



Walmart Forecasting

Problem Statement

- Attempting to predict weekly sales



Data Collection and Data Wrangling

- Kaggle

-

	Store	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	stock_price	sp500
0	1	2010-02-05	1643690.90	0	42.31	2.572	211.096358	8.106	53.45000	1066.19
1	1	2010-02-12	1641957.44	1	38.51	2.548	211.242170	8.106	52.89999	1075.51
2	1	2010-02-19	1611968.17	0	39.93	2.514	211.289143	8.106	53.49001	1109.17
3	1	2010-02-26	1409727.59	0	46.63	2.561	211.319643	8.106	54.07001	1104.49
4	1	2010-03-05	1554806.68	0	46.50	2.625	211.350143	8.106	54.14000	1138.69

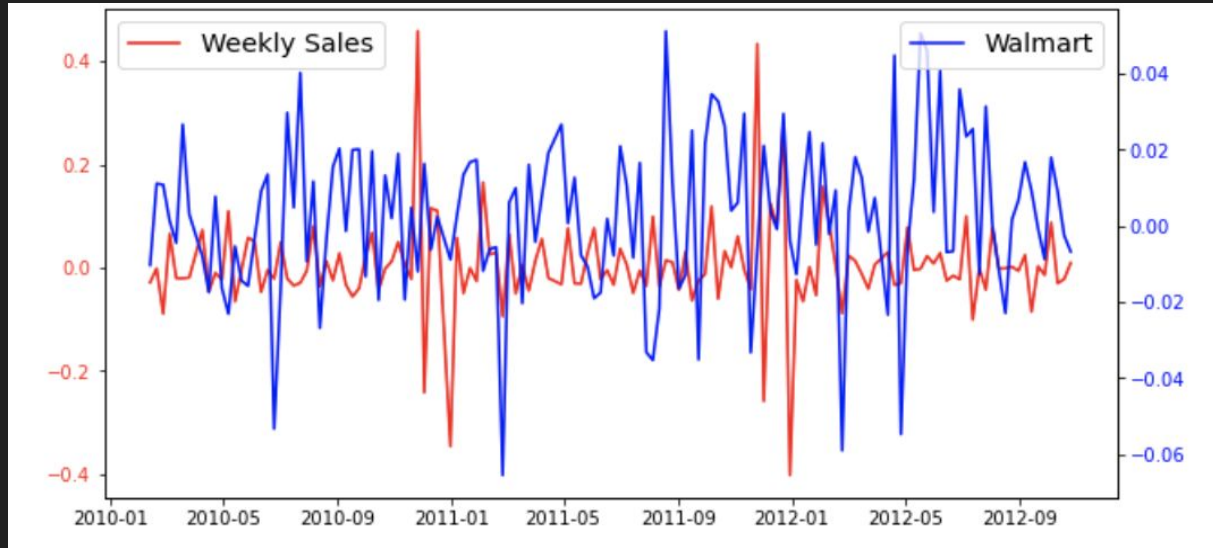
Exploratory Data Analysis

Exploratory Data Analysis

Weekly Sales Vs. Walmart Stock Prices (Percentage Change)

Exploratory Data Analysis

Weekly Sales Vs. Walmart Stock Prices (Percentage Change)

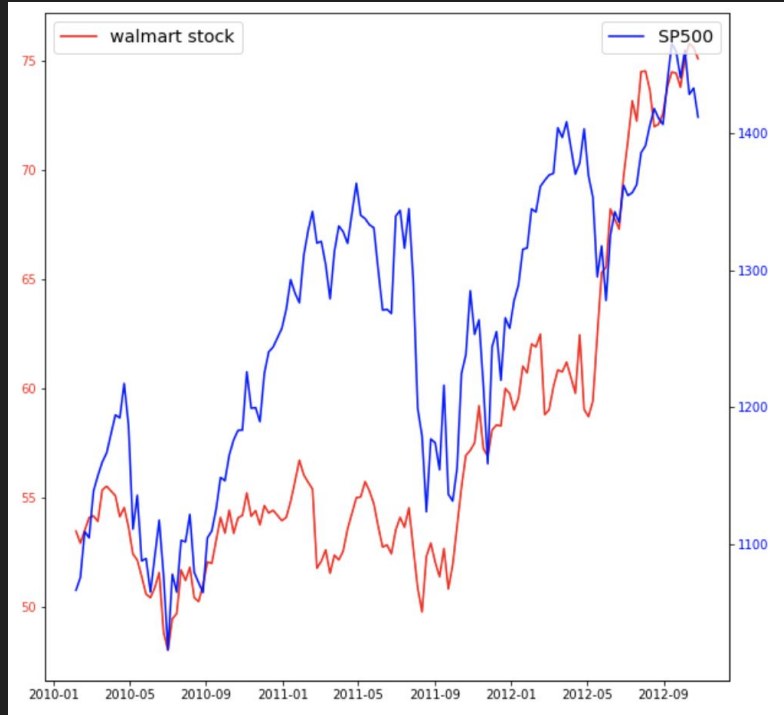


Exploratory Data Analysis

Walmart Stocks vs SP500

Exploratory Data Analysis

Walmart Stocks vs SP500

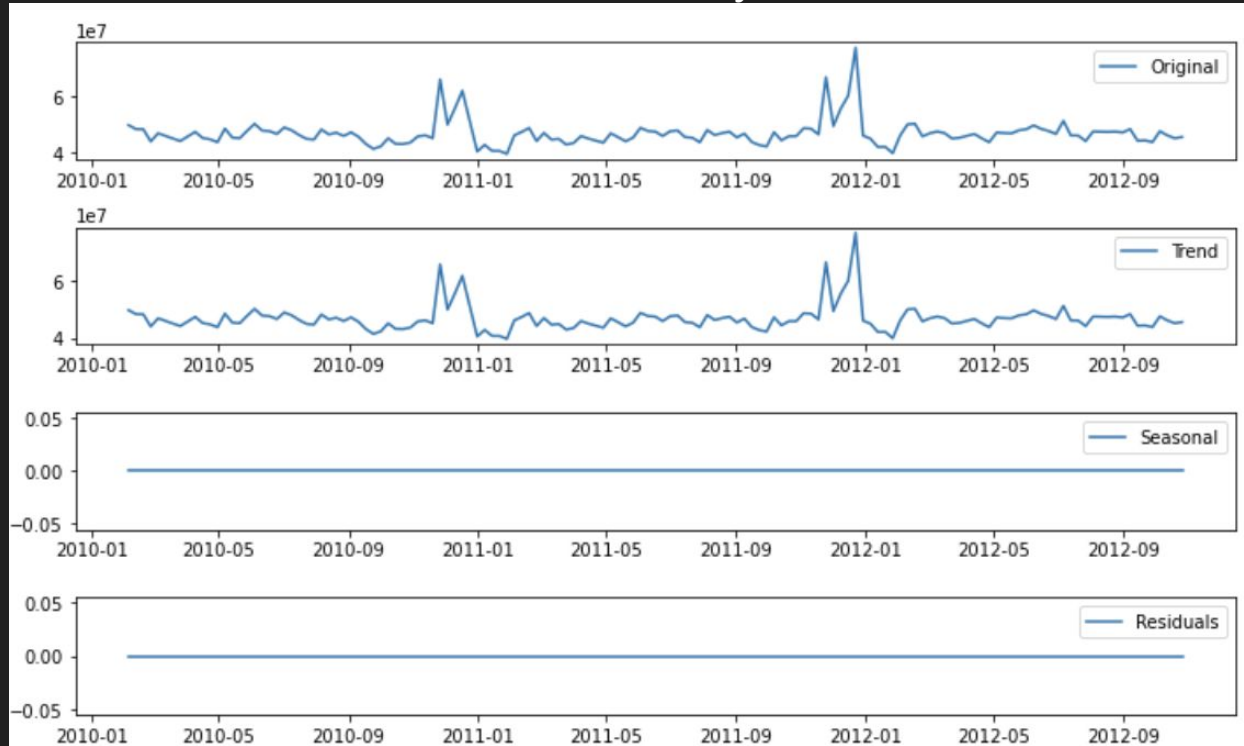


Exploratory Data Analysis

Seasonality

Exploratory Data Analysis

Seasonality

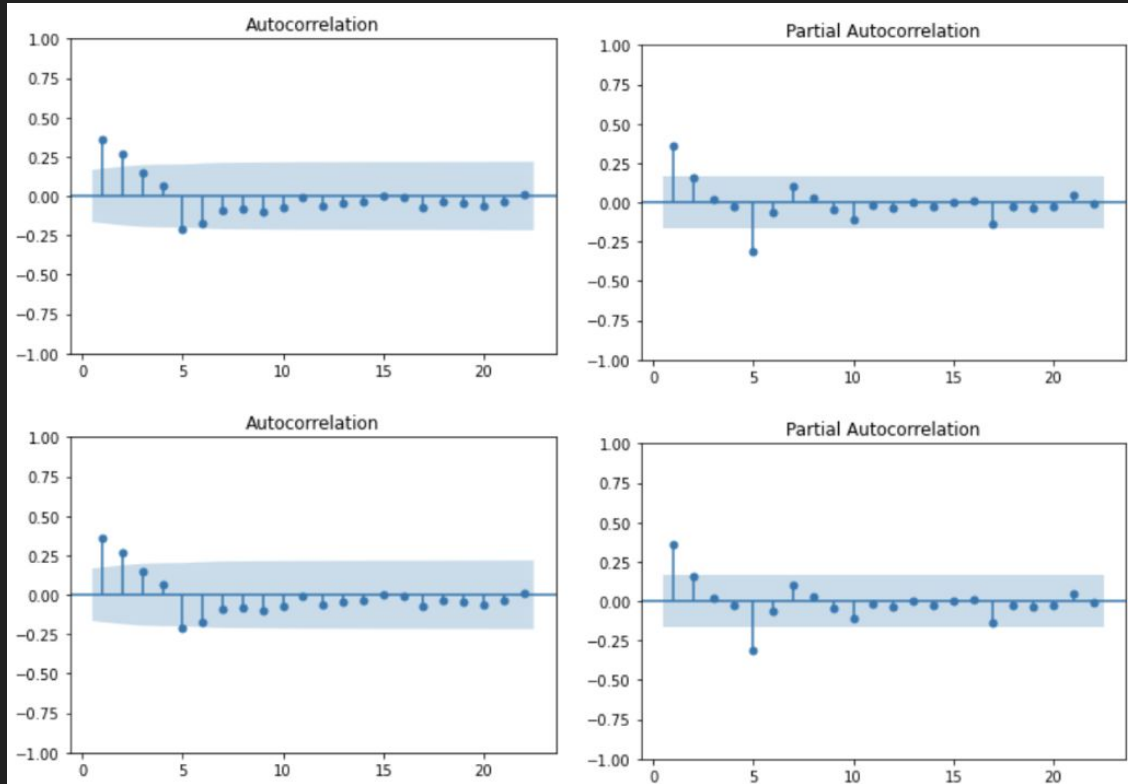


Exploratory Data Analysis

Autocorrelation vs Partial Autocorrelation

Exploratory Data Analysis

Autocorrelation vs Partial Autocorrelation

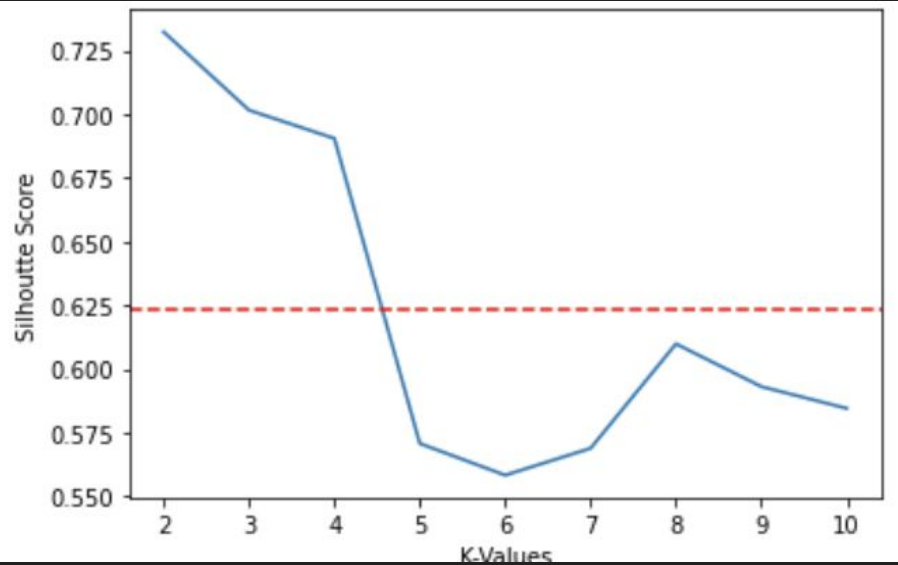


Pre-Processing

- Location -> Clustering based on Temperature and Gas Price
-

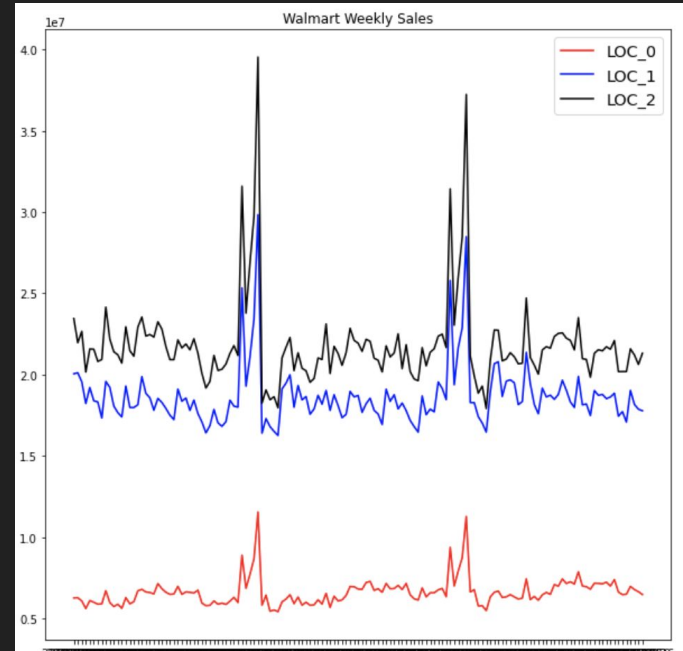
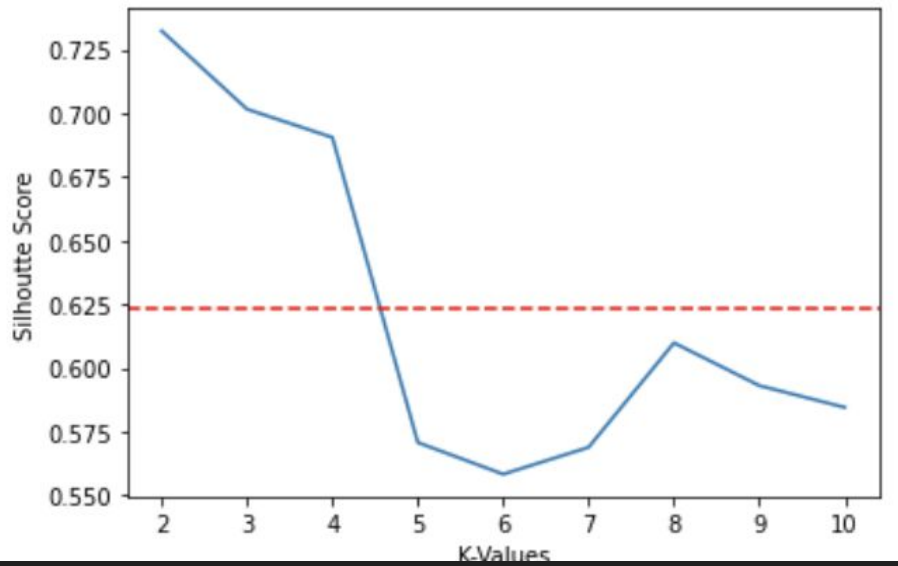
Pre-Processing

- Location -> Clustering based on Temperature and Gas Price



Pre-Processing

- Location -> Clustering based on Temperature and Gas Price



Pre-Processing

Alternative Dataset with Locations

One hot encoding

	Date	Weekly_Sales	Holiday_Flag	Temperature	Fuel_Price	CPI	Unemployment	stock_price	sp500	loc_0	loc_1	loc_2
0	2010-02-05	1643690.90	0	42.31	2.572	211.096358	8.106	53.45000	1066.19	0	1	0
1	2010-02-12	1641957.44	1	38.51	2.548	211.242170	8.106	52.89999	1075.51	0	1	0
2	2010-02-19	1611968.17	0	39.93	2.514	211.289143	8.106	53.49001	1109.17	0	1	0
3	2010-02-26	1409727.59	0	46.63	2.561	211.319643	8.106	54.07001	1104.49	0	1	0
4	2010-03-05	1554806.68	0	46.50	2.625	211.350143	8.106	54.14000	1138.69	0	1	0

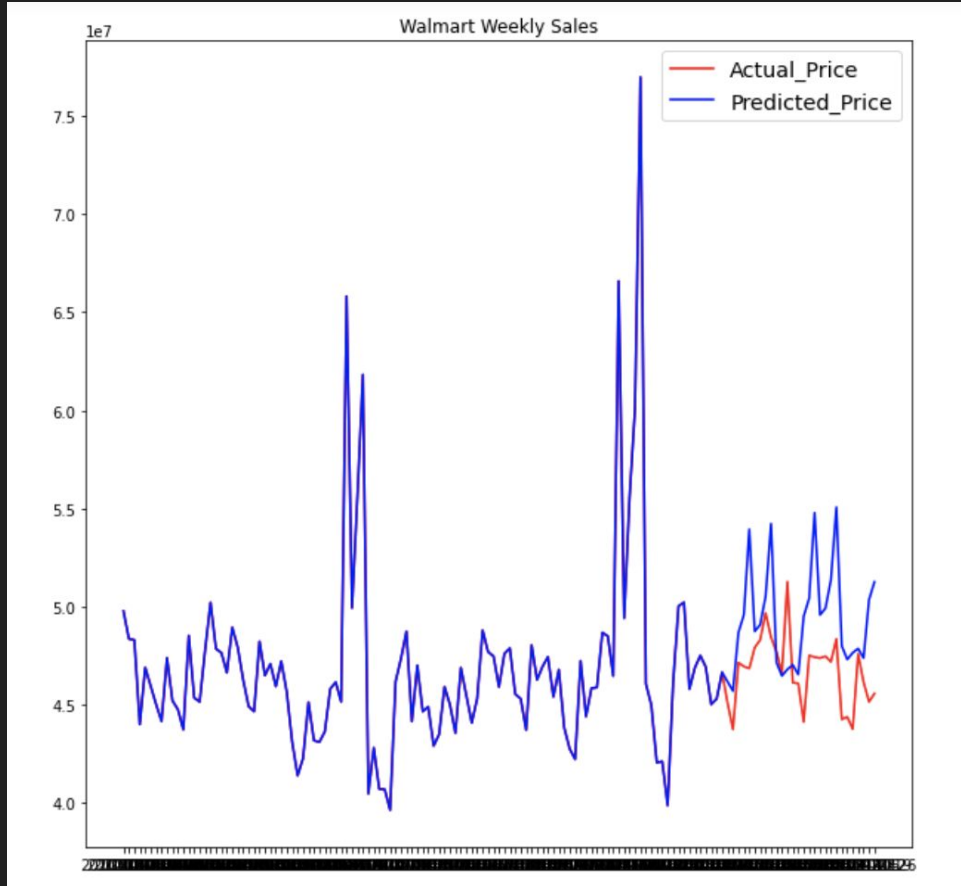
Pre-Processing

Dataset with Target Variables

	time	Weekly_Sales	Walmart	sp500	Temperature	Fuel_Price	CPI	Unemployment	Holiday_Flag	Target
0	2010-02-05	49750740.50	53.45000	1066.19	34.037333	2.717844	167.730885	8.619311	0	-0.028423
1	2010-02-12	48336677.63	52.89999	1075.51	34.151333	2.694022	167.825608	8.619311	1	-0.001235
2	2010-02-19	48276993.78	53.49001	1109.17	37.719778	2.672067	167.871686	8.619311	0	-0.089244
3	2010-02-26	43968571.13	54.07001	1104.49	39.243556	2.683933	167.909657	8.619311	0	0.066022
4	2010-03-05	46871470.30	54.14000	1138.69	42.917333	2.731200	167.947628	8.619311	0	-0.020184
...
133	2012-09-21	44354547.11	74.45000	1460.15	67.924889	3.907911	176.242124	7.237333	0	-0.013970
134	2012-09-28	43734899.40	73.80000	1440.67	68.754444	3.854578	176.373588	7.237333	0	0.087613
135	2012-10-05	47566639.31	75.13000	1460.93	65.973111	3.845222	176.505052	6.953711	0	-0.030234
136	2012-10-12	46128514.25	75.81000	1428.59	58.342667	3.896733	176.636515	6.953711	0	-0.021811
137	2012-10-19	45122410.57	75.62000	1433.19	60.705333	3.880000	176.652613	6.953711	0	0.009346

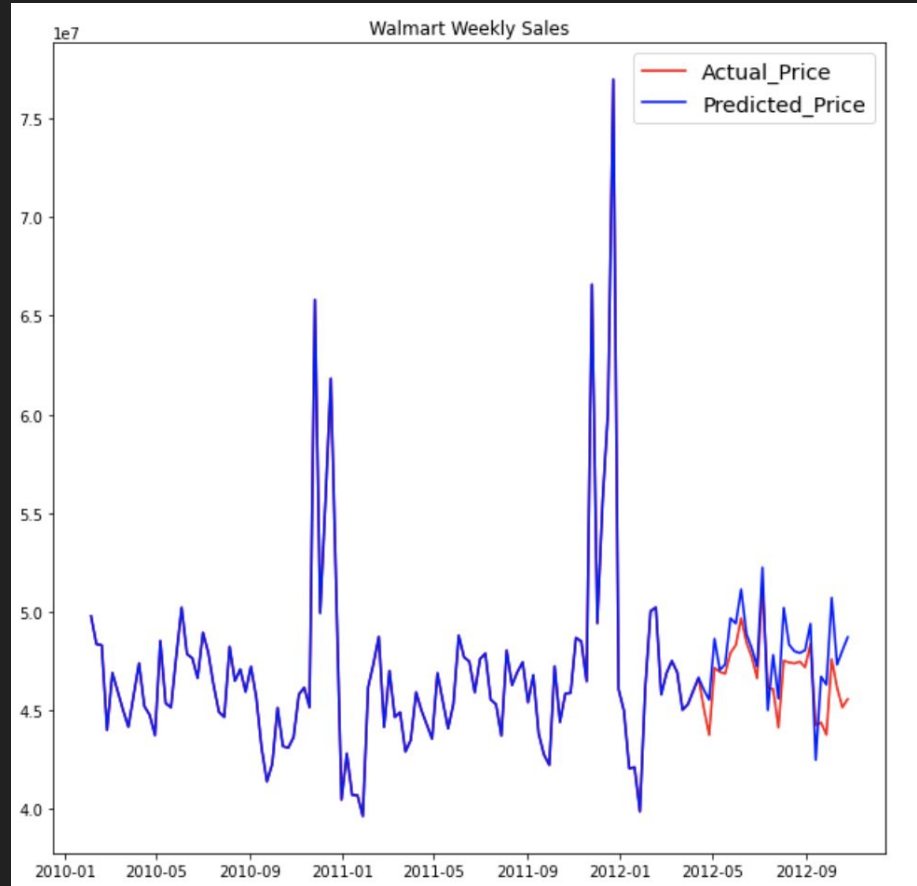
Modeling

- ARIMA



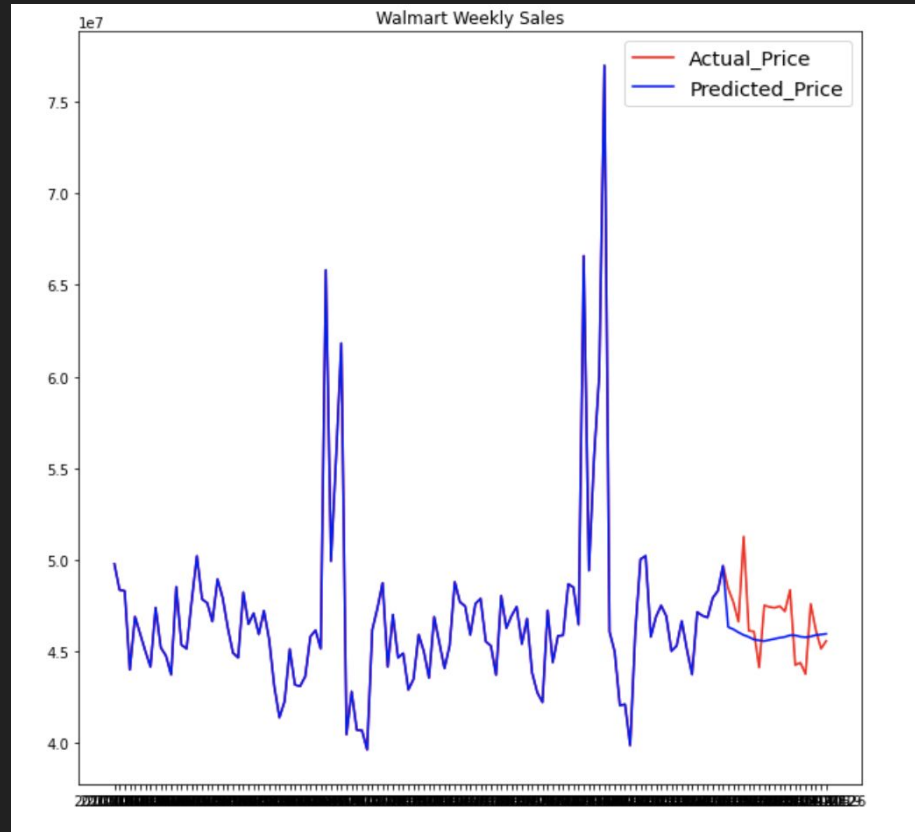
Modeling

- SARIMAX with Exogenous Variables



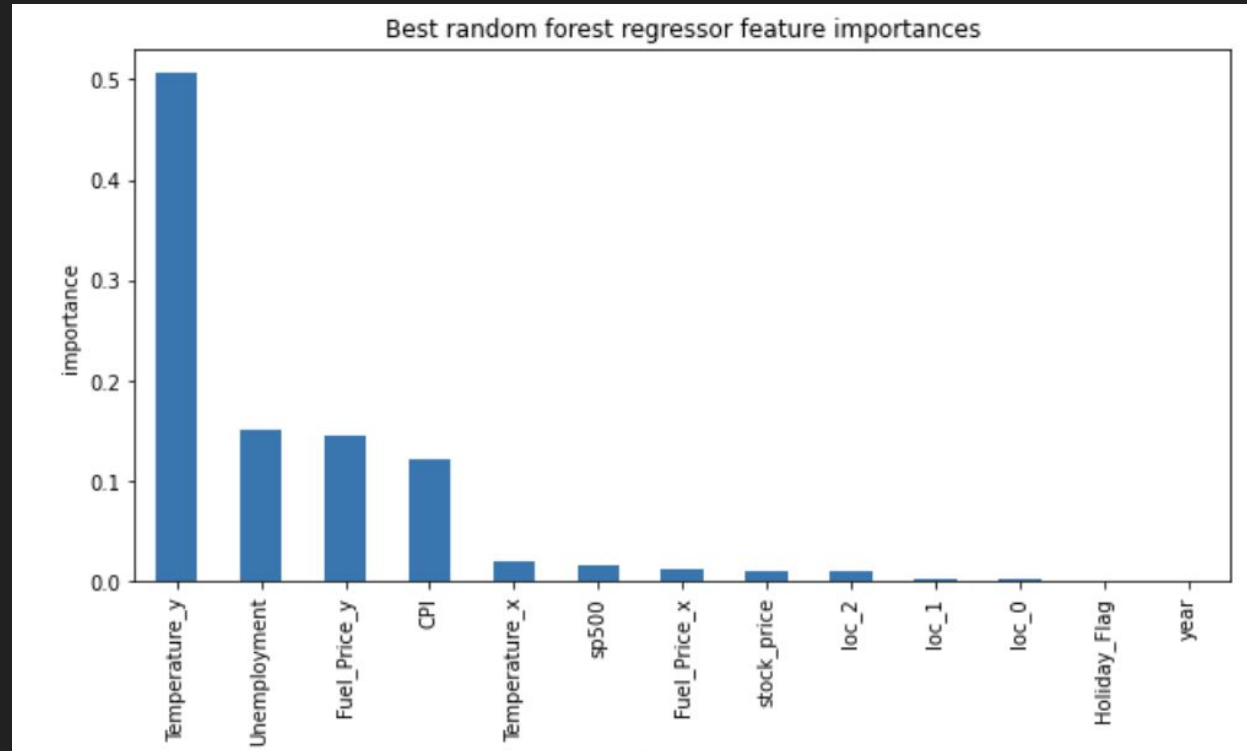
Modeling

- LSTM

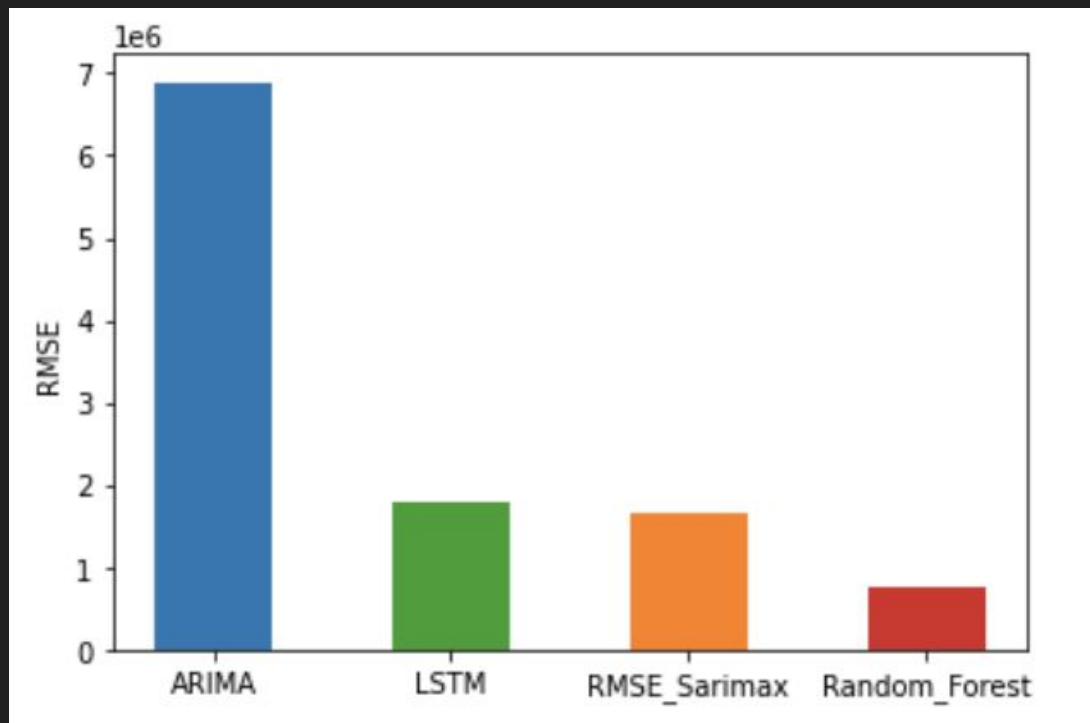


Modeling

- Random Forest



Model Selection



Future Improvement

- Acquire More Data
- Parallel Training
- Multiple Inputs LSTM