

# Comcast Telecom Consumer Complaints. Project 2

Swati

20/01/2021

- Comcast Telecom Consumer Complaints-Project 2
  - **Brief Description of the Project**
  - **Objective 1:Import data into R environment**
  - **Objective 2:Provide the trend chart for the number of complaints at monthly and daily granularity levels**
  - **Objective3:Provide a table with the frequency of complaint types**
  - **Objective4:Provide state wise status of complaints in a stacked bar chart**
  - **Objective5:Which state has the highest percentage of unresolved complaints**
  - **Objective 6:Provide the percentage of complaints resolved till date**

## Comcast Telecom Consumer Complaints-Project 2

### Brief Description of the Project

#### DESCRIPTION

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

## Objective 1:Import data into R environment

```
telecom <- read.csv("F:/data science/R/Datasets/Project_Simplilearn/Comcast Telecom Consumer Complaints/Comcast T  
elecom Consumer Complaints/Comcast Telecom Complaints data.csv",  
stringsAsFactors = FALSE )
```

## Objective 2:Provide the trend chart for the number of complaints at monthly and daily granularity levels

```
head(telecom, 5)
```

```
## Ticket.. Customer.Complaint  
## 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  
## Date Time Received.Via City State Zip.code Status  
## 1 22-04-2015 3:53:50 PM Customer Care Call Abingdon Maryland 21009 Closed  
## 2 4/8/2015 10:22:56 AM Internet Acworth Georgia 30102 Closed  
## 3 18-04-2015 9:55:47 AM Internet Acworth Georgia 30101 Closed
```

```
## 4 5/7/2015 11:59:35 AM Internet Acworth Georgia 30101 Open
## 5 26-05-2015 1:25:26 PM Internet Acworth Georgia 30101 Solved
## Filing.on.Behalf.of.Someone
## 1 No
## 2 No
## 3 Yes
## 4 Yes
## 5 No
```

Since attribute Date is not in a single format(4/08/2015 and 26-05-2015), it is required to present the Date Column in one single format for better understanding.

### formatting date

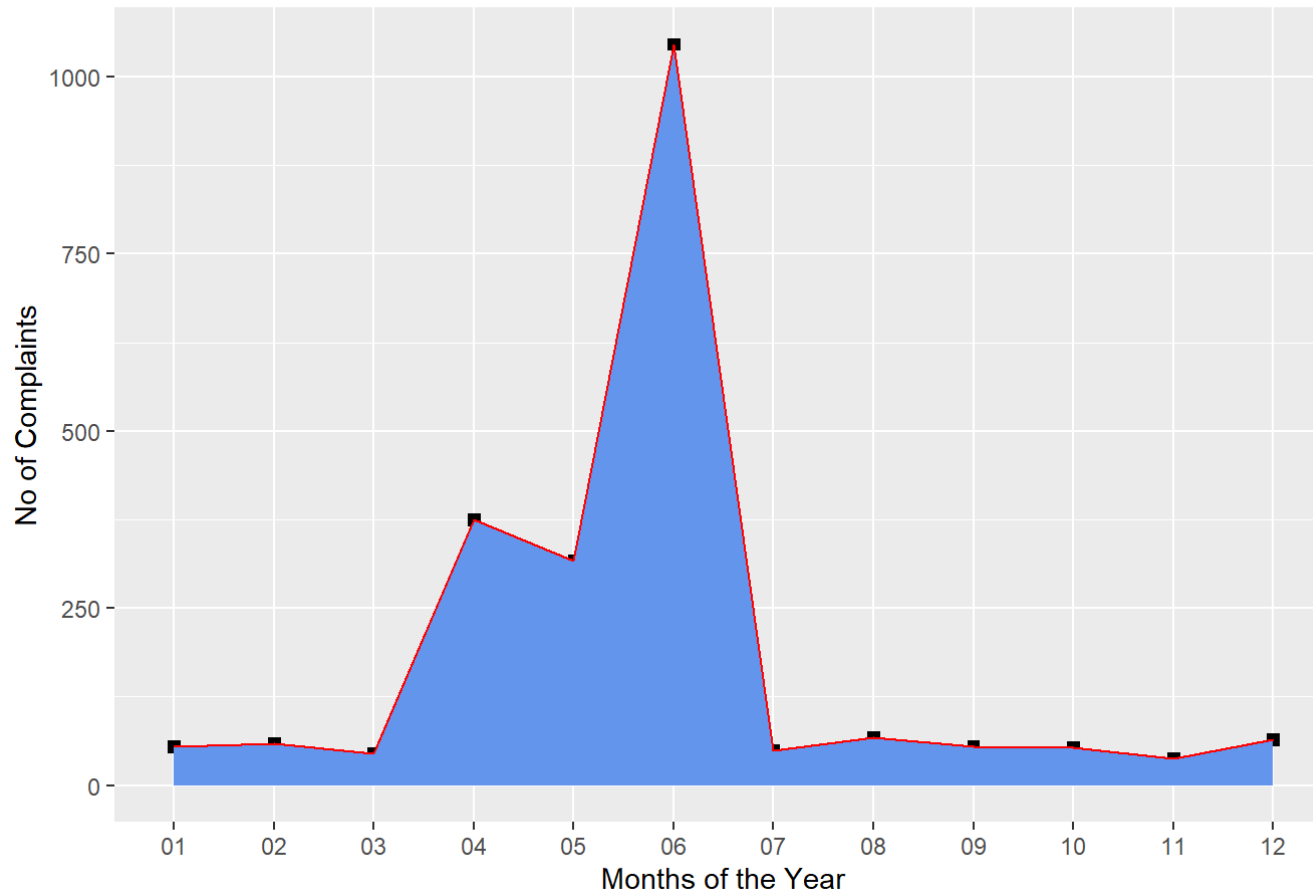
```
date_format1 <- strptime(telecom$Date, format = "%d-%m-%Y")
date_format2 <- strptime(telecom$Date, format = "%d/%m/%Y" )
date_format1[is.na(date_format1)] <- date_format2[!is.na(date_format2)]
date_format1 <- as.Date(date_format1)

telecom$Date <- date_format1
```

```
months <- format(telecom$Date, "%m")
data=data.frame(table(months))

monthlytrend <- ggplot(data=data , aes(x=months, y=Freq, group=1))+
  geom_path(color = "red")+
  ggtitle("Monthly Trend Of Customer Complaints")+geom_point(size=2, shape=22, fill="black")+geom_area(color="red", fill="cornflowerblue")+xlab("Months of the Year")+
  ylab("No of Complaints")
monthlytrend
```

### Monthly Trend Of Customer Complaints



According to Monthly Trend Of Customer Complaints Graph, maximum complaints were received in the months April, May and June. The financial year starts from April and a lot of pending works are completed during this period and hence more utilization of Network bandwidth so we can infer that this could be the reason of maximum complaints registered during this period.

```
day <- format(telecom$Date, "%a")
table(day)
```

```
## day
## Fri Mon Sat Sun Thu Tue Wed
```

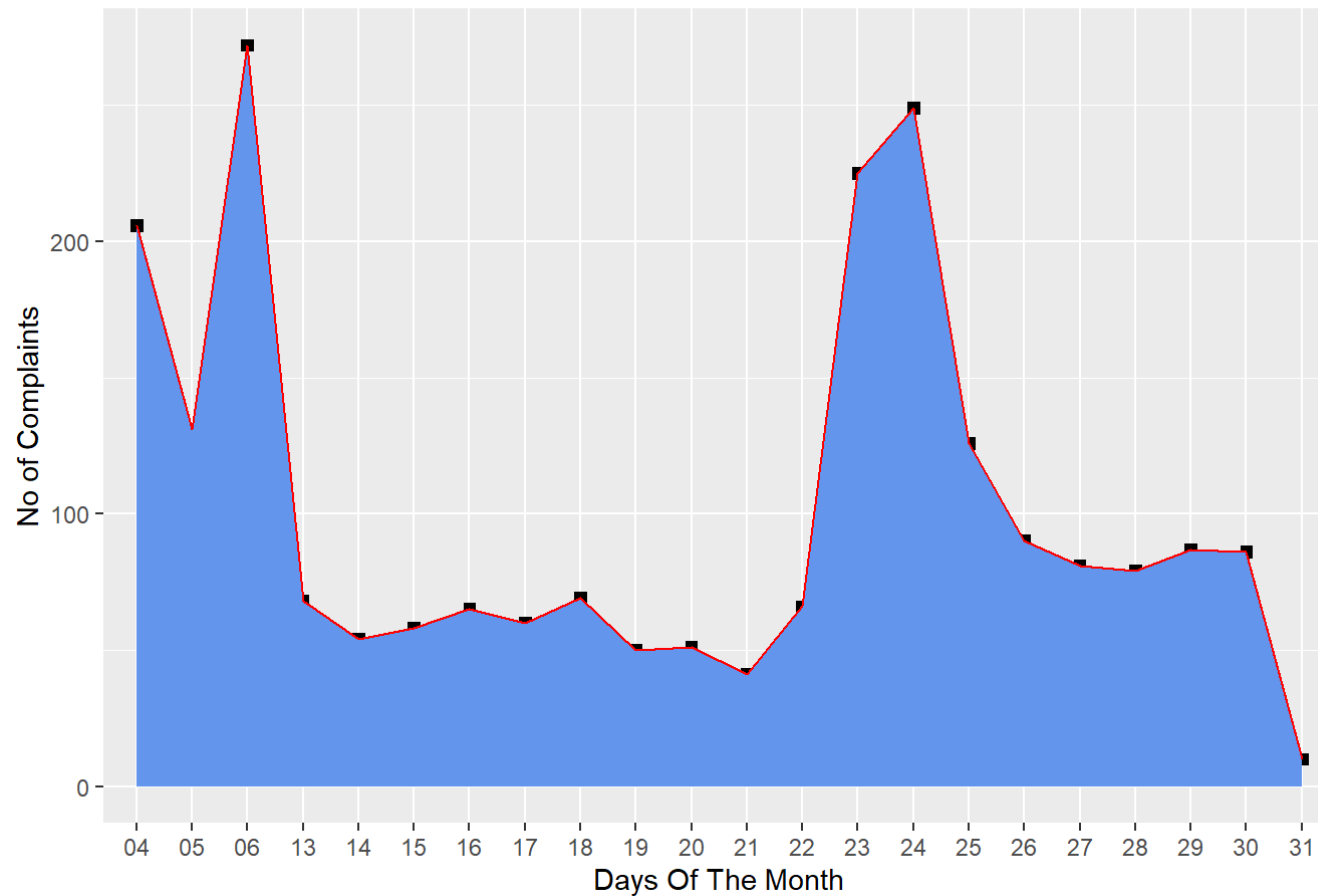
```
## 290 256 214 228 331 464 441
```

There is one more trend which can be noted here i.e, most of the complaints are registered on Tuesday and Wednesday and least on Saturday which is the family day of the week. However the trend is not so prominent.

```
date <- format(telecom$Date, "%d")
x=data.frame(table(date))

daytrend <- ggplot(data=x , aes(x=date, y=Freq, group=5))+
  geom_path(color = "red")+
  ggtitle("Day Wise Trend Of Customer Complaints")+geom_point(size=2, shape=22, fill="black")+geom_area(color="red", fill="cornflowerblue")+xlab("Days Of The Month")+
  ylab("No of Complaints")
daytrend
```

Day Wise Trend Of Customer Complaints



From the above graph we can depict that maximum number of complaints are registered in the starting(4th,5th,6th of the months) and end(23rd and 24th of the months)of the months.

## Objective3:Provide a table with the frequency of complaint types

It is very important to figure out the major issues for which the complaints are being registered in order to improve Comcast Overall Performance and Market Image.

The registered complaints are not divided in specific categories howsoever the nature of the Customer.Complaint column is available in Random, i.e for the service issue the registered complaints are “Throttling service and unreasonable data caps”, “Horrible Internet Service”, “Comcast not working and no service to boot” etc. It is required to divide Complaints in major categories like Service, Internet and Billing. There could be other categories as well howsoever majority of the complaints are registered around these issues.

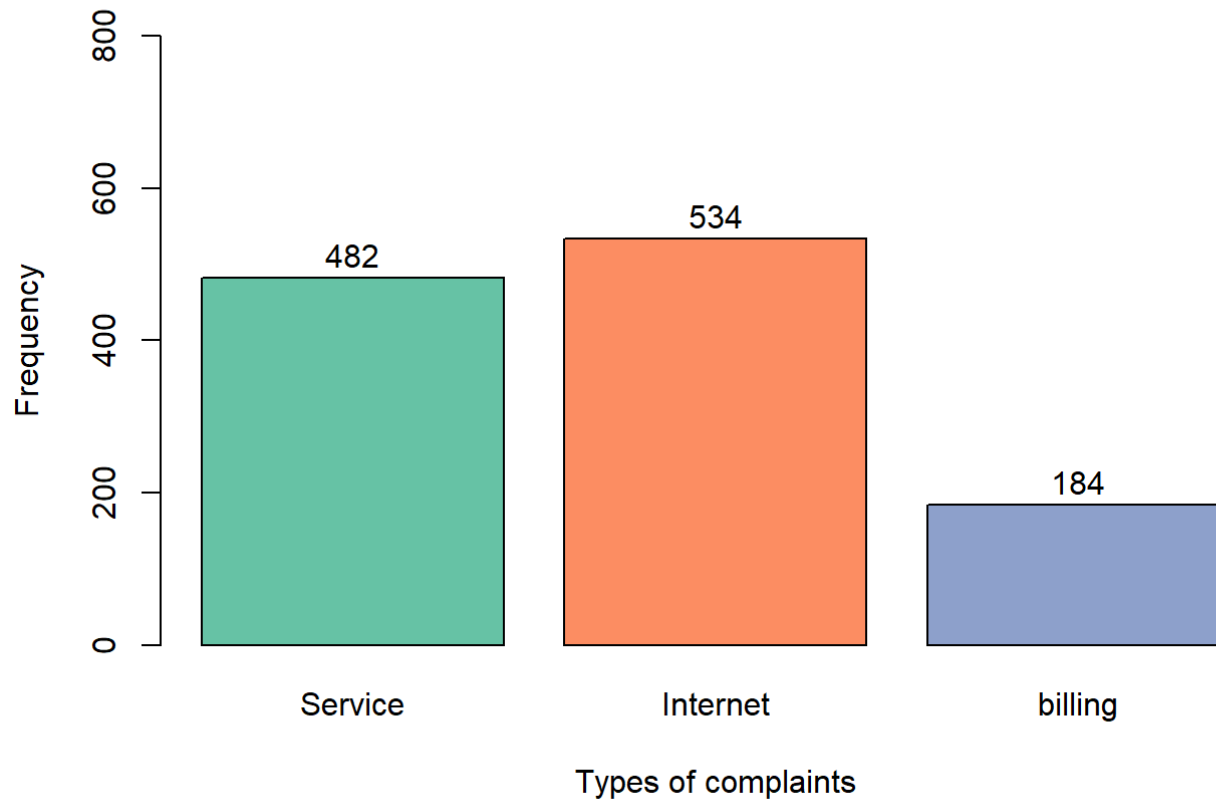
```
Service_Issue <- grep("ervice", telecom$Customer.Complaint, ignore.case = FALSE)
Internet_Issue <- grep("net", telecom$Customer.Complaint, ignore.case = FALSE)
Overbilling_Issue <- grep("bill", telecom$Customer.Complaint, ignore.case = FALSE)

complaint_types <- c(length(Service_Issue), length(Internet_Issue), length(Overbilling_Issue))
table(complaint_types)
```

```
## complaint_types
## 184 482 534
##    1    1    1
```

Here 1200 of the total complaints belong to 3 categories i.e, Service, Internet and Billing.

```
complaint.types <- barplot(complaint_types,
  col = brewer.pal(3, "Set2"),
  xlab = "Types of complaints",
  ylab = "Frequency",
  names.arg = c("Service", "Internet", "billing"),
  ylim = c(0, 800))
text(x = complaint.types, y = complaint_types + 30, labels = complaint_types)
```



## Objective4:Provide state wise status of complaints in a stacked bar chart

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.

```
telecom[telecom$Status=="Solved", "Status"] <- "Closed"  
telecom[telecom$Status=="Pending", "Status"] <- "Open"
```



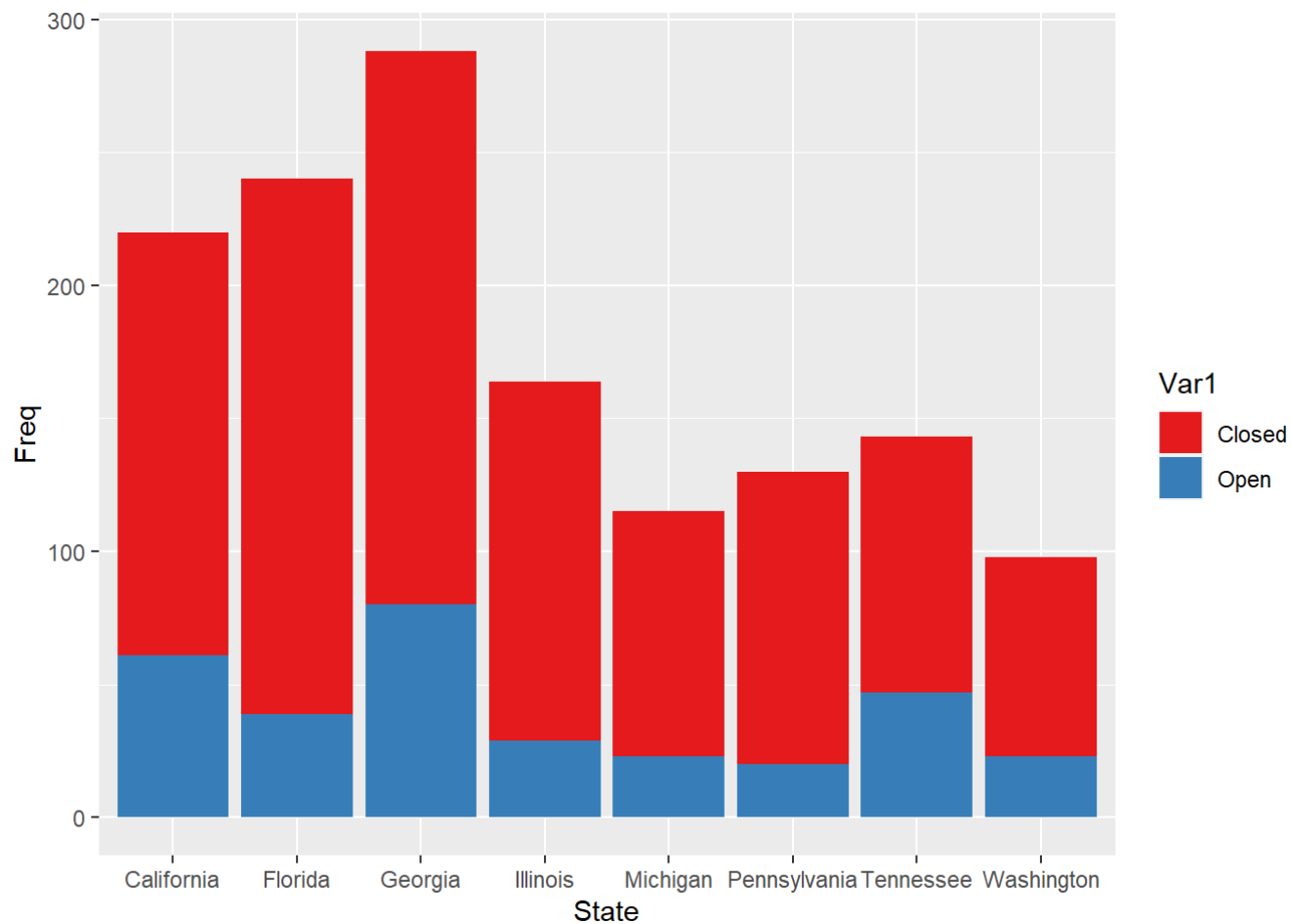
```
Status.State <-table(telecom$Status, telecom$State)
Status.State
```

```
##
##      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
## Closed       75          8
## Open        23          3
```

There are states with huge number of complaints. On the contrary there are states with very few registered complaints. Let us see the states from which huge number of complaints are coming

```
m<-data.frame(table(telecom$State))
m1<-m %>% filter(Freq>80)
colnames(m1)[colnames(m1)=="Var1"] <- "State"
colnames(m1)[colnames(m1)=="Freq"] <- "Total"
Status.State.df <-data.frame(Status.State)
colnames(Status.State.df)[colnames(Status.State.df)=="Var2"] <- "State"
data_set <- merge(m1, Status.State.df,by = "State" , all.x = T)

ggplot(data = data_set, mapping = aes(x=State, y=Freq, fill=Var1)) +geom_col()+scale_fill_brewer(palette ="Set1"
)
```



There are many states for which the total number of Open cases are less than 20, the company needs to solve these issues on priority. On the other hand there are countries with huge number of open complaints.

```
State_OpenCases <- filter(data.frame(Status.State), Var1=="Open" & Freq > 20)
State_OpenCases
```

```
##   Var1    Var2 Freq
## 1 Open California  61
## 2 Open  Colorado  22
```

```
## 3 Open    Florida    39
## 4 Open    Georgia    80
## 5 Open    Illinois   29
## 6 Open    Michigan   23
## 7 Open    Tennessee  47
## 8 Open    Texas      22
## 9 Open    Washington 23
```

There are total 9 States with huge number of complaints.

```
State_name <- State_OpenCases$Var2[State_OpenCases$Freq == max(State_OpenCases$Freq)]
Max_Value  <- max(State_OpenCases$Freq)
cat("The state with maximum number of open cases is", as.character(State_name), "with total", Max_Value, "open cases")
```

```
## The state with maximum number of open cases is Georgia with total 80 open cases
```

## Objective5:Which state has the highest percentage of unresolved complaints

```
State_OpenCases["Percentage"] <- (State_OpenCases$Freq/sum(State_OpenCases$Freq))*100
State_Percentage <- State_OpenCases$Var2[State_OpenCases$Percentage == max(State_OpenCases$Percentage)]
Max_Percentage <- max(State_OpenCases$Percentage)

cat(as.character(State_Percentage), "is the state with maximum", Max_Percentage, "% of total open cases")
```

```
## Georgia is the state with maximum 23.12139 % of total open cases
```

## Objective 6:Provide the percentage of complaints resolved till date

```
Closed_CustomerCare_percent<-(sum(telecom$Status=="Closed" & telecom$Received.Via=="Customer Care Call")/sum(telecom$Status=="Closed"))*100
```

```
Closed_Internet_percent <-(sum(telecom$Status=="Closed" & telecom$Received.Via=="Internet")/sum(telecom$Status=="Closed"))*100
```

```
cat("The total percentage of resolved complaint via Customer Care Call is ", Closed_CustomerCare_percent,"%","\n",  
,"The total percentage of resolved complaints vis Internet is ", Closed_Internet_percent)
```

```
## The total percentage of resolved complaint via Customer Care Call is  50.61511 %  
## The total percentage of resolved complaints vis Internet is  49.38489
```