

# Capstone Project

Stock Price Prediction with ARIMA, SARIMAX, GRU, LSTM

# The problem

The objective is to create a predictive model to forecast future stock closing prices.

*For the purpose of this presentation I will be showcasing the AMD stock ticker.*

# Overview



Obtain Data

EDA

Pre-Processing

Modeling &  
Evaluation

Evaluation &  
Future Works

## Pandas Datareader

- Tiingo
- IEX
- Alpha Vantage
- Econdb
- Enigma
- Quandl
- St.Louis FED (FRED)
- Kenneth French's data library
- World Bank
- OECD
- Eurostat
- Thrift Savings Plan
- Nasdaq Trader symbol definitions
- Stooq
- MOEX
- Naver Finance
- Yahoo Finance

Obtain Data

EDA

Pre-Processing

Modeling &  
Evaluation

Evaluation &  
Future Works



Obtain Data

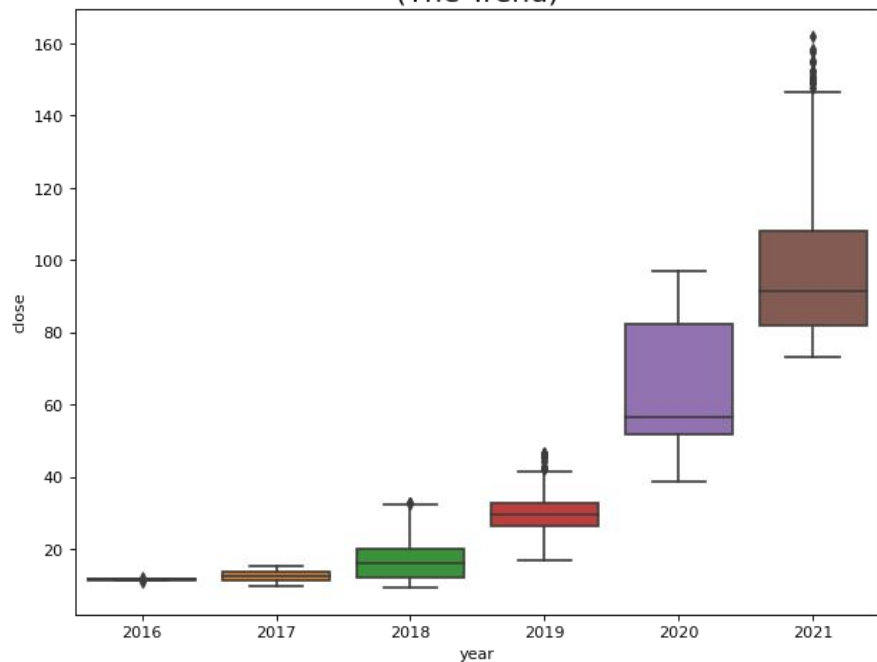
EDA

Pre-Processing

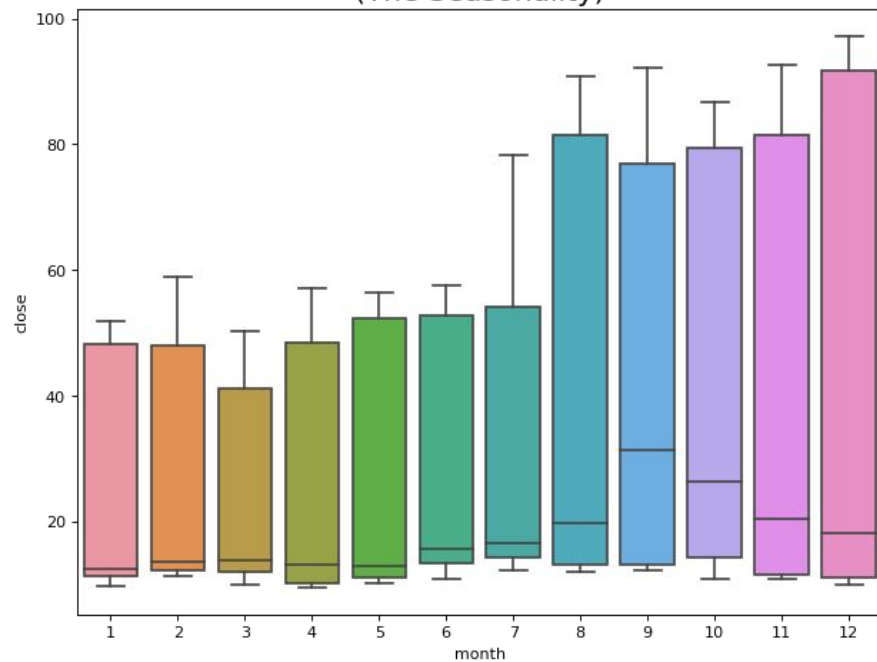
Modeling &  
Evaluation

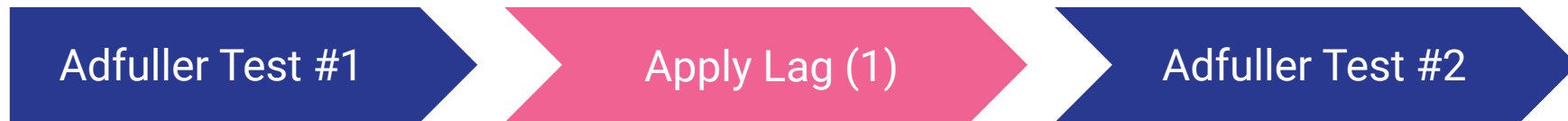
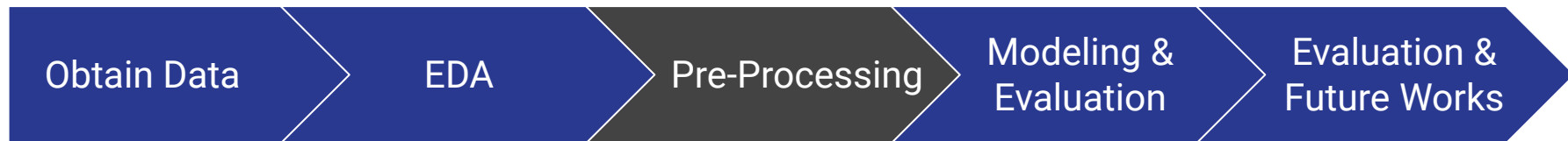
Evaluation &  
Future Works

Year-wise Box Plot  
(The Trend)



Month-wise Box Plot  
(The Seasonality)





**ADF Test Statistic :**

1.3842659609766095

**p-value :** 0.997041479296224

**Number of Observations Used :** 1237

**Comment :** Time series is non-stationary

**Procedure**

Day #1 = N/A

Day #2 = Day #2 - Day #1

Day #3 = Day #3 - Day #2

Day #4 = Day #4 - Day #3

.

.

**ADF Test Statistic :**

-9.059840432443018

**p-value :** 4.594794188707755e-15

**Number of Observations Used :**  
1242

**Comment :** Time series is stationary

Obtain Data

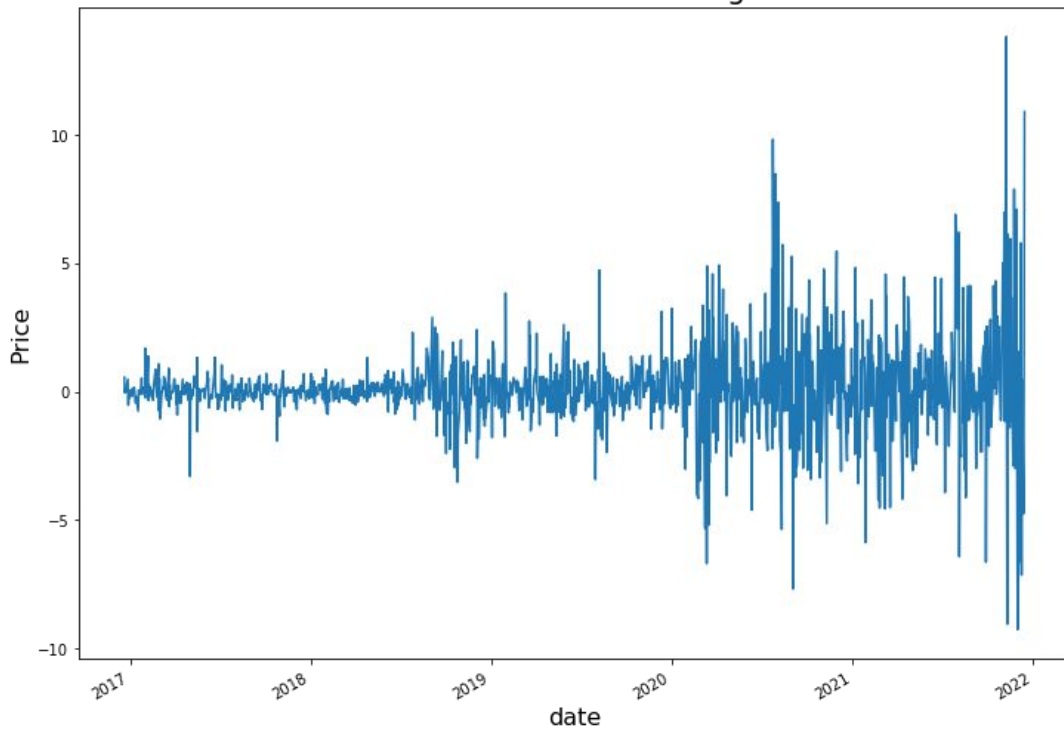
EDA

Pre-Processing

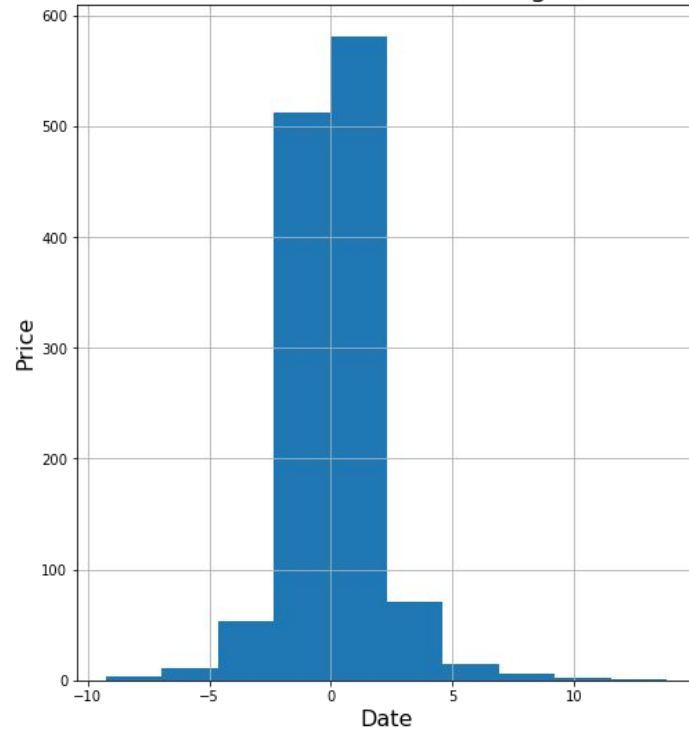
Modeling &  
Evaluation

Evaluation &  
Future Works

Close First Differencing



Close First Differencing





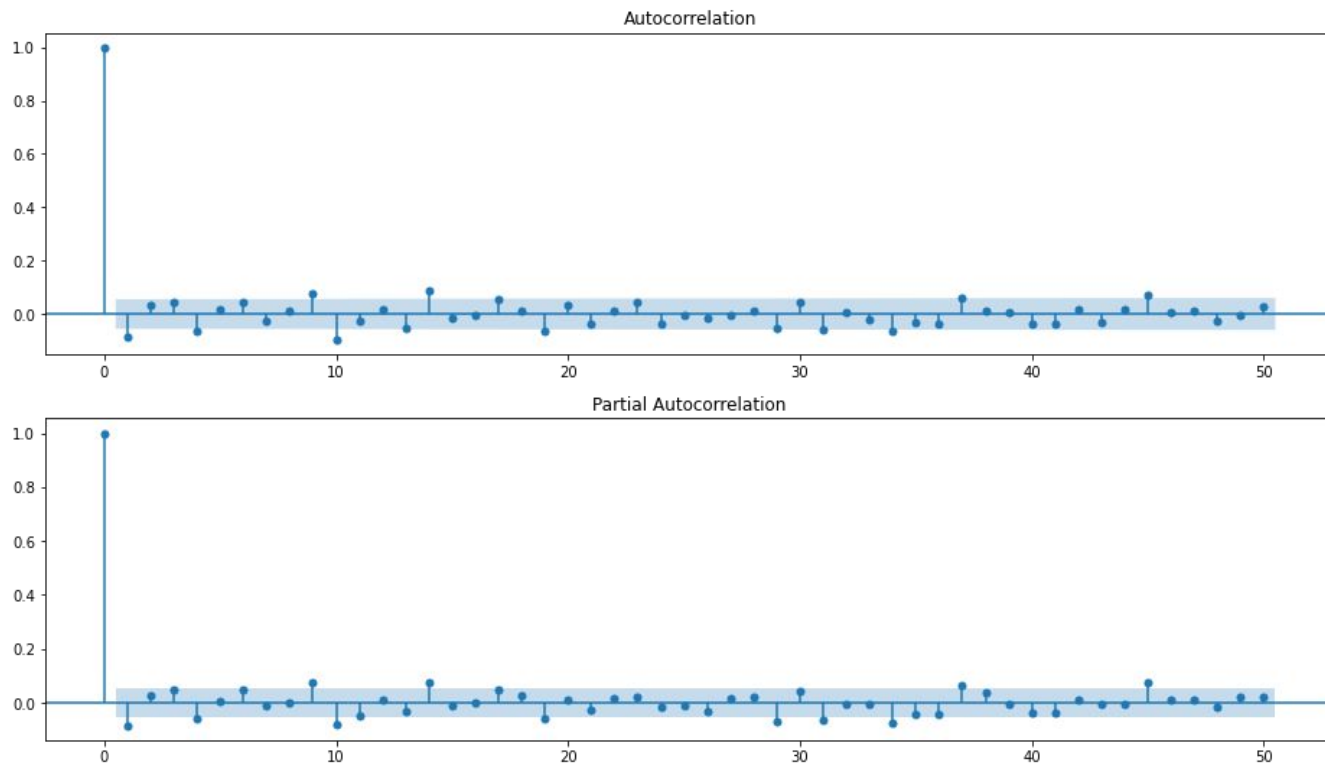
Obtain Data

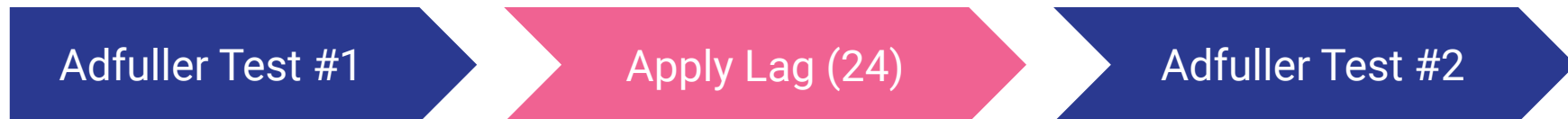
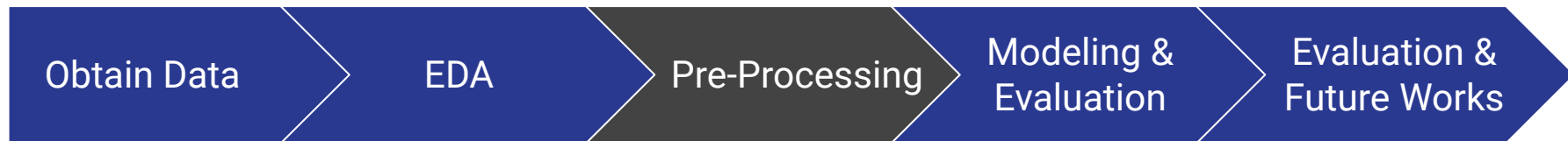
EDA

Pre-Processing

Modeling &  
Evaluation

Evaluation &  
Future Works





**ADF Test Statistic :**

1.3842659609766095

**p-value :** 0.997041479296224

**Number of Observations Used :** 1237

**Comment :** Time series is non-stationary

**Procedure**

Day #1 to Day #24 = N/A

Day #25 = Day #25 - Day #1

Day #26 = Day #26 - Day #2

Day #27 = Day #27 - Day #3

.

.

**ADF Test Statistic :**

-8.610593374916578

**p-value :** 6.491890430629663e-14

**Number of Observations Used :**  
1209

**Comment :** Time series is stationary

Obtain Data

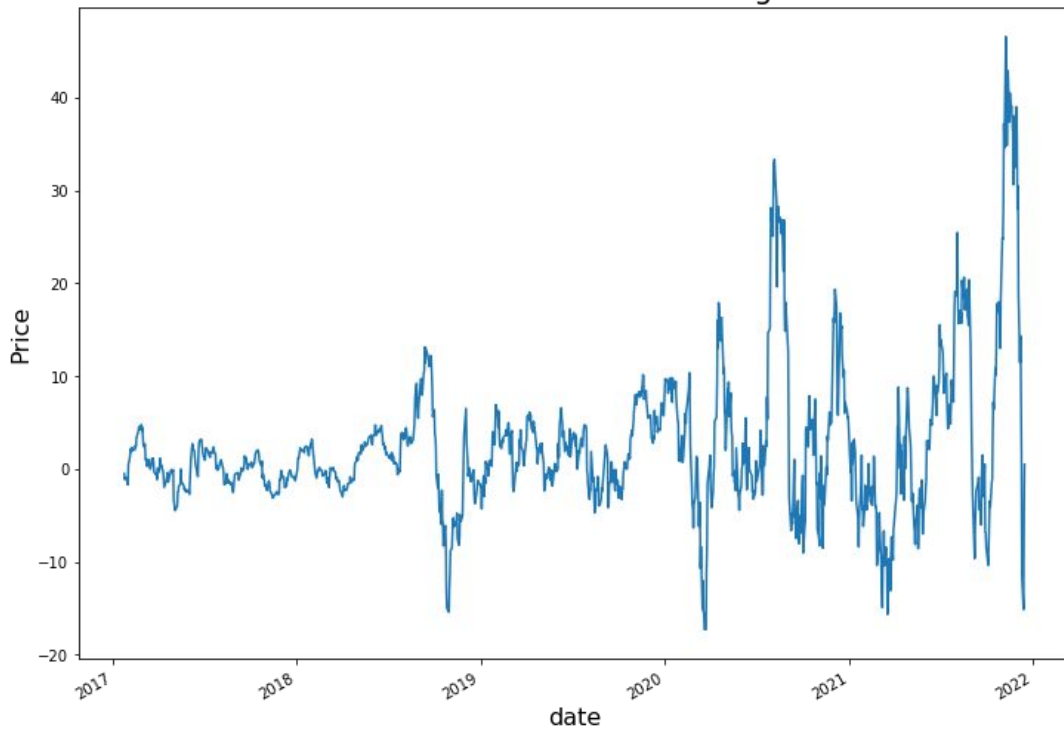
EDA

Pre-Processing

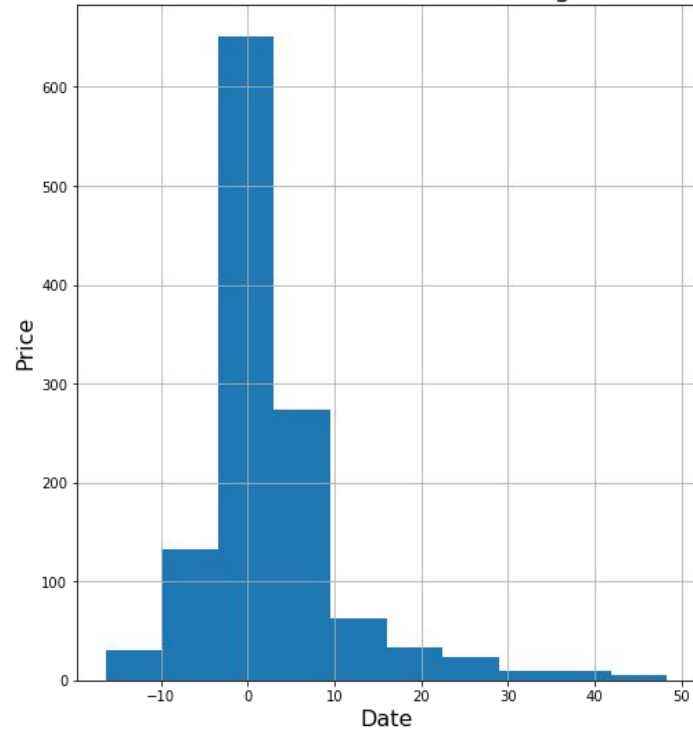
Modeling &  
Evaluation

Evaluation &  
Future Works

Seasonal First Differencing



Seasonal First Differencing



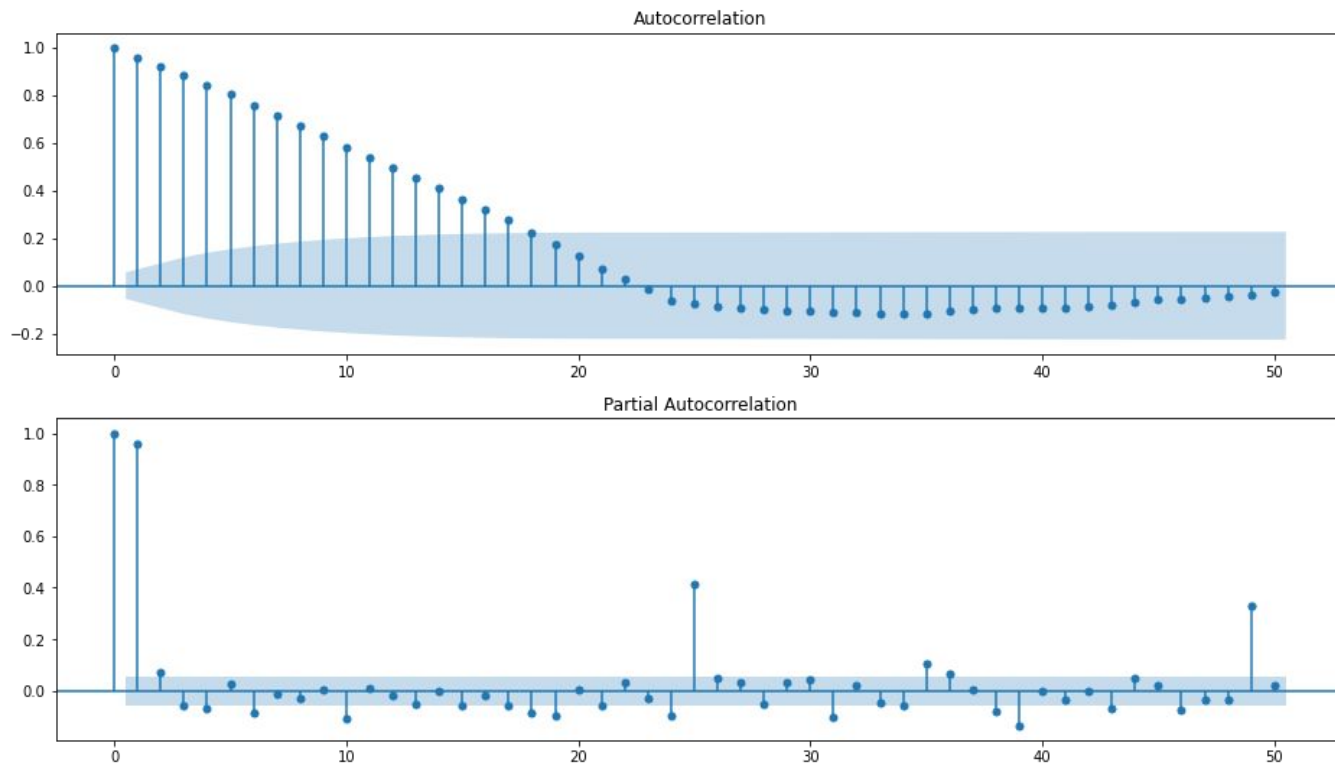
Obtain Data

EDA

Pre-Processing

Modeling &  
Evaluation

Evaluation &  
Future Works



Obtain Data

EDA

Pre-Processing

Modeling &  
Evaluation

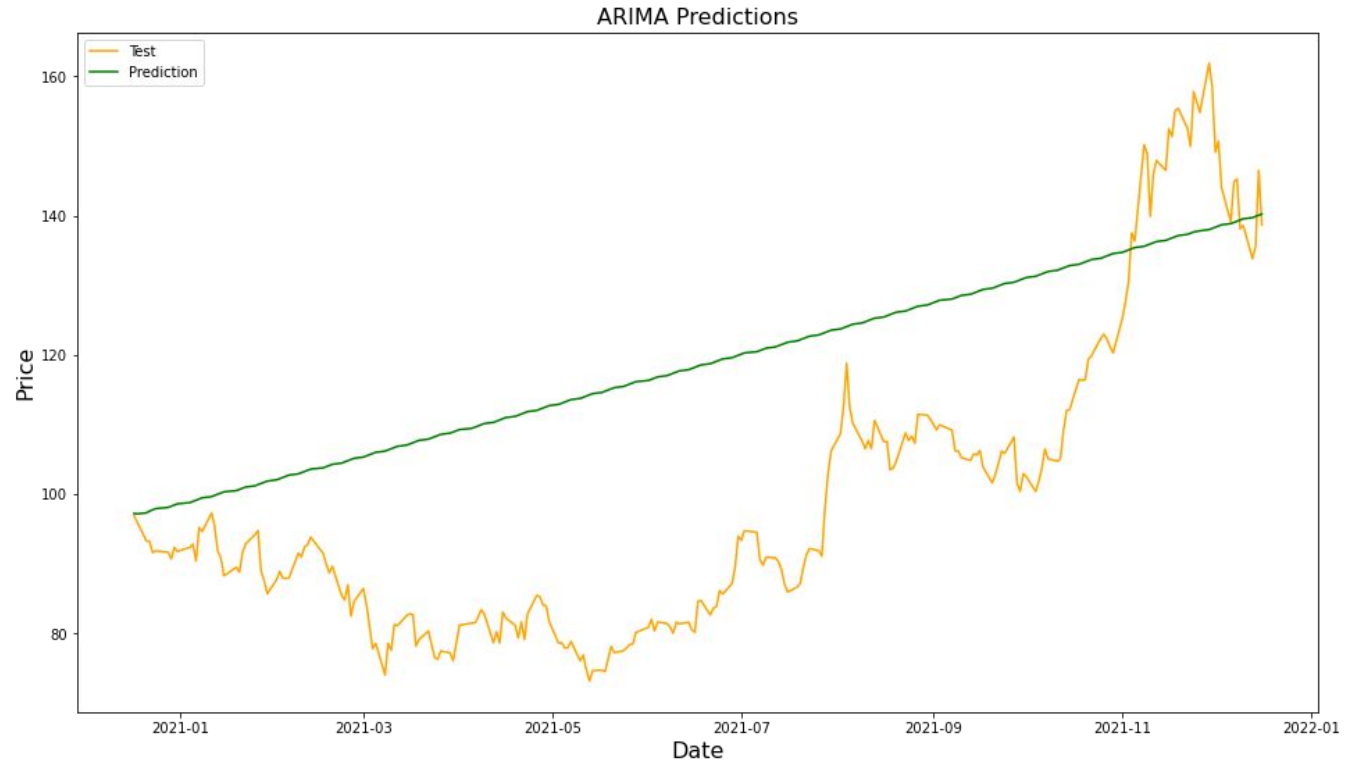
Evaluation &  
Future Works

## ARIMA Architecture

80 - 20 Train Test Split

Order = (5, 2, 1), # (p, d, q)

*p* : Trend Autoregression Order  
*d* : Trend difference order  
*q* : Trend moving average order



Obtain Data

EDA

Pre-Processing

Modeling &  
Evaluation

Evaluation &  
Future Works

## SARIMAX Architecture

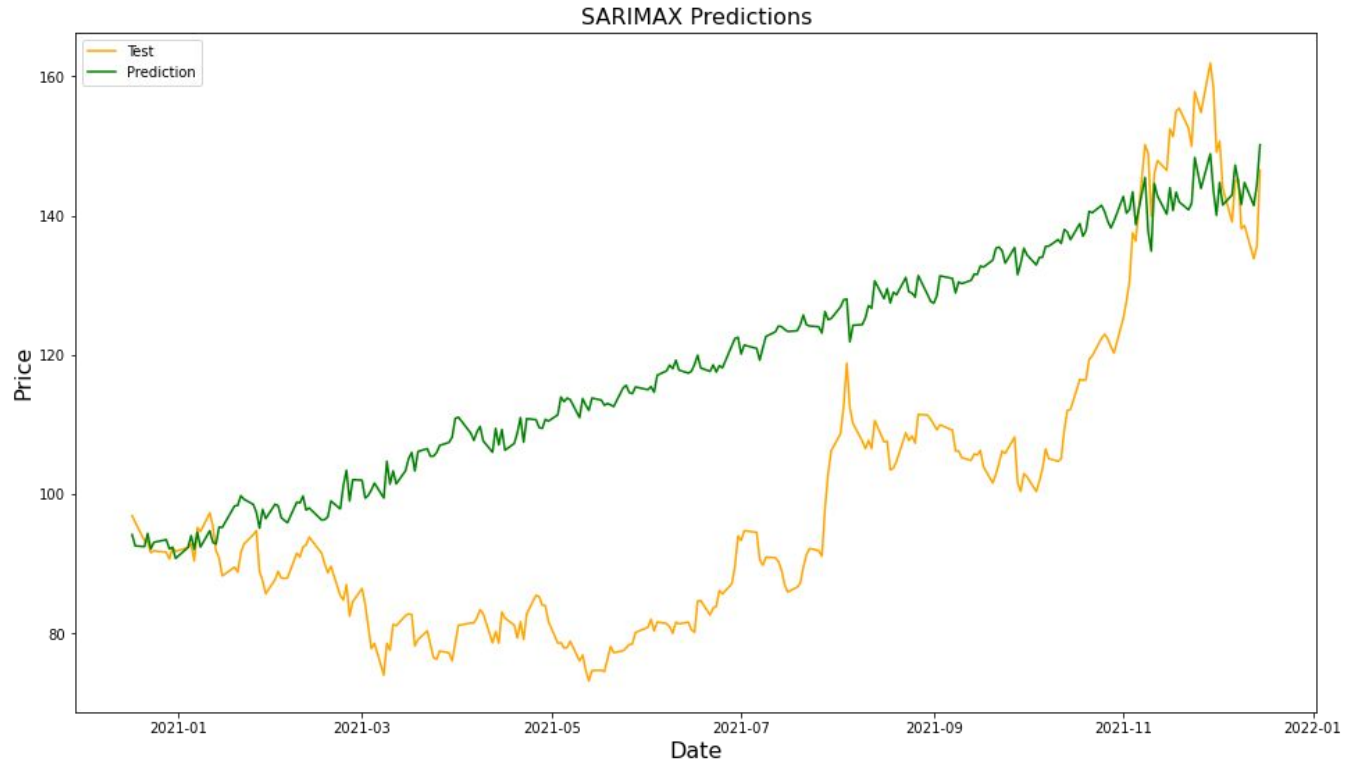
80 - 20 Train Test Split

Order = (8, 1, 0), # (p, d, q)

Seasonal\_order = (2, 1, 0,  
24), # (p, d, q, m)

Exog = Lag(1)

*p* : Trend Autoregression Order  
*d* : Trend difference order  
*q* : Trend moving average order  
*m* : The number of time steps for a  
single seasonal period  
Exog: OLS Regressor



Obtain Data

EDA

Pre-Processing

Modeling &  
Evaluation

Evaluation &  
Future Works

## GRU Architecture

*80 - 20 Train Test Split*

### Layer #1 - GRU

Units : 128  
Activation : Tanh  
Recurrent Activation : Sigmoid

### Layer #2 - GRU

Units : 64  
Activation : Tanh  
Recurrent Activation : Sigmoid

### Layer #3 - Dropout

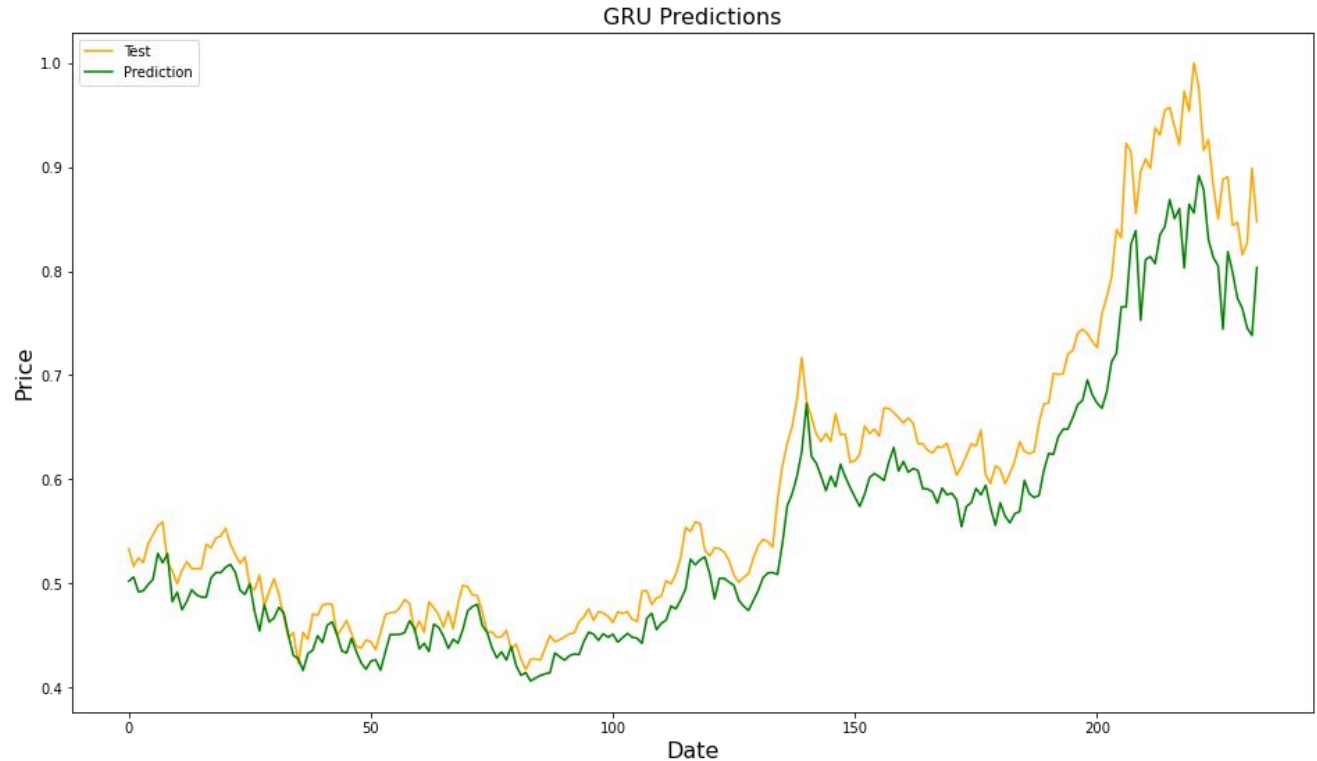
Rate : 0.2

### Layer #4 - Dense

Units : 1

### Evaluator

Loss : Mean Squared Error  
Optimizer : Adam



Obtain Data

EDA

Pre-Processing

Modeling &  
Evaluation

Evaluation &  
Future Works

## LSTM Architecture

80 - 20 Train Test Split

### Layer #1 - LSTM

Units : 128  
Activation : Tanh  
Recurrent Activation : Sigmoid

### Layer #2 - LSTM

Units : 64  
Activation : Tanh  
Recurrent Activation : Sigmoid

### Layer #3 - Dropout

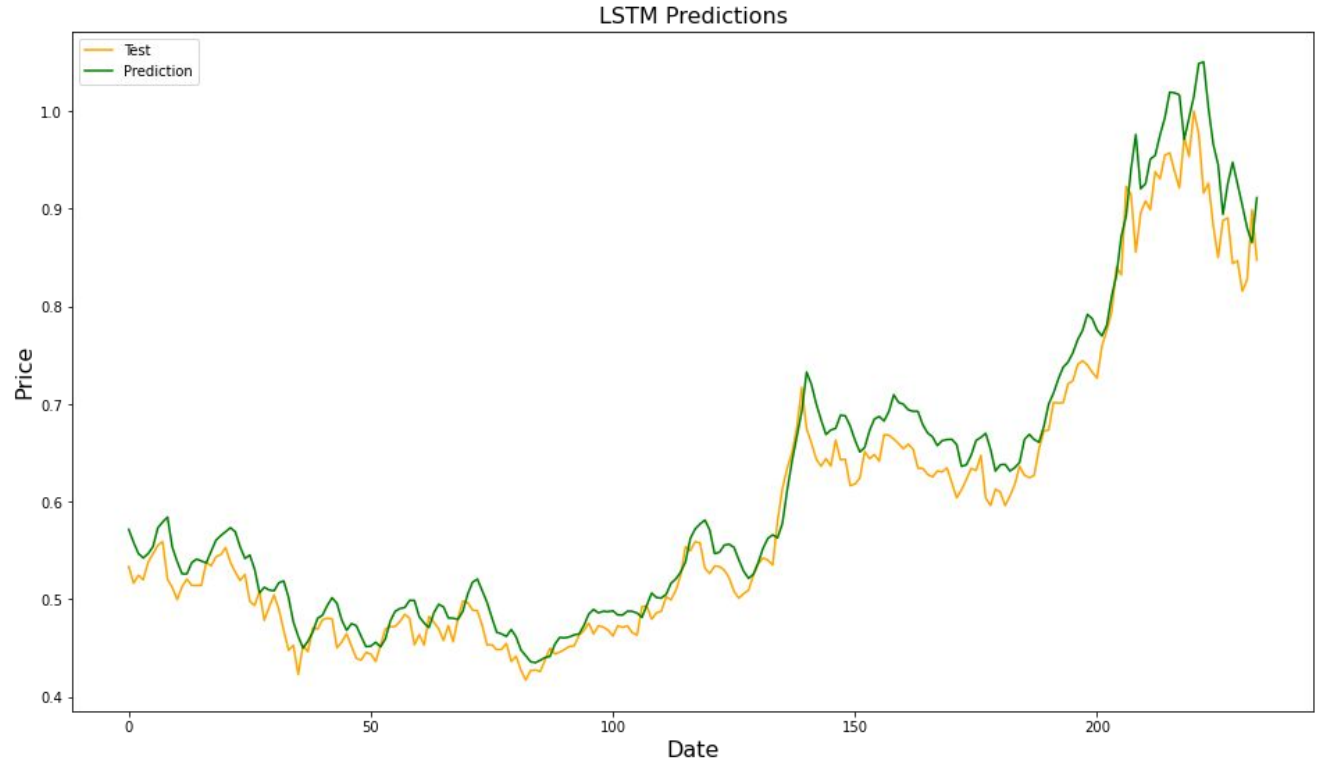
Rate : 0.2

### Layer #4 - Dense

Units : 1

### Evaluator

Loss : Mean Squared Error  
Optimizer : Adam







**Metric Used : MAPE (Mean absolute percentage error)**

The average difference between the predicted value and the actual value is

24.29%

**Metric Used : MAPE (Mean absolute percentage error)**

The average difference between the predicted value and the actual value is

26.19%

**Metric Used : MAPE (Mean absolute percentage error)**

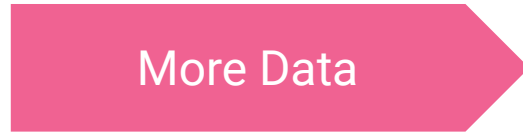
The average difference between the predicted value and the actual value is

6.3%

**Metric Used : MAPE (Mean absolute percentage error)**

The average difference between the predicted value and the actual value is

4.58%



- Include News Sentiment data from
  - News Articles
  - Reddit posts
  - Tweets
  - Discord channels
- Add statistical weights on certain price movements based on specific chart indicators



- Provide data on any stock ticker that is search
  - Bullish/Bearish
  - Increase/Decrease (with %)
- Personal Screener
- Recommended stocks
  - Industry
  - High Volume
  - Momentum



# Thank You

