```
FILENAME REFFILE '/home/u43532469/Project/Comcast_telecom_complaints_data.csv';

PROC IMPORT DATAFILE=REFFILE /*importing the dataset from retail analysis_dataset from file location*/
    DBMS=CSV
    OUT=WORK.Comcast;
    GETNAMES=YES;
RUN;

PROC CONTENTS DATA=WORK.Comcast; RUN;


%web_open_table(WORK.Comcast);

libname mylib '/home/u43532469/Project';

data mylib.comdata; /*to create a library comdata under mylib*/
set comcast;
format Date_month_year mmddyy10.; /*Converting datatype from Date9 to mmddyy10*/
run;

Proc Means Data=mylib.comdata;   /*Performing descriptive statistics on the dataset*/
Run;

Proc Means Data=mylib.comdata N Mean Std Min Max Median Mode p25 p75;  /*Performing median mode and quantile operations on the dataset*/
Run;

/*Finding frequency statistics of Date_month_year column*/
Proc freq Data=mylib.comdata  order=freq;
Tables Date_month_year;
run;

/*Creating ComplaintsByDate table with the frequency values of Date_month_year column*/
proc summary Data=mylib.comdata nway order=freq;
   class Date_month_year;
   output out=mylib.ComplaintsByDate(drop=_type_ rename=(_freq_=ComlaintsFreq));
run;

/*Creating new column named month_name in ComplaintsByMonth table*/
data mylib.ComplaintsByMonth;
set mylib.ComplaintsByDate;
month_name=put(Date_month_year, monname.);
run;

/*Daily Trends of Complaints*/
 ods graphics on / width=16in;
 ods graphics on / height=6in;
PROC SGPLOT data=mylib.ComplaintsByDate;
 VBAR Date_month_year / RESPONSE = ComlaintsFreq;
 TITLE 'Daily Trends of Complaints';
RUN;
ods graphics off;
/* We found that 23rd and 24th June 2015 has most number of complaints reported */

/*Monthly Trends of Complaints*/
PROC SGPLOT data=mylib.ComplaintsByMonth;
 VBAR month_name / RESPONSE = ComlaintsFreq;
 TITLE 'Monthly Trends of Complaints';
RUN;
/* We found that June 2015 has most number of complaints reported */

/*Frequency of complaint types*/
/*Creating Word dictionary from complaint types column*/
data mylib.complaintsType / view=mylib.complaintsType;
length word $12;
set mylib.comdata;
do i = 1 by 1 until(missing(word));
    word = upcase(scan('Customer Complaint'n, i));
    if not missing(word) then output;
    end;
keep word;
run;
/*Creating complaintsWordsFreq table with the frequency of used words in Complaint types Column*/
proc summary Data=mylib.complaintsType nway order=freq;
   class word;
   output out=mylib.complaintsWordsFreq(drop=_type_ rename=(_freq_=ComlaintsFreq));
run;
/* Assigning unique values to Complaint types column to make it categorical */
data mylib.comdataupdated;
   set mylib.comdata;
   if find('Customer Complaint'n, 'Internet', 'i') then 'Customer Complaint'n='Internet Issues';
   else if find('Customer Complaint'n, 'Speed', 'i') then 'Customer Complaint'n='Internet Issues';
   else if find('Customer Complaint'n, 'Data', 'i') then 'Customer Complaint'n='Internet Issues';
   else if find('Customer Complaint'n, 'Service', 'i') then 'Customer Complaint'n='Service issues';
   else if find('Customer Complaint'n, 'Customer', 'i') then 'Customer Complaint'n='Service issues';
   else if find('Customer Complaint'n, 'Billing', 'i') then 'Customer Complaint'n='Billing Issues';
   else if find('Customer Complaint'n, 'Bill', 'i') then 'Customer Complaint'n='Billing Issues';
   else if find('Customer Complaint'n, 'Charges', 'i') then 'Customer Complaint'n='Billing Issues';
   else 'Customer Complaint'n='Other Issues';
run;
/* we have divided all Customer complaints into 4 types Internet issues, Billing issues, Service Issues, and Other Issues */
/* finding most reported complaints */
Proc freq Data=mylib.comdataupdated  order=freq;   /*to find the frequency statistics of the dataset*/
Tables 'Customer Complaint'n;
run;
/*We found that the Internet issues are most*/

/*Converting "Status" column data to categorical data with Open and closed value*/
data mylib.comdatastatus;
   set mylib.comdata;
   if find(Status, 'Pending', 'i') then Status='Open';
   else if find(Status, 'Solved', 'i') then Status='Closed';
```

```
run;

Proc freq Data=mylib.comdatastatus  order=freq;
Tables Status;
run;

/*to find the Maximum Number of complaints State-wise*/
Proc freq Data=mylib.comdatastatus  order=freq;
Tables State;
run;
/*We found that the Georgia State Reported 288 issues which is highest among other States*/

/*To find the state having the highest percentage of unresolved complaints  */
Proc tabulate  Data=mylib.comdatastatus  order=freq;
class State;
class Status;
Tables Status*( colpctn n), State ;
run;
/*We found that the Kansas State has highest unresolved issues percentage of 50% which is highest among other States*/

/*To find out  the percentage of complaints resolved till date, which were received through the Internet and customer care calls. */
Proc freq Data=mylib.comdatastatus  order=freq;
Tables 'Received Via'n;
run;

/* we have found that all issues are reported via Internet and customer care calls only, So we will find the resolved issues with the total issues */
Proc tabulate  Data=mylib.comdatastatus  order=freq;
class 'Received Via'n;
class Status;
Tables Status, (all='All Status' 'Received Via'n)*( colpctn);
run;
/*We found that the total percentage of Closed issues is 76.75% which were received through the Internet and customer care calls. */
```