

WALMART SALES FORECAST

OPIM-5671-Data Mining and Business Intelligence
- Prof Sudip Bhattacharjee

Team 5 :

- 1.Anand Shiv Sharma
- 2.Manas Joshi
- 3.Sri Varshini Chava
- 4.Anuja Sunil Kamble
- 5.Mitul Krishnasamy Sivakumar

Problem Statement:

- Retail industry in the US has faced with growing concern of intense competition and increasingly tighter margins.
- Future becomes imperative to forecast future sales.

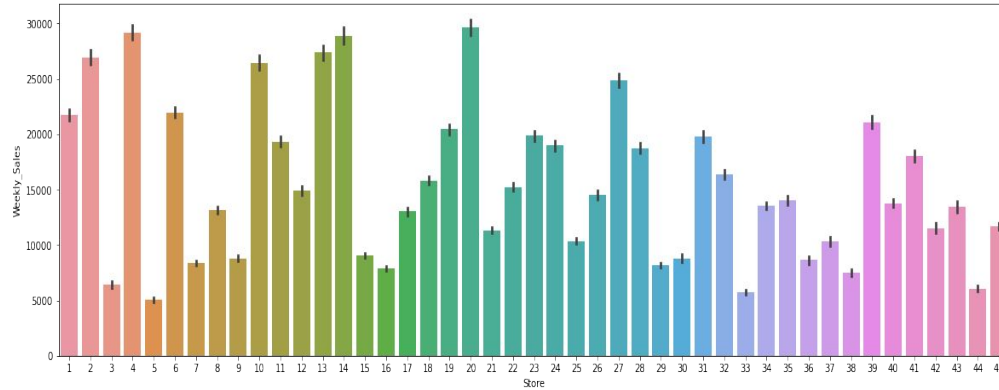
Solution:

- Forecast the weekly sales in Walmart Stores
- Use time Series Forecasting methodology with SAS Studio 3.8

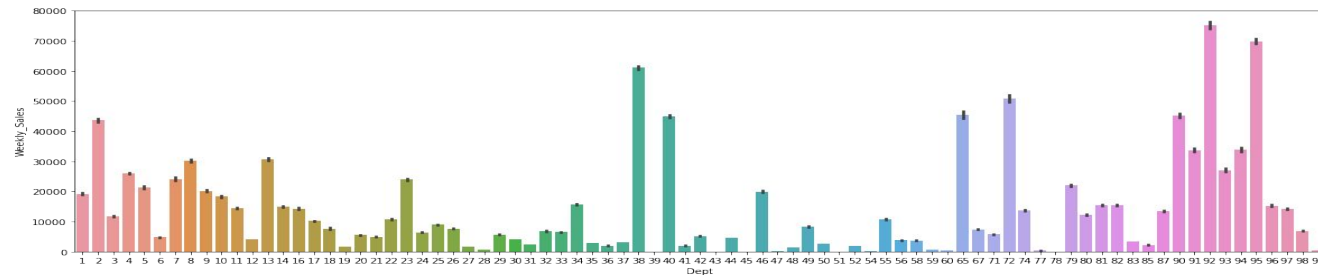
Introduction

- Weekly Sales forecasting of Walmart Sales data using time series analysis.
- 45 stores and 81 departments with 16 columns and 420,212 rows
- 4 different stores and department 92 to analyse the data.
- Dependent Variable: Weekly Sales
- Independent Variables: Is Holiday, Markdown 1, Markdown 2, Markdown 3, Markdown 4, Markdown 5
- Time Variable: Date
- Other Variables - Temperature, Fuel-Price, Unemployment, Size, Type, CPI, Store, Dept

Exploratory Data Analysis



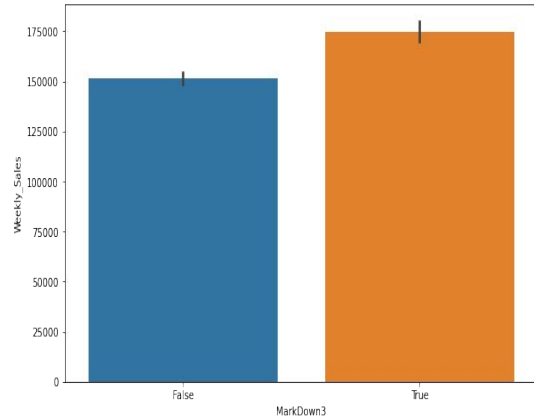
Store vs Weekly_Sale



Dept vs Weekly_Salees

Data Preprocessing and Modification

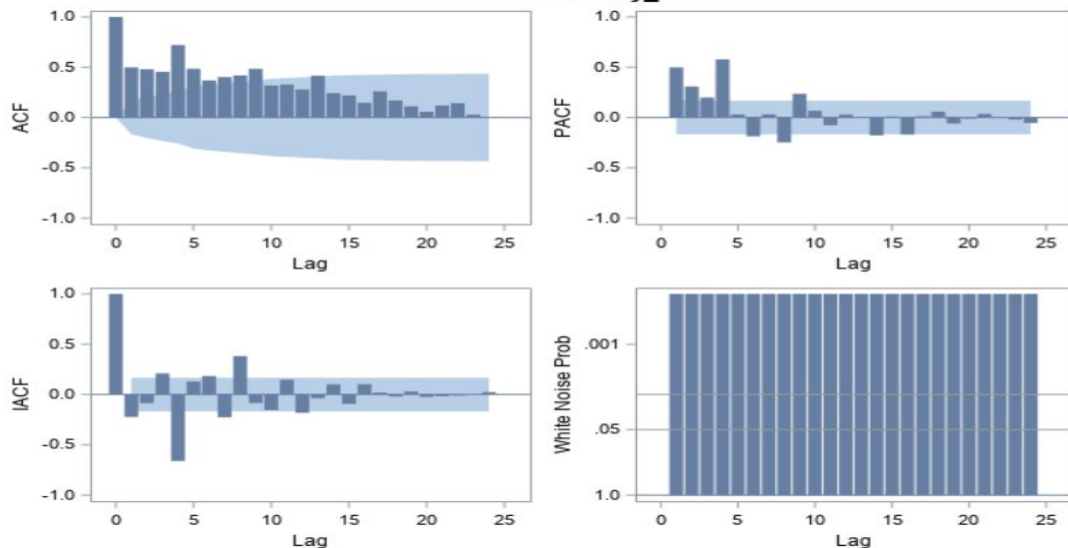
- Outliers and Missing Values
- Data Imputation
- Data Integration



```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 420212 entries, 0 to 421569
Data columns (total 16 columns):
 #   Column              Non-Null Count  Dtype  
---  -
 0   Store               420212 non-null  int64  
 1   Dept               420212 non-null  int64  
 2   Date               420212 non-null  object  
 3   Weekly_Sales       420212 non-null  float64 
 4   IsHoliday          420212 non-null  int64  
 5   Temperature        420212 non-null  float64 
 6   Fuel_Price         420212 non-null  float64 
 7   MarkDown1          420212 non-null  float64 
 8   MarkDown2          420212 non-null  float64 
 9   MarkDown3          420212 non-null  float64 
10   MarkDown4          420212 non-null  float64 
11   MarkDown5          420212 non-null  float64 
12   CPI                420212 non-null  float64 
13   Unemployment        420212 non-null  float64 
14   Type               420212 non-null  object  
15   Size               420212 non-null  int64
```

Time Series Exploration

Correlations for Weekly_Sales



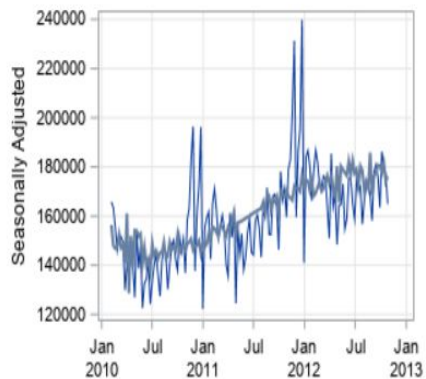
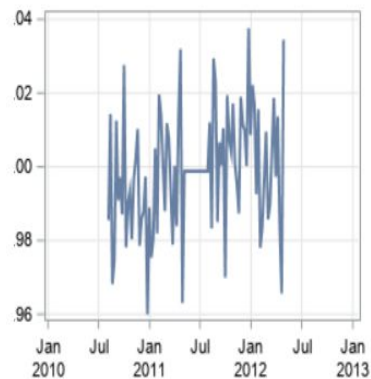
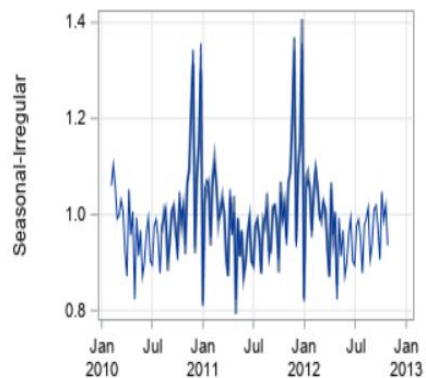
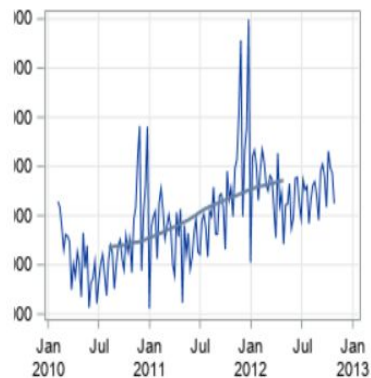
Dept=92

Variable Information

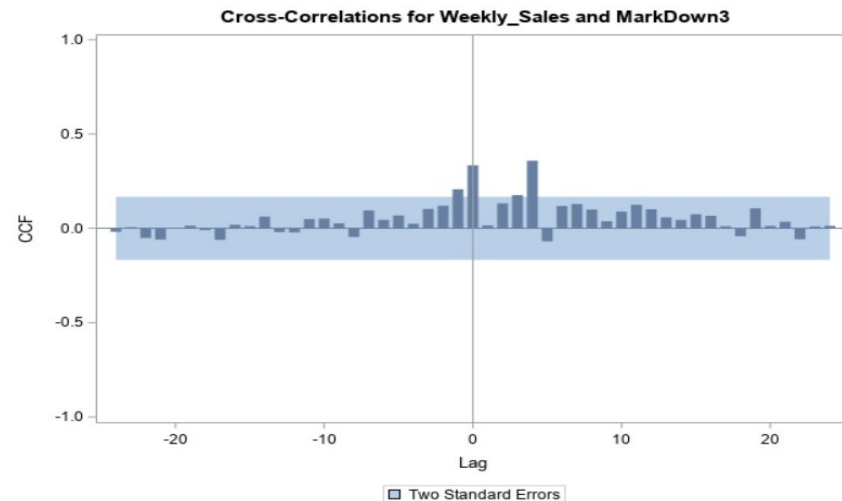
Name	Weekly_Sales
Label	Weekly_Sales
First	Fri, 5 Feb 2010
Last	Fri, 26 Oct 2012
Number of Observations Read	143

Time Series Descriptive Statistics

Variable	Weekly_Sales
Number of Observations	143
Number of Observations Used	143
Number of Missing Observations	0
Minimum	122263.2
Median	158924
Maximum	239759.3
Mean	159365.1
Standard Deviation	19283.27



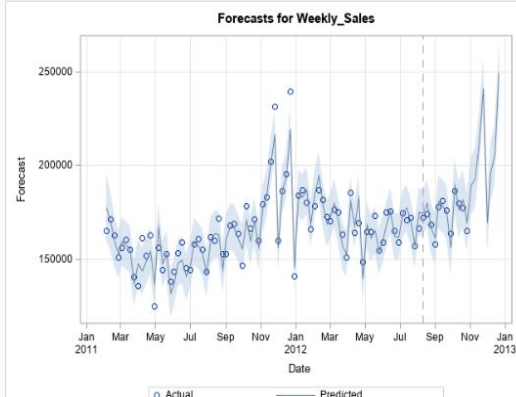
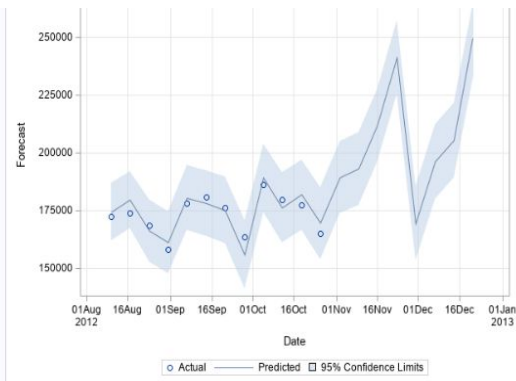
Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-1.0328	0.4660	-0.72	0.4042		
	1	-0.3276	0.6072	-0.38	0.5438		
	2	-0.1190	0.6546	-0.19	0.6170		
Single Mean	0	-71.2061	0.0012	-6.85	<.0001	23.43	0.0010
	1	-37.0903	0.0012	-4.25	0.0008	9.03	0.0010
	2	-23.0596	0.0042	-3.19	0.0230	5.09	0.0360
Trend	0	-107.102	0.0001	-9.24	<.0001	42.69	0.0010
	1	-77.2871	0.0005	-6.24	<.0001	19.52	0.0010
	2	-59.8574	0.0005	-5.00	0.0004	12.51	0.0010



Modelling and Forecasting

```
proc arima data=Work.preProcessedData plots
  (only)=(series(corr crosscorr) residual(corr normal)
    forecast(forecast forecastonly) ) out=STSM.OUTPUT;
  identify var=Weekly_Sales(52);
  estimate p=(1 2 3 4) q=(1 2 3 4) method=ML outest=work.outest
    outstat=work.outstat;
  forecast lead=20 back=12 alpha=0.05 id=Date interval=week.6;
  by Dept;
  run;
quit;
```

	Dept	_TYPE_	_STAT_	_VALUE_
1	92	ML	AIC	1862.4772924
2	92	ML	SBC	1885.0750279
3	92	ML	LOGLIK	-922.2386462
4	92	ML	SSE	3297549123.5



*Forecasting model type:

ARIMA

▼ Model Settings

▼ ARIMA

Autoregressive order (p): 4

Differencing order (d): 1

Moving average order (q): 5

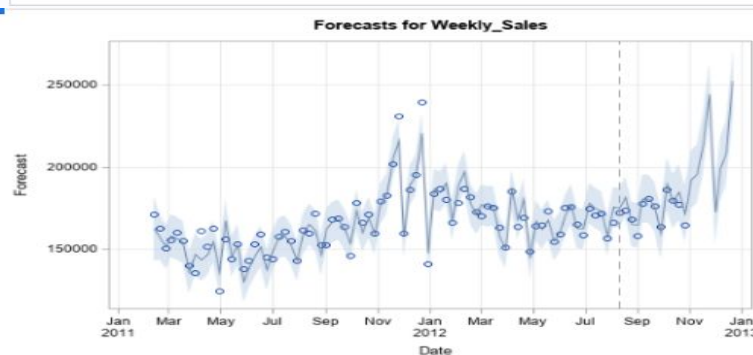
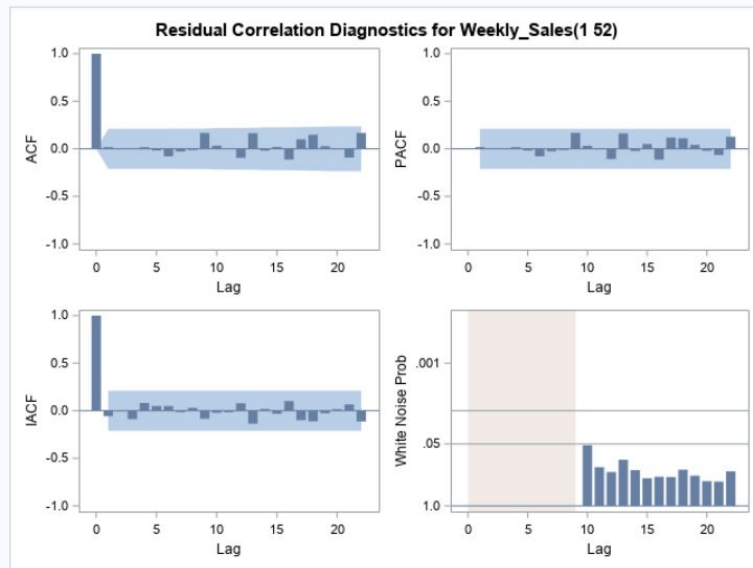
▼ Seasonal ARIMA

Autoregressive order (P): 0

Differencing order (D): 1

Moving average order (Q): 0

	Dept	_TYPE_	_STAT_	_VALUE_
1	92	ML	AIC	1841.7766891
2	92	ML	SBC	1866.7747858
3	92	ML	LOGLIK	-910.8883446
4	92	ML	SSE	3073522932.6
5	02	MI	NIIMDECIF	00



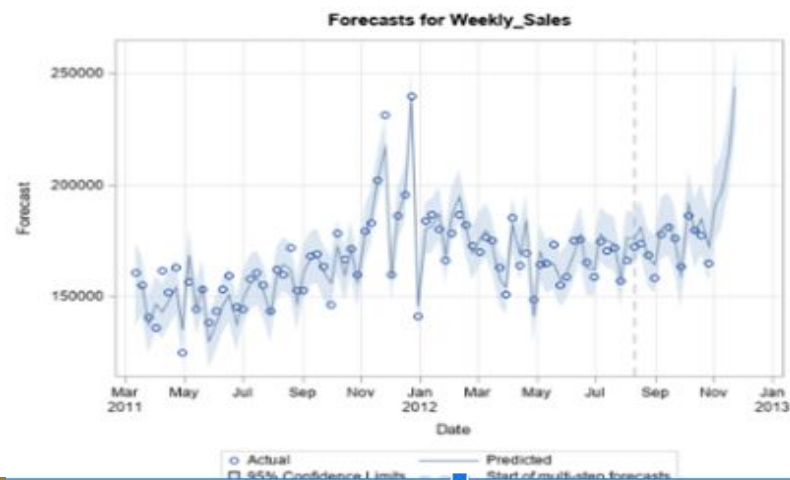
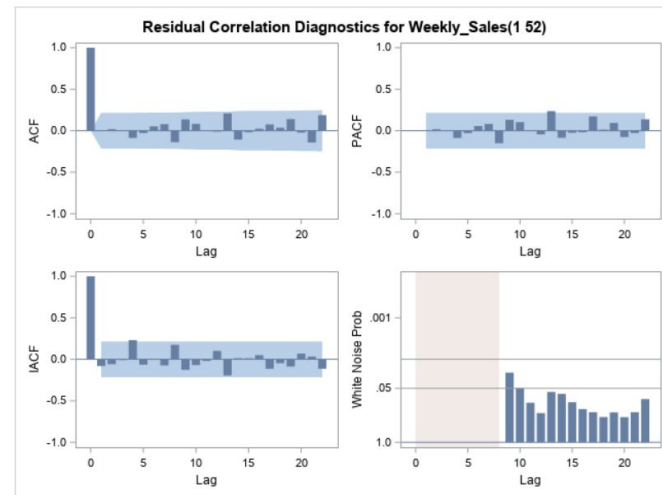
ARIMAX

```

1
2 proc arima data=Work.preProcessedData plots
3     (only)=(series(corr crosscorr) residual(corr normal)
4         forecast(forecast forecastonly) ) out=STSM.OUTPUT;
5     identify var=Weekly_Sales(1 52) crosscorr=(MarkDown3(1 52) );
6     estimate p=(1 2 3 4) q=(1 2 3 4) input=(4 $ MarkDown3) method=ML
7         outest=work.outest0003 outstat=work.outstat0003;
8     forecast lead=20 back=12 alpha=0.05 id=Date interval=week.6 printall;
9 run;
10 quit;

```

	TYPE	_STAT_	_VALUE_
1	ML	AIC	1756.4477392
2	ML	SBC	1780.9912122
3	ML	LOGLIK	-868.2238696
4	ML	SSE	2840226278.2



Model Comparison & Business Insights

Model	p	d	q	P	D	Q	AIC	SBC	SSE	MAPE(%)
ARIMAX	4	1	4	0	1	0	1756	1780	2.84E+09	1.72
ARIMA	4	1	5	0	1	0	1841	1866	3.074E+09	1.75
ARIMA	4	0	4	0	1	0	1862	1885	3.298E+09	1.82

- Managing Inventory
- Setting Sales Target
- Improving financial planning
- Responding to market changes

References

Data - <https://www.kaggle.com/datasets/aslanahmedov/walmart-sales-forecast>

