Data Science with SAS Project

Retail Analysis with Walmart Data

DESCRIPTION

One of the leading retail stores in the US, Walmart, would like to predict the sales and demand accurately. There are certain events and holidays which impact sales on each day. There are sales data available for 45 stores of Walmart. The business is facing a challenge due to unforeseen demands and runs out of stock some times, due to the inappropriate machine learning algorithm. An ideal ML algorithm will predict demand at different points of time covering seasonality and ingest factors like economic conditions including CPI, Unemployment Index, etc.

Walmart runs several promotional markdown events throughout the year. These markdowns precede prominent holidays, the four largest of all, which are the Super Bowl, Labour Day, Thanksgiving, and Christmas. The weeks including these holidays are weighted five times higher in the evaluation than non-holiday weeks. Part of the challenge presented by this competition is modeling the effects of markdowns on these holiday weeks in the absence of complete/ideal historical data. Historical sales data for 45 Walmart stores located in different regions are available

Analysis:

Basic Statistics tasks

- Which store has maximum sales
- Which store has maximum standard deviation i.e., the sales vary a lot. Also, find out the coefficient of mean to standard deviation
- Which store/s has good quarterly growth rate in Q3'2012
- Some holidays have a negative impact on sales. Find out holidays which have higher sales than the mean sales in non-holiday season for all stores together
- Provide a monthly and semester view of sales in units and give insights

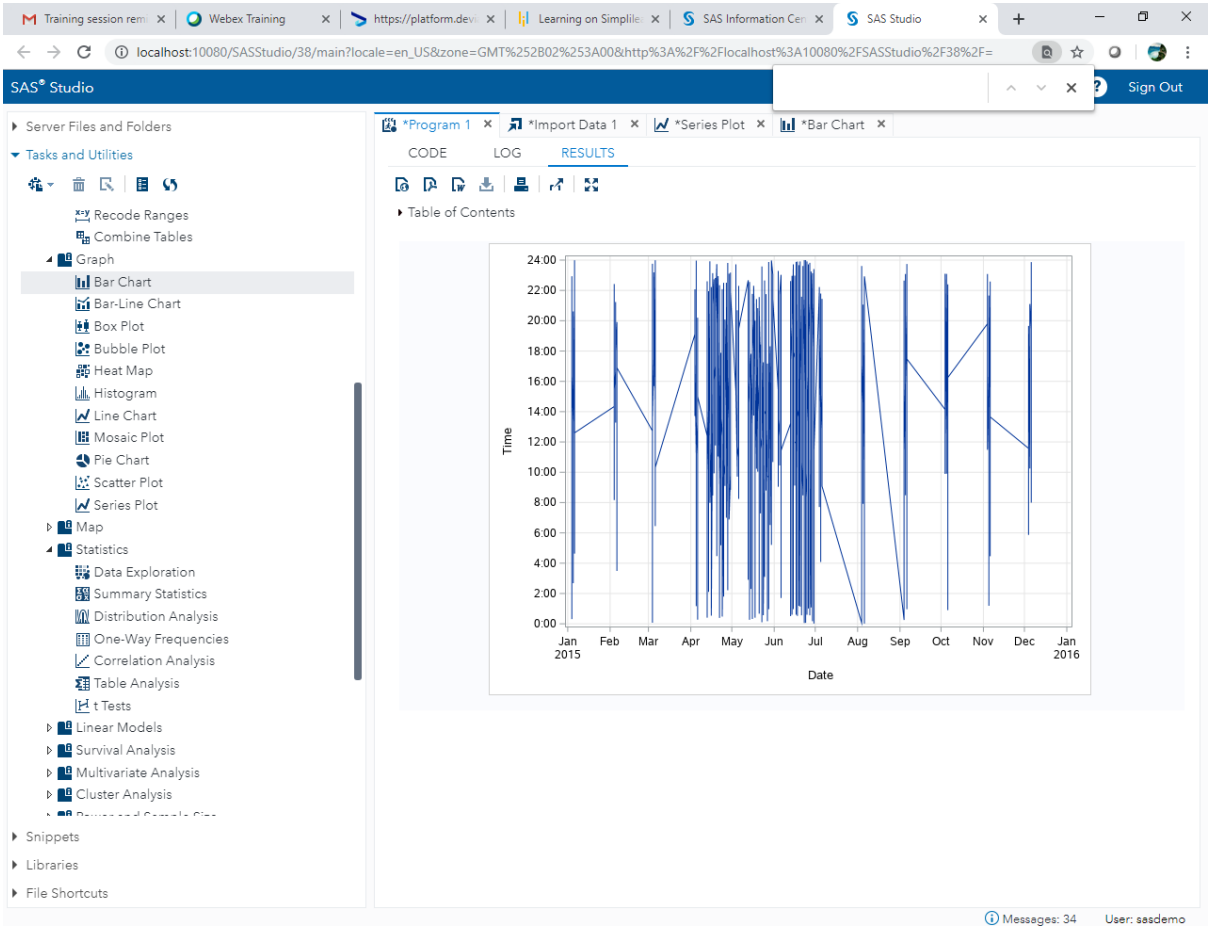
Statistical Model

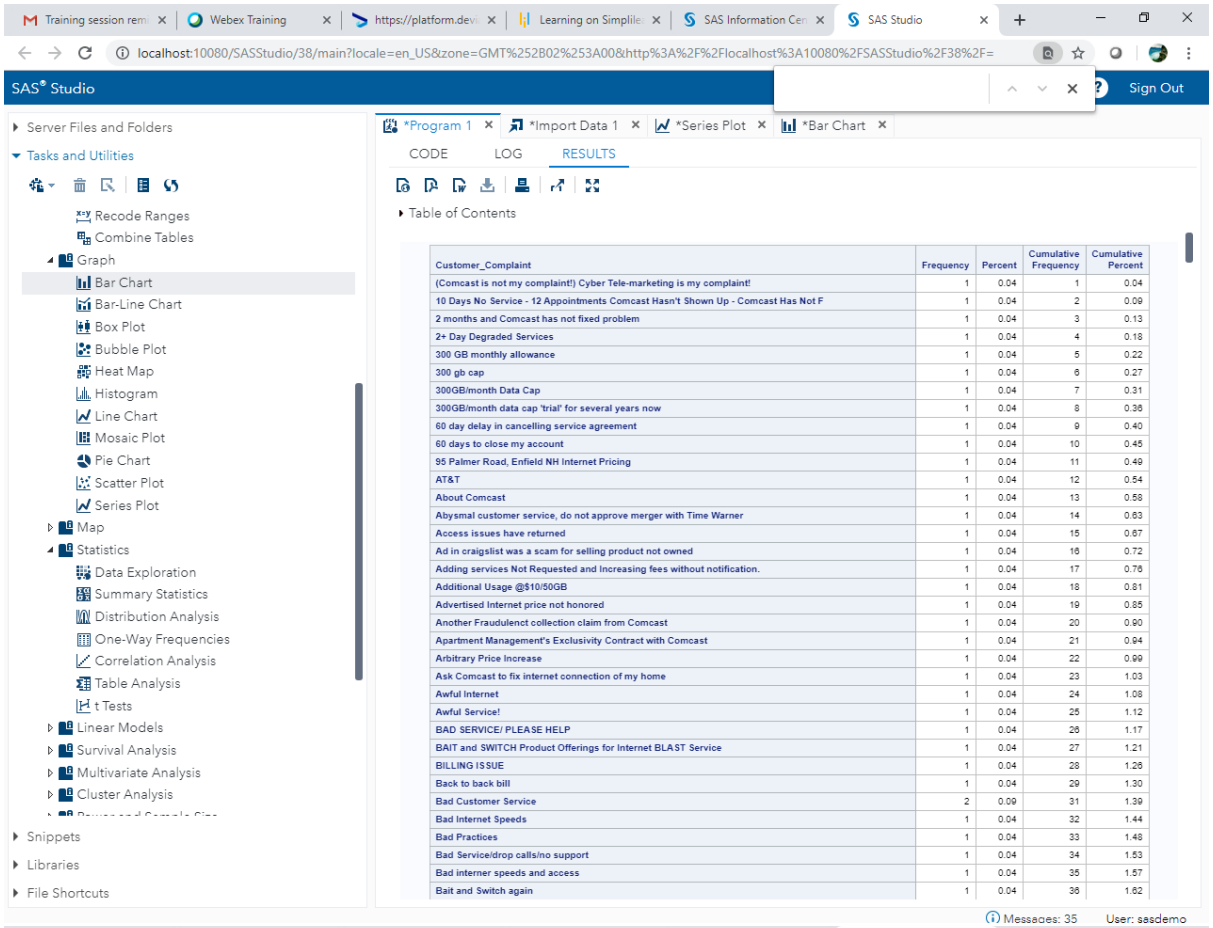
For Store 1 – Build prediction models to forecast demand

- Linear Regression – Utilize variables like date and restructure dates as 1 for 5 Feb 2010(starting from the earliest date in order). Hypothesize if CPI, unemployment, and fuel price have any impact on sales.
- Time series forecasting model –
    - Hypothesize if the data is fit for time series analysis – check for white noise probability test
    - Make adjustments in historical data for events like holidays, if applicable
    - Build ARIMA model to forecast 6 months i.e., input utilize only till April 2012.

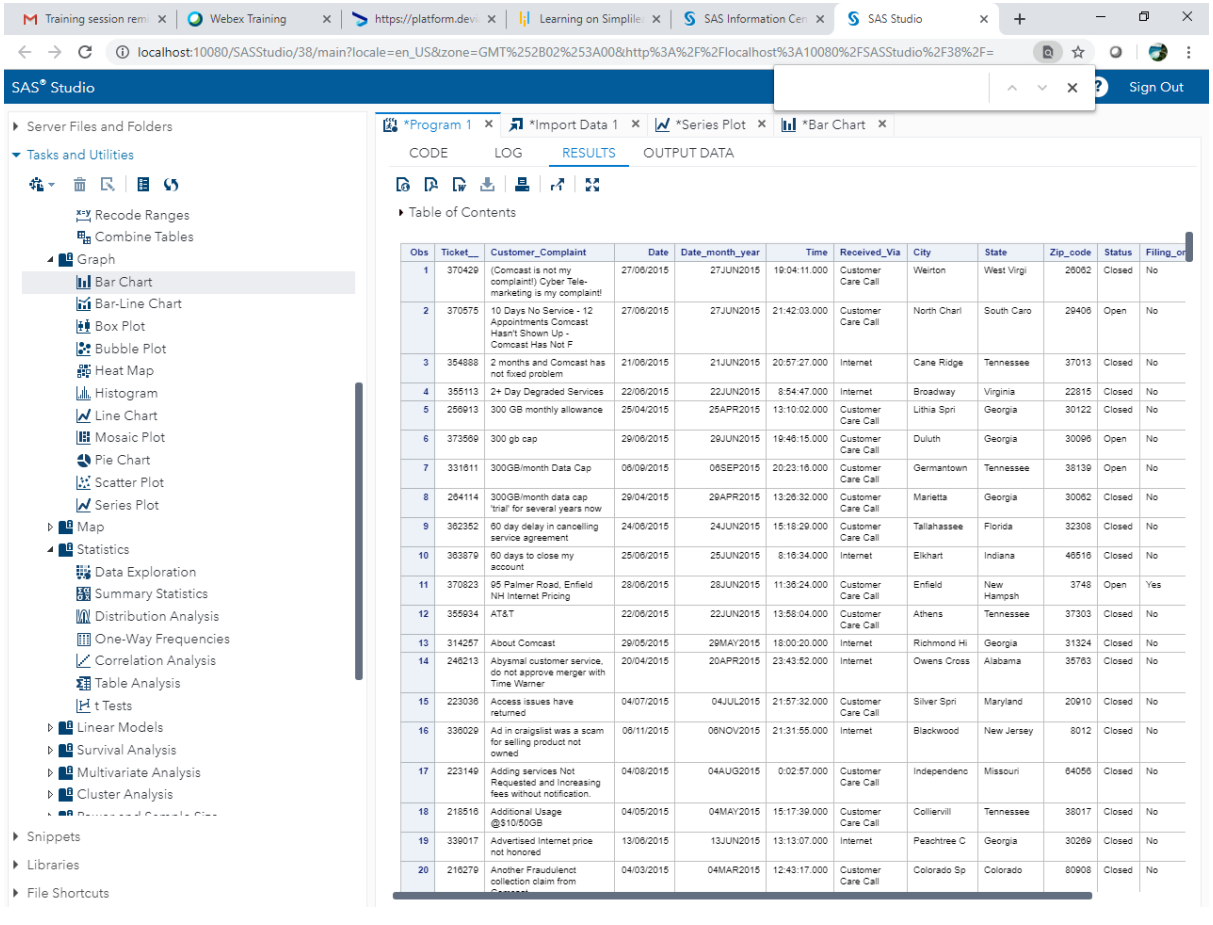
Predict next 6 months i.e., June to Oct 2010. Check for MAPE.

Screenshots

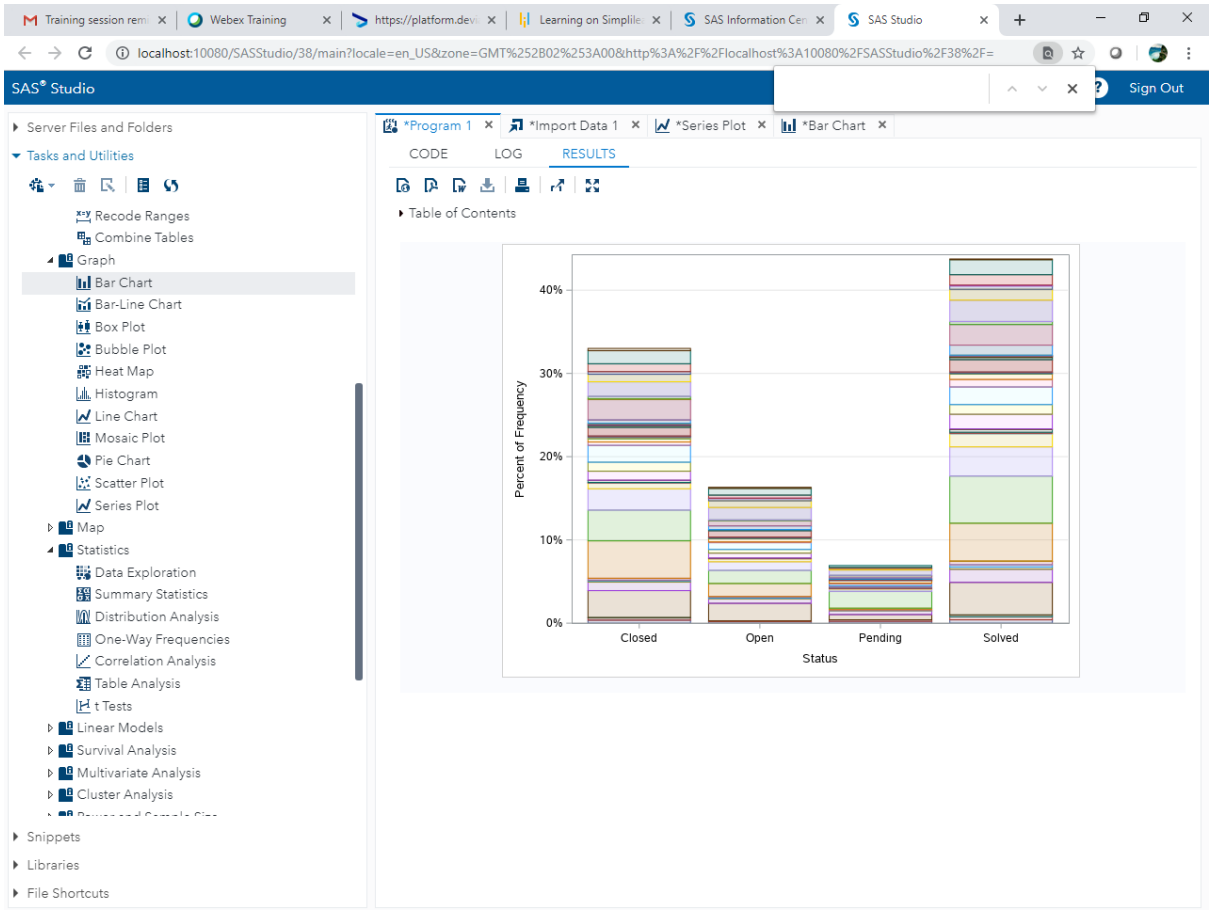


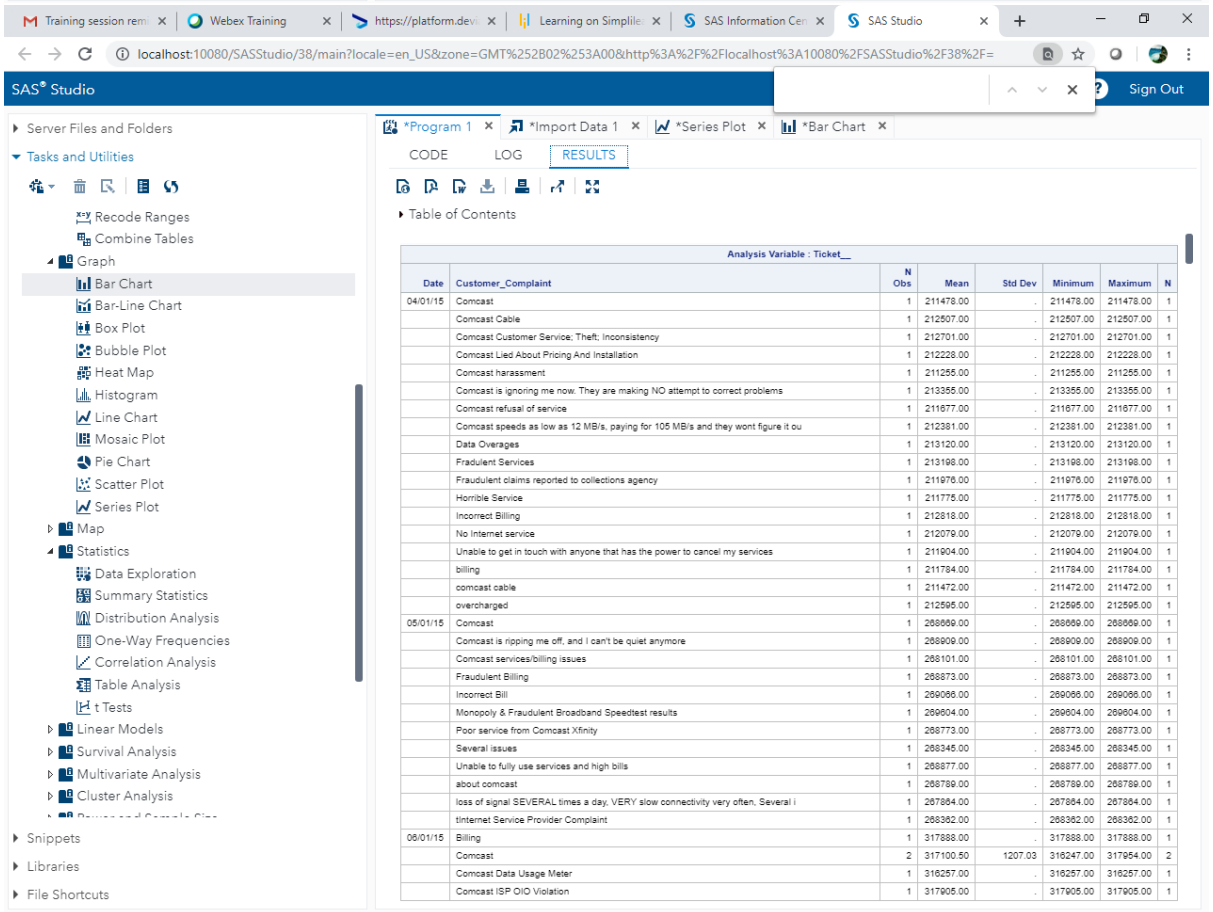


Customer_Complaint	Frequency	Percent	Cumulative Frequency	Cumulative Percent
(Comcast is not my complaint!) Cyber Tele-marketing is my complaint!	1	0.04	1	0.04
10 Days No Service - 12 Appointments Comcast Hasn't Shown Up - Comcast Has Not F	1	0.04	2	0.09
2 months and Comcast has not fixed problem	1	0.04	3	0.13
2+ Day Degraded Services	1	0.04	4	0.18
300 GB monthly allowance	1	0.04	5	0.22
300 gb cap	1	0.04	6	0.27
300GB/month Data Cap	1	0.04	7	0.31
300GB/month data cap 'trial' for several years now	1	0.04	8	0.36
60 day delay in cancelling service agreement	1	0.04	9	0.40
60 days to close my account	1	0.04	10	0.45
95 Palmer Road, Enfield NH Internet Pricing	1	0.04	11	0.49
AT&T	1	0.04	12	0.54
About Comcast	1	0.04	13	0.58
Abysmal customer service, do not approve merger with Time Warner	1	0.04	14	0.63
Access issues have returned	1	0.04	15	0.67
Ad in craigslist was a scam for selling product not owned	1	0.04	16	0.72
Adding services Not Requested and Increasing fees without notification.	1	0.04	17	0.76
Additional Usage @$10/50GB	1	0.04	18	0.81
Advertised Internet price not honored	1	0.04	19	0.85
Another Fraudulenct collection claim from Comcast	1	0.04	20	0.90
Apartment Management's Exclusivity Contract with Comcast	1	0.04	21	0.94
Arbitrary Price Increase	1	0.04	22	0.99
Ask Comcast to fix internet connection of my home	1	0.04	23	1.03
Awful Internet	1	0.04	24	1.08
Awful Service!	1	0.04	25	1.12
BAD SERVICE/ PLEASE HELP	1	0.04	26	1.17
BAIT and SWITCH Product Offerings for Internet BLAST Service	1	0.04	27	1.21
BILLING ISSUE	1	0.04	28	1.26
Back to back bill	1	0.04	29	1.30
Bad Customer Service	2	0.09	31	1.39
Bad Internet Speeds	1	0.04	32	1.44
Bad Practices	1	0.04	33	1.48
Bad Service/drop calls/no support	1	0.04	34	1.53
Bad interner speeds and access	1	0.04	35	1.57
Bait and Switch again	1	0.04	36	1.62



Obs	Ticket_	Customer_Complaint	Date	Date_month_year	Time	Received_Via	City	State	Zip_code	Status	Filing_or
1	370429	(Comcast is not my complaint!) Cyber Tele-marketing is my complaint!	27/06/2015	27JUN2015	19:04:11.000	Customer Care Call	Weirton	West Virgi	26062	Closed	No
2	370575	10 Days No Service - 12 Appointments Comcast Hasn't Shown Up - Comcast Has Not F	27/06/2015	27JUN2015	21:42:03.000	Customer Care Call	North Charl	South Caro	29406	Open	No
3	354888	2 months and Comcast has not fixed problem	21/06/2015	21JUN2015	20:57:27.000	Internet	Cane Ridge	Tennessee	37013	Closed	No
4	355113	2+ Day Degraded Services	22/06/2015	22JUN2015	8:54:47.000	Internet	Broadway	Virginia	22815	Closed	No
5	256913	300 GB monthly allowance	25/04/2015	25APR2015	13:10:02.000	Customer Care Call	Lithia Spri	Georgia	30122	Closed	No
6	373569	300 gb cap	29/06/2015	29JUN2015	19:46:15.000	Customer Care Call	Duluth	Georgia	30096	Open	No
7	331611	300GB/month Data Cap	06/09/2015	06SEP2015	20:23:16.000	Customer Care Call	Germantown	Tennessee	38139	Open	No
8	264114	300GB/month data cap 'trial' for several years now	29/04/2015	29APR2015	13:26:32.000	Customer Care Call	Marietta	Georgia	30062	Closed	No
9	362352	60 day delay in cancelling service agreement	24/06/2015	24JUN2015	15:18:29.000	Customer Care Call	Tallahassee	Florida	32308	Closed	No
10	363879	60 days to close my account	25/06/2015	25JUN2015	8:16:34.000	Internet	Elkhart	Indiana	46516	Closed	No
11	370823	95 Palmer Road, Enfield NH Internet Pricing	28/06/2015	28JUN2015	11:36:24.000	Customer Care Call	Enfield	New Hampsh	3748	Open	Yes
12	355934	AT&T	22/06/2015	22JUN2015	13:58:04.000	Customer Care Call	Athens	Tennessee	37303	Closed	No
13	314257	About Comcast	29/05/2015	29MAY2015	18:00:20.000	Internet	Richmond Hi	Georgia	31324	Closed	No
14	246213	Abysmal customer service, do not approve merger with Time Warner	20/04/2015	20APR2015	23:43:52.000	Internet	Owens Cross	Alabama	35763	Closed	No
15	223036	Access issues have returned	04/07/2015	04JUL2015	21:57:32.000	Customer Care Call	Silver Spri	Maryland	20910	Closed	No
16	336029	Ad in craigslist was a scam for selling product not owned	06/11/2015	06NOV2015	21:31:55.000	Internet	Blackwood	New Jersey	8012	Closed	No
17	223149	Adding services Not Requested and Increasing fees without notification.	04/08/2015	04AUG2015	0:02:57.000	Customer Care Call	Independenc	Missouri	64056	Closed	No
18	218516	Additional Usage @$10/50GB	04/05/2015	04MAY2015	15:17:39.000	Customer Care Call	Colliervill	Tennessee	38017	Closed	No
19	339017	Advertised Internet price not honored	13/06/2015	13JUN2015	13:13:07.000	Internet	Peachtree C	Georgia	30269	Closed	No
20	216279	Another Fraudulenct collection claim from	04/03/2015	04MAR2015	12:43:17.000	Customer Care Call	Colorado Sp	Colorado	80908	Closed	No





Analysis Variable : Ticket__

Date	Customer_Complaint	N Obs	Mean	Std Dev	Minimum	Maximum	N
04/01/15	Comcast	1	211478.00	.	211478.00	211478.00	1
	Comcast Cable	1	212507.00	.	212507.00	212507.00	1
	Comcast Customer Service; Theft; Inconsistency	1	212701.00	.	212701.00	212701.00	1
	Comcast Lied About Pricing And Installation	1	212228.00	.	212228.00	212228.00	1
	Comcast harassment	1	211255.00	.	211255.00	211255.00	1
	Comcast is ignoring me now. They are making NO attempt to correct problems	1	213355.00	.	213355.00	213355.00	1
	Comcast refusal of service	1	211677.00	.	211677.00	211677.00	1
	Comcast speeds as low as 12 MB/s, paying for 105 MB/s and they wont figure it ou	1	212381.00	.	212381.00	212381.00	1
	Data Overages	1	213120.00	.	213120.00	213120.00	1
	Fradulent Services	1	213198.00	.	213198.00	213198.00	1
	Fraudulent claims reported to collections agency	1	211976.00	.	211976.00	211976.00	1
	Horrible Service	1	211775.00	.	211775.00	211775.00	1
	Incorrect Billing	1	212818.00	.	212818.00	212818.00	1
	No Internet service	1	212079.00	.	212079.00	212079.00	1
	Unable to get in touch with anyone that has the power to cancel my services	1	211904.00	.	211904.00	211904.00	1
	billing	1	211784.00	.	211784.00	211784.00	1
	comcast cable	1	211472.00	.	211472.00	211472.00	1
	overcharged	1	212595.00	.	212595.00	212595.00	1
05/01/15	Comcast	1	268869.00	.	268869.00	268869.00	1
	Comcast is ripping me off, and I can't be quiet anymore	1	268909.00	.	268909.00	268909.00	1
	Comcast services/billing issues	1	268101.00	.	268101.00	268101.00	1
	Fraudulent Billing	1	268873.00	.	268873.00	268873.00	1
	Incorrect Bill	1	269066.00	.	269066.00	269066.00	1
	Monopoly & Fraudulent Broadband Speedtest results	1	269604.00	.	269604.00	269604.00	1
	Poor service from Comcast Xfinity	1	268773.00	.	268773.00	268773.00	1
	Several issues	1	268345.00	.	268345.00	268345.00	1
	Unable to fully use services and high bills	1	268877.00	.	268877.00	268877.00	1
	about comcast	1	268789.00	.	268789.00	268789.00	1
	loss of signal SEVERAL times a day, VERY slow connectivity very often, Several i	1	267864.00	.	267864.00	267864.00	1
	tInternet Service Provider Complaint	1	268362.00	.	268362.00	268362.00	1
06/01/15	Billing	1	317888.00	.	317888.00	317888.00	1
	Comcast	2	317100.50	1207.03	316247.00	317954.00	2
	Comcast Data Usage Meter	1	316257.00	.	316257.00	316257.00	1
	Comcast ISP OIO Violation	1	317905.00	.	317905.00	317905.00	1

Source Code

```
FILENAME REFFILE '/folders/myfolders/Walmart_Store_sales.csv';

PROC IMPORT DATAFILE=REFFILE
       DBMS=CSV
       OUT=WORK.Walmart_Store_sales;
       GETNAMES=YES;
RUN;
/*Basic Statistics tasks*/
PROC UNIVARIATE Data = Walmart_Store_sales;
run;

/*Linear Regression – Utilize variables like date and restructure
dates*/
PROC SORT Data = Walmart_Store_sales;
By Date;
run;

PROC REG Data = Walmart_Store_sales;
Model Date = CPI Unemployment Fuel_Price;
run;

/*Time series forecasting model*/
PROC ARIMA Data = Walmart_Store_sales;
identify var = date(1) crosscorr=(CPI Unemployment Fuel_Price);
estimate p=1;
forecast lead=12 interval=month out= Walmart_Store_sales;
run;
```