Capstone Project

Stock Price Prediction with ARIMAX, SARIMAX, GRU, LSTM

Henri Oei - DSIF2

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
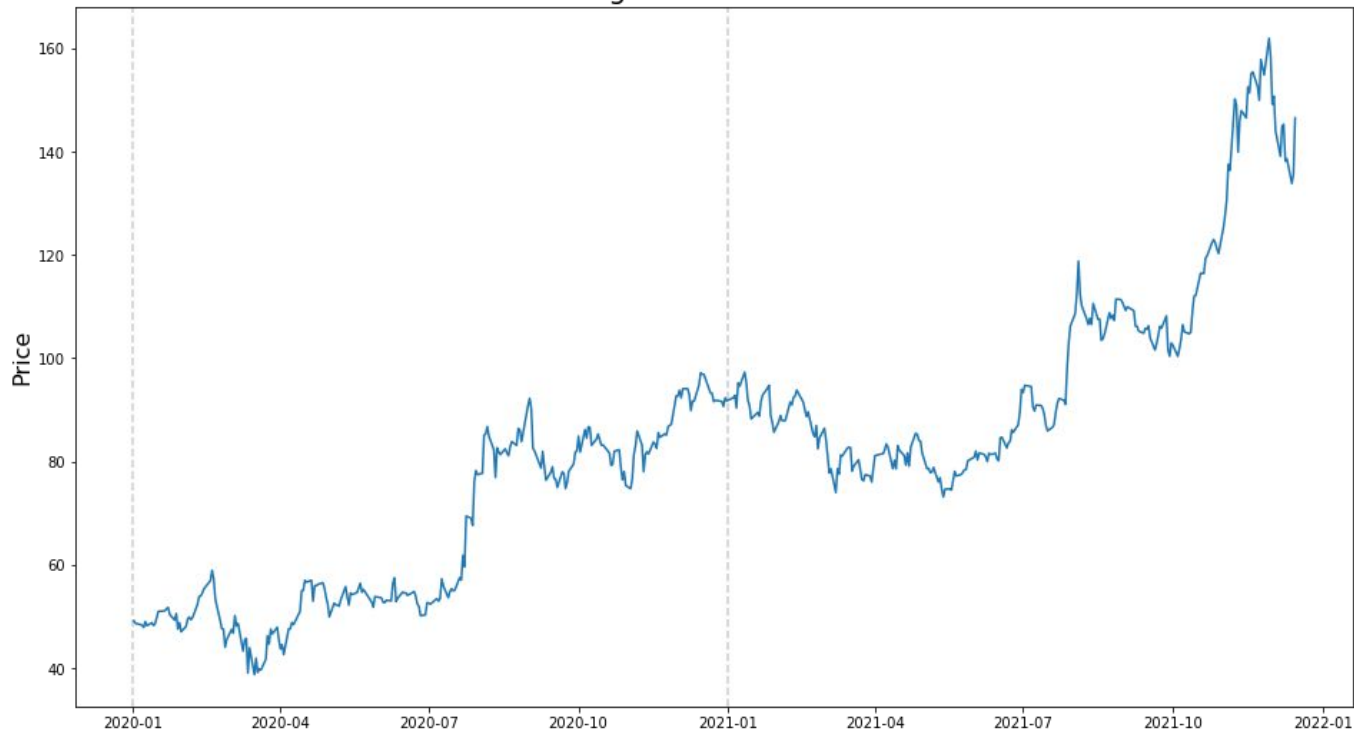- Stooq
- MOEX
- Naver Finance
- Yahoo Finance

Closing Price of AMD Stock

Year-wise Box Plot (The Trend)
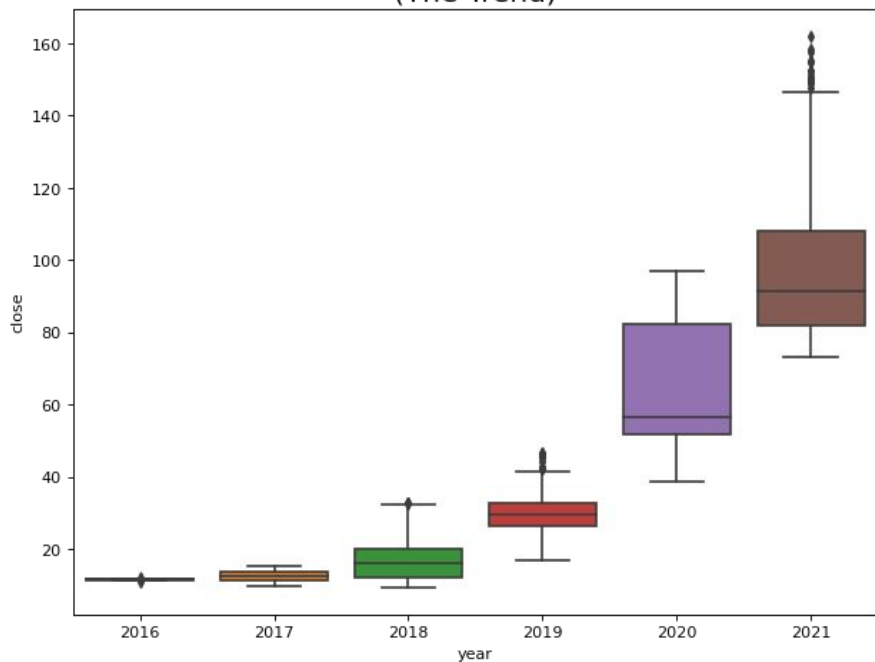
Month-wise Box Plot (The Seasonality)
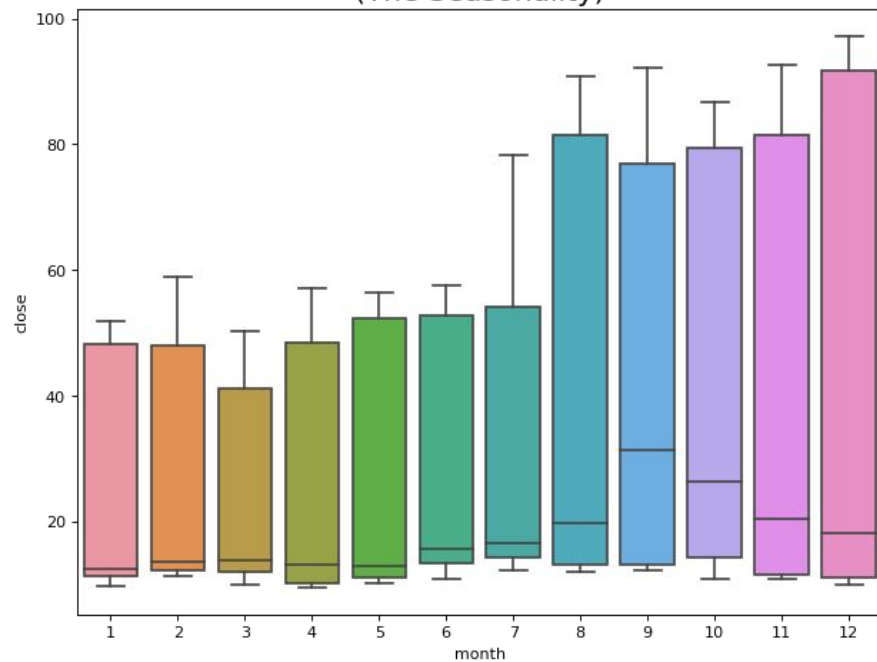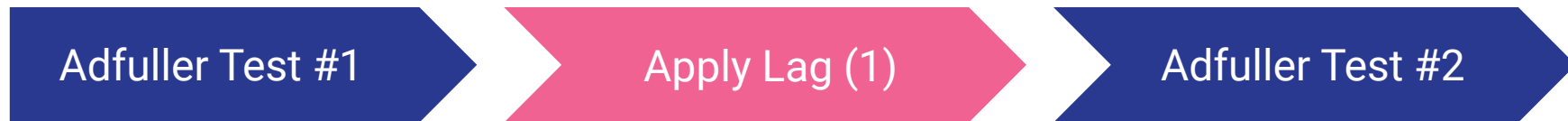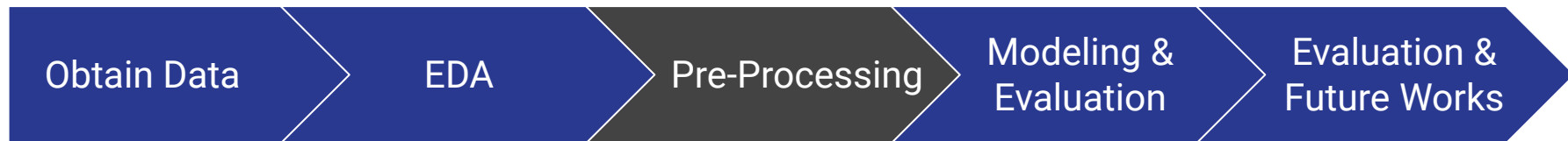
Obtain Data → EDA → Pre-Processing → Modeling & Evaluation → Evaluation & Future Works

Adfuller Test #1 → Apply Lag (1) → Adfuller Test #2

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

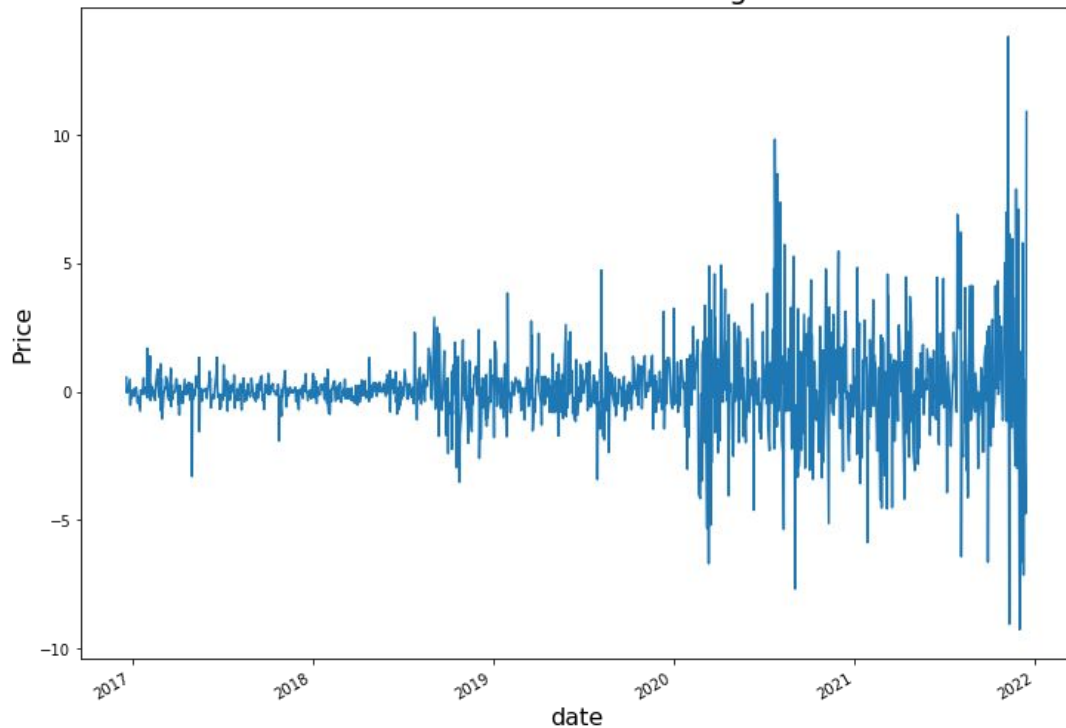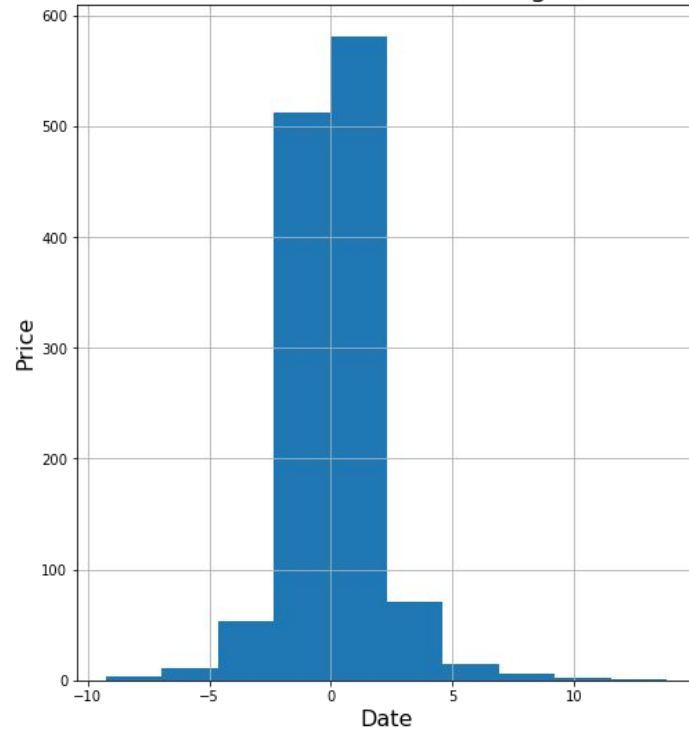**Comment** : Time series is stationary

Close First Differencing

Close First Differencing

Autocorrelation

Partial Autocorrelation

Obtain Data → EDA → Pre-Processing → Modeling & Evaluation → Evaluation & Future Works

Adfuller Test #1 → Apply Lag (24) → Adfuller Test #2
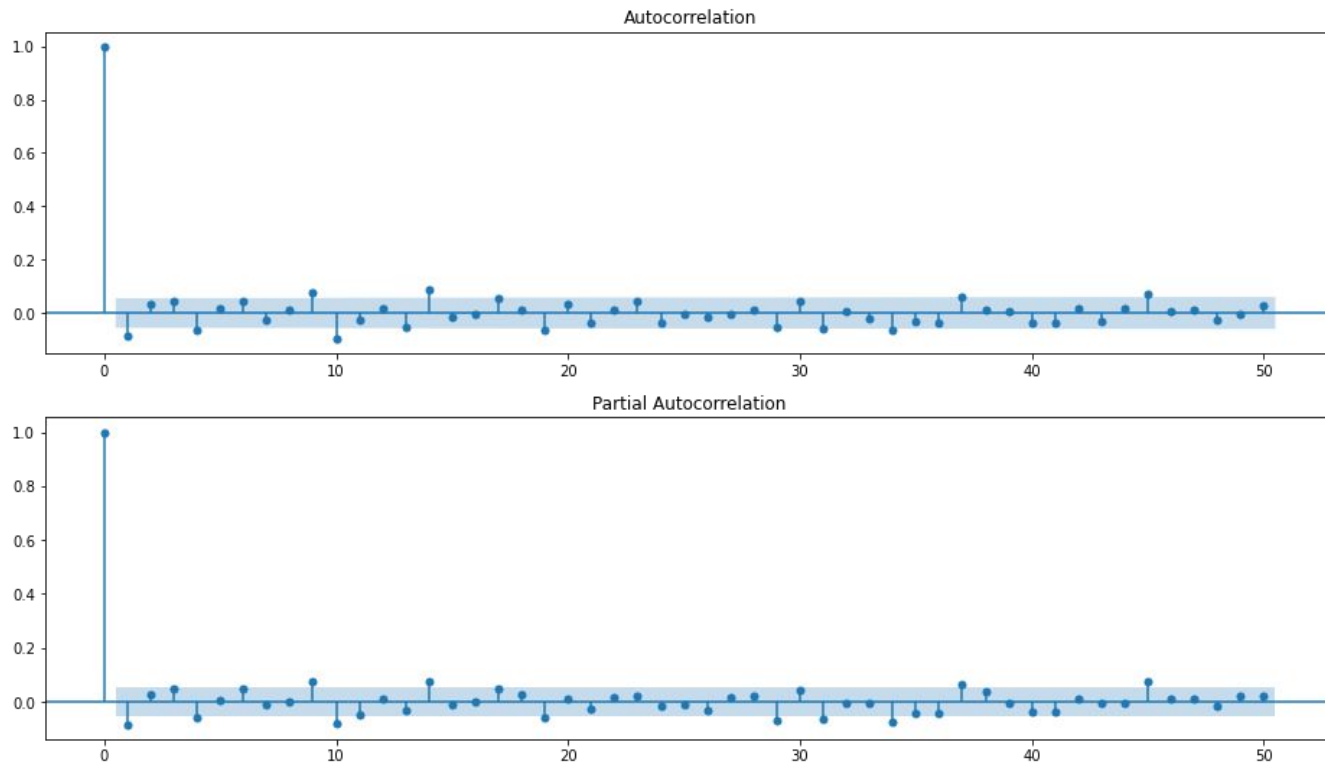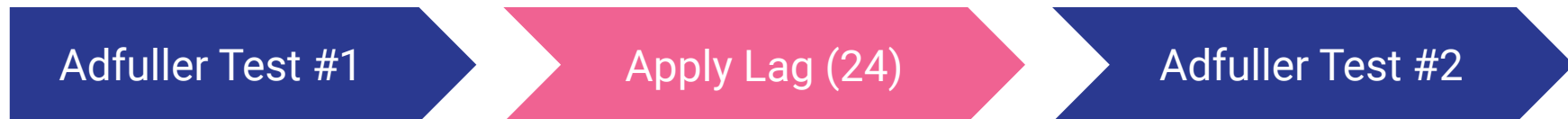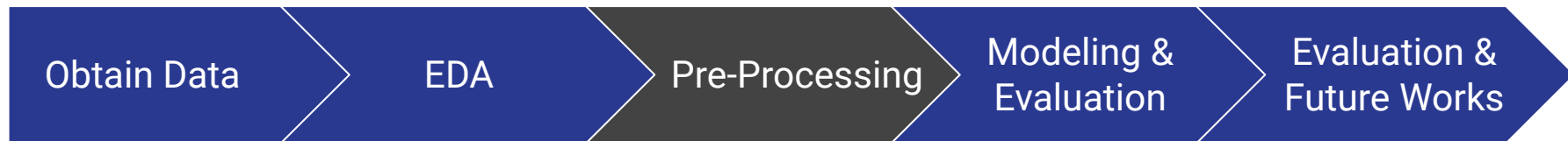
ADF Test Statistic :

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
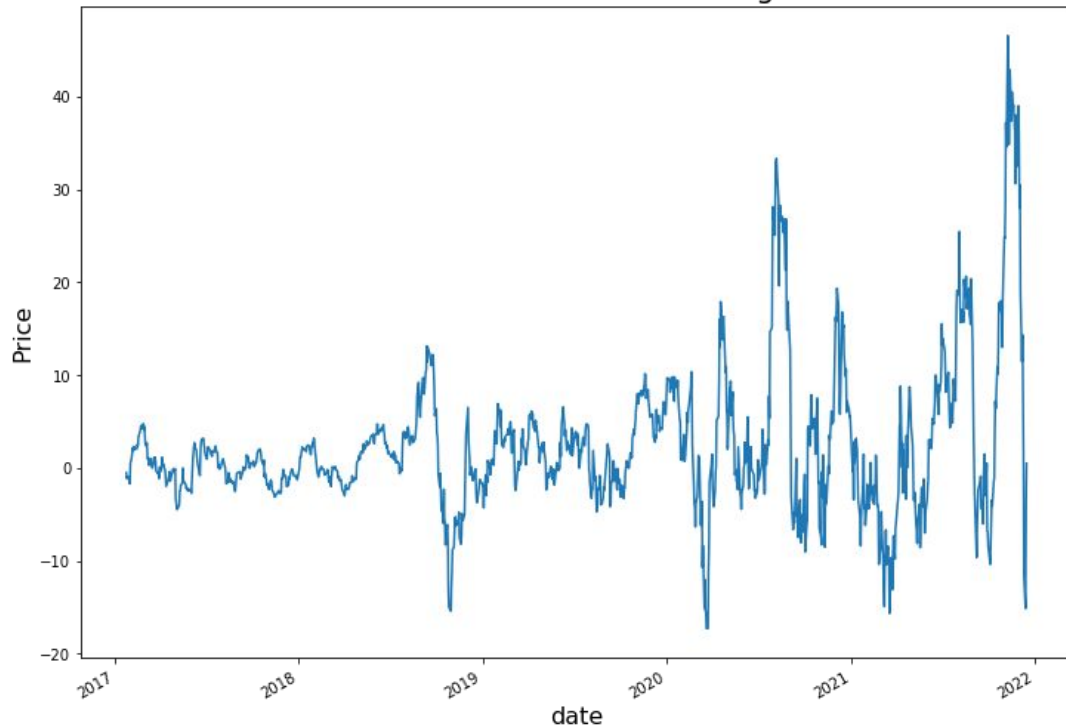
Comment : Time series is stationary

Seasonal First Differencing

Seasonal First Differencing
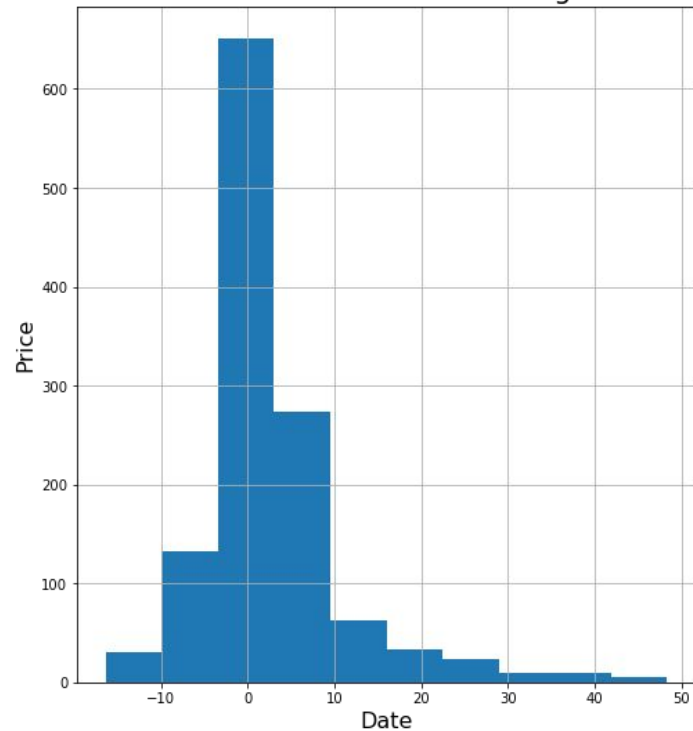
Autocorrelation

Partial Autocorrelation
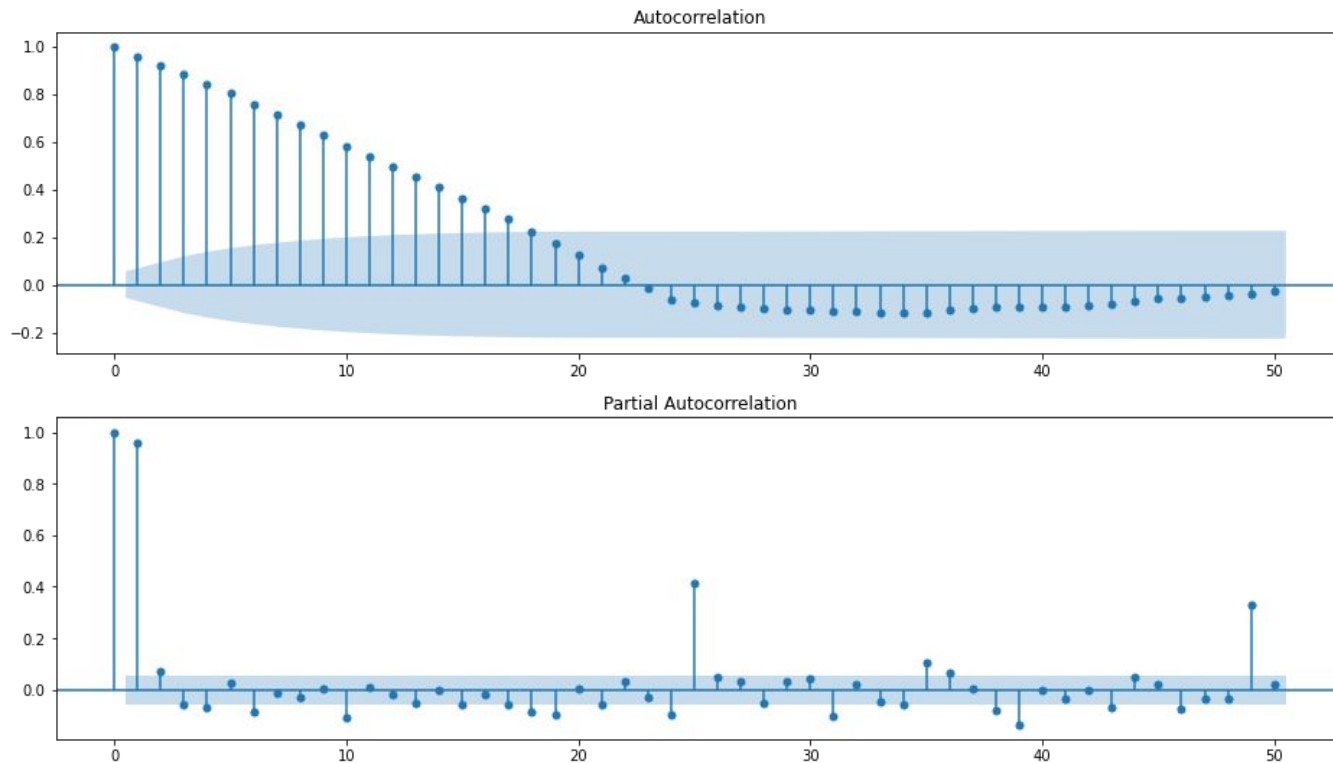
ARIMAX Architecture (5Y)

80 - 20 Train Test Split

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

Exog = Lag(24)

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

ARIMAX Predictions

ARIMAX Architecture (2Y)

80 - 20 Train Test Split

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

Exog = Lag(24)

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

ARIMA Predictions

SARIMAX Architecture (5Y)
80 - 20 Train Test Split

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

Seasonal_order = (1, 1, 1, 24), # (p, d, q, m)

Exog = Lag(24)

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

SARIMAX Predictions

SARIMAX Architecture (2Y)

80 - 20 Train Test Split
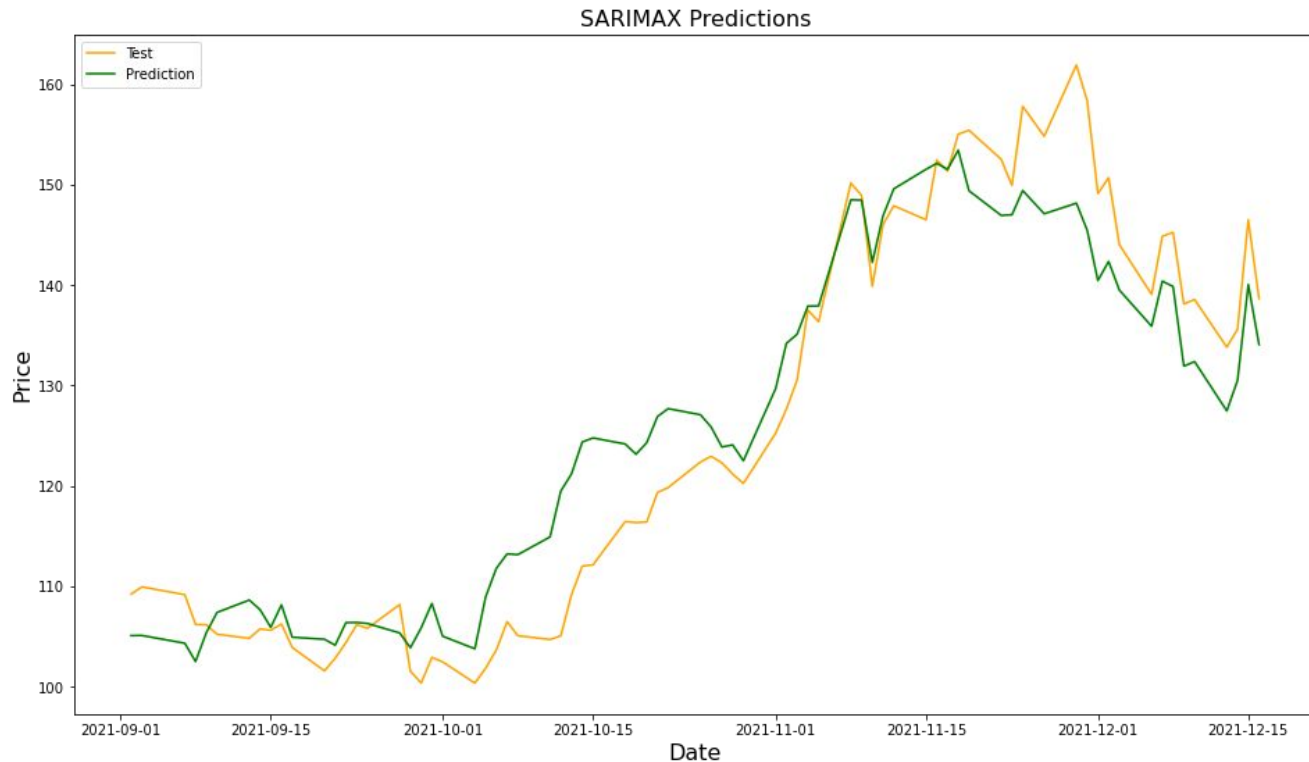
Order = (1, 1, 1), # (p, d, q)

Seasonal_order = (1, 1, 1, 24), # (p, d, q, m)

Exog = Lag(24)

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

SARIMAX Predictions

GRU Loss Interaction

Epochs = 100

Batch Size = 32

Validation Split = 0.2

Verbose = 1

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

GRU Predictions

LSTM Loss Interaction

Epochs = 100

Batch Size = 32

Validation Split = 0.2

Verbose = 1
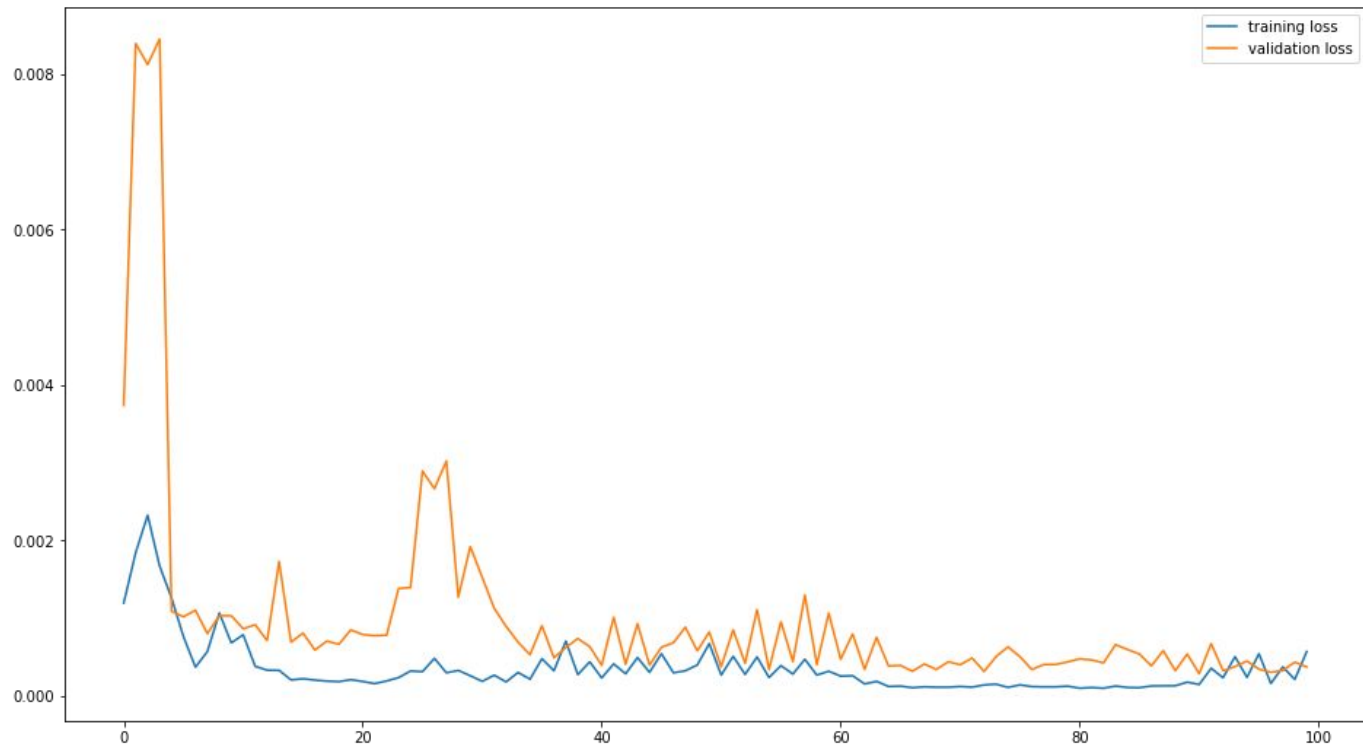
LSTM Architecture

80 - 20 Train Test Split

Layer #1 - LSTM

Units : 128
Activation : Tanh
Recurrent Activation : Sigmoid

Layer #2 - LSTM

Units : 64
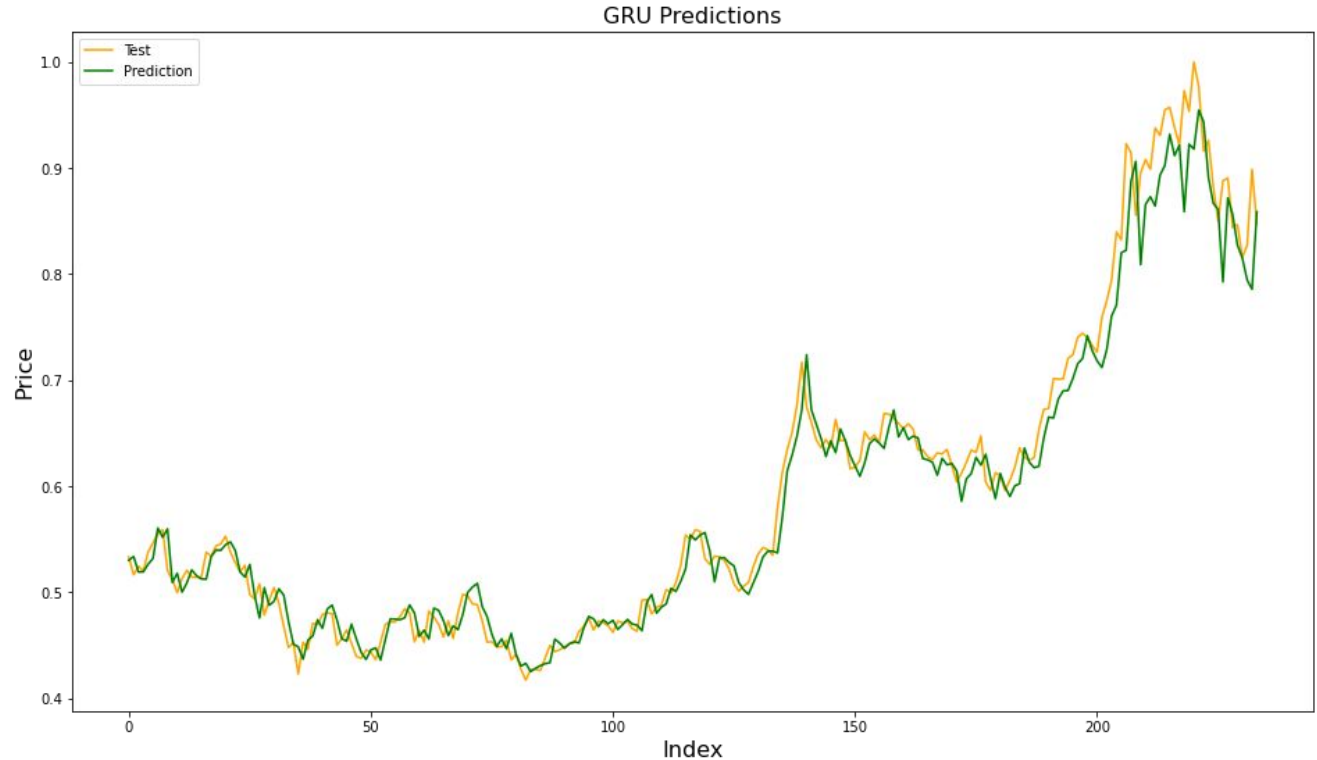Activation : Tanh
Recurrent Activation : Sigmoid

Layer #3 - Dropout

Rate : 0.2

Layer #4 - Dense

Units : 1

Evaluator

Loss : Mean Squared Error
Optimizer : Adam

LSTM Predictions

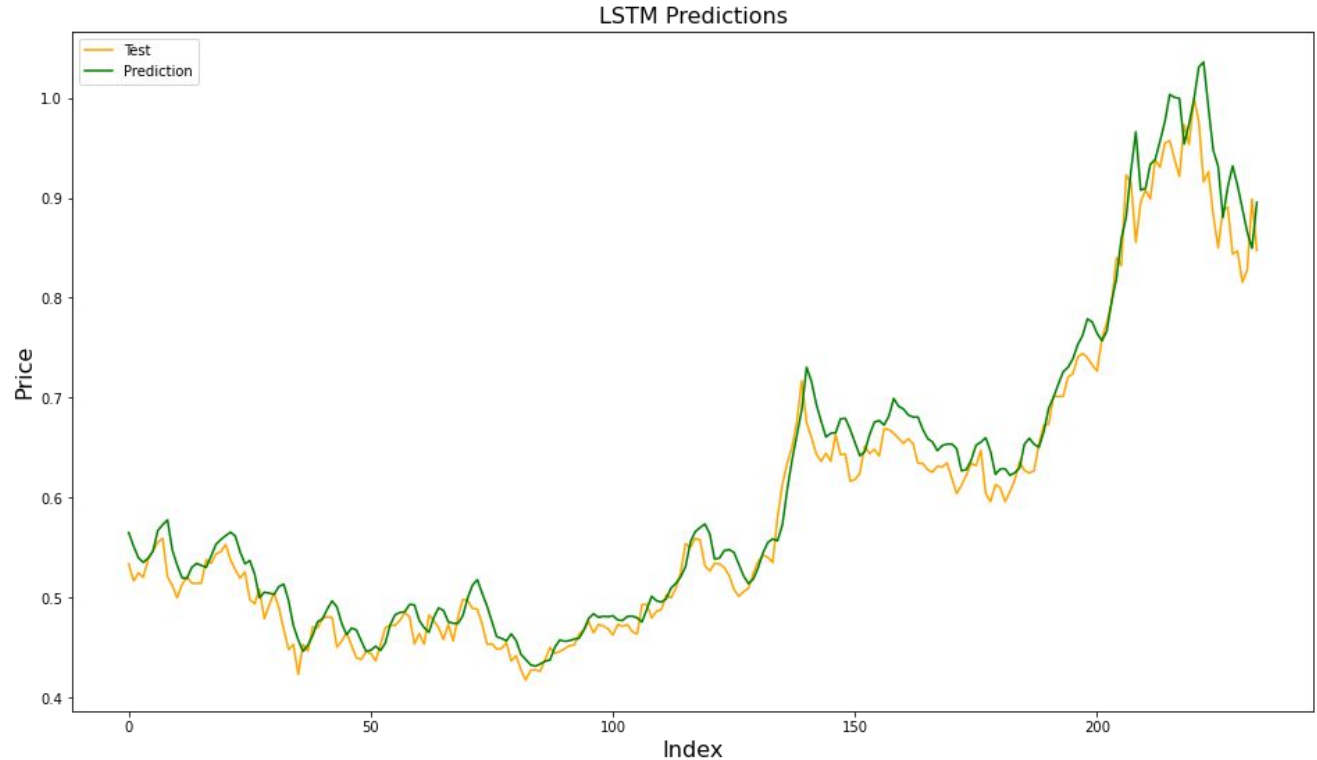| Obtain Data	EDA	Pre-Processing	Modeling & Evaluation	Evaluation & Future Works

ARIMA	SARIMAX	GRU	LSTM
Metric Used : MAPE (Mean absolute percentage error)	**Metric Used : MAPE (Mean absolute percentage error)**	**Metric Used : MAPE (Mean absolute percentage error)**	**Metric Used : MAPE (Mean absolute percentage error)**
The average difference between the predicted value and the actual value is	The average difference between the predicted value and the actual value is	The average difference between the predicted value and the actual value is	The average difference between the predicted value and the actual value is
(5Y) 12.61%	(5Y) 13.88%	2.6%	3.53%
(2Y) 21.5%	**(2Y) 3.97%**		

| Obtain Data | EDA | Pre-Processing | Modeling & Evaluation | Evaluation & Future Works |

| More Data | Create a WebApp | AI Trading Bot |

- 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