



STOCK FORECASTING

BY ZHUOQUAN CHEN

BACKGROUND

Technology has become a part of finance. Many financial firms are now growing into technology companies instead of staying in the traditional aspect. The frequency of financial trading generates large data volumes, which make financial firms paid more and more attention to technology over the years.

**Today, technology is one of the main
pushers in finance.**

PROBLEM STATEMENT

This is a project about stock price prediction, which fully reflects how the difference between using technology and traditional methods for financial analysis. In this project, I am using apple stock data for the past 5 years, where contains market analysis and close price prediction. In the term of market analysis, I calculated the stock return in daily, weekly, monthly and even quarterly, and also used a trading strategy. In the term of price prediction, I used the time series model ARIMA and tried to find the best p, d, q to fit the model for predicting the price movements, and used mean squared error as evaluation function of this model.

EDA

Closing Price

Using different time interval for more flexibility on the frequency conversion of times series.

Trends:

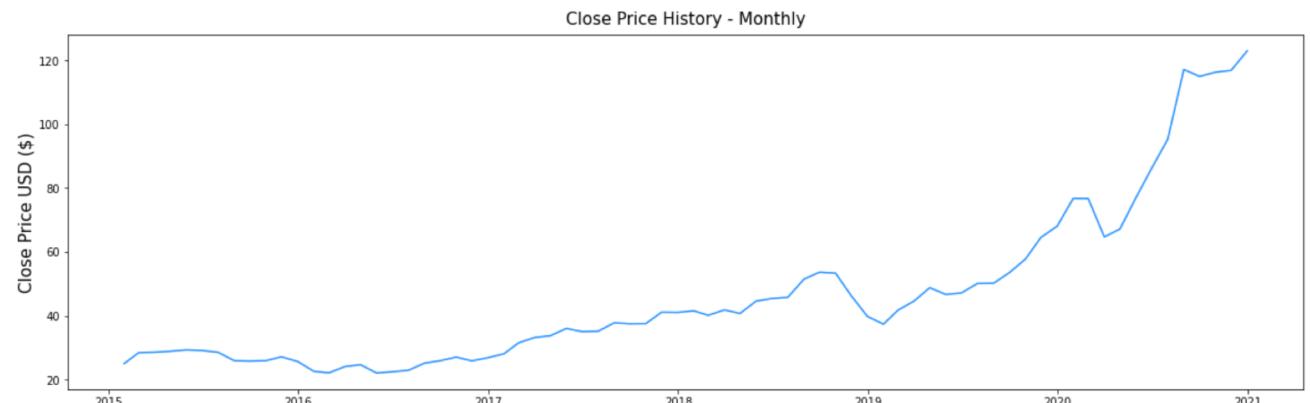
Across the full range of time, the close price increases, especially in 2020.

Seasonality:

There are fluctuations, but can't see a fixed and known frequency.

Autocorrelation:

The data are correlated with each other.

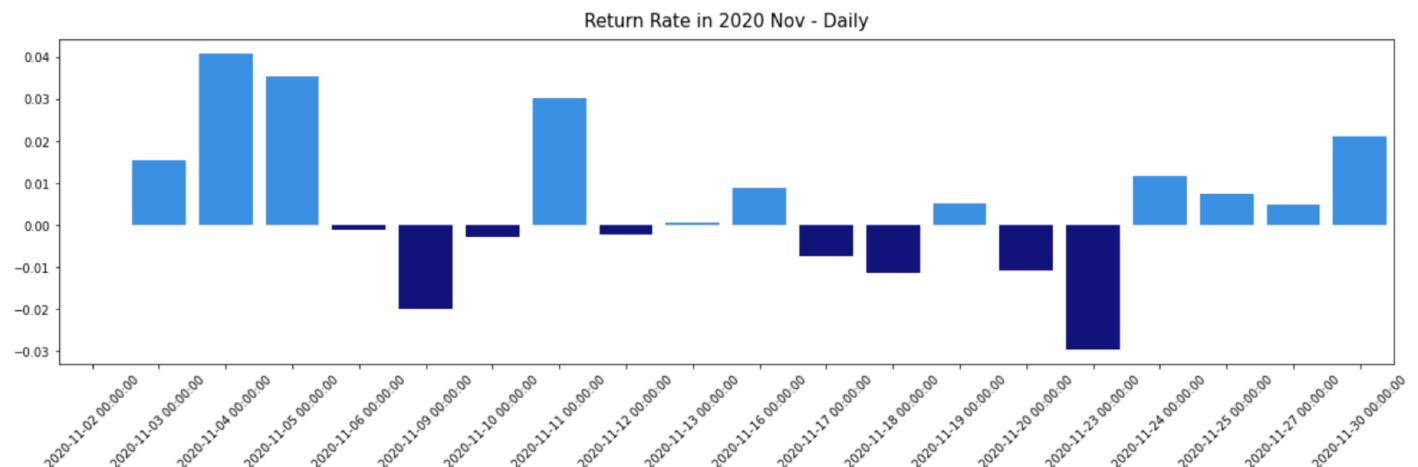
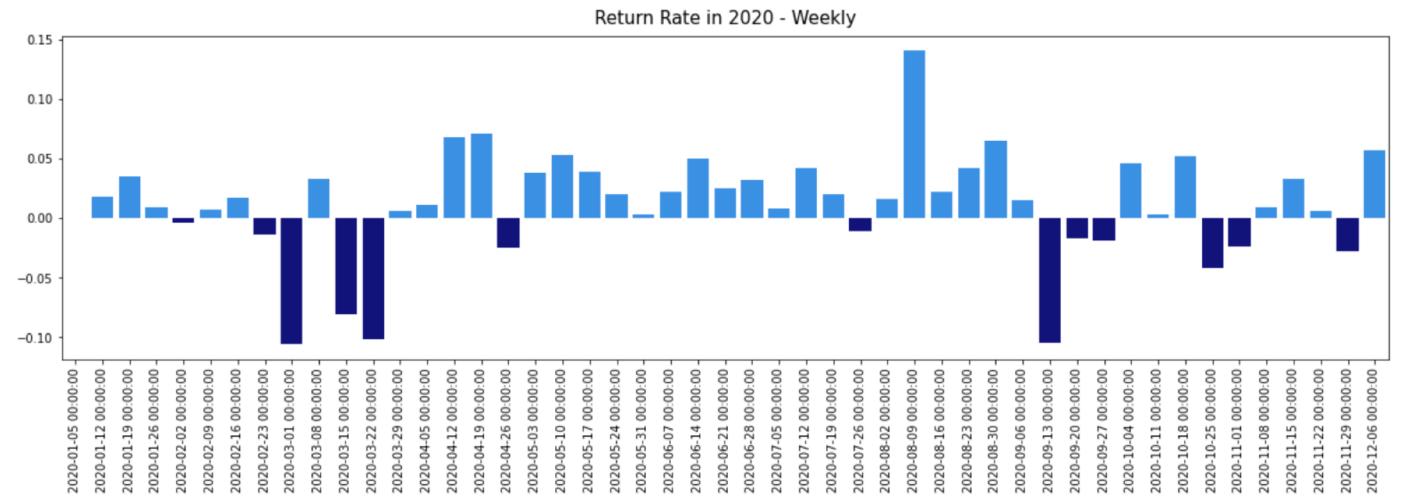


EDA

Return Calculation

Close price / Close price . shift(1) - 1

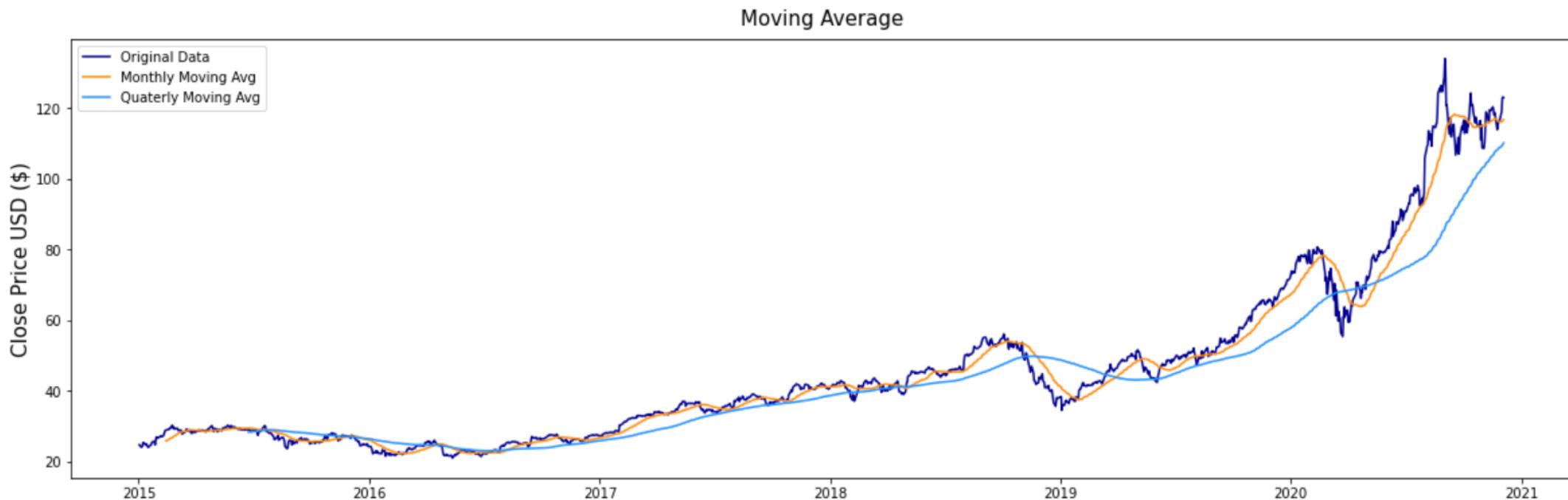
- Weekly return in 2020
- Daily return in 2020 Nov



EDA

Moving Windows (Moving/Rolling Average)

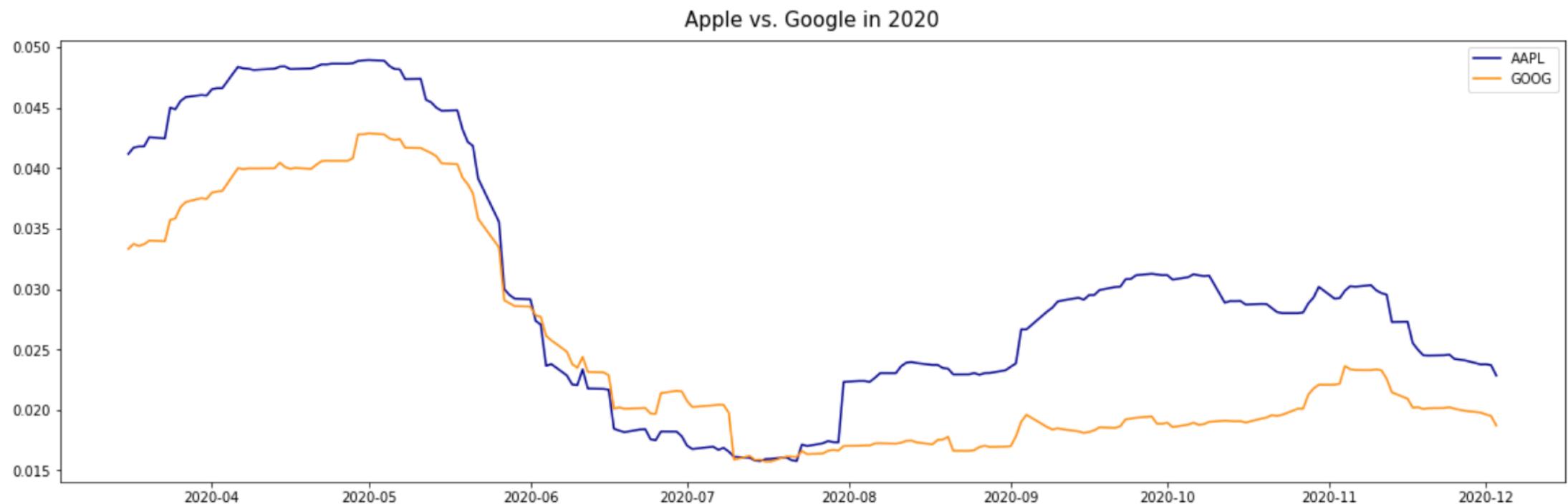
- To compute the statistic data by a particular period of time and then slide the window across the data by a specified interval.
- A rolling mean is smooth out short-term fluctuations and highlight longer-term trends in data.



EDA

Volatility

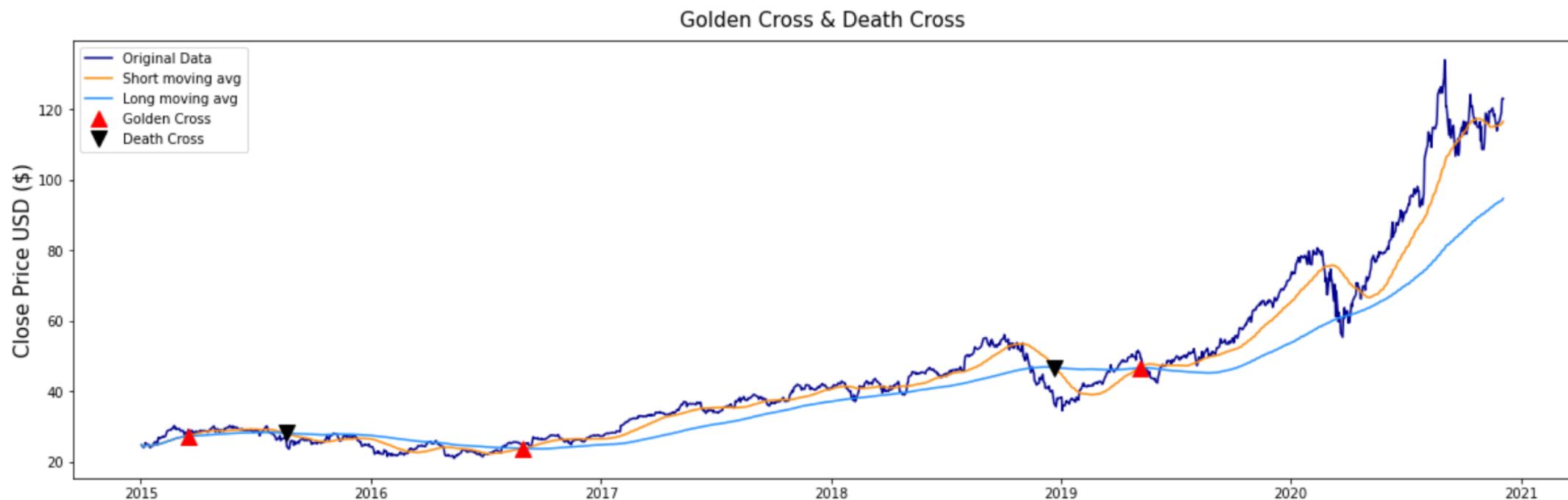
The volatility is the variance of the returns, which is a measurement of the change of a stock over a specific period. Generally, the higher the volatility, the riskier the investment in that stock. The volatility is calculated by taking a rolling window standard deviation on the percentage change in a stock.



EDA

Trading Strategy - The momentum strategy

- The 50-day and 200-day moving average are considered to be important trading signals.
- A buy signal is generated when the short-term average crosses the long-term average and rises above it, people also called it a Golden Cross.
- A sell signal is triggered by a short-term average crossing long-term average and falling below it, people also called it a Death Cross.



more information: <https://www.investopedia.com/terms/m/movingaverage.asp>

ARIMA

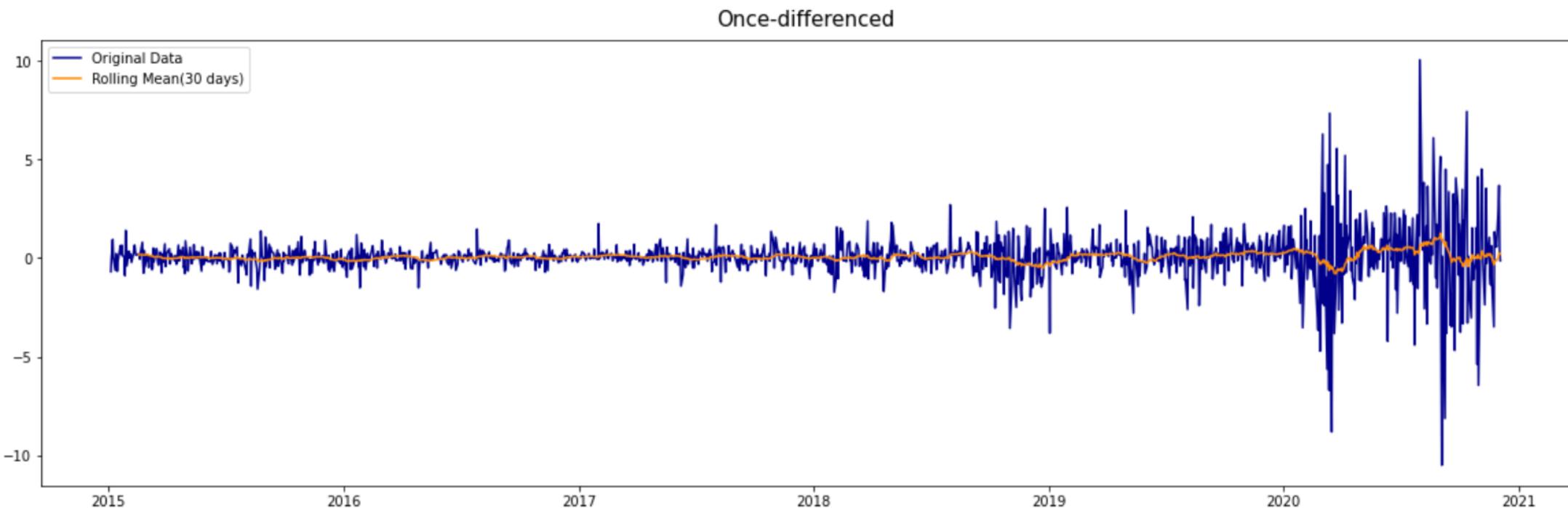
The ARIMA model consists of three parts: AR+ I + MA

- For non-stationary time series, d should be carried out difference first to transform it into stationary time series.
- Find out the appropriate p and q values by using ACF and PACF or AICAR.
- Using p, d and q as parameters in ARIMA model.

ARIMA

Difference d of Time Series

The requirement of ARIMA model for time series is stationary type. So, when you get a non-stationary time series, the first thing you do is you do a difference in the time series until you get a stationary time series.

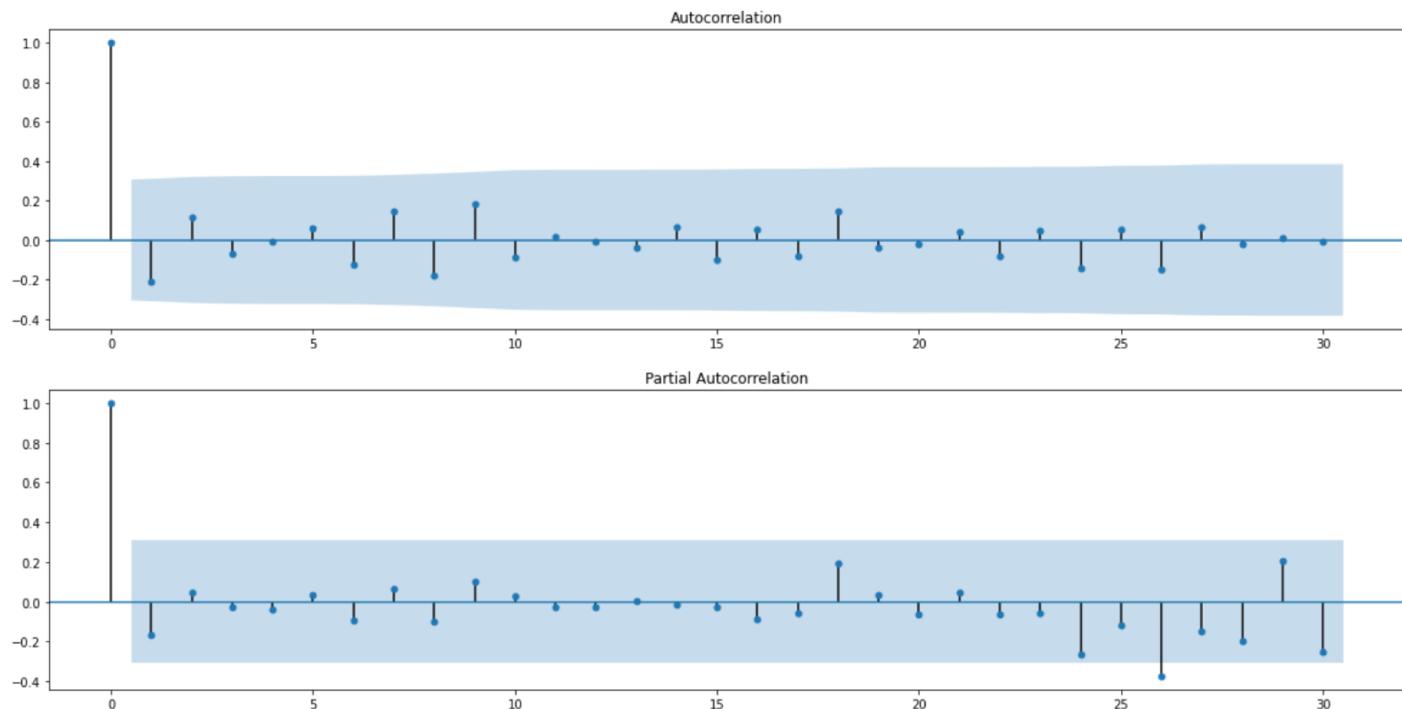


ARIMA

The appropriate p, q

Option 1: Using ACF & PACF

$p = 1, q = 0$



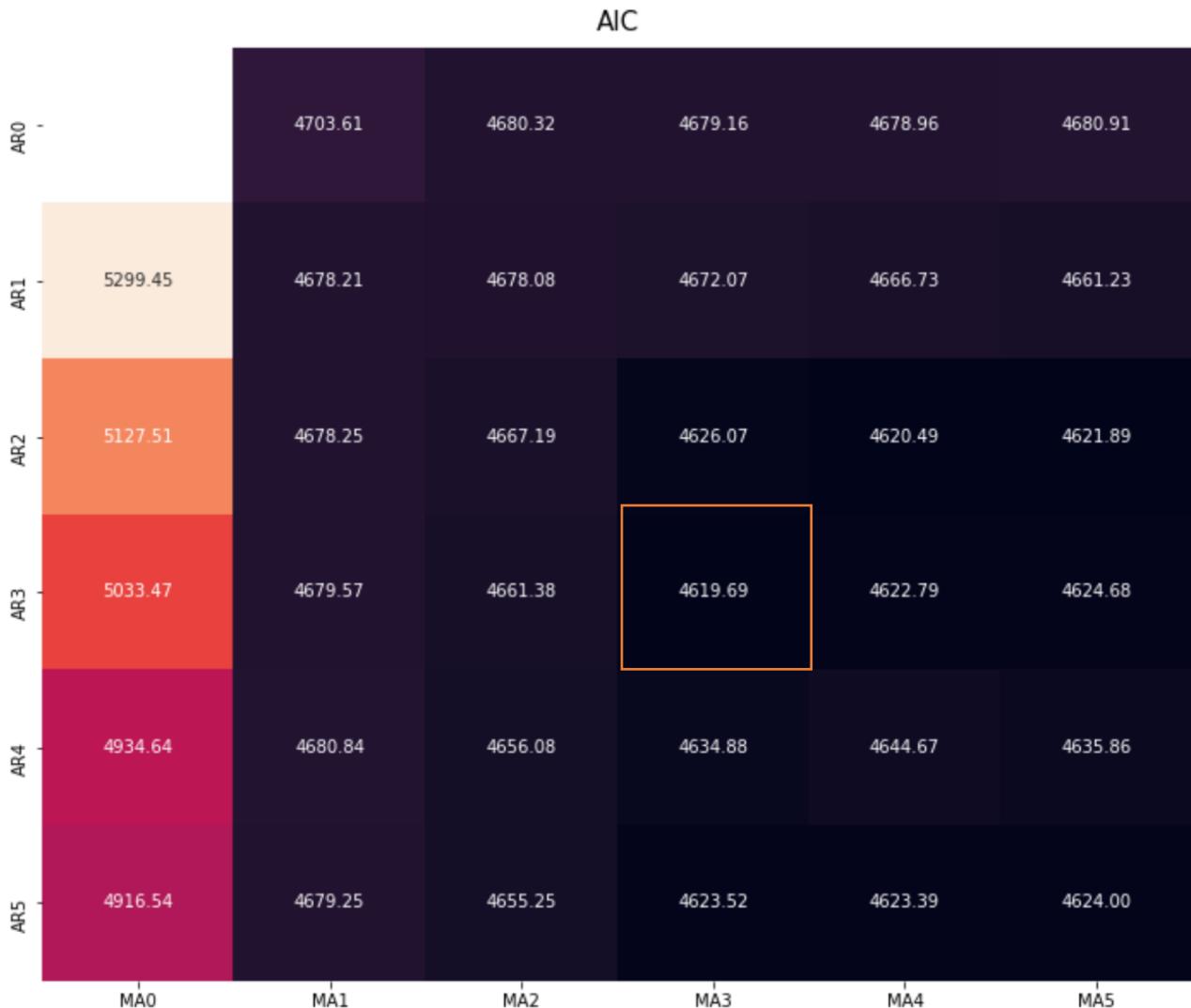
ARIMA

The appropriate p, q

Option 2: Using AIC

Using this method can create a AIC heatmap, which find the x, y correspond the smallest number.

p = 3, q = 4

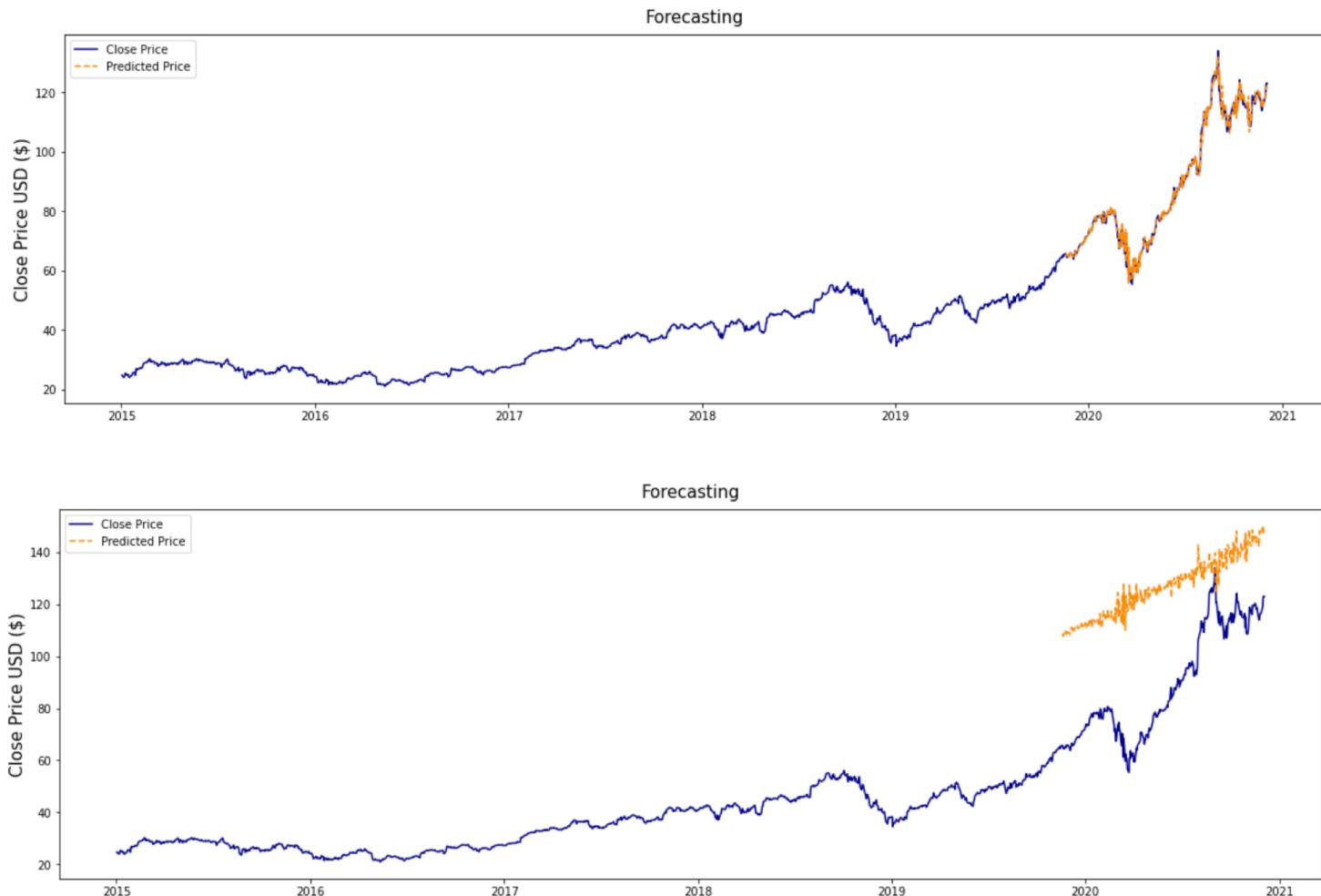


ARIMA Modeling

Predicting in 2020.

$P, d, q = 1, 1, 0$

$P, d, q = 3, 1, 4$



ARIMA

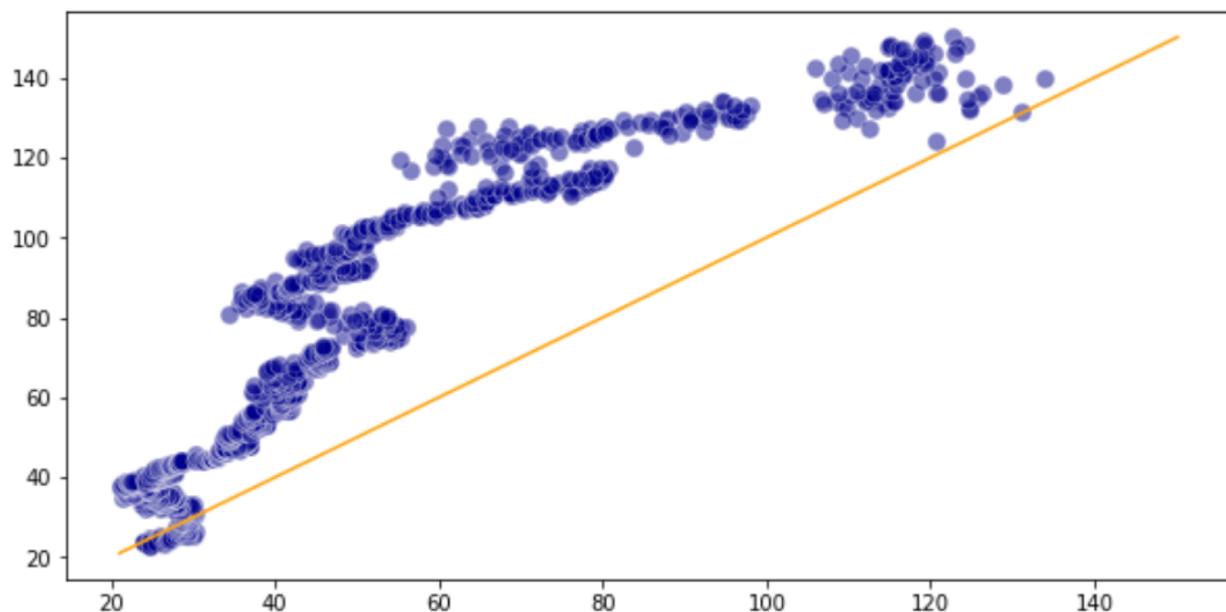
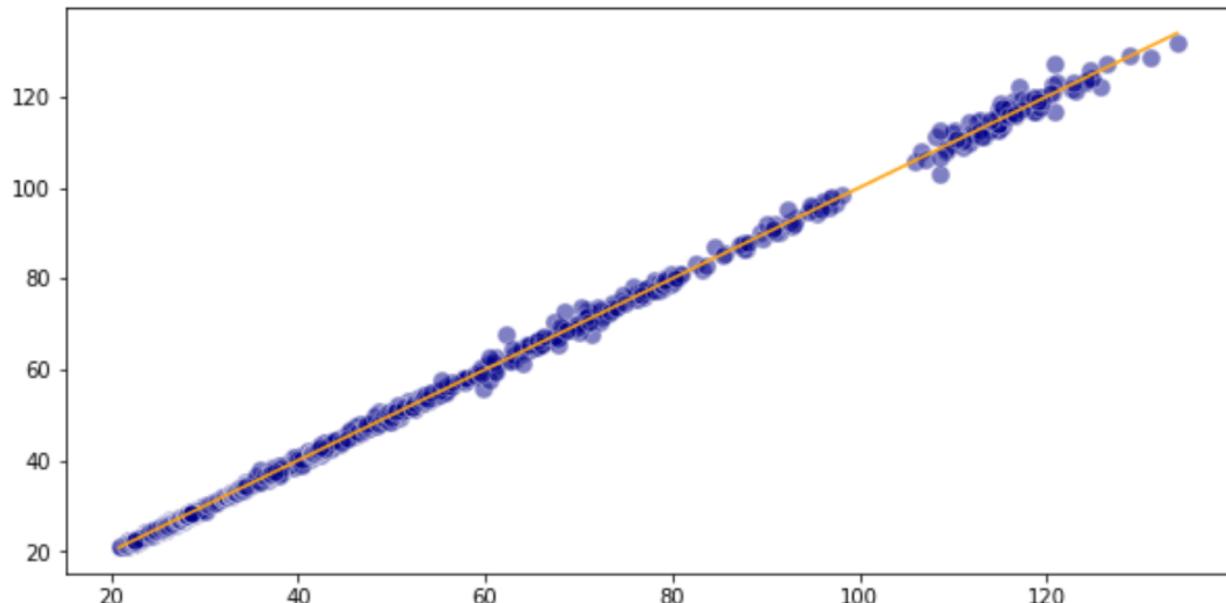
Model Evaluation

P, d, q = 1, 1, 0

rmse: 0.7067031040119878

P, d, q = 3, 1, 4

rmse: 28.4553423052126



ARIMA Forecasting

The 10 business days will predict out in future.

	Date	Close	Preds	ARIMA_Forecast_10D
0	2020-11-05	118.824997	116.550079	NaN
1	2020-11-06	118.690002	116.665791	NaN
2	2020-11-09	116.320000	116.748941	NaN
3	2020-11-10	115.970001	117.702144	NaN
4	2020-11-11	119.489998	120.045282	NaN
5	2020-11-12	119.209999	117.510147	NaN
6	2020-11-13	119.260002	119.775513	NaN
7	2020-11-16	120.300003	120.623684	NaN
8	2020-11-17	119.389999	119.137167	NaN
9	2020-11-18	118.029999	118.914252	NaN
10	NaT	NaN	NaN	119.325669 2020-12-04
11	NaT	NaN	NaN	118.041195 2020-12-07
12	NaT	NaN	NaN	116.784790 2020-12-08
13	NaT	NaN	NaN	117.711698 2020-12-09
14	NaT	NaN	NaN	118.425742 2020-12-10
15	NaT	NaN	NaN	118.413601 2020-12-11
16	NaT	NaN	NaN	118.638143 2020-12-14
17	NaT	NaN	NaN	118.569488 2020-12-15
18	NaT	NaN	NaN	116.953294 2020-12-16
19	NaT	NaN	NaN	117.523460 2020-12-17

Comparing ARIMA vs. Linear Regression vs. Real Close Price

	Date	Close	Preds	ARIMA_Forecast_10D	LR_Forecast_10D	
0	2020-11-05	118.824997	116.550079		NaN	NaN
1	2020-11-06	118.690002	116.665791		NaN	NaN
2	2020-11-09	116.320000	116.748941		NaN	NaN
3	2020-11-10	115.970001	117.702144		NaN	NaN
4	2020-11-11	119.489998	120.045282		NaN	NaN
5	2020-11-12	119.209999	117.510147		NaN	NaN
6	2020-11-13	119.260002	119.775513		NaN	NaN
7	2020-11-16	120.300003	120.623684		NaN	NaN
8	2020-11-17	119.389999	119.137167		NaN	NaN
9	2020-11-18	118.029999	118.914252		NaN	NaN
10	NaT	NaN	NaN	119.325669	121.072899	2020-12-04 122.25
11	NaT	NaN	NaN	118.041195	119.742124	2020-12-07 123.75
12	NaT	NaN	NaN	116.784790	116.169513	2020-12-08 124.38
13	NaT	NaN	NaN	117.711698	117.520759	2020-12-09 121.78
14	NaT	NaN	NaN	118.425742	118.401117	2020-12-10 123.24
15	NaT	NaN	NaN	118.413601	118.974370	2020-12-11
16	NaT	NaN	NaN	118.638143	121.492608	2020-12-14
17	NaT	NaN	NaN	118.569488	125.249479	2020-12-15
18	NaT	NaN	NaN	116.953294	125.618002	2020-12-16
19	NaT	NaN	NaN	117.523460	125.474688	2020-12-17

CONCLUSION

Even ARIMA model performs pretty well in time series, such as stock, a variety of calculations, analysis, and trading strategy seem reasonable. However, many other factors can impact stocks' price, such as national policies, corporation internal issues, etc. and this eventually can not be predictable and computable. What we can do is improving the trading strategies, algorithms, or models.

NEXT STEP

In the term of analysis, I will try to do more trading strategies for a further insight or analysis, such as inversion strategy, forecasting strategy and High-Frequency Trading (HFT) strategy. In the term of model and prediction, I will try to use another technology, such as Long Short Term Memory, to see how they perform different, then find out a best solution.

THANK YOU!