CODE:

```
#Author: U. SAI RAVI TEJA
#Set the working directory
setwd(choose.dir())

#Import the dataset
dataset=read.csv('D:/Simplilearn/DS with R/Comcast Telecom Complaints data.csv')

#include necessary libraries
library(ggplot2)
library(dplyr)

#Extract date column and preprocess
dates=dataset$Date
class(dates)

#Getting the date format correctly
dates=gsub("/","-",dates)

#Transform to Date type
dates=as.Date(dates,"%d-%m-%Y")

#Verify the orignal date with transformed date
head(dataset$Date)
dates[1:6]
```

```
#Updating the date column with actual date
dataset$Date=dates

#Unique days from the dataset
mydates=as.factor(dataset$Date)

#filtering records per day
dataset$Date=as.character(dataset$Date)
perday =dataset %>% group_by(dataset['Date']) %>% tally()
perday=as.data.frame(perday)
perday$complaints=perday$n

#removing n column and keeping complaints column
perday=perday[,-2]

#permonth spilts
library(lubridate)
dataset$Date=dates
permonth =dataset %>% group_by(month=floor_date(Date, "month")) %>% tally()
permonth=as.data.frame(permonth)
permonth$complaints=permonth$n
permonth=permonth[,-2]

#plot the trend of complaints on monthly basis
ggplot(data = permonth, aes(x = month, y = complaints)) +
  geom_bar(stat = "identity",fill="red")+
  labs(x = "Month",
       y = "No.of complaints",
       title = "Trend chart for a monthly complaints",
       subtitle = "Year 2015")

#MAX per Month complaints
max(permonth$complaints)

#plot the trend of complaints on daily basis
perday$Date=as.Date(perday$Date)
class(perday$d)
ggplot(data = perday, aes(x =as.POSIXct(Date), y = complaints)) +
  geom_line(color="red")+geom_point(size = 0.5)+
  theme(axis.text.x = element_text(angle = 90))+
  labs(x = "Day",
       y = "No.of complaints",
       title = "Trend chart for a Daily complaints",
       subtitle = "Year 2015") +scale_x_datetime(breaks = "1 weeks",date_labels = "%d/%m")


#MAX Per day complaints
max(perday$complaints)


#creating a new categorical variable
```

```
dataset$TicketStatus[which(dataset["Status"]=="Open" |
dataset["Status"]=="Pending")]="Open"

dataset$TicketStatus[which(dataset["Status"]=="Closed" |
dataset["Status"]=="Solved")]="Closed"


#Adding a col based on complaint types

network_issue<- contains(dataset$Customer.Complaint,match =
'network',ignore.case = T)

internet_issue<- contains(dataset$Customer.Complaint,match =
'internet',ignore.case = T)

dataset$ComplaintType[internet_issue]= "Internet"

dataset$ComplaintType[network_issue]="Network"

dataset$ComplaintType[-c(network_issue,internet_issue)]="Other Domains"


#create a table based on frequency of comp tyypes

FreqOfCompTypes=table(dataset$ComplaintType)


#States spilt

ByState =dataset %>% group_by(dataset['State'],dataset["TicketStatus"]) %>%
tally()


#state wise complaints in stacked bar chart

ggplot(ByState,aes(fill=TicketStatus,y=n,x=State))+geom_bar(position =
"stack",stat = "identity",width = 0.8)+

    labs(title = "State wise complaints with status") +

    theme(axis.text.x = element_text(angle = 90)) +

    ylab(label = "Complaints") +

    scale_fill_manual("legend", values = c("Closed" = "green", "Open"="red"))
```

```
#State with max complaints
df=as.data.frame(dataset %>% group_by(dataset['State']) %>% tally())
state_withMaxComplaints=df$State[which(df$n==max(df$n))]

#state with high unresolved tickets
unresolvedStates=filter(ByState,TicketStatus=="Open")
state_withHighestUnresolved=unresolvedStates$State[which(unresolvedStates$n==max(unresolvedStates$n))]

#create required Objects by Filtering
levels(factor(dataset$Received.Via))
FilterData=as.data.frame(dataset %>% group_by(dataset['Received.Via'],dataset['TicketStatus']) %>% tally())
filter(FilterData,TicketStatus=="Closed")
comp_by=as.data.frame(dataset %>% group_by(dataset['Received.Via']) %>% count())

#percentage of complaints resolved through calls
total1=comp_by$n[which(comp_by['Received.Via']=="Customer Care Call")]
closedTotal1=FilterData$n[which(FilterData['Received.Via']=="Customer Care Call" & FilterData['TicketStatus']=='Closed' )]
percentage_call=closedTotal1/total1*100

#percentage of complaints resolved through Internet
total2=comp_by$n[which(comp_by['Received.Via']=="Internet")]
closedTotal2=FilterData$n[which(FilterData['Received.Via']=="Internet" & FilterData['TicketStatus']=='Closed' )]
```

```
percentage_Internet=closedTotal2/total2*100
```