Comcast EDA REPORT - By Srijan Gupta

Date - 14/03/2020

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.

Data Dictionary

1. Ticket #: Ticket number assigned to each complaint
2. Customer Complaint: Description of complaint
3. Date: Date of complaint
4. Time: Time of complaint
5. Received Via: Mode of communication of the complaint
6. City: Customer city
7. State: Customer state
8. Zipcode: Customer zip
9. Status: Status of complaint
10. Filing on behalf of someone

Tasks Performed Are:-

1. **Import data into the R environment**

We will be Loading the Data into the R environment by using readXl library, after which by making the data ready for the analysis and so as to form conclusions where Comcast is lacking in serving their customers.

Code : -

```
library(readxl)
Data <- read.csv("../input/Comcast Telecom Complaints data.csv")
head(Data,4) #Printing Few Rows Of Data
str(Data) #checking the structure of the dataset.
```

	Ticket..	Customer.Complaint	Date	Time	Received.Via	City	State	Zip.code	Status	Filing.on.Behalf.of.Someone
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<int>	<fct>	<fct>
1	250635	Comcast Cable Internet Speeds	22-04-2015	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No
2	223441	Payment disappear - service got disconnected	4/8/2015	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No
3	242732	Speed and Service	18-04-2015	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes
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	Yes

2. **Provide the trend chart for the number of complaints at monthly and daily granularity levels.**

Now, to perform this analysis we will extract the month from the given “Date” column and create a new column called “Month” where we store all the months when complaints are filed. So, before extracting month we need to put the dates in the “Date” column in the correct format. By using the Lubridate library we will transform all the dates in a particular format which will make it easy for us to extract month and make analysis later.

Code to clean the “Date” column and put in the same format:-

```
#Loading The Date Into Single Format
#Use Lubridate Library to Format the Date Column
library(lubridate)
li<-parse_date_time(x = Data$Date,
                    orders = c("d m y", "d B Y", "m/d/y"),
```

```
                    locale = Sys.getlocale("LC_TIME"))
data2<-Data
data2$Date <- li
#Dates Loaded In the Same Format in the new Dataframe
#str(data2$Date)

#Extracting Month Column and Converting to The labels.
data2$Month <- format(as.Date(data2$Date), "%m")
data2$Month<- month.abb[as.integer(data2$Month)]
head(data2)
```

Which give you a data Frame With Column “Month” and cleans up the “Date” Column

	Ticket..	Customer.Complaint	Date	Time	Received.Via	City	State	Zip.code	Status	Filing.on.Behalf.of.Someone	Month
	<fct>	<fct>	<dttm>	<fct>	<fct>	<fct>	<fct>	<int>	<fct>	<fct>	<chr>
1	250635	Comcast Cable Internet Speeds	2015-04-22	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	Apr
2	223441	Payment disappear - service got disconnected	2015-08-04	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	Aug
3	242732	Speed and Service	2015-04-18	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	Apr
4	277946	Comcast Imposed a New Usage Cap of 300GB that punishes streaming.	2015-07-05	11:59:35 AM	Internet	Acworth	Georgia	30101	Open	Yes	Jul
5	307175	Comcast not working and no service to boot	2015-05-26	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	May
6	338519	ISP Charging for arbitrary data limits with	2015-12-06	9:59:40 PM	Internet	Acworth	Georgia	30101	Solved	No	Dec

So, Now To Plot Trend Charts We Group Date and Complaints Column, to plot it on daily granularity level.

Code:-

```
library(dplyr)
data_date<-data2 %>% group_by(Date) %>% dplyr::summarise(frequency = n())
df <-data_date[order(-data_date$frequency),]
dff<-head(df)
dff
```

The Few Rows of DataFrame Showing On which Date how many numbers of Complaints was Filed.

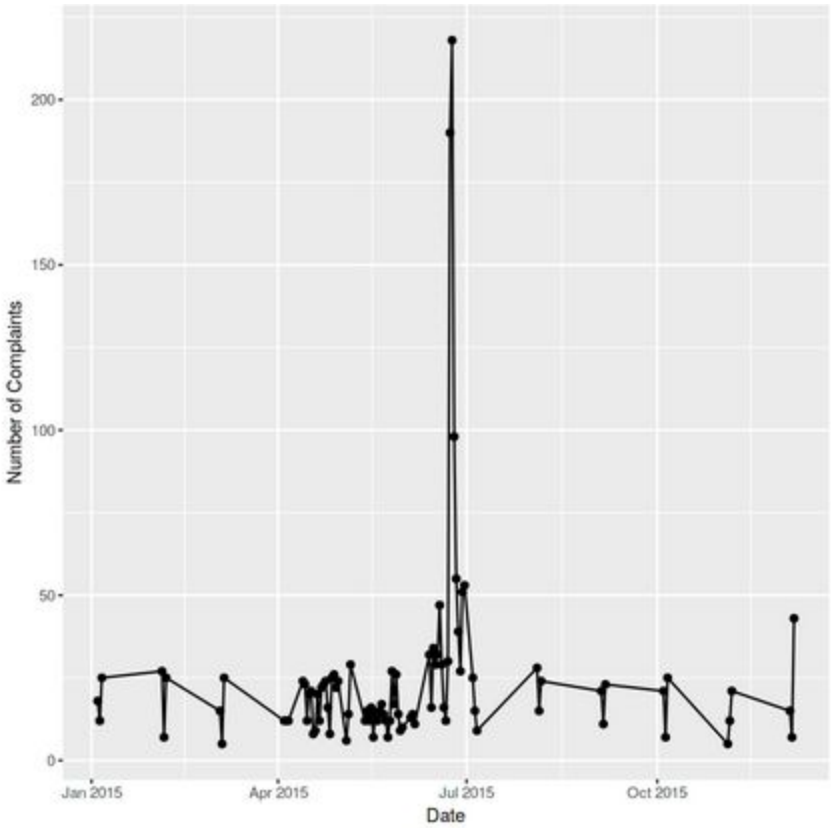
Date	frequency
<dttm>	<int>
2015-06-24	218
2015-06-23	190
2015-06-25	98
2015-06-26	55
2015-06-30	53
2015-06-29	51

Analysis:- This Above Data Frame Tells me that, on June 24 Comcast was reported with 218 complaints, particularly indicating a doomsday for the company to fix up many issues popping up on that day.

Let's Plot them to get better visualizations -

Code -

```
library(ggplot2)
ggplot(data_date, aes(Date, frequency, group = 1)) +
  geom_point() +
  geom_line() +
  xlab("Date") +
  ylab("Number of Complaints")
```

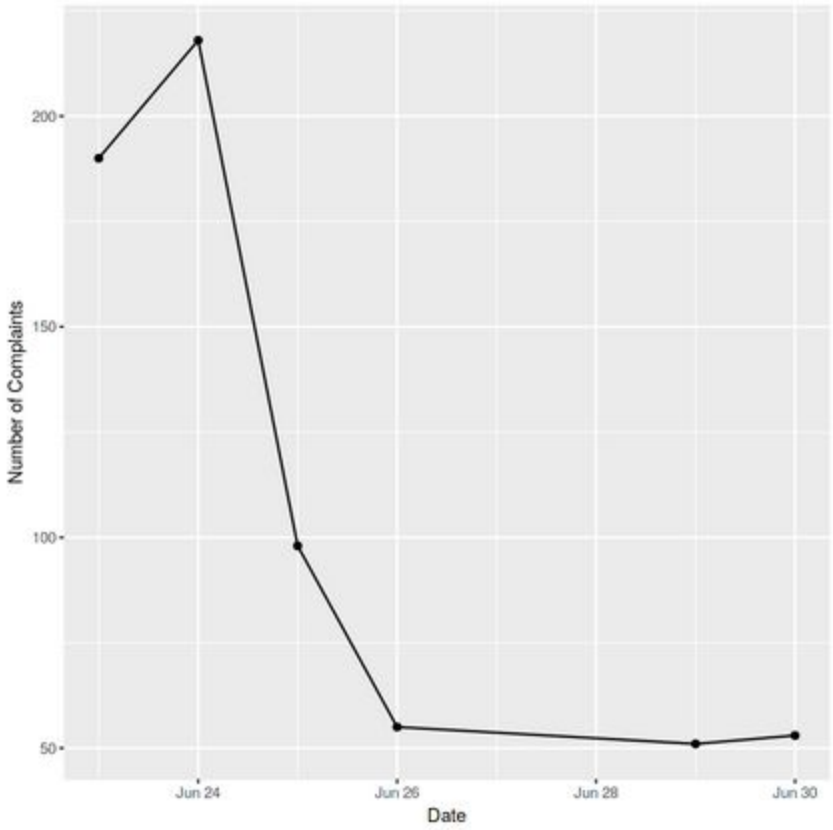


Analysis:- Clearly, from the above Trend Graph, we can easily say that in the month of JUNE 2015, Comcast got reported with Maximum Number of complaints.

Let's Plot the DateWise/ DailyWise Trend Graph:-

Code:-

```
library(ggplot2)
ggplot(dff, aes(Date, frequency, group = 1)) +
  geom_point() +
  geom_line() +
  xlab("Date") +
  ylab("Number of Complaints")
```



Analysis:- Clearly you can see on June 24, Company got reported with many complaints. This is the following trend for a few observations from the month of June

Let's Plot Trend Graphs for Month Granularity Level:-

Code -

```
data_month<-data2 %>%
group_by(Month) %>% dplyr :: summarise(frequency = n())
data_month
```

Month	frequency
<chr>	<int>
Apr	375
Aug	67
Dec	65
Feb	59
Jan	55
Jul	49
Jun	1046
Mar	45
May	317
Nov	38
Oct	53
Sep	55

To Print What Kind Of Levels, We are dealing with, which is very obvious -
Code -

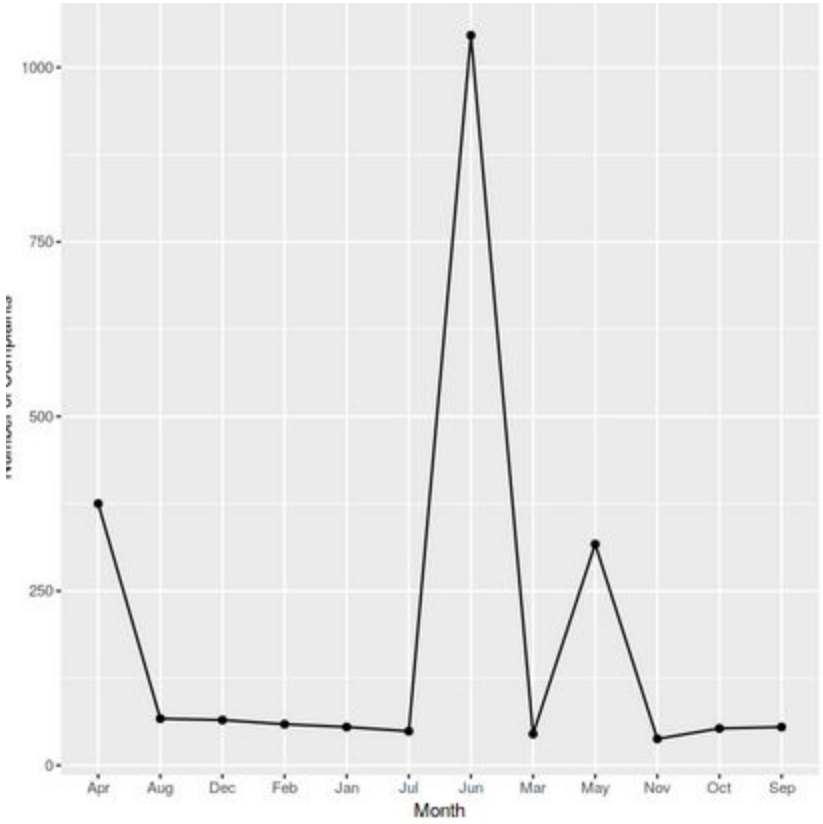
```
data2$Month <- as.factor(data2$Month)
levels(data2$Month)
```

'Apr' · 'Aug' · 'Dec' · 'Feb' · 'Jan' · 'Jul' · 'Jun' · 'Mar' · 'May' · 'Nov' · 'Oct' · 'Sep'

The Trend Plot -
Code -

```
library(ggplot2)
ggplot(data_month, aes(Month, frequency, group = 1)) +
  geom_point() +
  geom_line() +
  xlab("Month") +
  ylab("Number of Complaints")

#You only have to add group = 1 into the ggplot or geom_line ases().
#For line graphs, the data points must be grouped so that it knows which points to connect.
#In this case, it is simple -- all points should be connected, so group=1.
```

#When more variables are used and multiple lines are drawn, the grouping for lines is usually done by variable.



By, Now we have quite a good amount of insights by which we can clearly say that in the Month of June, date 24 company was filed with a maximum number of complaints.

3. Frequency Table For Customer Complaints During the Year 2015 - 2016 Period

Let's Fetch the Frequency Table for Complaints, To Know Which Kind of Issue is rising Up more to Provide more staff and Infrastructure to improve the customer service.

Code -

```
library(dplyr)
#Converting All String Values to Lower, so as to Eliminate Duplication of Any Complaint
data3<-data2%>% mutate(Customer.Complaint = tolower(Customer.Complaint))
CustTable <- table(data3$Customer.Complaint)
CustTable <- data.frame(CustTable)
filtered<-CustTable %>%
  rename(
        CustomerComplaintType = Var1,
        Frequency = Freq
        )
final <- filtered %>% arrange(desc(Frequency))

#Fetching The Top 20 complaints filed by customers on different days.
final_most<-head(final,20)
final_most
```

DataFrame -

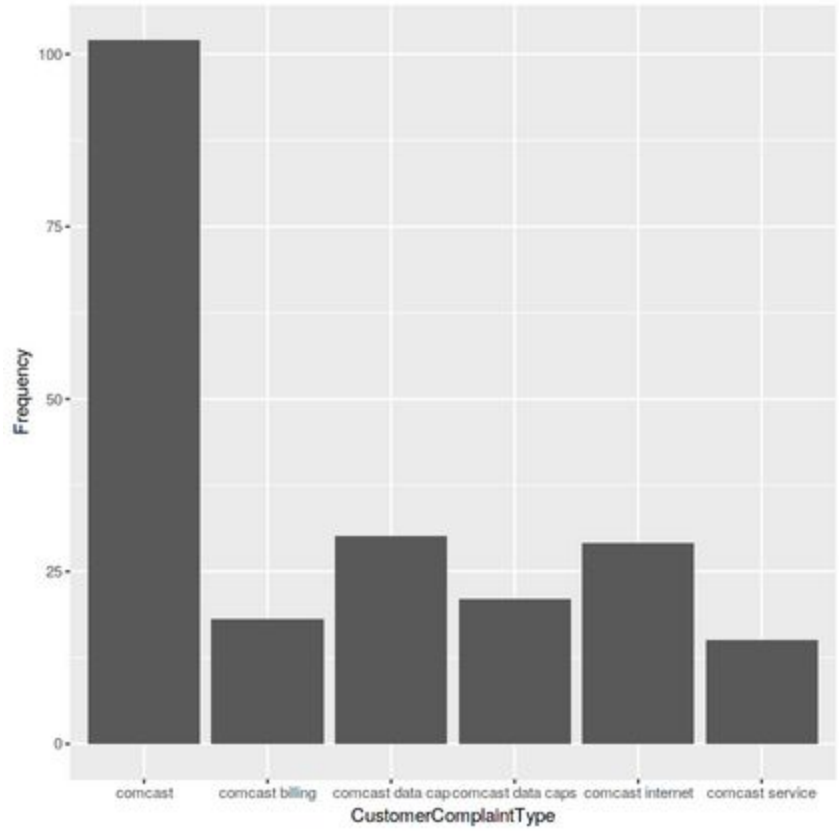
	CustomerComplaintType	Frequency
	<fct>	<int>
1	comcast	102
2	comcast data cap	30
3	comcast internet	29
4	comcast data caps	21
5	comcast billing	18
6	comcast service	15
7	internet speed	15
8	data caps	13
9	unfair billing practices	13
10	data cap	12
11	comcast complaint	11
12	comcast/xfinity	11
13	comcast internet service	10
14	billing	9
15	billing issues	8

Plotting the Above Data -

Code -

```
library(ggplot2)
ggplot(head(final_most,6), aes(CustomerComplaintType, Frequency)) +
  geom_bar(stat = "identity")
```

Plot -



Analysis:- Looking at the Above Table And Bar Plot:-

Customer Are Mainly complaining about the Data Caps, Internet Speed, Billing Methods and Services that Comcast is Providing and Very few Cases were registered against Comcast Cable Services.

4. **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.**

Code-

```
library(plyr)
Data$Status_New<-revalue(Data$Status, c(Pending = "Open", Solved = "Closed"))
head(Data)
```

DataFrame -

A data.frame: 6 × 11

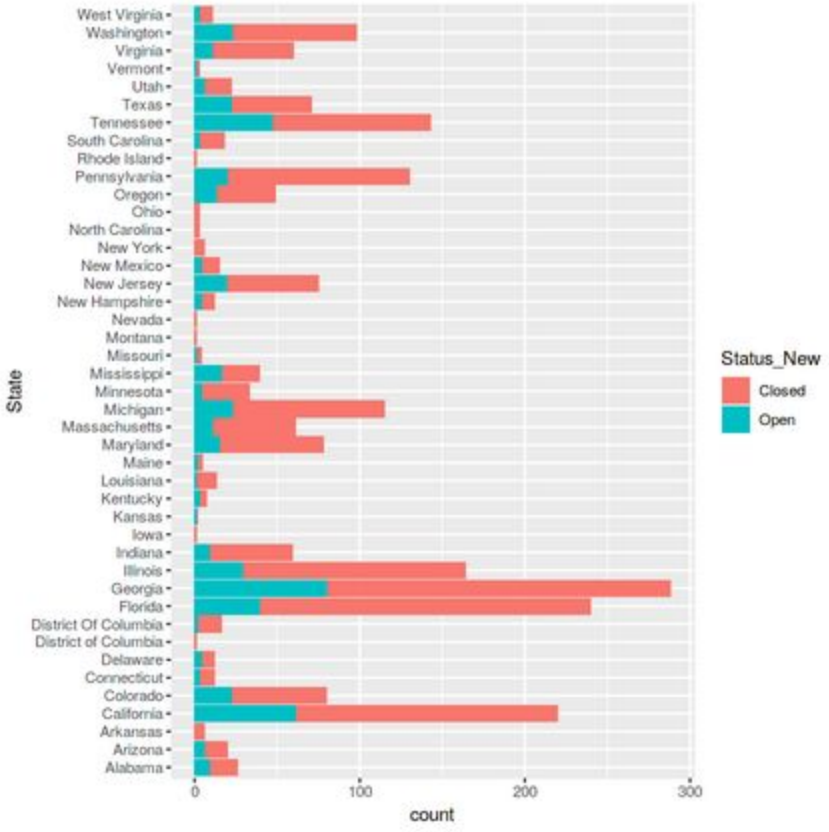
	Ticket..	Customer.Complaint	Date	Time	Received.Via	City	State	Zip.code	Status	Filing.on.Behalf.of.Someone	Status_New
	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<fct>	<int>	<fct>	<fct>	<fct>
1	250635	Comcast Cable Internet Speeds	22-04-2015	3:53:50 PM	Customer Care Call	Abingdon	Maryland	21009	Closed	No	Closed
2	223441	Payment disappear - service got disconnected	4/8/2015	10:22:56 AM	Internet	Acworth	Georgia	30102	Closed	No	Closed
3	242732	Speed and Service	18-04-2015	9:55:47 AM	Internet	Acworth	Georgia	30101	Closed	Yes	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	Yes	Open
5	307175	Comcast not working and no service to boot	26-05-2015	1:25:26 PM	Internet	Acworth	Georgia	30101	Solved	No	Closed
6	338519	ISP Charging for arbitrary data limits with overage fees	6/12/2015	9:59:40 PM	Internet	Acworth	Georgia	30101	Solved	No	Closed

We Can see that in the New Column as Status_New is created, now we will perform further analysis using this new Column.

CODE- To Plot Stacked Bar chart

```
library(gridExtra)
ggplot(Data, aes(y = State)) + geom_bar(aes(fill = Status_New))
```

Plot -



Analysis:- Clearly Looking at the chart we can say, Georgia and Florida are the Two where Comcast has a good number of Happy customers by solving the issues in bulk.

5. Percentage of Resolved Issues Till Date Received Via either Calls or Internet.

Code -

```
df1 <- table(Data$Received.Via, Data$Status_New)
df1 <- cbind(df1, Total = rowSums(df1))
Df1
```

A matrix: 2 × 3 of type dbl

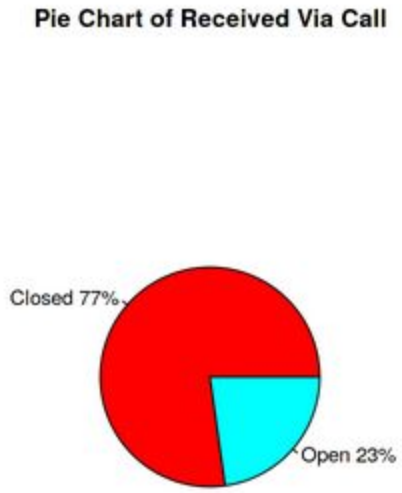
	Closed	Open	Total
Customer Care Call	864	255	1119
Internet	843	262	1105

Let’s Plot this Data To get a Clear picture of this -

Code -

```
# Pie Chart with Percentages

slices <- c(864, 255)

lbls <- c("Closed", "Open")

pct <- round(slices/sum(slices)*100)

lbls <- paste(lbls, pct) # add percents to labels

lbls <- paste(lbls,"%",sep="") # ad % to labels

pie(slices,labels = lbls, col=rainbow(length(lbls)),

   main="Pie Chart of Received Via Call")
```

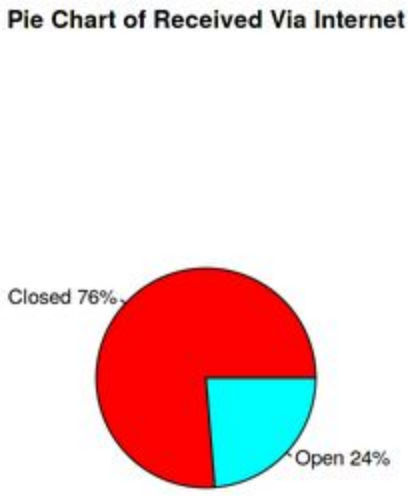
Plot _ For Calls _



Plot for via Internet -

```
# Pie Chart with Percentages
slices <- c(843, 262)
lbls <- c("Closed", "Open")
pct <- round(slices/sum(slices)*100)
lbls <- paste(lbls, pct) # add percents to labels
lbls <- paste(lbls,"%",sep="") # ad % to labels
pie(slices,labels = lbls, col=rainbow(length(lbls)),
   main="Pie Chart of Received Via Internet")
```

plot_ for_via_Internet



Solutions -

1. The Company should Focus more on resolving complaints - Customer Are Mainly complaining about the Data Caps, Internet Speed, Billing Methods and Services that Comcast is Providing and Very few Cases were registered against Comcast Cable Services.
2. In Georgia and Florida company services are already Improving but, in States - California, Colorado and Illinois company should extend their resources in terms of the above-mentioned issues in order to improve their customer servicing.
3. During the month of June and the start of July, the Company reported lots of complaints, so as to for future reference they can keep this in check already so as to provide better services during these months. While working with their BPO clients to extend the staff during such days. Which ensures proper feedback for the particular arisen issue.