

Walmart Sales Data

Prediction of Sales for Walmart Stores

Group #10

Arpit AJMERA
Corentin CORMAN
Thomas KREMER
Asmita NIRAULA
Jérémy PINTO
Julien WARNERY

Case Background

Provided

- ❖ 45 Walmart stores in USA



- ❖ Anonymized sales data of each department



To Calculate

- ❖ Understand sales pattern



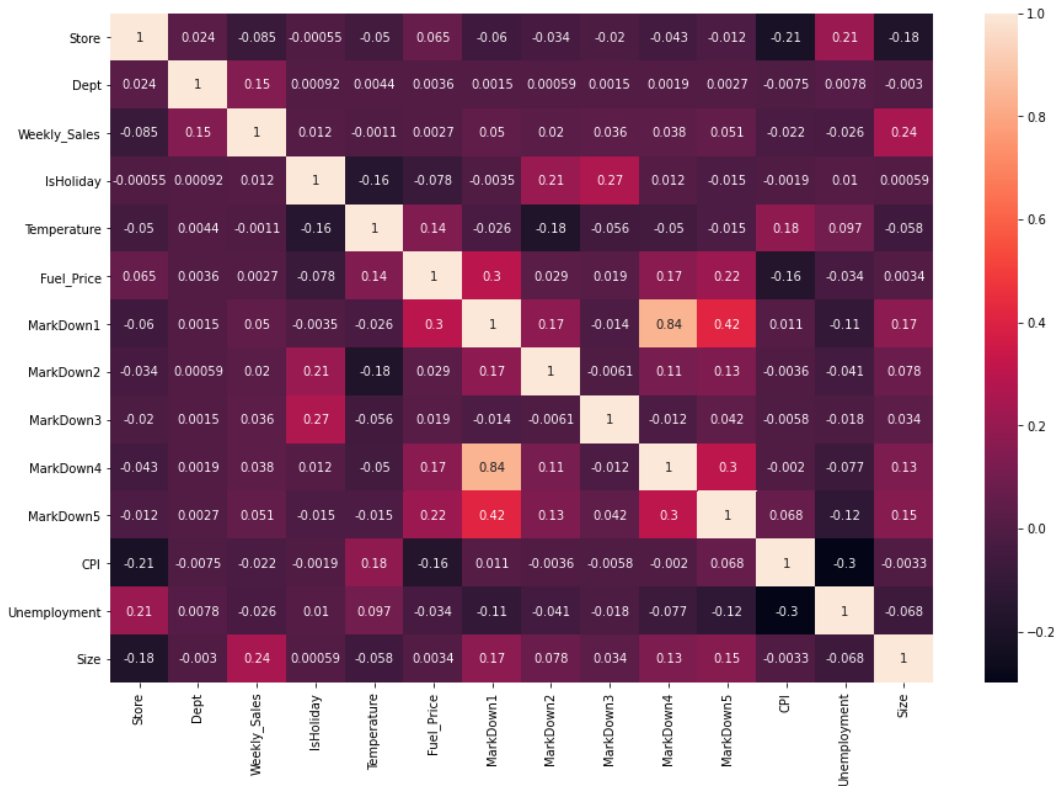
- ❖ Forecast future sales

Data Description

- Time scope: 2.5 years (Starting from 2010)
- Features type: store, date, weekly sales, temperature, fuel price, ...
- Stores type: A, B, C

	Store	Weekly_Sales	Temperature	Fuel_Price	MarkDown1	MarkDown2	MarkDown3	MarkDown4	MarkDown5	CPI	Unemployment	Size
count	421570.000000	282451.000000	421570.000000	421570.000000	421570.000000	421570.000000	421570.000000	421570.000000	421570.000000	421570.000000	421570.000000	421570.000000
mean	22.200546	15983.429692	60.090059	3.361027	2590.074819	879.974298	468.087665	1083.132268	1662.772385	171.201947	7.960289	136727.915739
std	12.785297	22661.092494	18.447931	0.458515	6052.385934	5084.538801	5528.873453	3894.529945	4207.629321	39.159276	1.863296	60980.583328
min	1.000000	-4988.940000	-2.060000	2.472000	0.000000	-265.760000	-29.100000	0.000000	0.000000	126.064000	3.879000	34875.000000
25%	11.000000	2079.330000	46.680000	2.933000	0.000000	0.000000	0.000000	0.000000	0.000000	132.022667	6.891000	93638.000000
50%	22.000000	7616.550000	62.090000	3.452000	0.000000	0.000000	0.000000	0.000000	0.000000	182.318780	7.866000	140167.000000
75%	33.000000	20245.745000	74.280000	3.738000	2809.050000	2.200000	4.540000	425.290000	2168.040000	212.416993	8.572000	202505.000000
max	45.000000	693099.360000	100.140000	4.468000	88646.760000	104519.540000	141630.610000	67474.850000	108519.280000	227.232807	14.313000	219622.000000

Correlation Matrix



- Discounts are correlated, generate higher sales
- Higher department numbers have higher sales
- Larger store generate more sales
- Little relationship between holidays, temperatures, or fuel prices with weekly sales

Linear Regression Model

Linear Regression

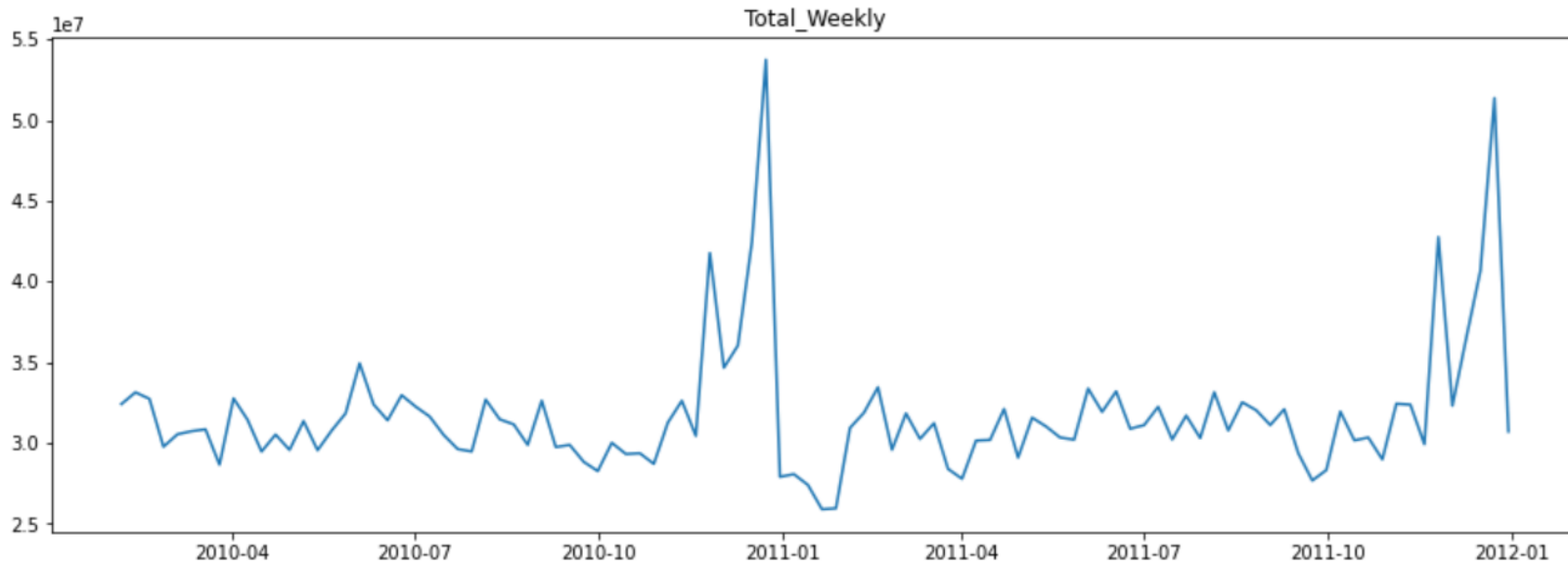
Predictors	Predicted
<ul style="list-style-type: none">• Temperature• Markdown 1-5• Fuel Price• CPI• Unemployment• Store type(A,B,C)• Store Size	<ul style="list-style-type: none">• Weekly Store/Department wise sales

```
Intercept:      [5318.13995152]
Coefficients:   [[ 3.20919973e+01 -7.44419827e+02  2.48390846e-02  1.31135440e-02
  1.19733848e-01 -5.01132189e-03  7.26738398e-02 -1.92687630e+01
 -2.94530147e+02  1.03317869e-01  1.60951115e+03]]
Root MSE test:  21512.740633601374
R^2 train:      0.06266184563738586
R^2 test:       0.06527778956334584
```

ARIMA Model

ARIMA Model

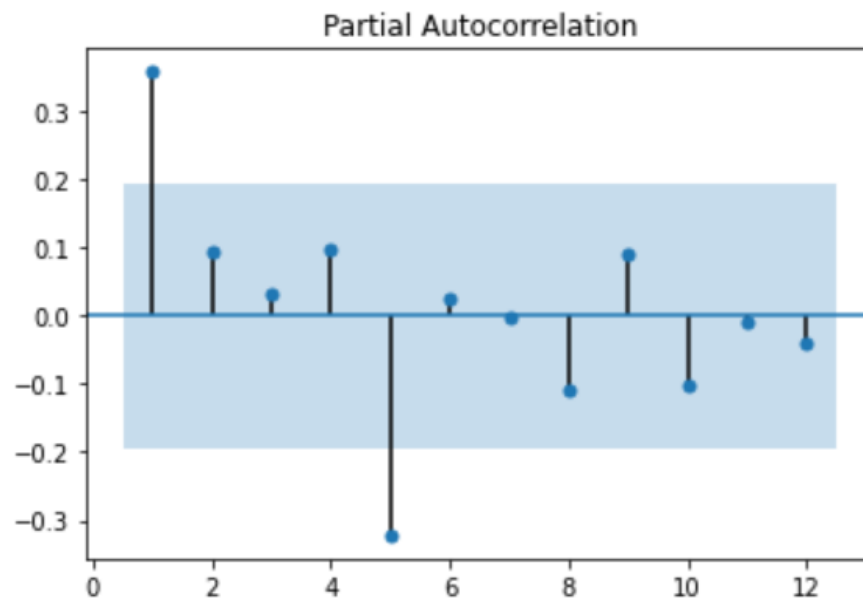
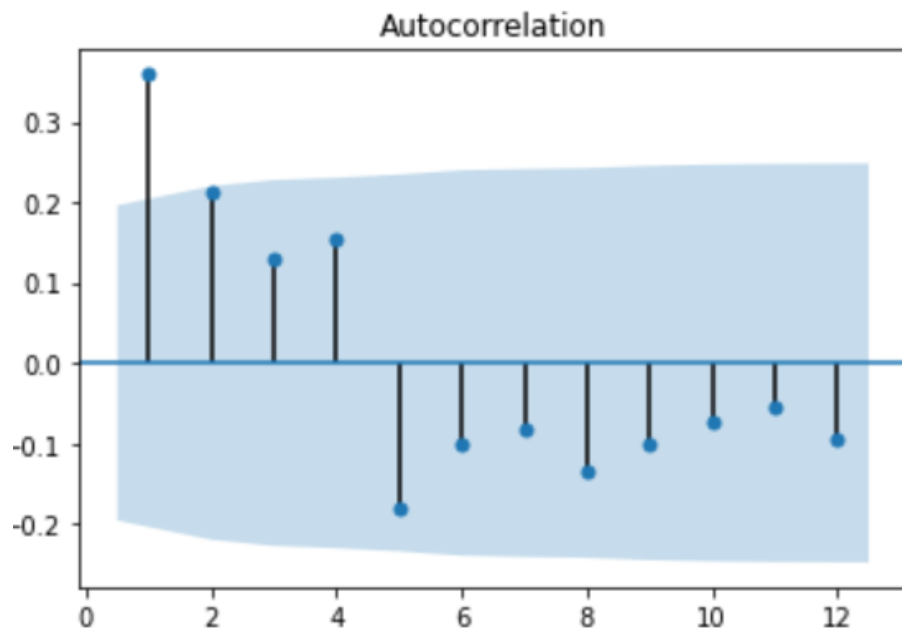
Total Weekly Walmart sales were forecasted instead of store/department wise sales.



The sales time series was observed to have some seasonality.

ARIMA Model - Parameter Selection

Total Walmart sales were forecasted instead of store/department wise sales.



Based on the ACF and PACF plots, ARIMA(2,0,1) was selected.

ARIMA Model - Results Summary

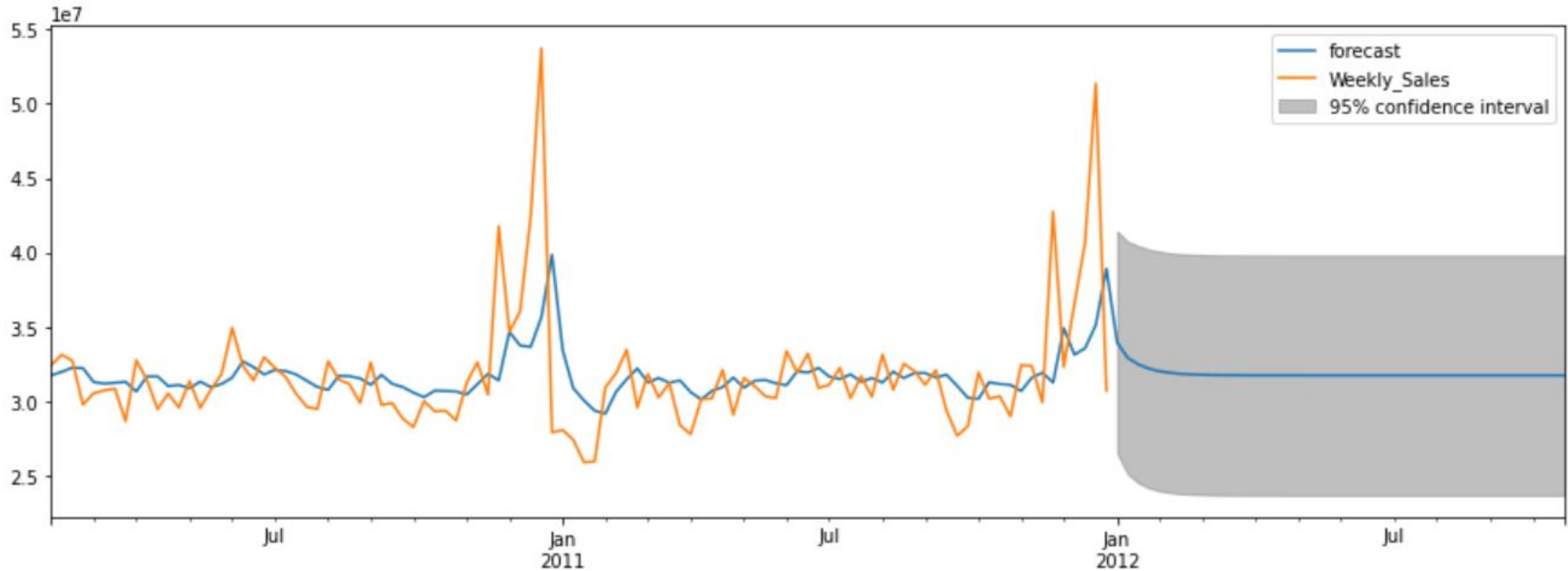
ARIMA Model Results						
=====						
Dep. Variable:	D.Weekly_Sales	No. Observations:	99			
Model:	ARIMA(2, 1, 1)	Log Likelihood	-1642.697			
Method:	css-mle	S.D. of innovations	3824270.328			
Date:	Tue, 23 Jun 2020	AIC	3295.395			
Time:	22:02:30	BIC	3308.370			
Sample:	02-12-2010	HQIC	3300.645			
	- 12-30-2011					
=====						
	coef	std err	z	P> z	[0.025	0.975]

const	2.431e+04	2.31e+04	1.053	0.295	-2.1e+04	6.96e+04
ar.L1.D.Weekly_Sales	0.3074	0.103	2.980	0.004	0.105	0.510
ar.L2.D.Weekly_Sales	0.1318	0.116	1.132	0.260	-0.096	0.360
ma.L1.D.Weekly_Sales	-0.9973	0.031	-32.440	0.000	-1.058	-0.937
Roots						
=====						
	Real	Imaginary	Modulus	Frequency		

AR.1	1.8251	+0.0000j	1.8251	0.0000		
AR.2	-4.1575	+0.0000j	4.1575	0.5000		
MA.1	1.0027	+0.0000j	1.0027	0.0000		

AR lag 1 and MA lag 1 coefficients have significant values.

ARIMA Model - Predictions

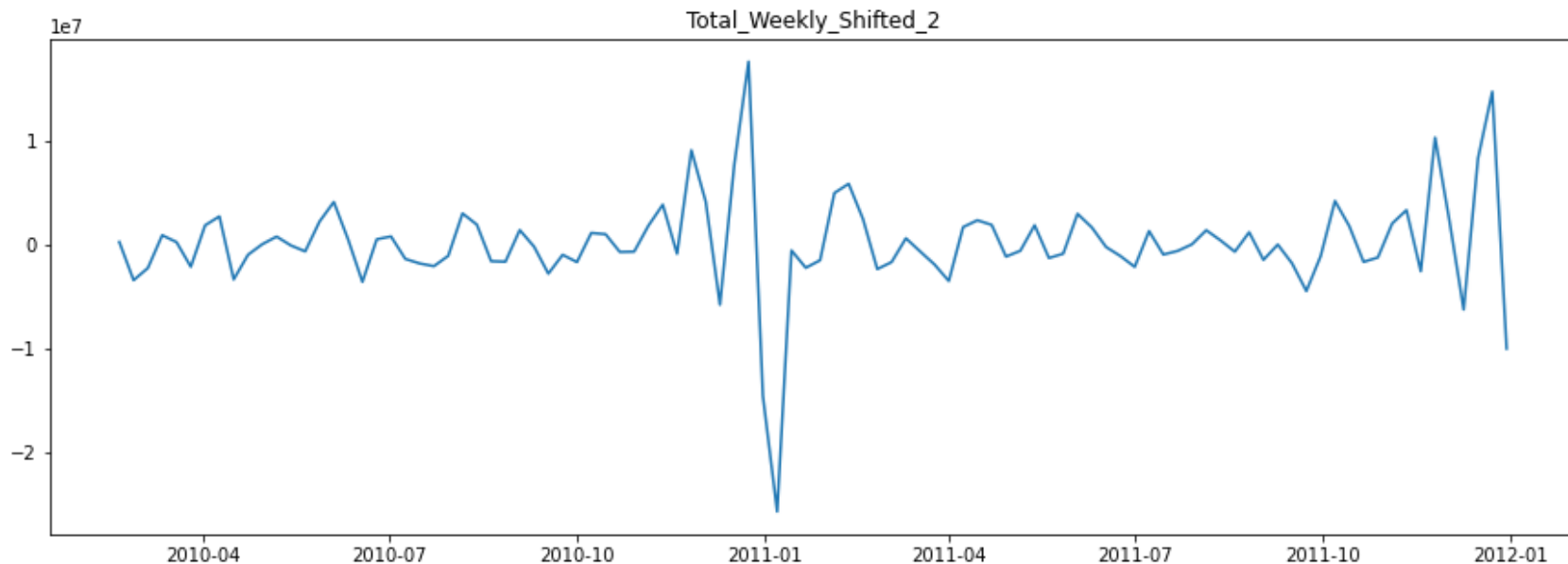


Total Weekly Sales for the next 6 months were predicted with an MAPE value of 0.99.

SARIMA Model

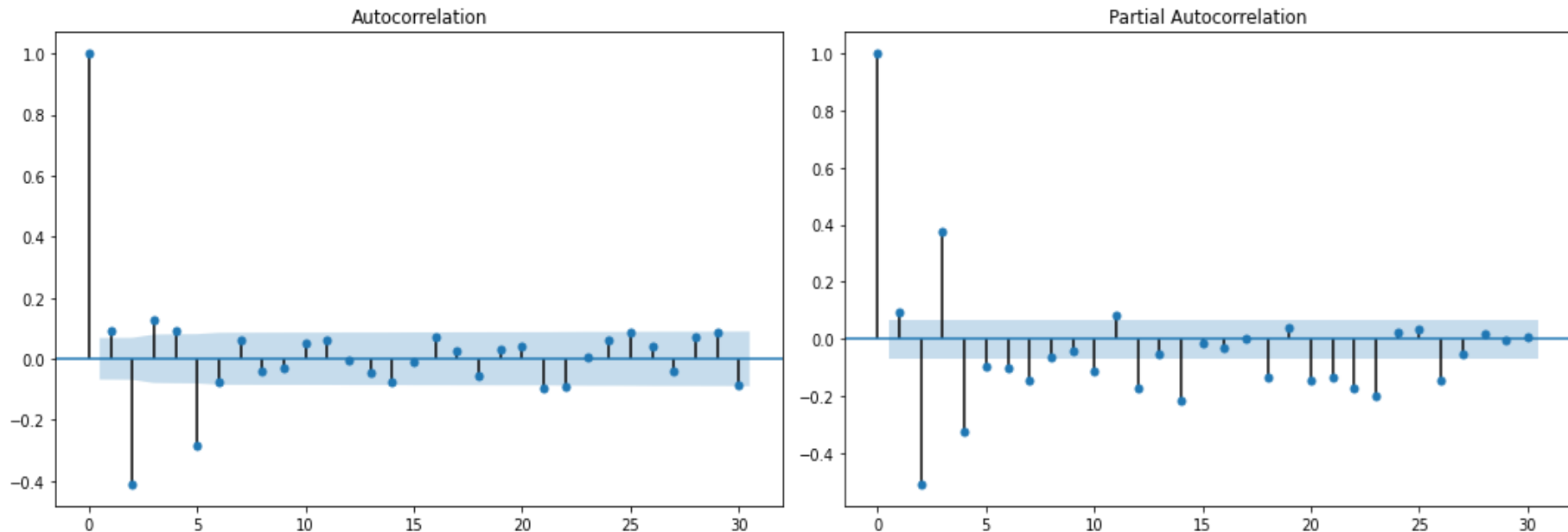
SARIMA Model

Total Weekly sales were taken instead of store/department wise sales.



The weekly sales was observed to be stationary.

SARIMA Model - Autocorrelation



There are large spikes and sharp drop-off in the graphs. The negative correlation suggests use of MA terms in relation to the lag.

SARIMA Model - Results Summary

Statespace Model Results

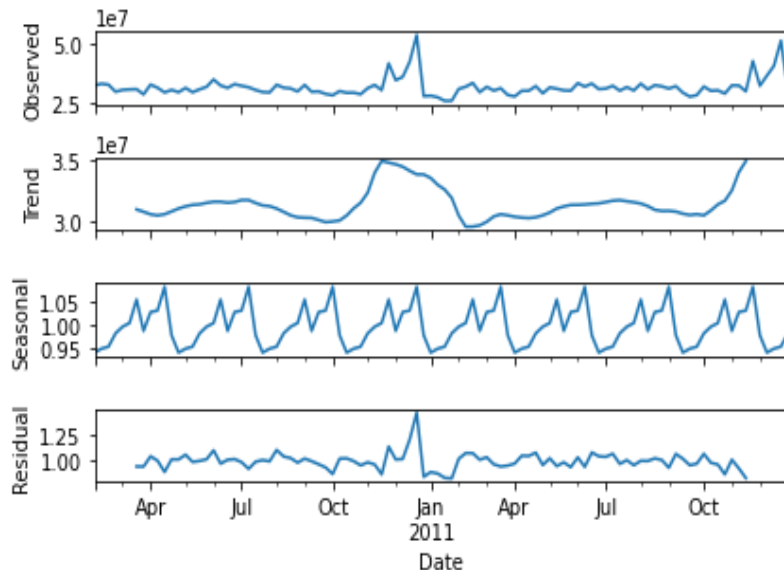
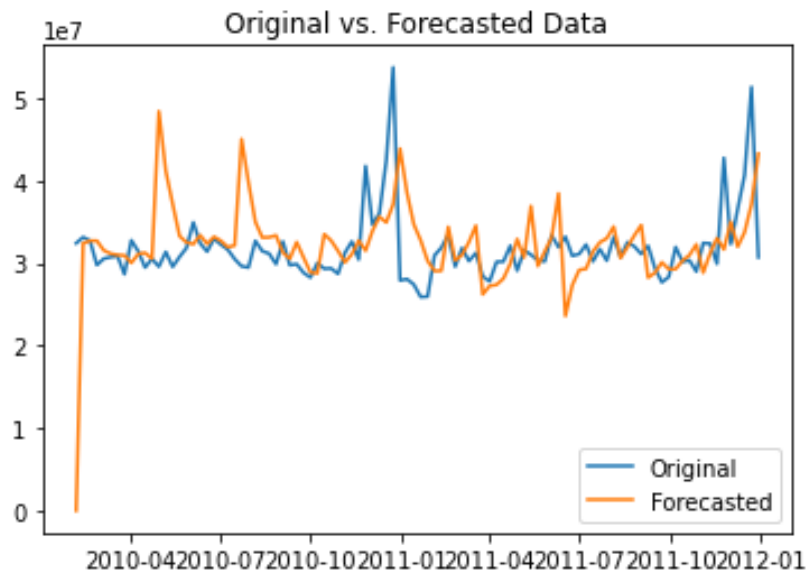
```
=====
Dep. Variable:          Weekly_Sales    No. Observations:          100
Model:                 SARIMAX(0, 1, 1)x(1, 1, 1, 12)    Log Likelihood          -1232.116
Date:                  Tue, 23 Jun 2020    AIC          2472.233
Time:                  18:49:00    BIC          2481.394
Sample:                02-05-2010    HQIC          2475.884
                  - 12-30-2011
Covariance Type:                opg
=====
```

	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.6047	0.062	-9.772	0.000	-0.726	-0.483
ar.S.L12	-0.3454	0.228	-1.513	0.130	-0.793	0.102
ma.S.L12	-0.3568	0.164	-2.171	0.030	-0.679	-0.035
sigma2	3.247e+13	1.75e-15	1.86e+28	0.000	3.25e+13	3.25e+13

```
=====
Ljung-Box (Q):          27.42    Jarque-Bera (JB):          19.42
Prob(Q):                0.93    Prob(JB):                0.00
Heteroskedasticity (H):  0.66    Skew:                    0.35
Prob(H) (two-sided):    0.31    Kurtosis:                5.43
=====
```

Both AIC and BIC are lower than ARIMA model. MAPE is calculated to be 0.42.

SARIMA Model - Prediction



The forecasted data looks fairly different from the actual data. There is also a major drop in the beginning. The decomposition model is clearer.

Conclusion

Conclusion

Best Model for the dataset

- Among our models, the MAPE is:
 - ARIMA - 0.99
 - SARIMA - 0.42
- Hence, SARIMA is the best model
- The linear regression model's RMSE can be used as a baseline for other regression models

Future Recommendation

- Use other techniques such as KNN regression or Random forest
- Attempt various regressor methods and get the average of the top models. The overfitting issue can be verified by checking RMSE or MAE
- Modify date feature into days, month, weeks

Thank You