Data Science Capstone project

Introduction

- To explore the effectiveness of using Exponential Moving Average (EMA) to predict the closing prices via a simple linear regression model.
- Independent Variable is the EMA of a stock and the Dependent Variable is its closing price.
- Simple linear regression is an analysis that assesses whether one predictor variables explain the dependent (criterion) variable.
- EMA is a type of moving average that gives more weighting or importance to recent price data.

Parameters and Data

- Effective time frame from around the onset of Covid (2020).
- Stock used is Apple Inc (AAPL).
- Data summary as follows:

DatetimeIndex: 564 entries, 2020-02-01 to 2022-03-28

Data columns (total 3 columns):

#	Column	Non-Null Count	Dtype
0	Open	564 non-null	float64
1	Close	564 non-null	float64
2	EMA_10	564 non-null	float64

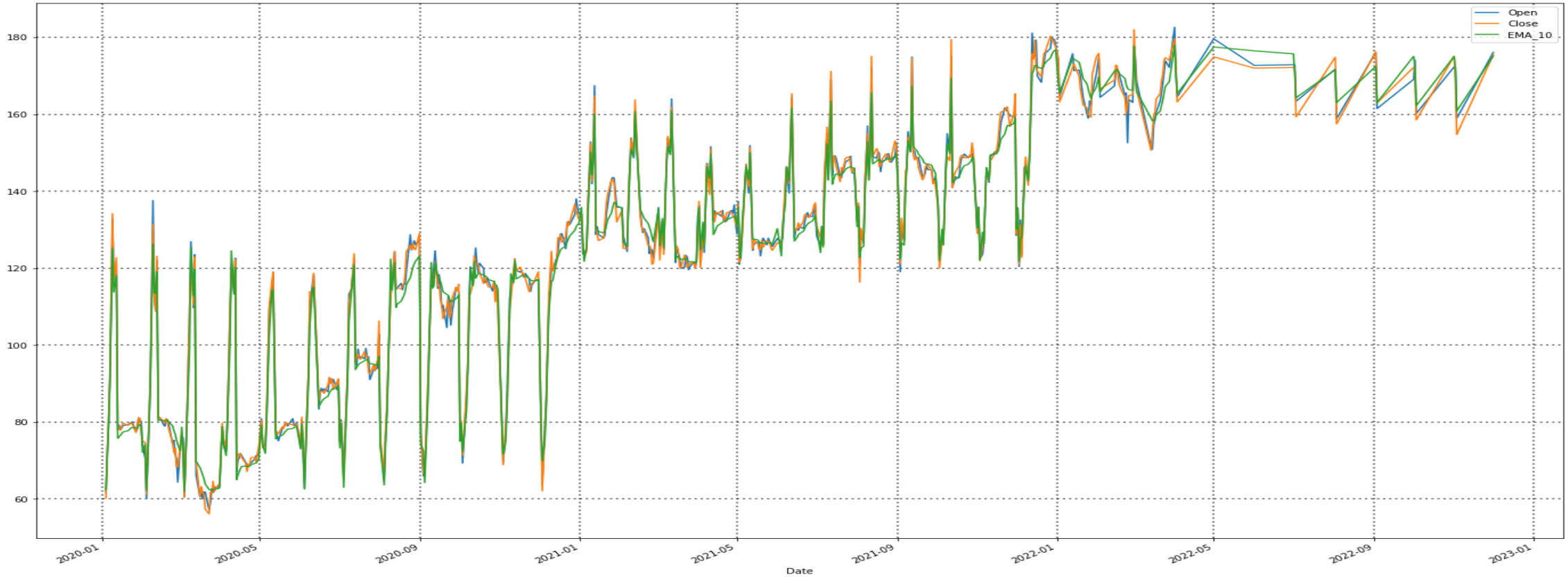
dtypes: float64(3)

memory usage: 17.6 KB

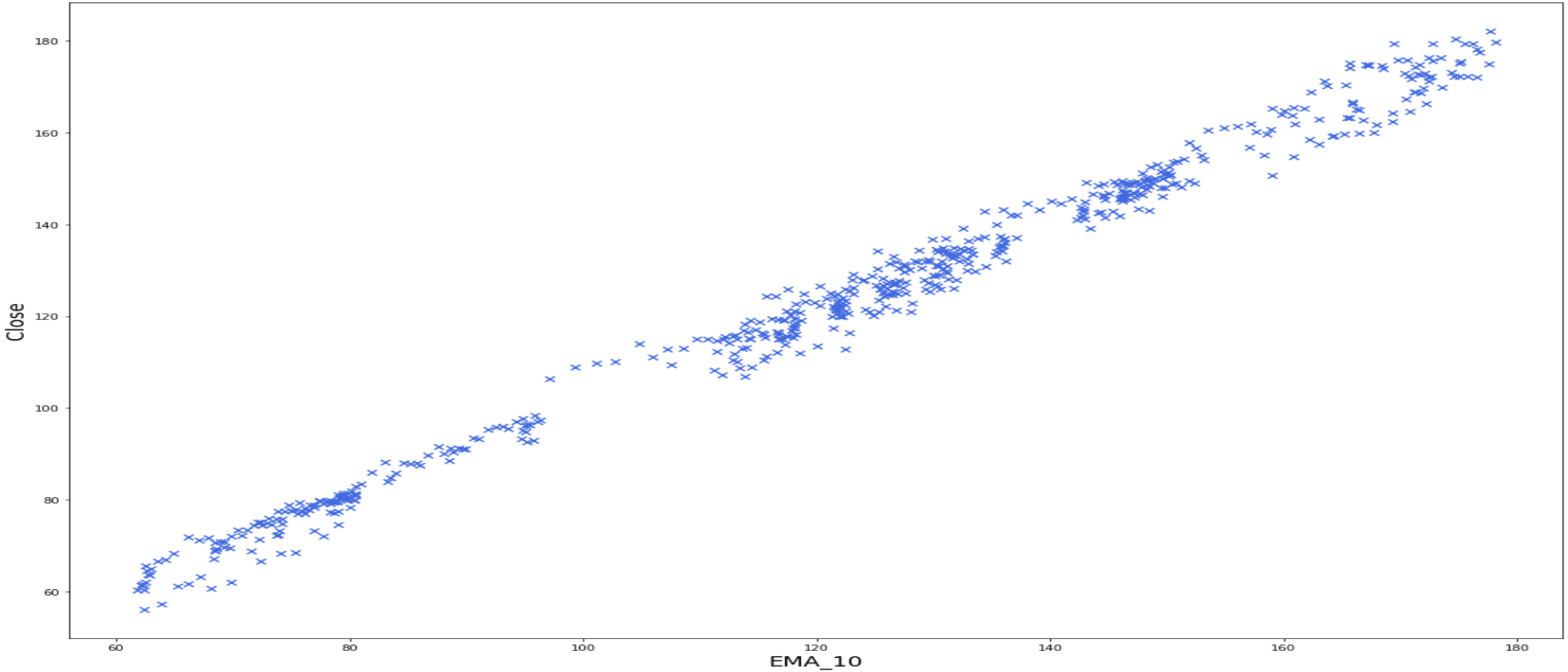
Duplicate:

- No duplicates or null values in data

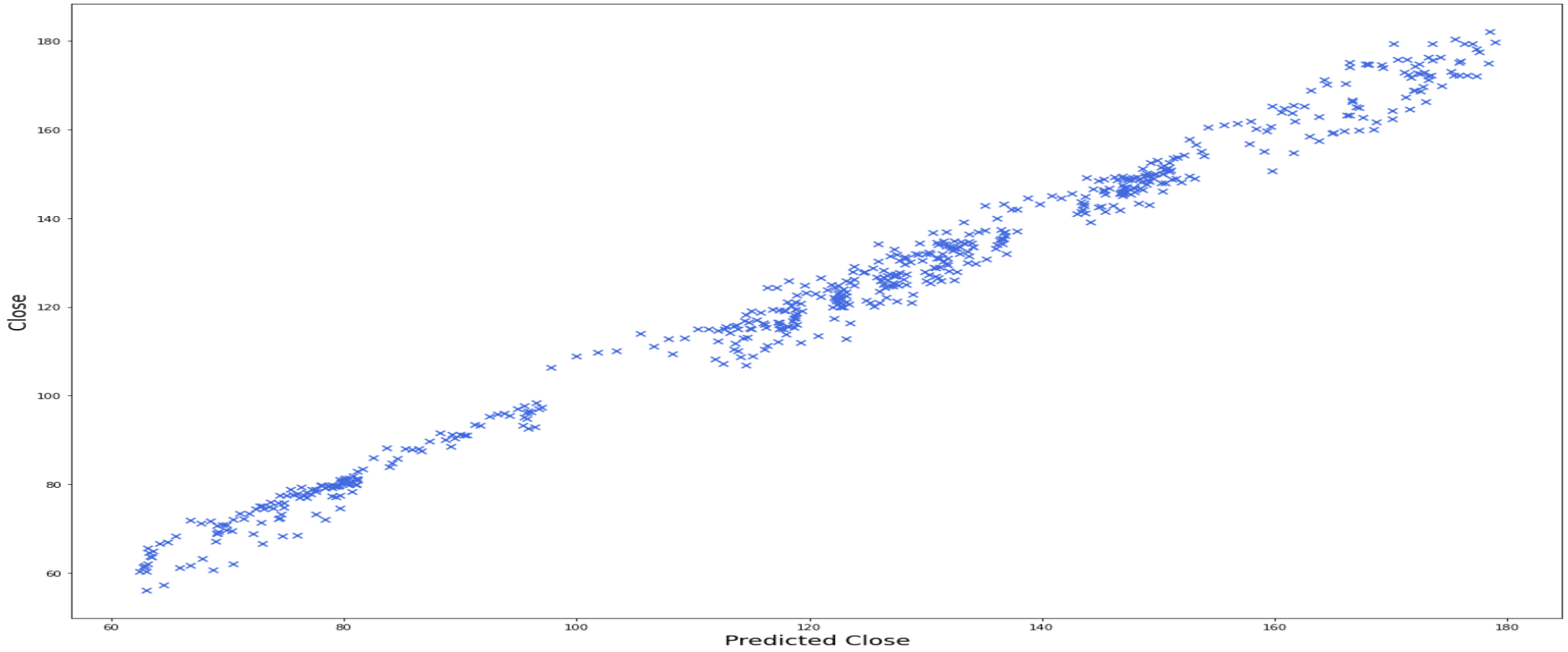
Plot denoting how the variables track with time



Plotting the 2 variables



Plotting Close against Predicted Close



Methodology

- Buy signal when Predicted Close > Opening price.
- Capital of $1000 for each trade where buy signal is triggered and trade is made
- Sell entirely at Close

Results

- 274 trades were made
- Based on $1000 per trade capital.
- Profits were $940.41 dollars which represents a return of roughly 94% of capital
- Underperformed compared to share price of the stock. A $1000 investment in the stock at the start of the period would have yielded a rounded return of $1349

Limitations of Model

- The assumptions that we are able to buy just at opening and sell just at close
- Brokerage fees were not taken into account which would have diminished profits further
- Only one stock used

Potential Improvements

- Adding Brokerage costs per trade
- Improve the model and profits by selling at prices above the predicted closing price instead of only at closing.
- Introduction of other TA to be used in conjunction (to predict the above mentioned point for example)
- Increasing the pool of datasets by testing the model on other stocks.