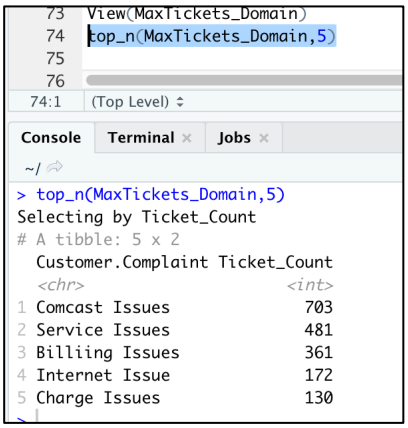
2. *Table to represent the frequency of complaint types.*





	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



```
73  View(MaxTickets_Domain)
74  top_n(MaxTickets_Domain,5)
75
76
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*

```
126 #State with Maximum Unresolved tickets for given year
127 comcast_status_open <- filter(comcast_status, Status == "Open")
128 MaxUnresolved_State <- comcast_status_open %>% group_by(City) %>% summarise(OpenStatus = n())
129 MaxUnresolved_State
130 top_n(MaxUnresolved_State,1)
131
132 # % of complaints received by Internet, Customer Calls and are Resolved Successfully
133
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

SUMMARY

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%

RECOMMENDATION TO COMCAST TELECOM

With the descriptive analysis done through the case study, it is recommended to the Comcast Telecom that the company should concentrate more on the common issues on the highest priority. They should understand the emotion of the customer who calls for the Complaints and also handle the calls in strategic way to address the problems in short span of time.