Walmart Stores Sales Forecasting

MSDS 596 - Regression & Time Series Analysis
Group D - Shourya Thatha Ravi, Rahul Rajesh Singh, Raviteja Arava, and Nish Patel
December 8, 2022

AGENDA

INTRODUCTION

INTRODUCTION

Objective: Predict weekly sales at Walmart using historical data provided for 45 stores and 81 departments

Importance: Through the prediction of future weekly sales across various stores, the company can manage their supply chain more efficiently, optimize inventory levels, accommodate higher volume of products, handle traffic, and provide markdowns accordingly.

Motivation: Forecasting sales is also a common task for data scientists so it can provide a good opportunity for the group to learn techniques and experiment with real-world data.

Dataset: Dataset was provided by Walmart for a recruiting competition on Kaggle.

DATASET

stores.csv

This file contains anonymized information about the 45 stores

- Store - the store number
- Type - the store type (A, B, C)
- Size - the size of the store

train.csv

This file contains the historical sales information by store and department

- Store - the store number
- Dept - the department number
- Date - the week
- Weekly_Sales - sales for the given department in the given store
- IsHoliday - whether the week is a special holiday week

DATASET

features.csv

This file contains information on additional factors that may impact weekly sales (such as promotional markdowns)

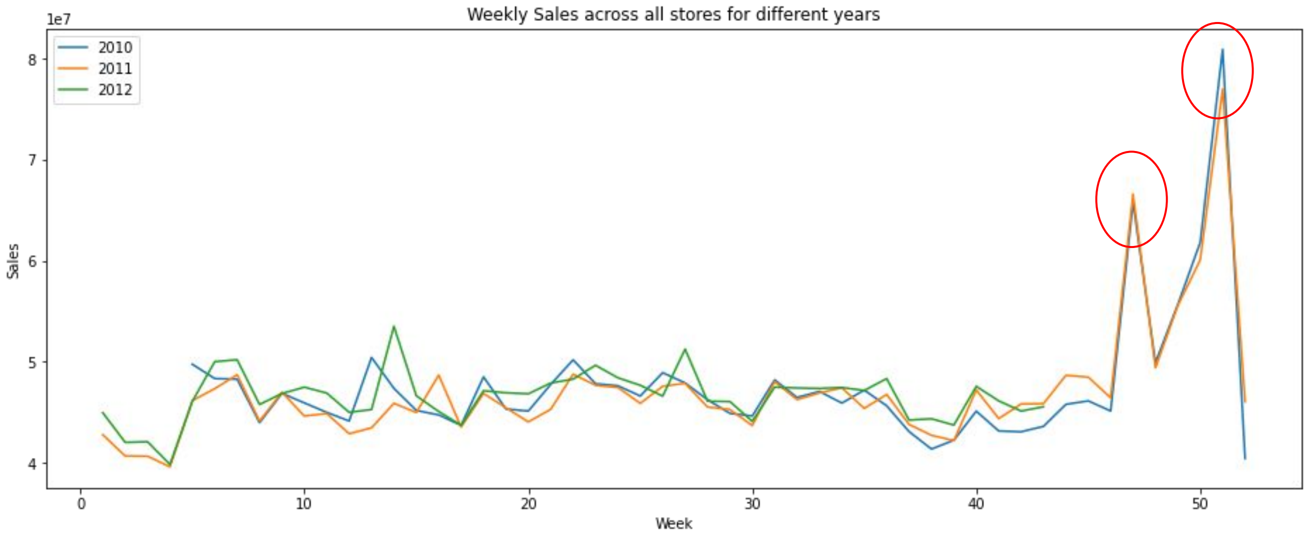
- Store - the store number
- Date - the week
- Temperature - average temperature in the region
- Fuel_Price - cost of fuel in the region
- MarkDown1-5 - anonymized data related to promotional markdowns that Walmart is running. MarkDown data is only available after Nov 2011, and is not available for all stores all the time. Any missing value is marked with an NA.
- CPI - the consumer price index
- Unemployment - the unemployment rate
- IsHoliday - whether the week is a special holiday week

EXPLORATORY DATA ANALYSIS

EXPLORATORY DATA ANALYSIS

Variable	# of nulls
Store	0
Date	0
Temperature	0
Fuel_Price	0
MarkDown1	4158
MarkDown2	5269
MarkDown3	4577
MarkDown4	4726
MarkDown5	4140
CPI	585
Unemployment	585
IsHoliday	0

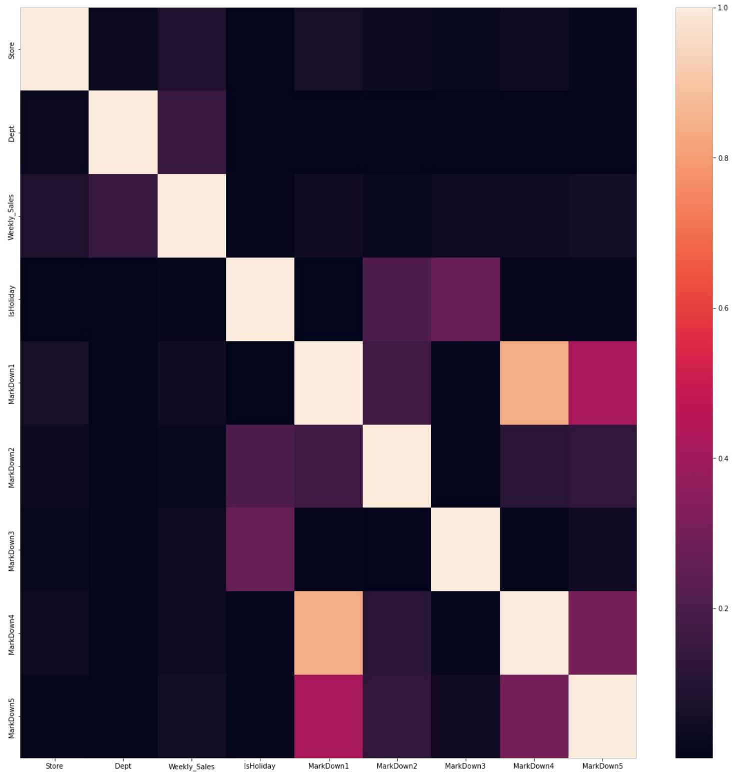
Replaced nulls with 0



Seasonality in the dataset shows spikes in weekly sales with holidays (Christmas and Thanksgiving especially).

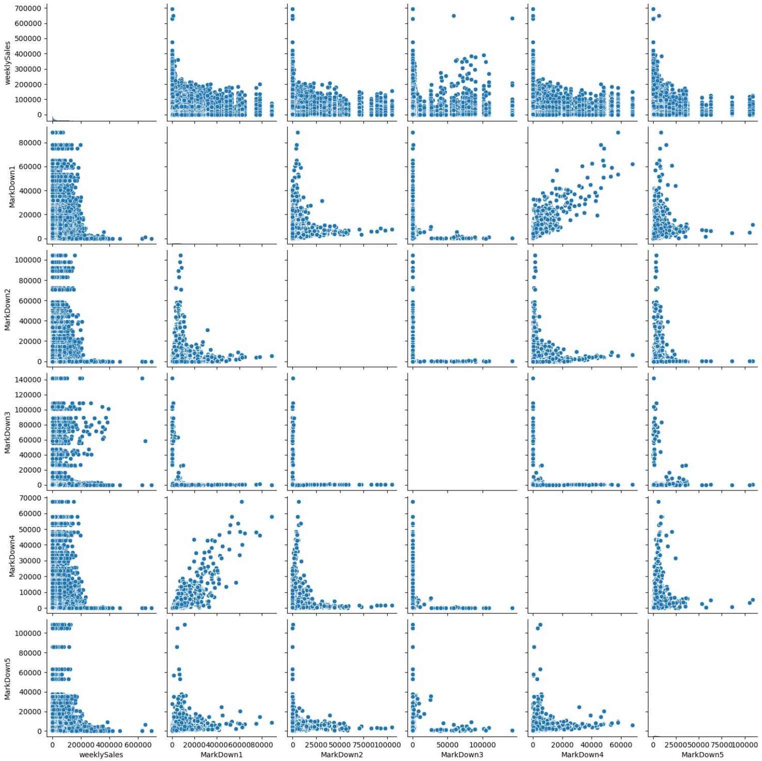
EXPLORATORY DATA ANALYSIS

Correlation Heat Map



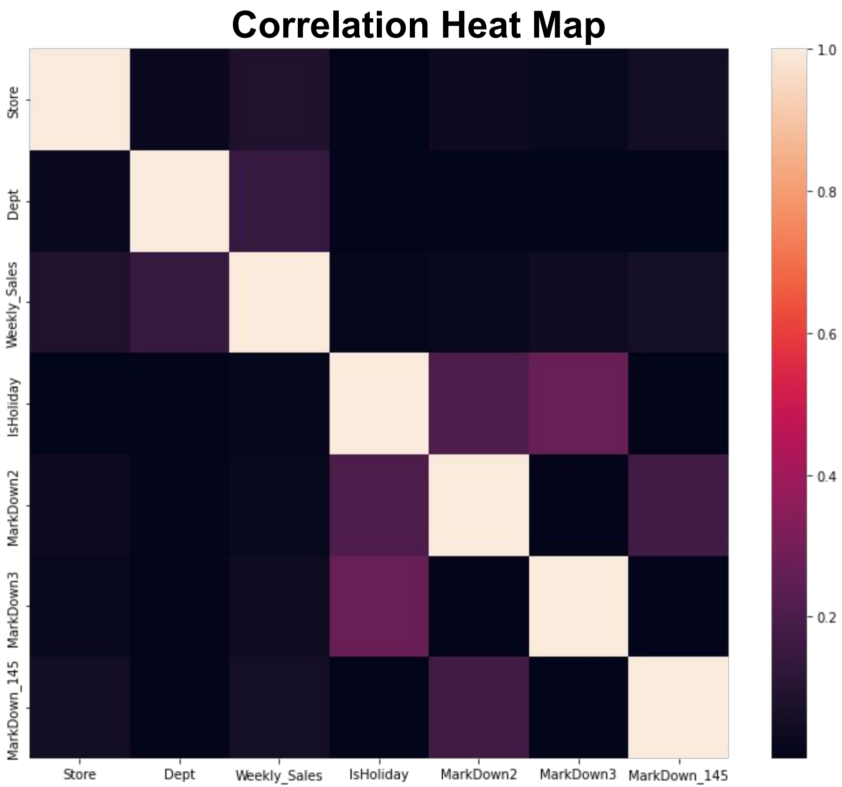
Some of the markdowns (namely 1, 4, & 5) show multicollinearity

Pairwise Plots

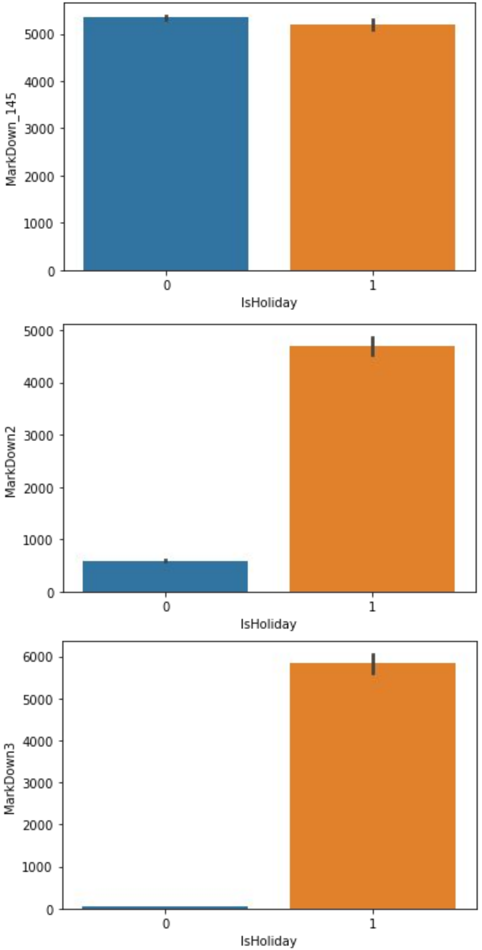


Pairwise plots clearly show multicollinearity between Markdown 1 and 4

EXPLORATORY DATA ANALYSIS

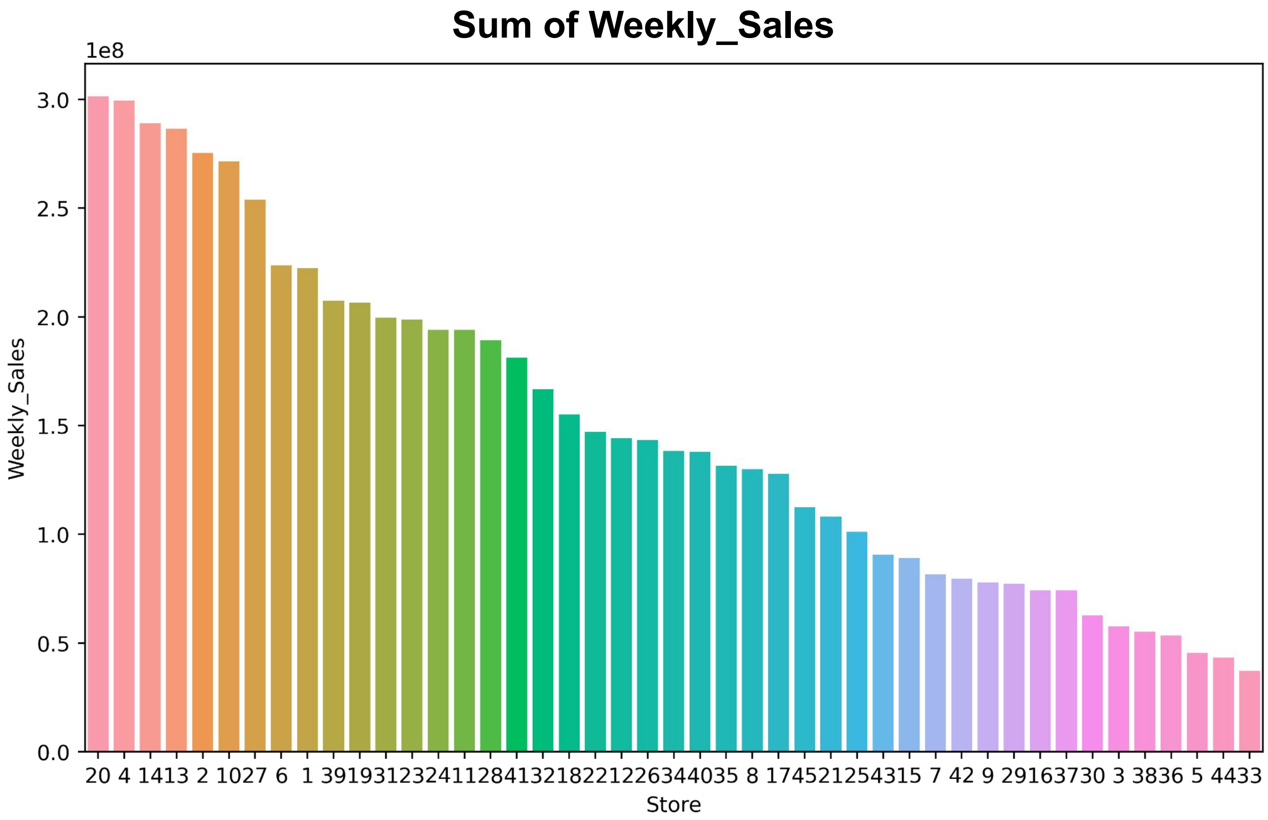


Markdowns 1, 4, and 5 were combined into one variable by summing them. This removed the multicollinearity issue.



isHoliday is correlated with Markdown2 and Markdown3 from the correlation heat map. When checking the flag vs Markdown2 and Markdown3, it is almost always true when these markdowns are active so we have dropped the isHoliday column.

EXPLORATORY DATA ANALYSIS



Stores 20 and 4 have the highest weekly sales

MODELING METHODOLOGY

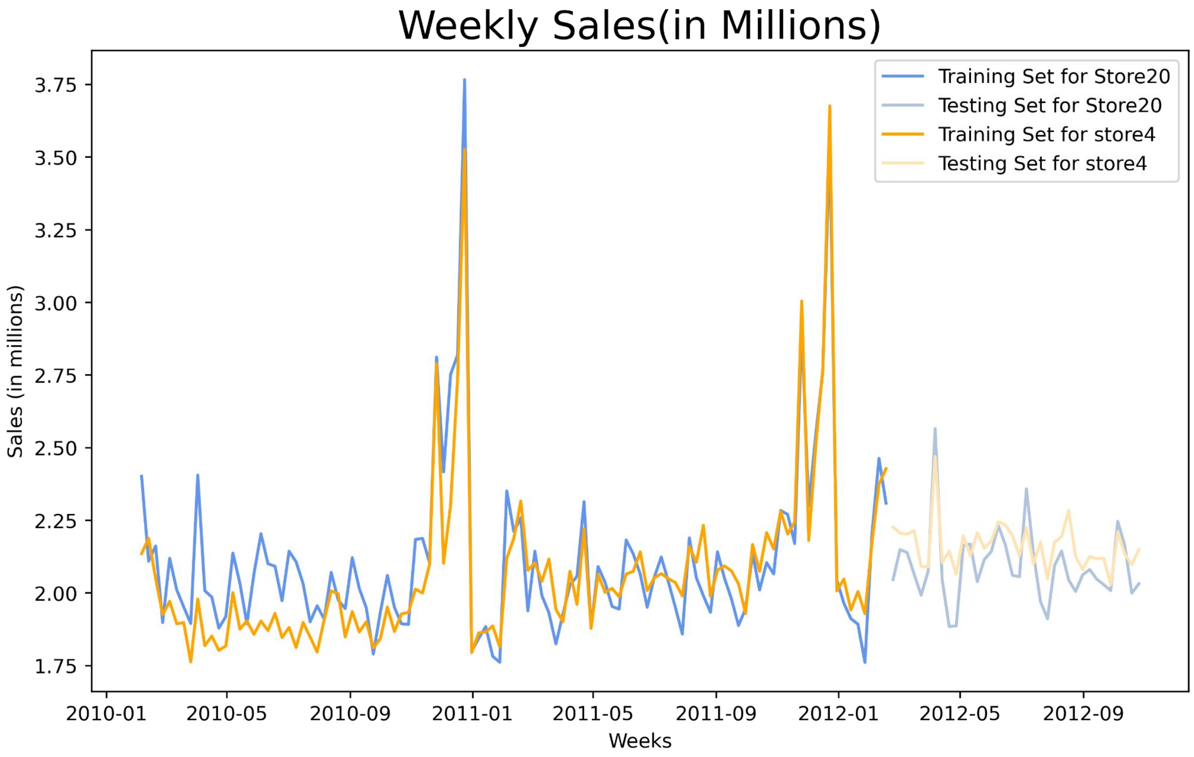
MODELING METHODOLOGY

- Attempt to use time-series modeling (univariate and multivariate) to predict Walmart weekly sales
- Focus on store-level predictions for the top 2 stores with the highest weekly sales (stores 20 and 4)
- Utilize 75% of data as the training set and the remaining 25% as the test set
- Use Mean Absolute Percentage Error (MAPE) as the evaluation metric to determine the best model
    - MAPE is selected because it is expressed as a percentage which makes it scale-independent which is useful because the magnitude of weekly_sales varies across time and across stores.
    - The formula for MAPE is below where A_t is the actual value, F_t is the forecasted value, and n is the sample size

$$\mathrm{M} = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{A_t - F_t}{A_t}\right|$$

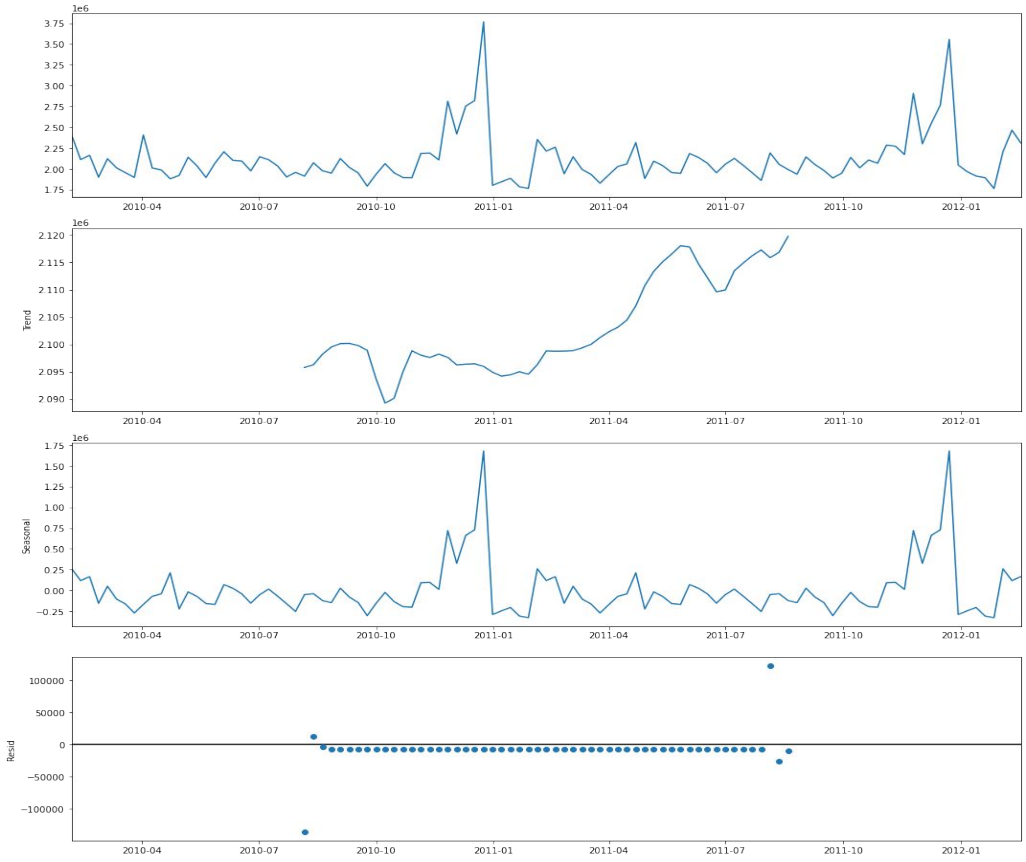
UNIVARIATE MODELING

Line Plot of the Data



- No upward/downward trend in general
- Seasonality observed (attributed to holidays)
- Additive seasonality
- Somewhat constant variance
- No outliers observed

Seasonal Decomposition



- An upward trend is noticed in the second graph
- The upward trend can be mitigated with differencing
- Seasonality is confirmed through the repeated peaks near the Thanksgiving and Christmas holidays in the third graph
- A few outliers are observed in the residual plot at the beginning and end of the dataset

Stationarity - Augmented Dickey Fuller (ADF)

H_0: Data is non-stationary **H_A**: Data is stationary

ADF Test Results

Weekly_Sales - Store 20

```
Results of Dickey-Fuller Test:
Test Statistic                  -4.719934
p-value                          0.000077
#Lags Used                       4.000000
Number of Observations Used    102.000000
Critical Value (1%)             -3.496149
Critical Value (5%)             -2.890321
Critical Value (10%)            -2.582122
```

ADF Test Results

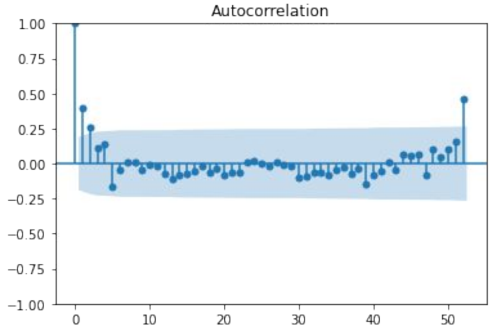
Weekly_Sales - Store 4

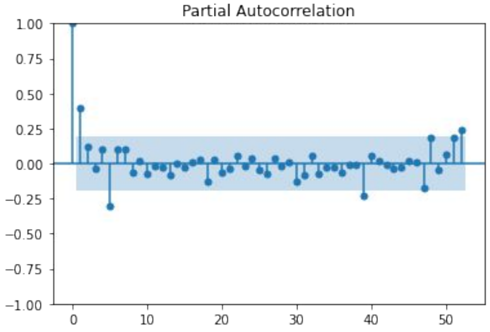
```
Results of Dickey-Fuller Test:
Test Statistic                  -3.509825
p-value                          0.007738
#Lags Used                       4.000000
Number of Observations Used    102.000000
Critical Value (1%)             -3.496149
Critical Value (5%)             -2.890321
Critical Value (10%)            -2.582122
```

The p-value for the ADF test for both stores is below 0.05 and the Test Statistic is less than the critical value at 5% so we can reject the null hypothesis, confirming that Weekly_Sales is stationary for these stores.

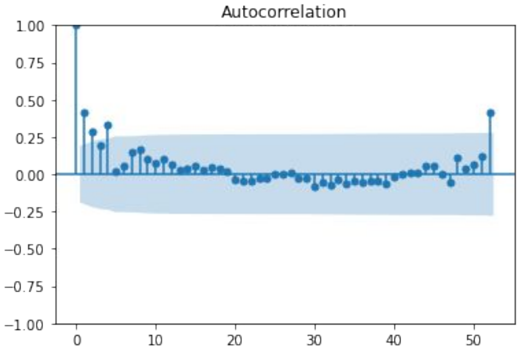
ACF & PACF Plot

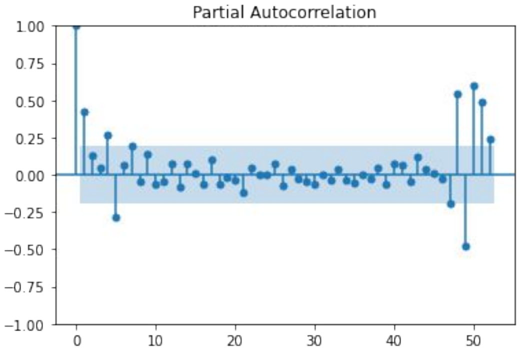
Store 20





Store 4





The ACF Plot reduces gradually, with lag up to 2 being significant

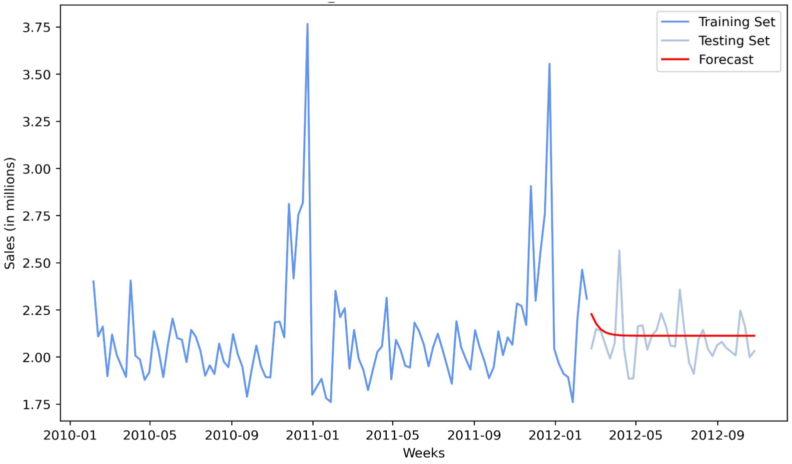
The PACF Plot is irregular with a single, significant lag of 1, before decreasing besides some lags spiking here and there

A lag of 1 for store 20 and lag of 4 for store 4 are chosen to begin modeling due to the PACF plots.

Auto Regression(AR)

Store 20

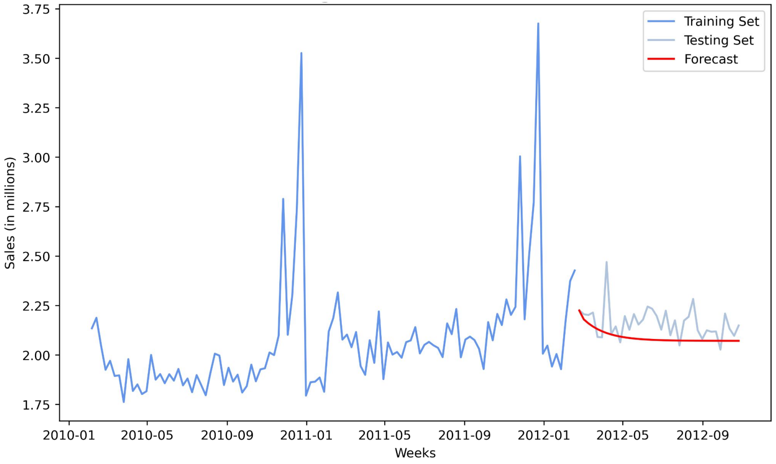
AR with lag 1



MAPE: 4.4%

Store 4

AR with lag 4



MAPE: 3.3%

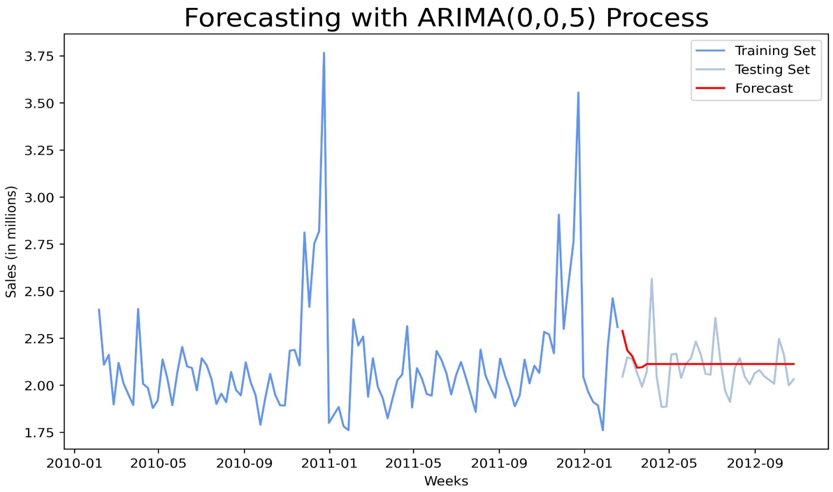
AR was chosen as the first model for univariate testing based on the ACF and PACF plots.

The forecast was mainly just the average of the testing dataset for all weeks, which is not correct and yields poor results.

Auto ARIMA

Store 20

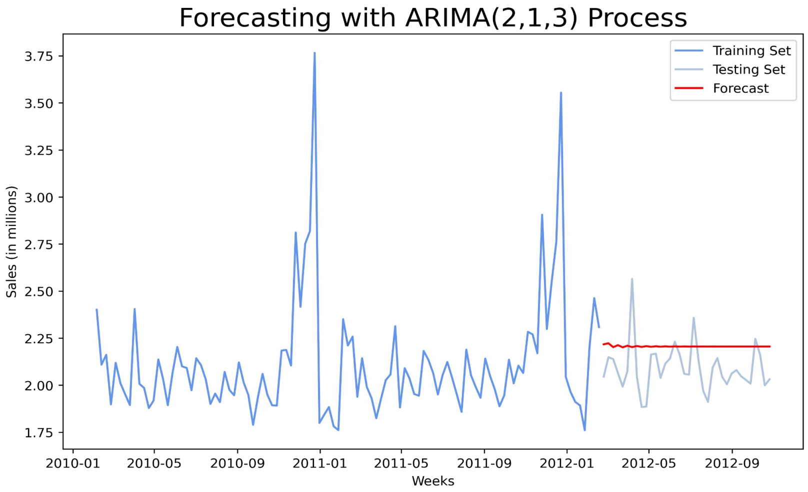
ARIMA(0,0,5)



MAPE: 4.6%

Store 4

ARIMA(2,1,3)



MAPE: 3.3%

Auto ARIMA was selected next to search for the best ARIMA coefficients to improve the forecast.

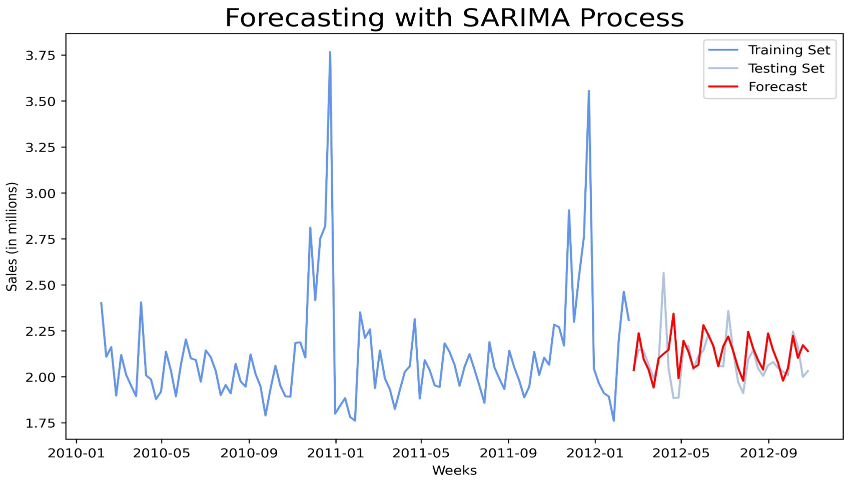
The forecast is slightly better than the AR model, but the forecast is still mainly just the average of the testing set which yields poor results.

SARIMA

Hyperparameters: (p, d, q)(P,D,Q)seasonality

Store 20

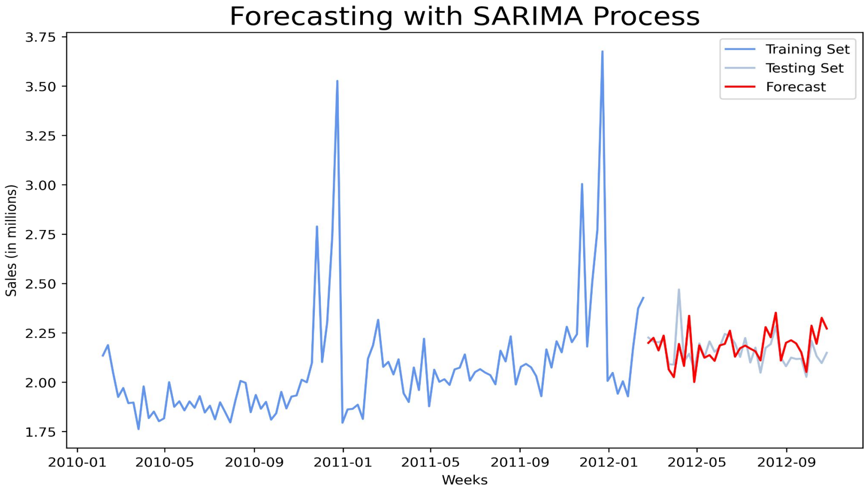
SARIMA(0,1,1)(0,1,1,52)



MAPE: 3.9%

Store 4

SARIMA(0,1,1)(0,1,1,52)



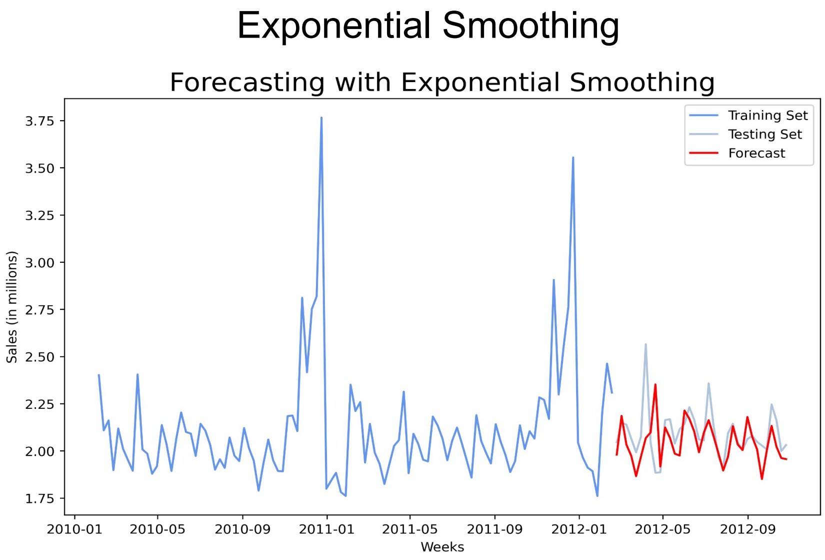
MAPE: 3.0%

Since the data has a clear seasonality component, SARIMA was chosen next to try to capture more than just the mean of the testing set.

The results are much better now that seasonality is considered.

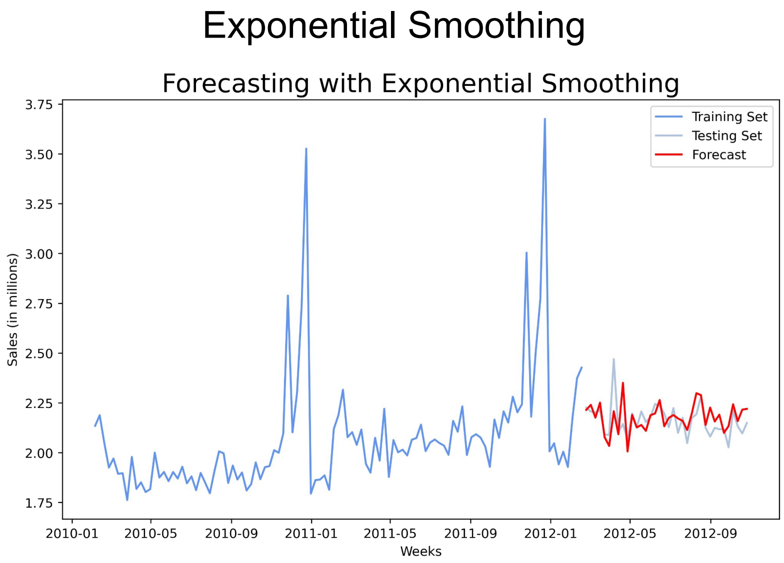
Exponential Smoothing

Store 20



MAPE: 4.4%

Store 4



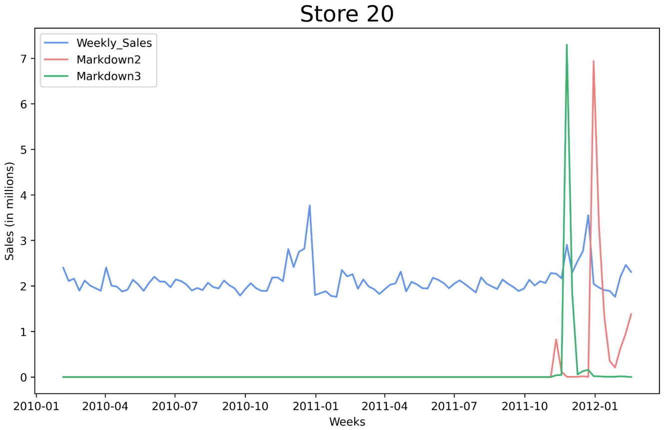
MAPE: 2.6%

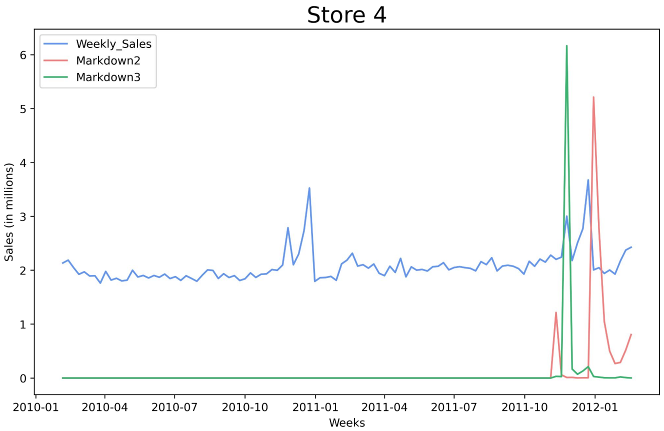
Similar to SARIMA, seasonality and trend are accounted for in exponential smoothing, so the results are much better compared to AR and ARIMA.

MULTIVARIATE MODELING

Multivariate Modeling

- Attempt to see if markdowns (discount data) have an effect on Weekly_Sales through multivariate modeling
- Utilize Date, Store, Weekly_Sales, Markdown_145, Markdown2, and Markdown3 to forecast Weekly_Sales in the test dataset





H_0: Data is non-stationary H_A: Data is stationary

Stationarity - Augmented Dicky Fuller (ADF)

ADF Test Results

Store 20

Markdown_145

Test Statistic	5.870487
p-value	1.000000
# Lags	13.000000
# Observations	93.000000
Critical Value (1%)	-3.502705
Critical Value (5%)	-2.893158
Critical Value (10%)	-2.583637

Markdown2

Test Statistic	-1.871161e+01
p-value	2.034595e-30
# Lags	1.300000e+01
# Observations	9.300000e+01
Critical Value (1%)	-3.502705e+00
Critical Value (5%)	-2.893158e+00
Critical Value (10%)	-2.583637e+00

Markdown3

Test Statistic	-8.036739e+00
p-value	1.893846e-12
# Lags	0.000000e+00
# Observations	1.060000e+02
Critical Value (1%)	-3.493602e+00
Critical Value (5%)	-2.889217e+00
Critical Value (10%)	-2.581533e+00

The p-value is greater than 0.05 and Test Statistic is greater than the Critical Value at 5% for Markdown_145 for both stores which fails to reject the null hypothesis indicating that Markdown_145 is non-stationary.

ADF Test Results

Store 4

Markdown_145

Test Statistic	5.752318
p-value	1.000000
# Lags	13.000000
# Observations	93.000000
Critical Value (1%)	-3.502705
Critical Value (5%)	-2.893158
Critical Value (10%)	-2.583637

Markdown2

Test Statistic	-5.968594e+00
p-value	1.960655e-07
# Lags	1.300000e+01
# Observations	9.300000e+01
Critical Value (1%)	-3.502705e+00
Critical Value (5%)	-2.893158e+00
Critical Value (10%)	-2.583637e+00

Markdown2 and Markdown3 have p-values less than 0.05 and Test Statistics lower than the Critical Value at 5% for both stores so we can reject the null hypothesis and conclude that these variables are stationary.

Markdown3

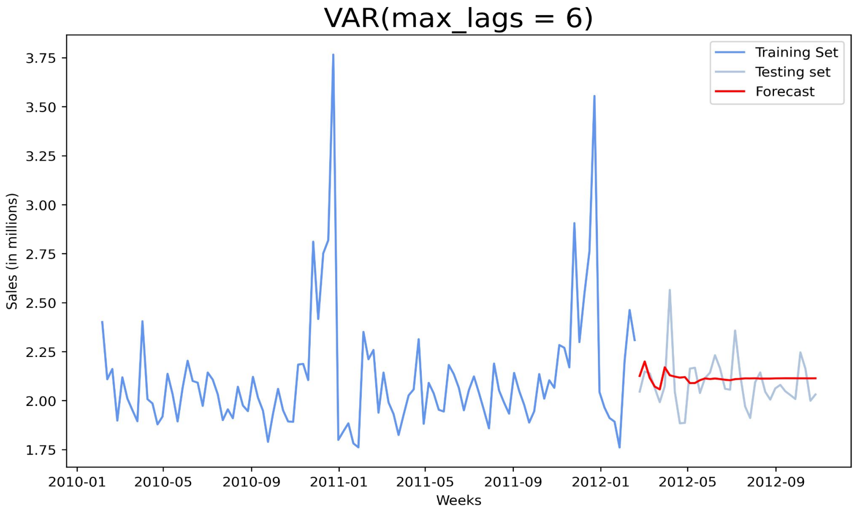
Test Statistic	-9.982978e+00
p-value	2.090912e-17
# Lags	0.000000e+00
# Observations	1.060000e+02
Critical Value (1%)	-3.493602e+00
Critical Value (5%)	-2.889217e+00
Critical Value (10%)	-2.581533e+00

Markdown_145 - Stationary Transformation

- Attempt to transform Markdown_145 to a stationary variable using:
    - Simple Moving Average
    - Cumulative Moving Average
    - Exponential Moving Average
    - Differencing
- This variable was still failing the ADF test for all of these methods so it was dropped from the dataset
- Date, Store, Weekly_Sales, Markdown2, and Markdown3 were ultimately used in the multivariate time series modeling

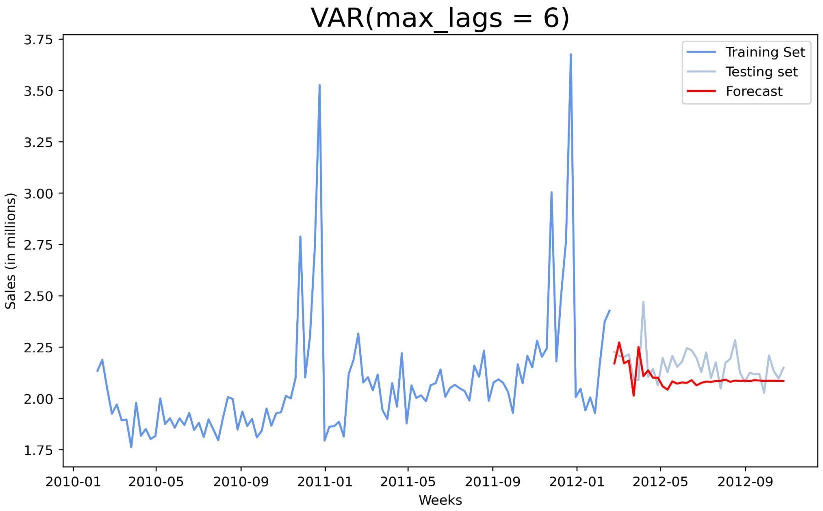
VAR

Store 20



MAPE: 4.5%

Store 4



MAPE: 3.8%

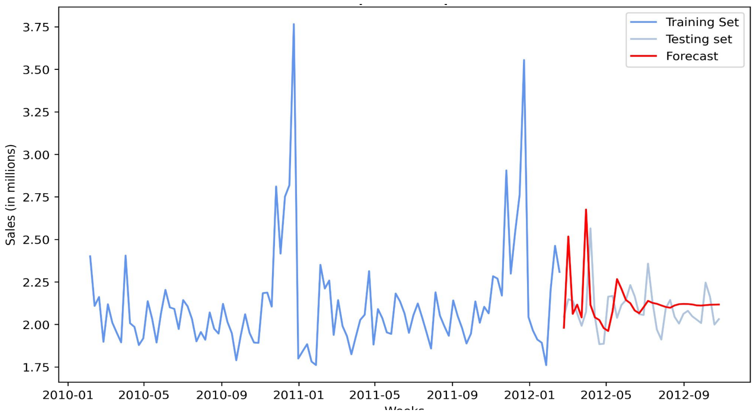
A grid search (with AIC as the metric) was conducted to determine the best value of max_lags for VAR modeling. The grid search resulted in an optimal max_lags value of 6.

Similar to the start of univariate modeling, the forecast accuracy is good but it is not capturing the patterns correctly

VARMA

Store 20

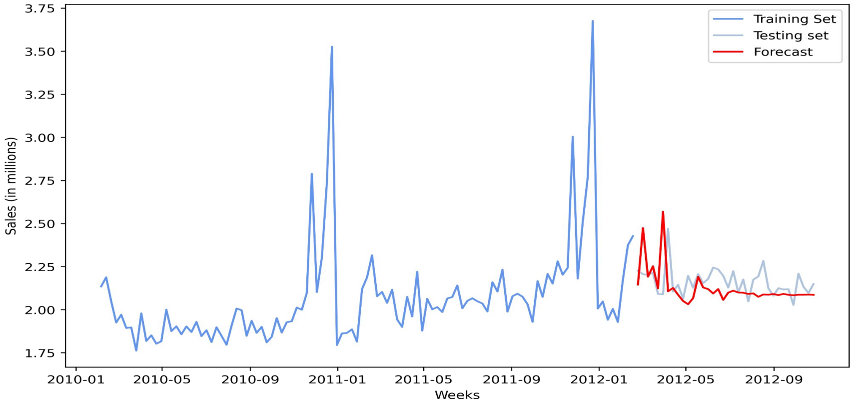
VARMA(6,0,6)



MAPE: 5.6%

Store 4

VARMA(6,0,6)



MAPE: 4.0%

Auto ARIMA was used to obtain optimal p and q values to use in VARMA.

The forecast is better than VAR with some of the patterns being captured but there is an increase in MAPE. Although, univariate models had consistently better results.

CONCLUSIONS

Granger Causality

Granger Causality Test Results

Store 20

Markdown2

```
Granger Causality
number of lags (no zero) 3
ssr based F test:         F=1.2810  , p=0.3946  , df_denom=4, df_num=3
```

Markdown3

```
Granger Causality
number of lags (no zero) 3
ssr based F test:         F=2.2592  , p=0.2237  , df_denom=4, df_num=3
```

Granger Causality Test Results

Store 4

Markdown2

```
Granger Causality
number of lags (no zero) 3
ssr based F test:         F=2.0898  , p=0.2443  , df_denom=4, df_num=3
```

Markdown3

```
Granger Causality
number of lags (no zero) 3
ssr based F test:         F=0.8110  , p=0.5505  , df_denom=4, df_num=3
```

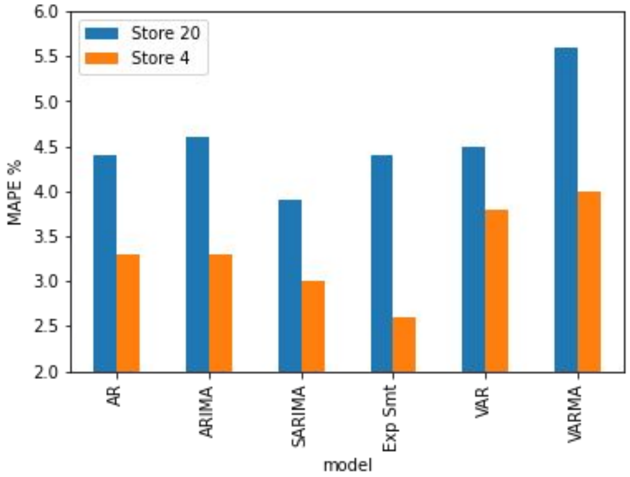
Hypotheses

H_0: Markdown2/3 does not Granger cause Sales
H_A: Markdown2/3 does Granger cause Sales

Based on the F test (due to smaller sample, F test is preferable) granger causality tests, we fail to reject the null hypotheses that Markdown2/Markdown3 do not Granger cause Weekly_Sales.

Results Comparison

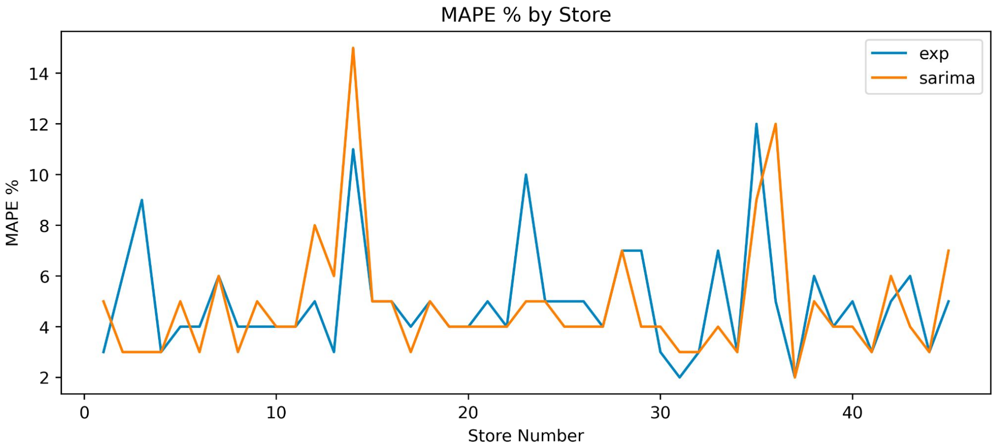
Model Type	Store 20 MAPE	Store 4 MAPE	Average MAPE
AR	4.4	3.3	3.85
ARIMA	4.6	3.3	3.95
SARIMA	3.9	3	3.45
Exponential Smoothing	4.4	2.6	3.5
VAR	4.5	3.8	4.15
VARMA	5.6	4	4.5



Based on MAPE and the Granger Causality test, univariate modeling seems to model the data better than multivariate. Within univariate, SARIMA and Exponential Smoothing perform the best.

Final Model Selection

- SARIMA is more difficult to test across stores as the parameters can vary by store. Performing a grid search for all stores is resource intensive and would take more time than is allocated to the project.
- SARIMA was run for all stores with a constant set of parameters since it was not feasible to optimize for all parameters for each store.
- When testing Exponential Smoothing across stores, it resulted in an average MAPE of 5.0 and a median MAPE of 5.0.
- Even without optimizing for parameters, SARIMA performed better with an average MAPE of 4.8 and median MAPE of 4.0 across stores.



QUESTIONS?