Data Science with SAS

Comcast Telecom Consumer Complaints

To Import the required dataset I follow the following steps:

- **Using the import data Utility in SAS Studio. It can be used to import Excel, csv, and other types of files in to SAS**
- **Choose the Comcast analysis project (Project 1) and checked the data**
- **Copy paste the file in to my folders file in SAS university edition and uploaded the file into my SAS environment**
- **Opened the SAS Studio and use the Import Data Utility in SAS Studio**
- **We can choose the drag and drop or select a file to import option. I used the select a file to import by navigating the files and folders and import the file into SAS for analysis.**
- **I specified the library and data table name**
- **SAS Automatically generates the Import Data. Within the import Data Tab, the default is the split view. We can see the settings and the code at the same time.**

SAS Code:

```
FILENAME REFFILE '/folders/myfolders/sasuser.v94/Comcast_telecom_complaints_data.csv';

PROC IMPORT DATAFILE=REFFILE
   DBMS=CSV
   OUT= comcast ;
   GETNAMES=YES;
```

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

SAS Code:

```
data comcast_1;

set comcast;
```

```
complain_year=year(Date_month_year);
complain_month=month(Date_month_year);
complain_ym=catx("-", of  complain_year complain_month);
run;


proc freq data= comcast_1;
tables complain_year ;
run;
```

Screenshot:

complain_year	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2015	2224	100.00	2224	100.00

- All records are from the year 2015

SAS Code:

```
proc freq data= comcast_1;
tables complain_month ;
run;
```

Screenshot:

complain_month	Frequency	Percent	Cumulative Frequency	Cumulative Percent
1	55	2.47	55	2.47
2	59	2.65	114	5.13
3	45	2.02	159	7.15
4	375	16.86	534	24.01
5	317	14.25	851	38.26
6	1046	47.03	1897	85.30
7	49	2.20	1946	87.50
8	67	3.01	2013	90.51
9	55	2.47	2068	92.99
10	53	2.38	2121	95.37
11	38	1.71	2159	97.08
12	65	2.92	2224	100.00

SAS Code:

```
proc freq data= comcast_1;

tables complain_ym;

run;
```

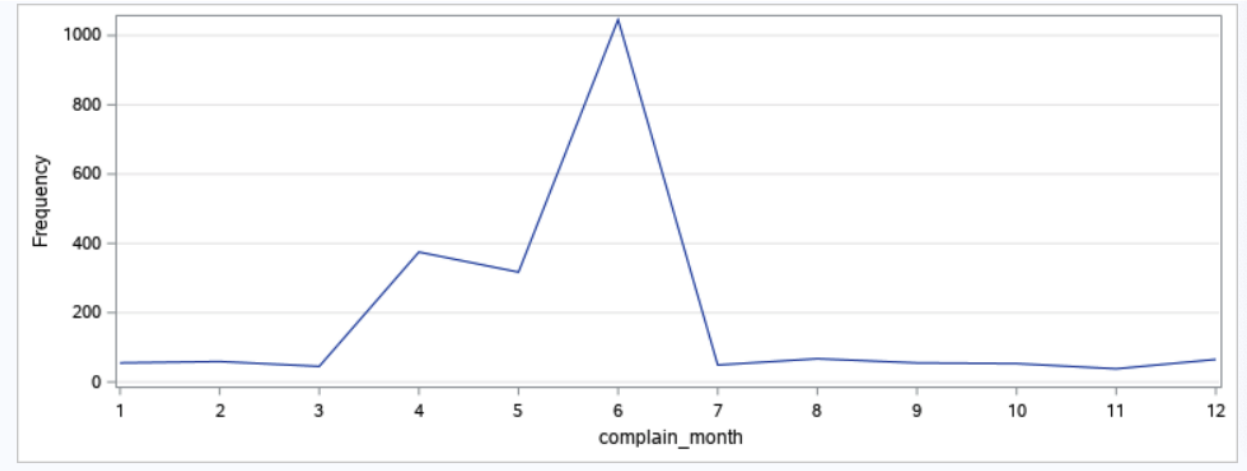
Screenshot:

complain_ym	Frequency	Percent	Cumulative Frequency	Cumulative Percent
2015-1	55	2.47	55	2.47
2015-10	53	2.38	108	4.86
2015-11	38	1.71	146	6.56
2015-12	65	2.92	211	9.49
2015-2	59	2.65	270	12.14
2015-3	45	2.02	315	14.16
2015-4	375	16.86	690	31.03
2015-5	317	14.25	1007	45.28
2015-6	1046	47.03	2053	92.31
2015-7	49	2.20	2102	94.51
2015-8	67	3.01	2169	97.53
2015-9	55	2.47	2224	100.00

SAS Code:

```
ods graphics / reset width=8in height=3in imagemap;

proc sgplot data= COMCAST_1;

   vline complain_month /;

   yaxis grid;

run;
```

Screenshot:



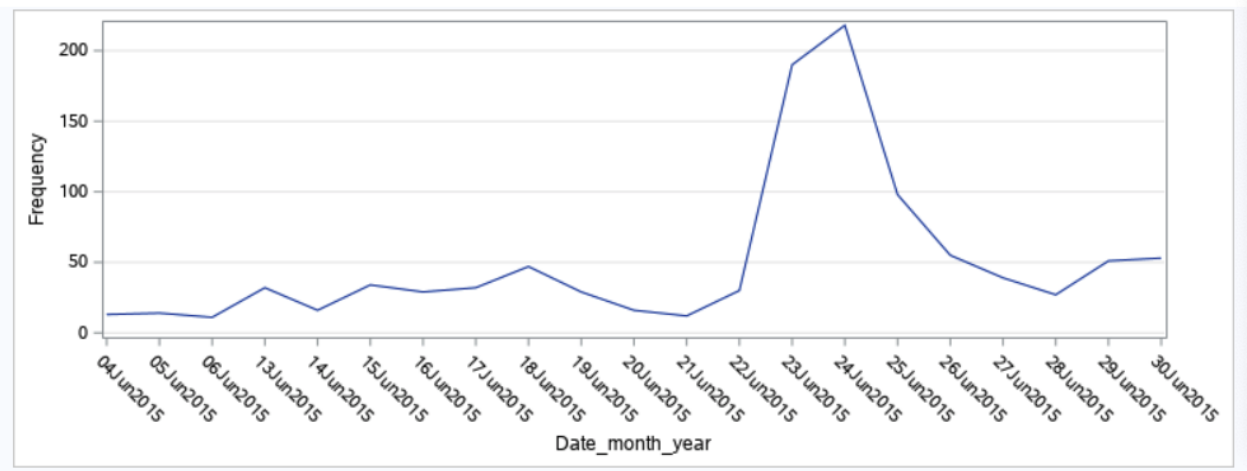
Interpretation:

- The month of June received the maximum number of complaints.

SAS Code:

```
Proc sgplot data=COMCAST_1;
   vline Date_month_year /;
   yaxis grid;
run;
proc sgplot data= COMCAST_1;
where complain_month=6;
   vline Date_month_year /;
   yaxis grid;
run;
```

Screenshot:



Interpretation:

- June 24, had the highest number of complaints .

3. Provide a table with the frequency of complaint types.

- **Which complaint types are maximum i.e., around internet, network issues, or across any other domains.**

  SAS Code:

```
data comcast_2;
set comcast;
format customer_complaint $60.;
customer_complaint=upcase(Customer_Complaint);
customer_complaint_v2=scan(customer_complaint, 1, 2,' ');
WORD = TRANSLATE(customer_complaint,' ',',.?!:&*$%@#_+-=/{}[]|()');
IF word in ('IS','ARE','WAS','WERE','THE','I','WE','THEY','THERE',
'THESE','IN','A','OF','TO','AND','IT','THAT','THIS', 'CAP', 'cap', 'caps', 'CAPS', 'COMCAST') THEN
DELETE;
if find(customer_complaint, "INTERNET")>0 then internet=1;
if find(customer_complaint, "NETWORK")>0 then network=1;
if find(customer_complaint, "PAYMENT" )>0 or
  find(customer_complaint, "BILLING")>0  then payment=1;
```

```
if find(customer_complaint, "DATA")>0 then data=1;
run;


proc freq data= comcast_2 noprint;
tables customer_complaint /out=compaint_types  ;
run;
proc sort data =compaint_types; by descending count; run;
proc freq data= comcast_2 noprint;
tables WORD / out=complaint_types  ;
run;
proc sort data =complaint_types; by descending count;
run;

PROC SQL NOPRINT;
CREATE TABLE WORDLIST AS SELECT WORD AS WORD, COUNT(*)  AS COUNT FROM
comcast_2 GROUP BY WORD ORDER BY COUNT DESCENDING;
QUIT;

proc means data= comcast_2 sum;
var internet network payment data;
output out=total_complaints;
run;
```

Screenshot:

	Customer_Complaint	COUNT
1	COMCAST DATA CAP	30
2	COMCAST INTERNET	29
3	COMCAST DATA CAPS	21
4	COMCAST BILLING	18
5	COMCAST SERVICE	15
6	INTERNET SPEED	15
7	DATA CAPS	13
8	UNFAIR BILLING PRACTICES	13
9	DATA CAP	12
10	COMCAST COMPLAINT	11

Interpretation:

- The top 10 types of complaints indicate that Data, Internet and Billing issues maybe the top complaint reasons**.**

Screenshot:

Variable	Sum
internet	529.0000000
network	2.0000000
payment	307.0000000
data	219.0000000

Interpretation:

- By observing the above word counts, we can infer that most common complaints are around: Internet, service, and payment related issues.

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.

SAS Code:

```
/*Create a new categorical variable with value as Open and Closed*/

proc freq data= comcast;

tables Status;

run;
```

Screenshot:

Status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Closed	734	33.00	734	33.00
Open	363	16.32	1097	49.33
Pending	154	6.92	1251	56.25
Solved	973	43.75	2224	100.00

SAS Code:

```
data comcast_3;
set comcast;
Format new_status $12.;
if status in ("Open", "Pending") then new_status="Open";
 else if status in ("Closed", "Solved") then new_status="Closed";
run;
proc freq data= comcast_3;
tables new_status;
run;
```

Screenshot:

new_status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Closed	1707	76.75	1707	76.75
Open	517	23.25	2224	100.00

Interpretation:

- We have categorized the variable status into new categories. According to the new categories of status, around 77% of the tickets are closed.

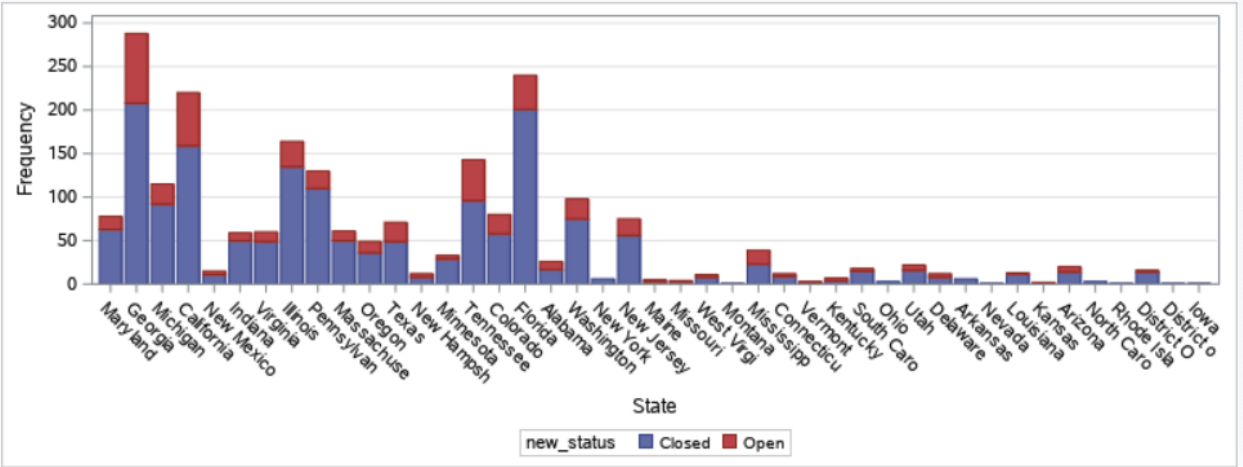
5. Provide state wise status of complaints in a stacked bar chart. Use the categorized variable from Q3. Provide insights on:

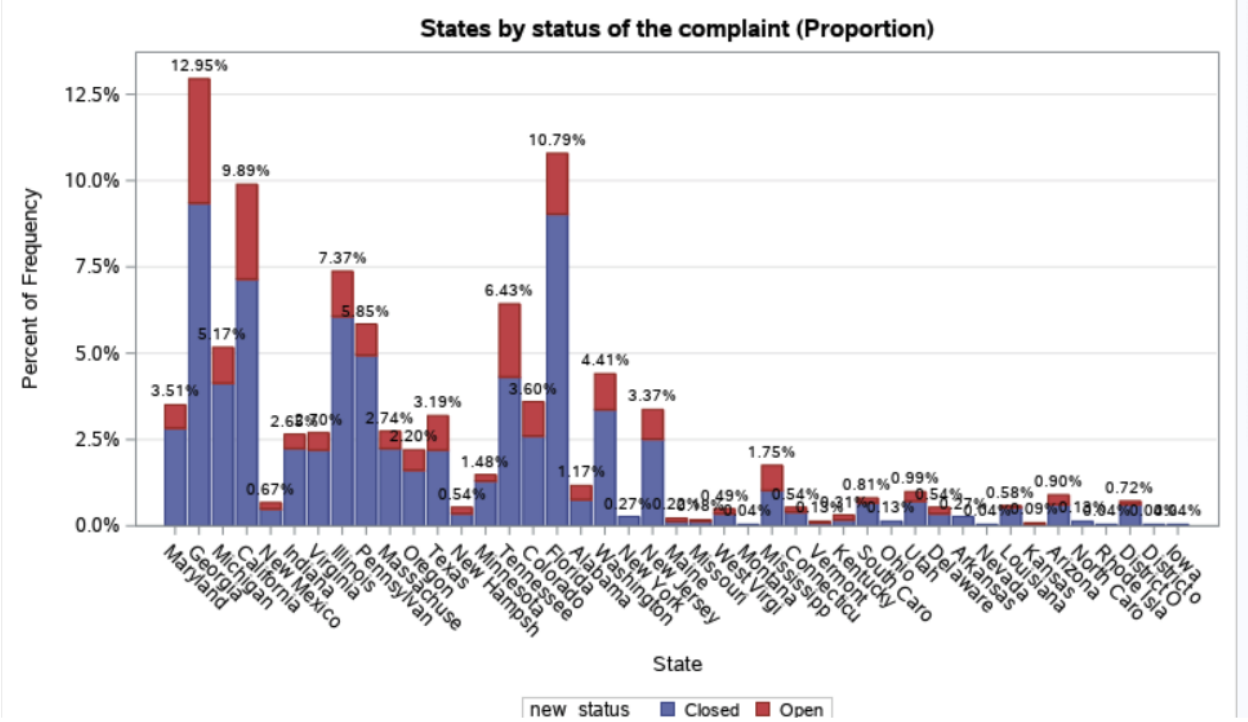
- **Which state has the maximum complaints**
- **Which state has the highest percentage of unresolved complaints**

SAS Code:

```
ods graphics / reset width=8in height=4.8in imagemap;

/*Which state has the maximum complaints*/

title 'States by status of the complaint (Frequency)';

proc sgplot data= comcast_3;

   vbar State / group=new_status groupdisplay=stack DATALABELFITPOLICY=NONE;

   yaxis grid discreteorder= data;

run;

title 'States by status of the complaint (Proportion)';

proc sgplot data=WORK.COMCAST_3;

   vbar State / group=new_status groupdisplay=stack datalabel stat=percent
DATALABELFITPOLICY=NONE;

   yaxis grid discreteorder= data ;

run;
```

Screenshot:





Interpretation:

Georgia has the highest number/proportion of complaints, around 12.9% of complaints came from the state of Georgia.

SAS Code:

```
proc freq data=comcast_3  ;
tables State*new_status   /   out=state_level_complains   OUTPCT     ;
run;
proc sort data=state_level_complains  ; by descending new_status descending PCT_ROW ;
run;
```

Screenshot:

	State	new_status	PCT_ROW
1	Kansas	Open	50
2	Kentucky	Open	42.857142857
3	Mississipp	Open	41.025641026
4	Maine	Open	40
5	Alabama	Open	34.615384615

Interpretation:

- Kansas has the highest number and proportion of unresolved cases.
- Around 50% of tickets originated from the state of Kansas are still open.

6. Provide the percentage of complaints resolved till date, which were received through the Internet and customer care calls.

SAS Code:

```
proc freq data=comcast_3  ;

tables   received_via          ;

run;

proc freq data=comcast_3  ;

tables  new_status   /   out=complains_status        ;

run;
```

Screenshot:

new_status	Frequency	Percent	Cumulative Frequency	Cumulative Percent
Closed	1707	76.75	1707	76.75
Open	517	23.25	2224	100.00

Interpretation:

- In the total of 2,224 tickets, around 1,707 (77%) tickets are closed.

SAS Code:

```
proc freq data=comcast_3  ;

tables  new_status*received_via          ;

run;
```

Screenshot:

Table of new_status by Received_Via			
	Received_Via		
new_status	Customer Care Call	Internet	Total
Closed	864 38.85 50.62 77.21	843 37.90 49.38 76.29	1707 76.75
Open	255 11.47 49.32 22.79	262 11.78 50.68 23.71	517 23.25
Total	1119 50.31	1105 49.69	2224 100.00

Interpretation:

- Resolution rate is 77% for the complaints received through Internet and customer care calls*/
- In the total of 1,119 tickets received via Customer Care Call, around 77% (n=864) were resolved.