# Team B Project

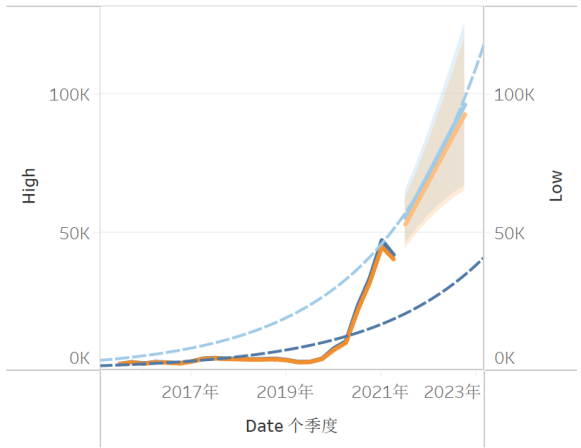
The Data Analysis about Tesla and Facebook Stock

## Members

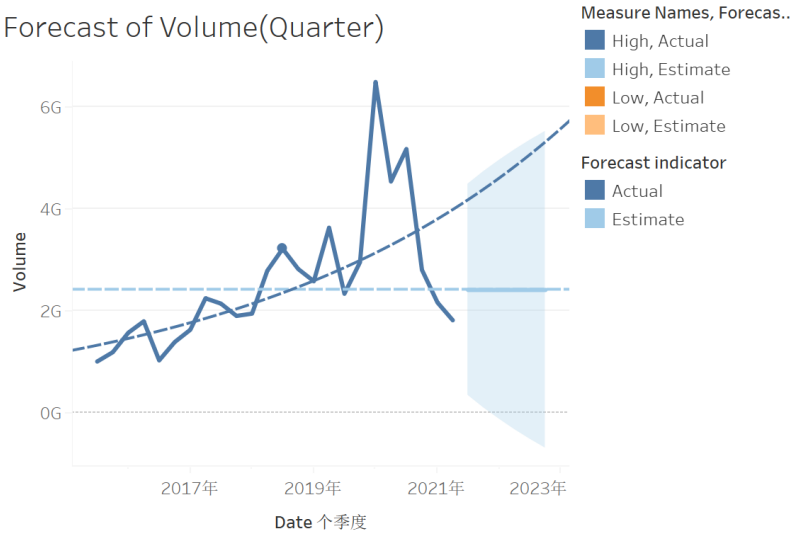
Dong Liu, Tianyang Liao,Zixuan Lu, Siyuan Li

TSLA

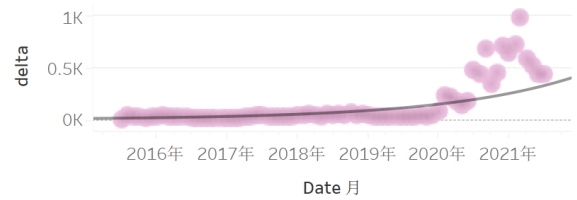
Quarter of Date (High&Low)




Forecast of Volume(Quarter)



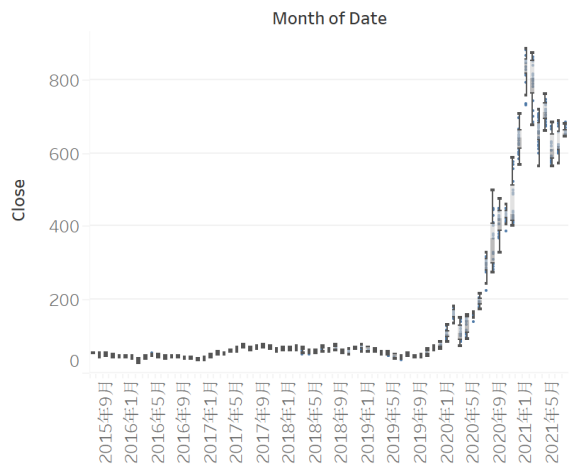

Forecast of Delta




VolatilityRatio

VR=delta/Close


BoxPlot Graph of Close



- Data Source

1. Data comes from https://finance.yahoo.com
2. Two stocks are Facebook and Tesla

- Data Manipulation

1. No data cleaning: no null values appears in data set.
2. Some attributes added: max_Delta(Max Price-Low Price), Delta(Close Price-Open Price), KPI(shows if the Delta is positive).

# KPI(IF Delta>0 THEN "Rise“ ELSE "Down“ END)

Tesla KPI text table

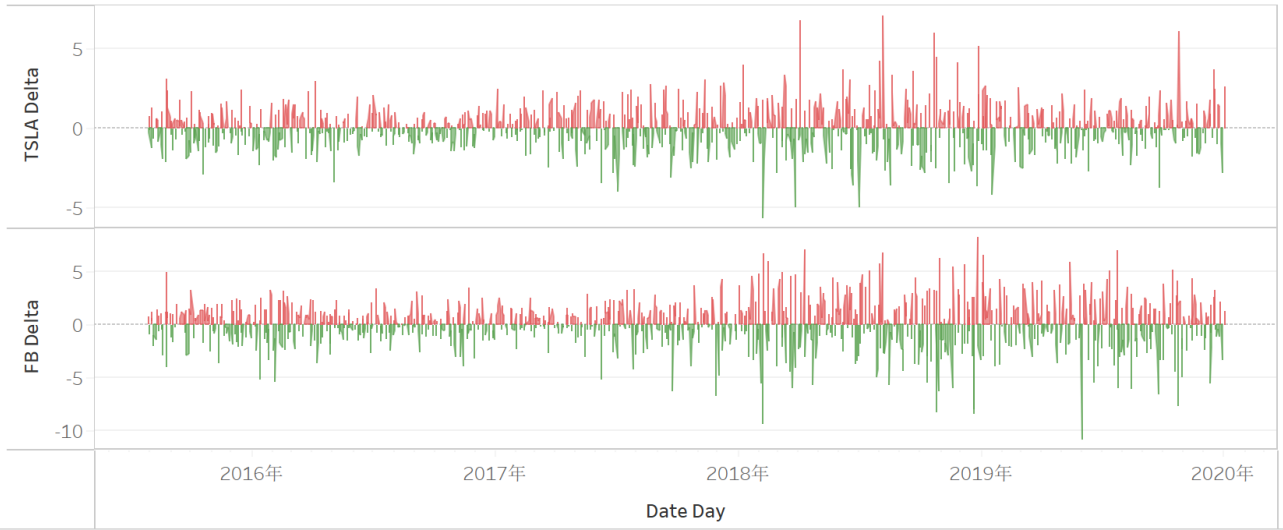
| Day of Date | | |
|---|---|---|
| 2015年7月29日 | ↓ | -0.09 |
| 2015年7月30日 | ↑ | 0.82 |
| 2015年7月31日 | ↓ | -0.29 |
| 2015年8月3日 | ↓ | -1.26 |
| 2015年8月4日 | ↑ | 1.25 |
| 2015年8月5日 | ↑ | 1.31 |
| 2015年8月6日 | ↓ | -0.68 |
| 2015年8月7日 | ↓ | -0.21 |
| 2015年8月10日 | ↑ | 0.60 |
| 2015年8月11日 | ↑ | 0.04 |
| 2015年8月12日 | ↑ | 0.63 |
| 2015年8月13日 | ↑ | 0.53 |
| 2015年8月14日 | ↓ | -0.82 |
| 2015年8月17日 | ↓ | -0.11 |
| 2015年8月18日 | ↑ | 1.07 |
| 2015年8月19日 | ↓ | -1.02 |

FaceBook KPI text table

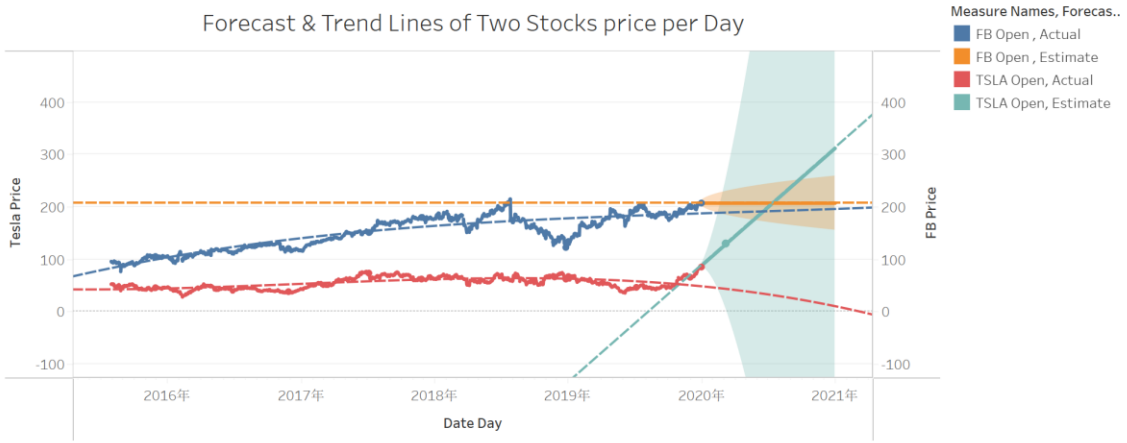
| Day of Date | | |
|---|---|---|
| 2015年8月3日 | ↑ | 0.61 |
| 2015年8月4日 | ↑ | 0.27 |
| 2015年8月5日 | ↑ | 1.19 |
| 2015年8月6日 | ↓ | -2.06 |
| 2015年8月7日 | ↓ | -1.08 |
| 2015年8月10日 | ↓ | -1.53 |
| 2015年8月11日 | ↓ | -0.11 |
| 2015年8月12日 | ↑ | 1.49 |
| 2015年8月13日 | ↓ | -0.62 |
| 2015年8月14日 | ↑ | 0.88 |
| 2015年8月17日 | ↓ | -0.49 |
| 2015年8月18日 | ↑ | 1.09 |
| 2015年8月19日 | ↑ | 0.68 |
| 2015年8月20日 | ↓ | -2.98 |
| 2015年8月21日 | ↓ | -1.46 |
| 2015年8月24日 | ↑ | 5.06 |

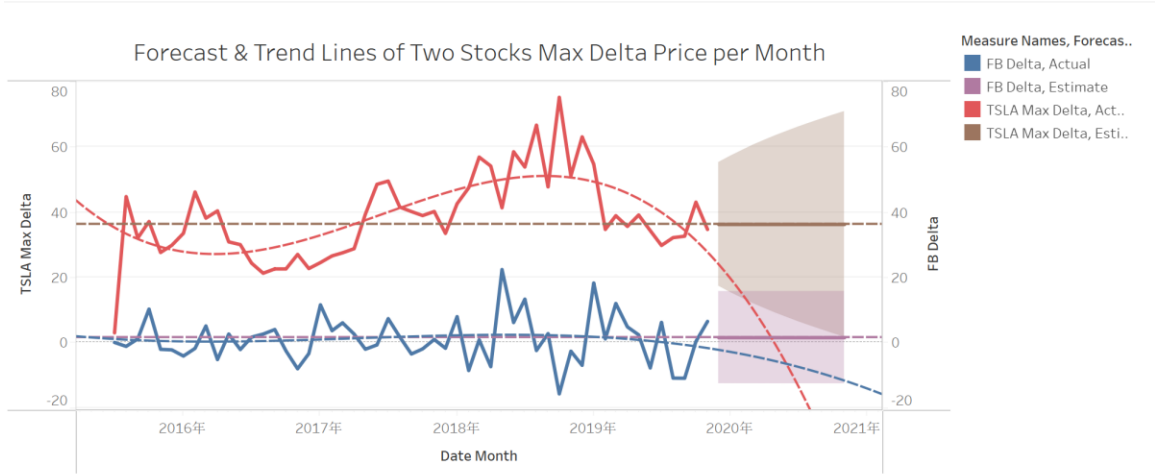
Tesla KPI
Down
Rise
Tesla KPI
↓ Down
↑ Rise
FB KPI
Down
Rise
FB KPI
↓ Down
↑ Rise

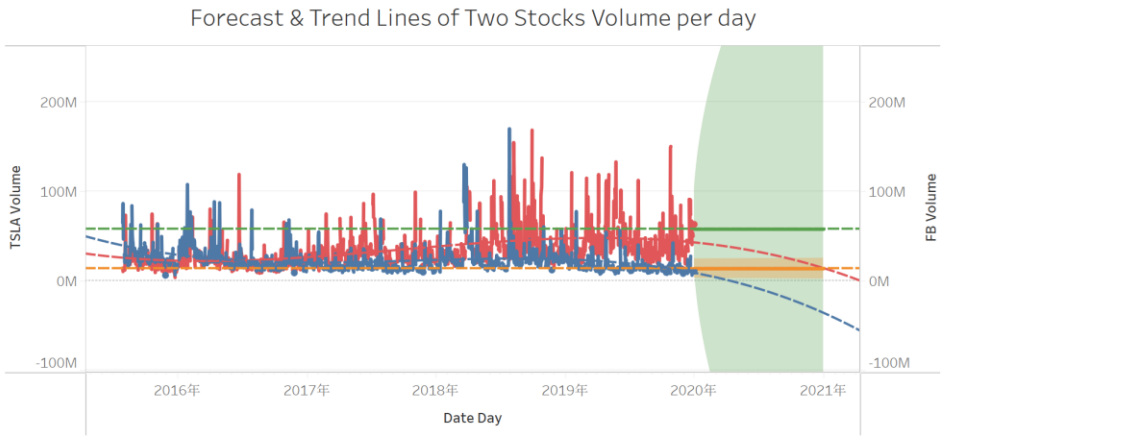
KPI per Day of Two Stocks

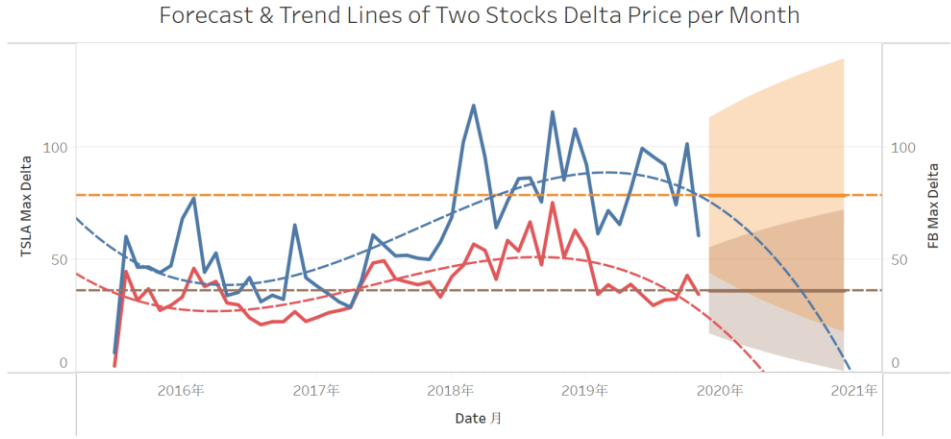


# Forecast & Trend Lines of Two Stocks

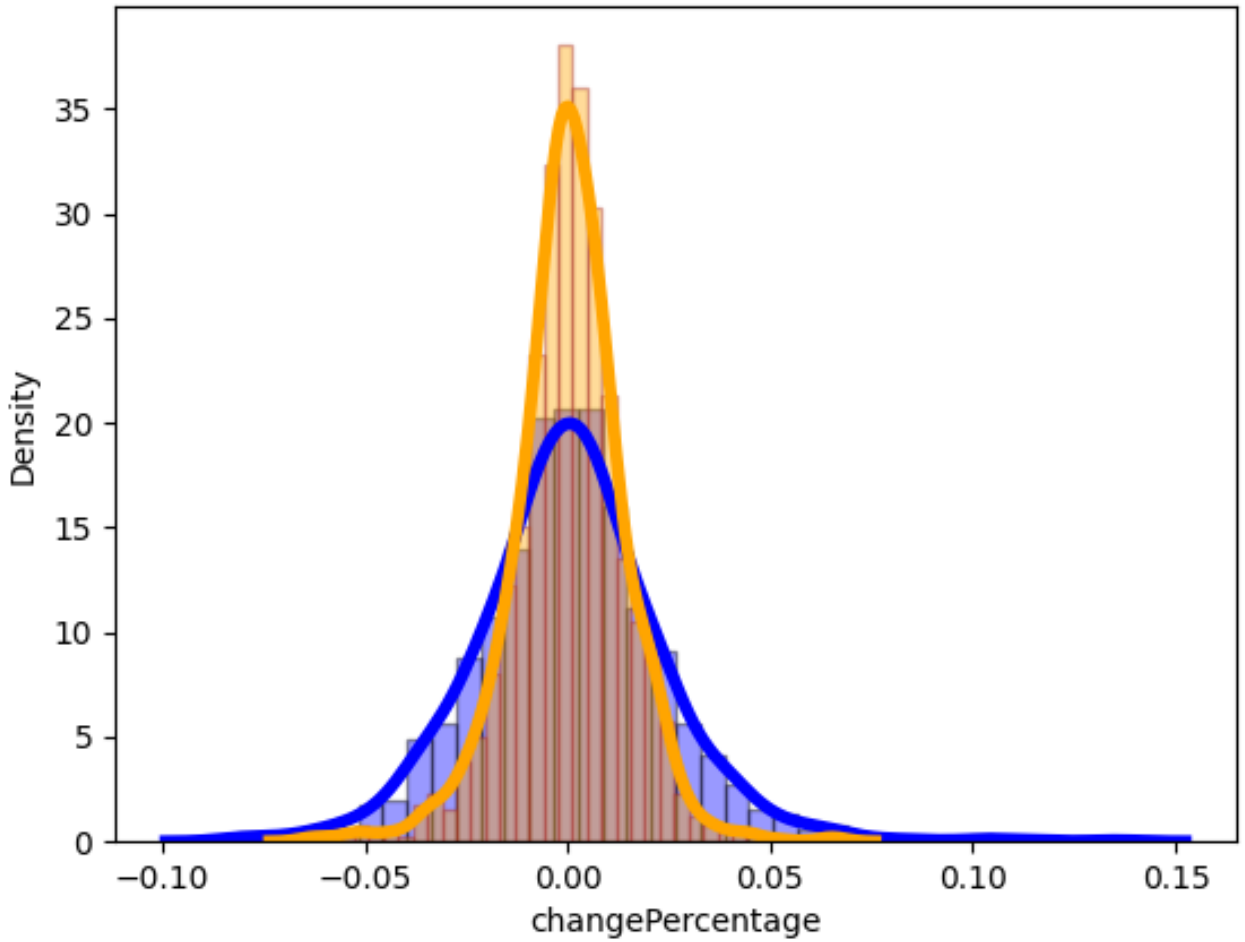


# Density (frequency) Plot of Change Percentage

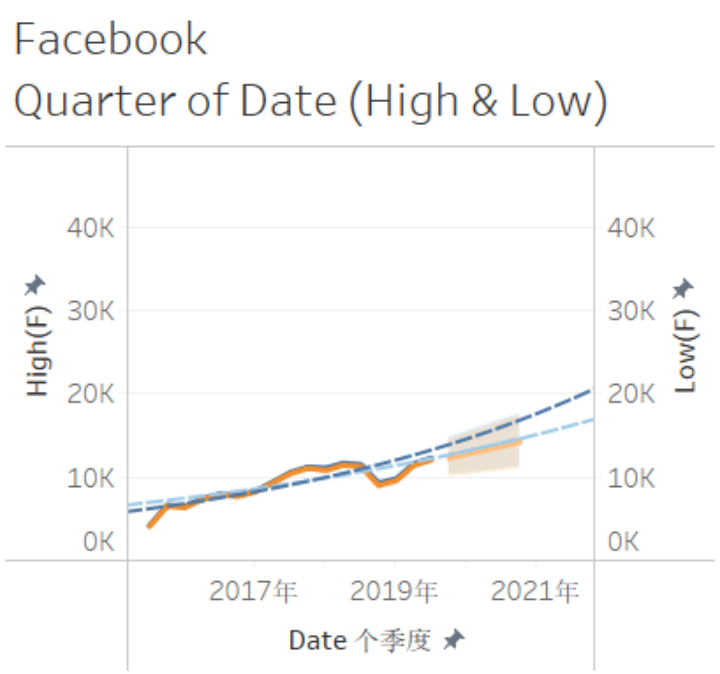


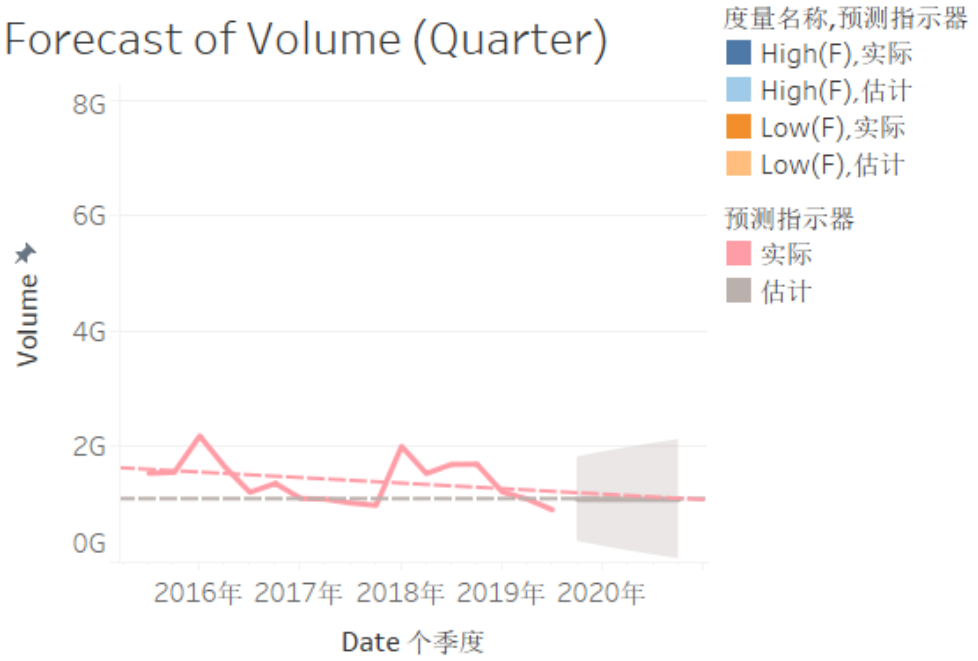

In order to estimate and compare this two stock's profit rate and risk, a new type of attribute, ChangePercentage, that present daily relatively change of price, is introduced for each stock. The formula that defines ChangePercentage is shown in below:

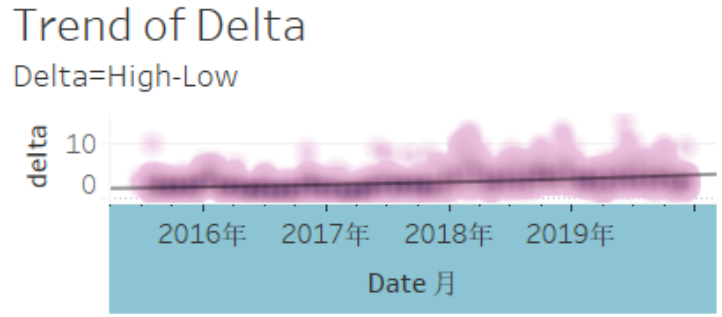
```
dftsla['changePercentage'] = dftsla['Close'] / dftsla['Open'] - dftsla['Open'] / dftsla['Open']
dffb['changePercentage'] = dffb['Close'] / dffb['Open'] - dffb['Open'] / dffb['Open']
```

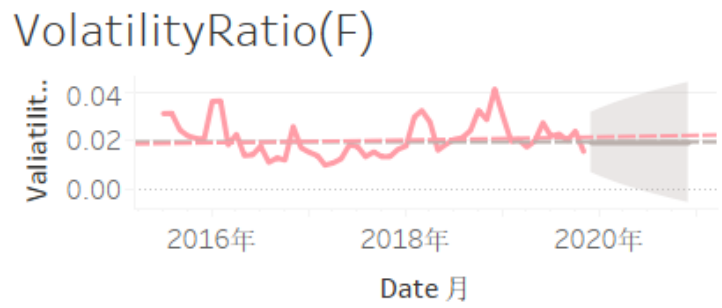
According to the change percentage of the two stock, we made this histogram, that can conveniently present and compare their risk and profit expectation.

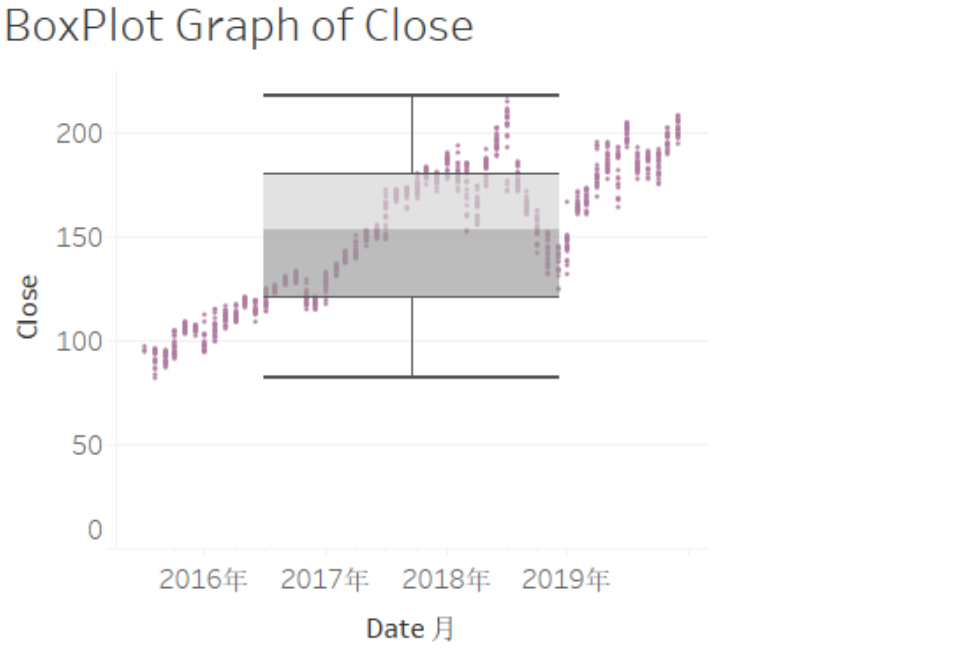
# Facebook Analysis



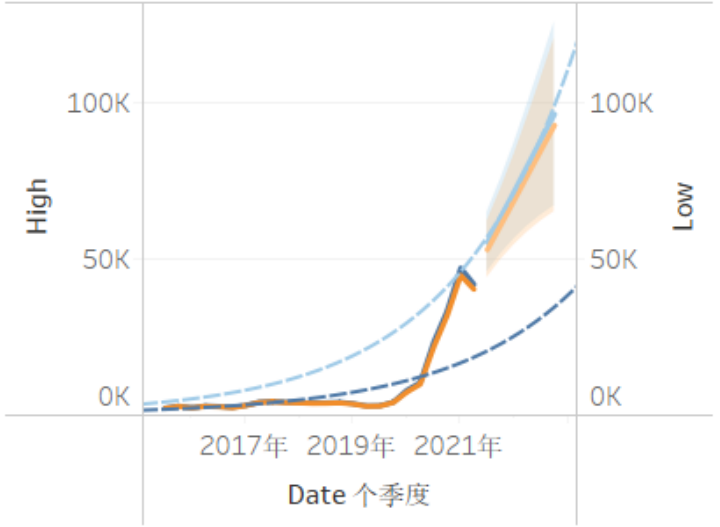



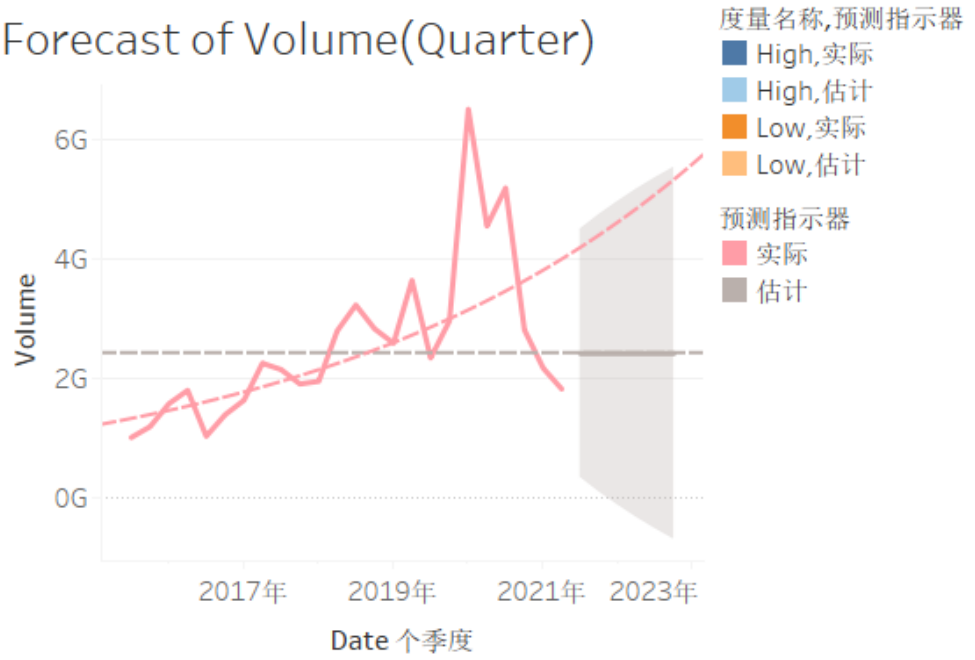
# TSLA Analysis

TSLA

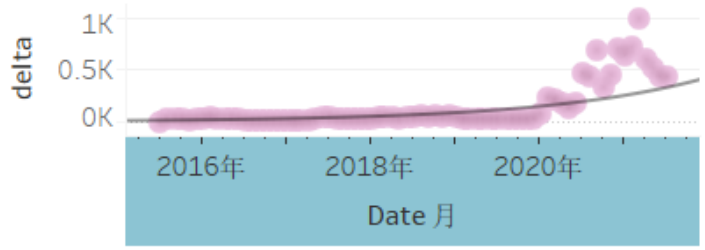
Quarter of Date (High&Low)



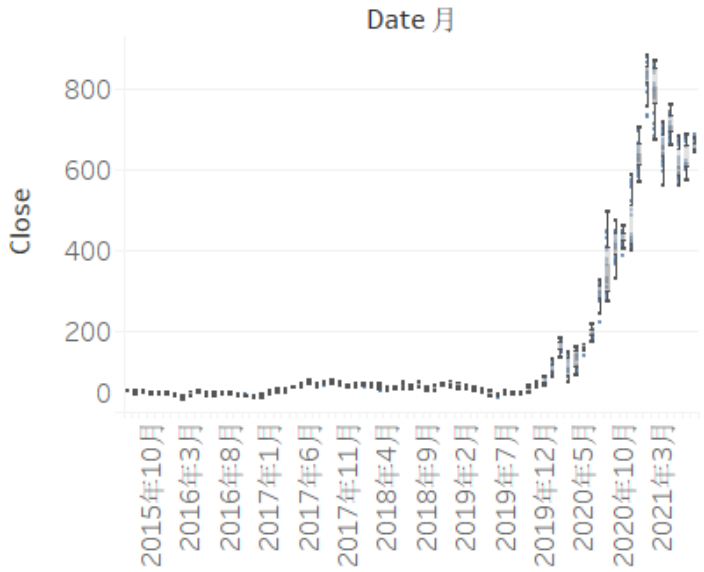

Forecast of Volume(Quarter)




Trend of Delta



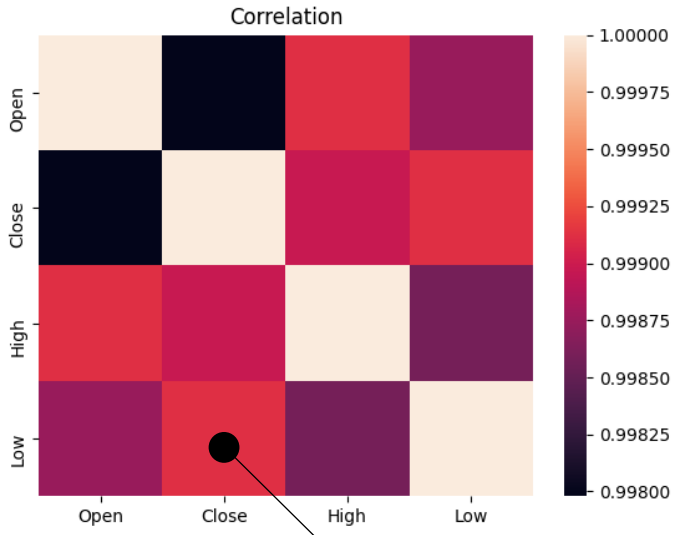

BoxPlot Graph of Close



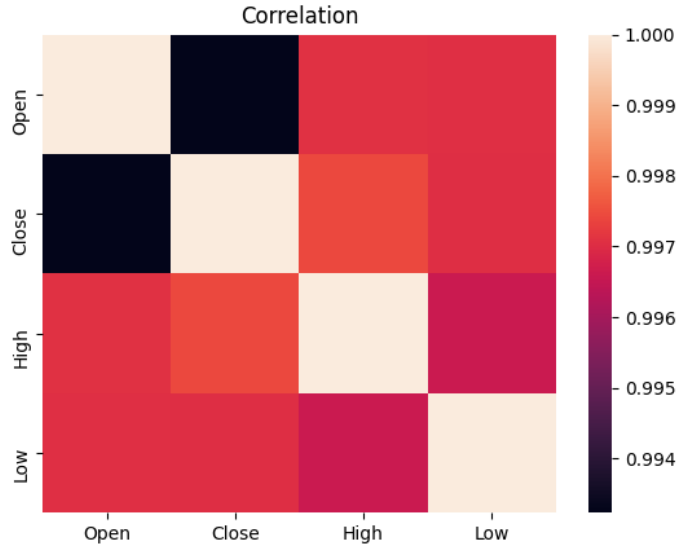

VolatilityRatio

VR=delta/Close

# The Correlation




FB Correlation Matrix

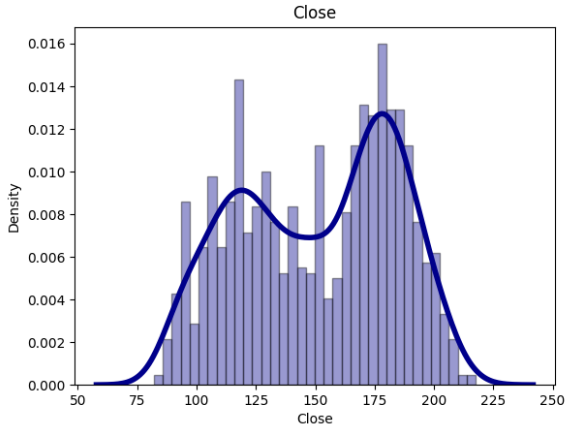



TSLA Correlation Matrix

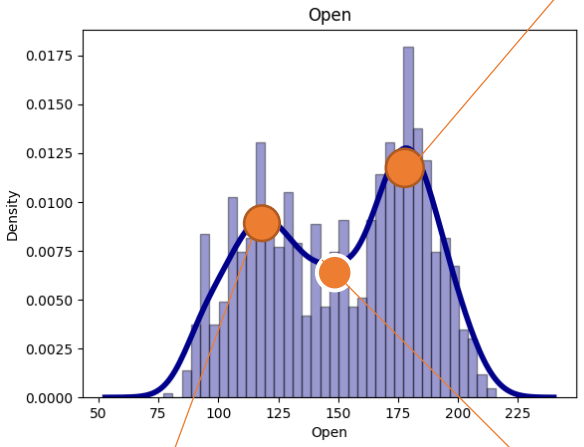
Strong Correlation

# Histogram & Density Plot

The density of the open price falling in this range is the highest

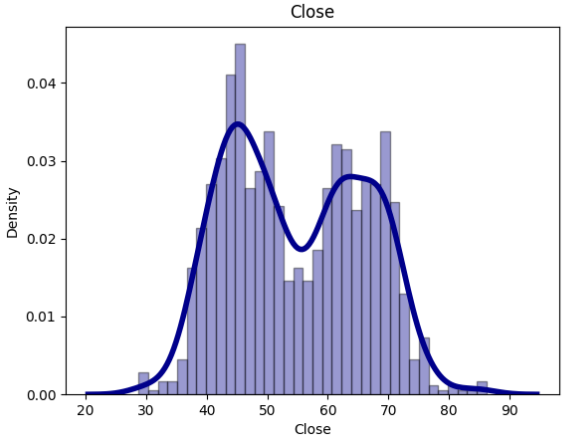



FB Close Price

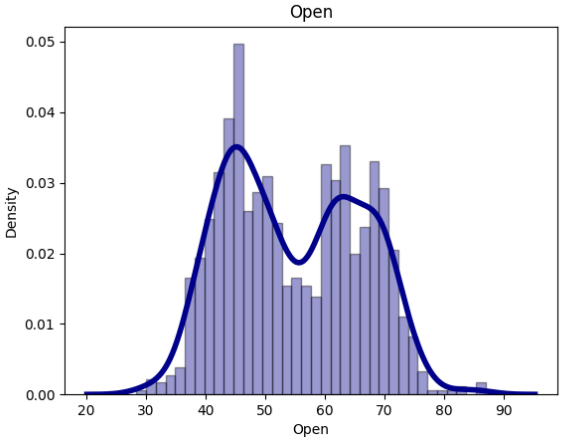



FB Open Price

Good Price



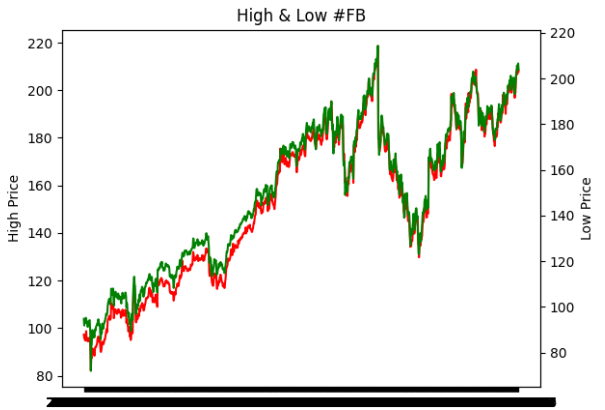

TSLA Close Price

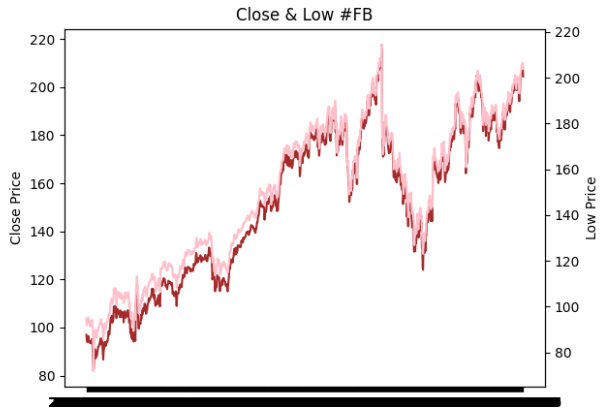



TSLA Open Price

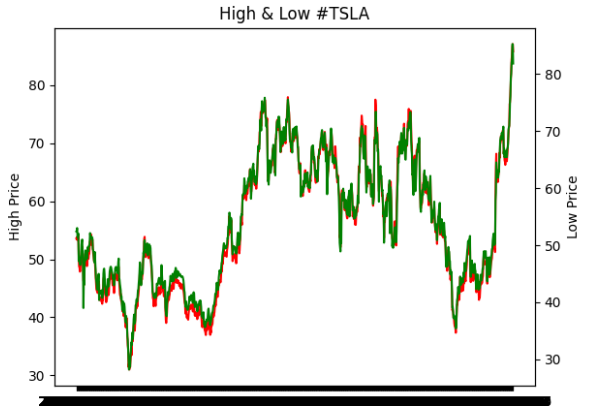
The first peak shows the consequences of the past.

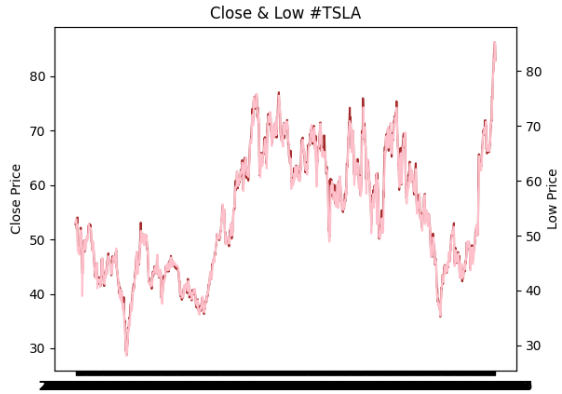
# Comparison Graph about High & Low Price, High & Open Price



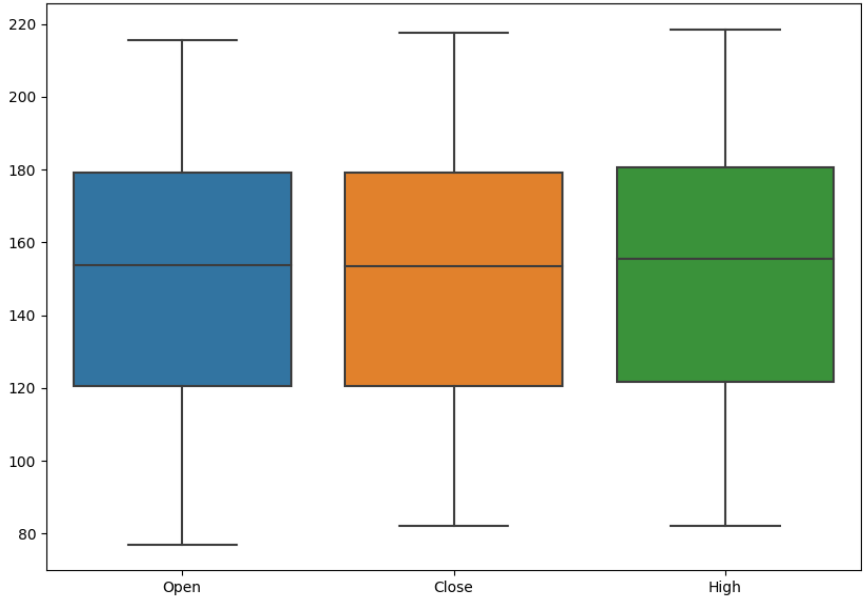

FB



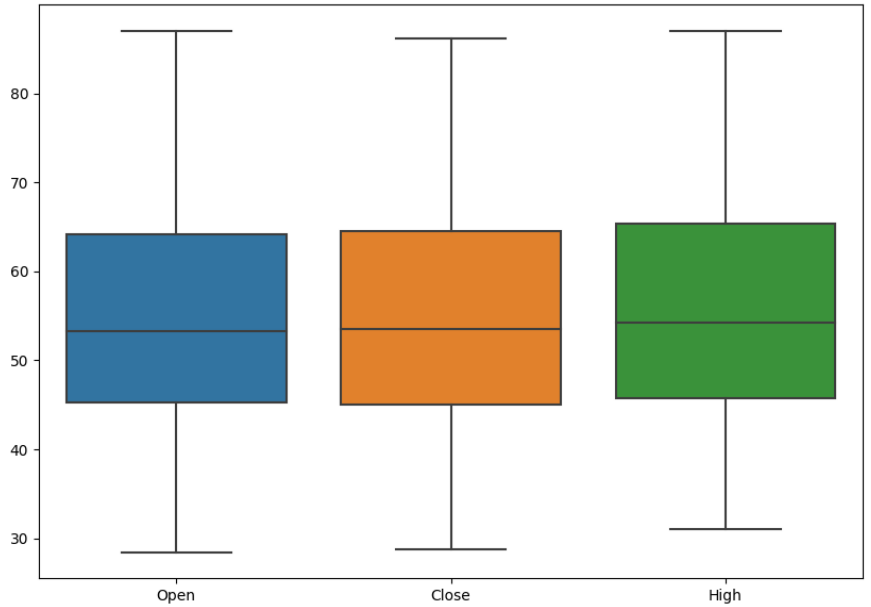




TSLA

# BoxPlot

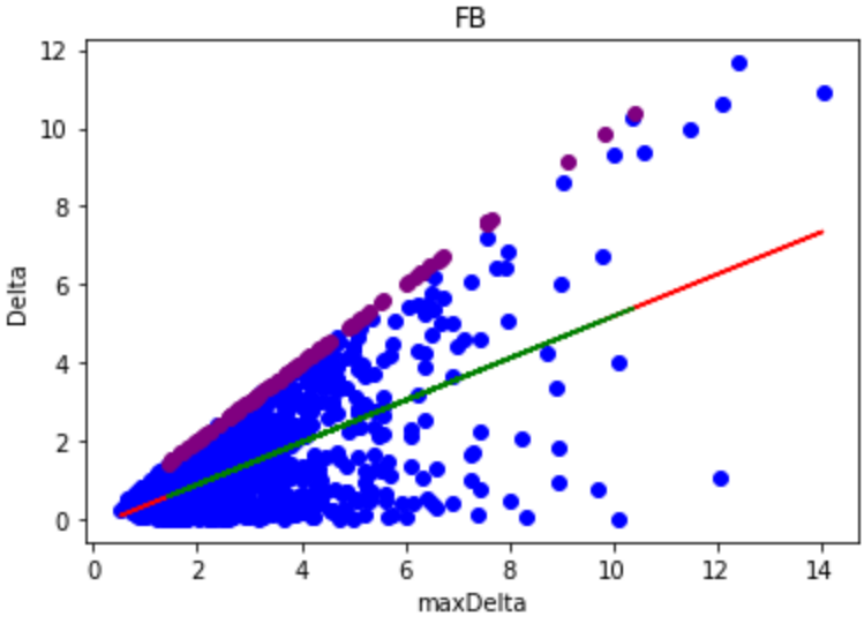


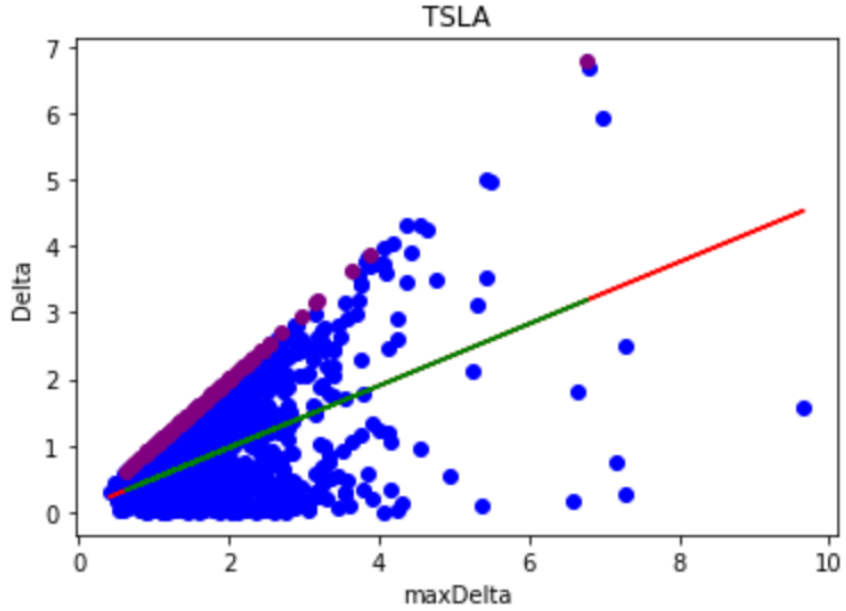

FB




TSLA

# Linear Regression Model on Delta and Max Delta

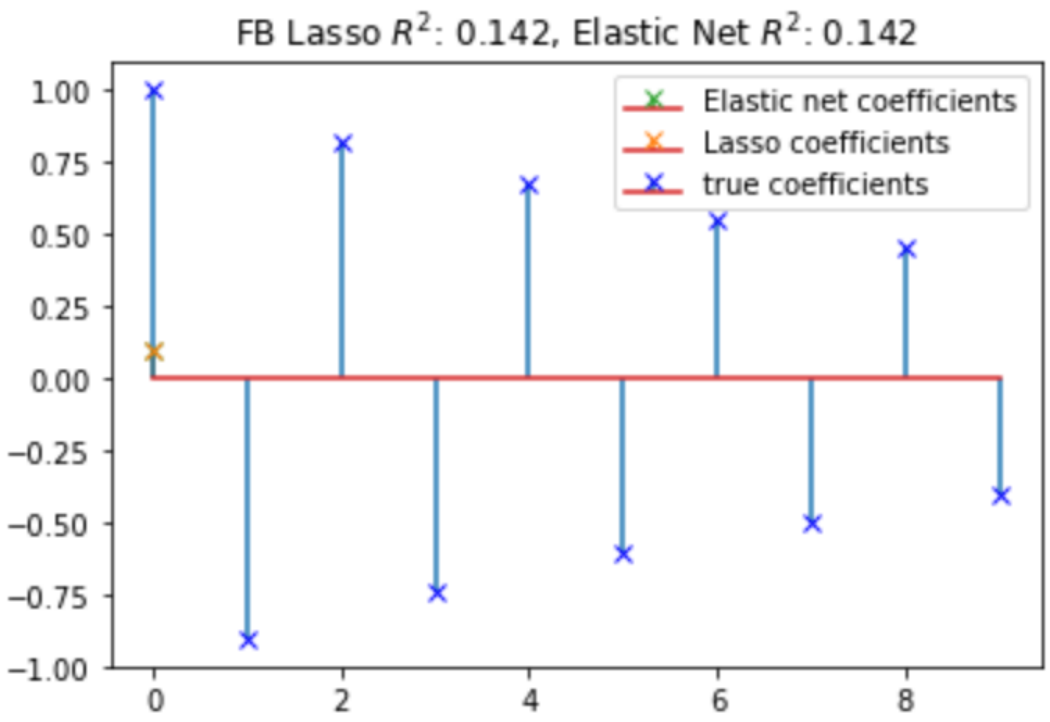


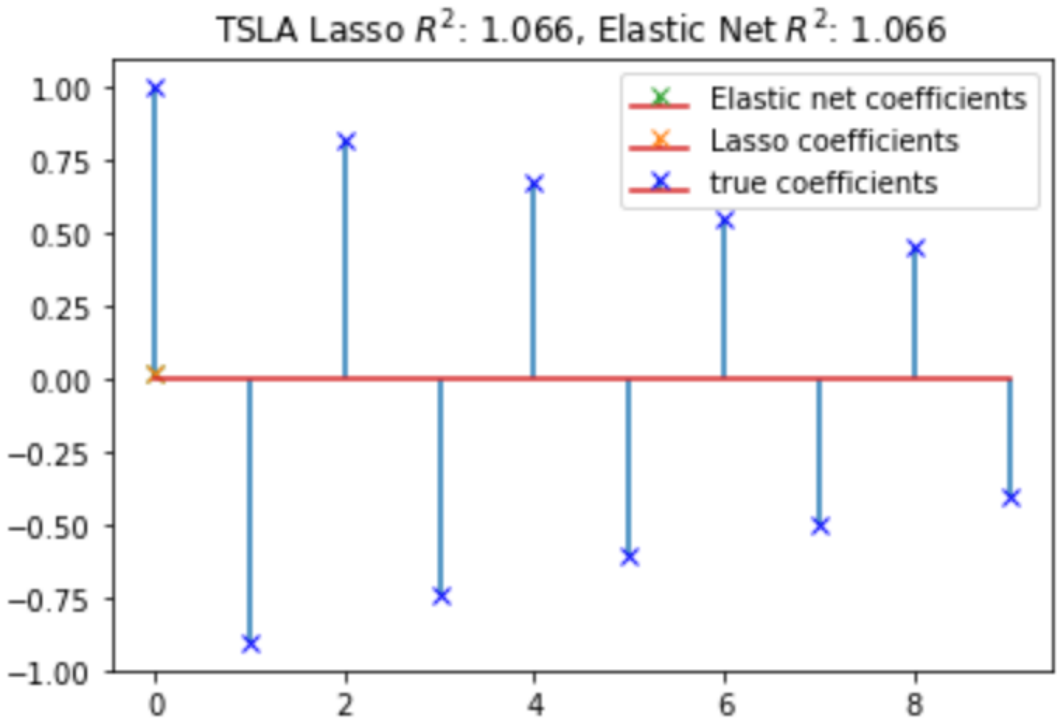



# Introduction to Lasso Regression & Elastic Net

- In statistics and machine learning, **lasso** (least absolute shrinkage and selection operator; also Lasso or LASSO) is a regression analysis method that performs both variable selection and regularization in order to enhance the prediction accuracy and interpretability of the resulting statistical model. It was originally introduced in geophysics, and later by Robert Tibshirani, who coined the term. Lasso was originally formulated for linear regression models. This simple case reveals a substantial amount about the estimator. These include its relationship to ridge regression and best subset selection and the connections between lasso coefficient estimates and so-called soft thresholding. It also reveals that (like standard linear regression) the coefficient estimates do not need to be unique if covariates are collinear.
- Reference: https://en.wikipedia.org/wiki/Lasso_(statistics)
- In statistics and, in particular, in the fitting of linear or logistic regression models, the **elastic net** is a regularized regression method that linearly combines the L1 and L2 penalties of the lasso and ridge methods.
- Reference: https://en.wikipedia.org/wiki/Elastic_net_regularization

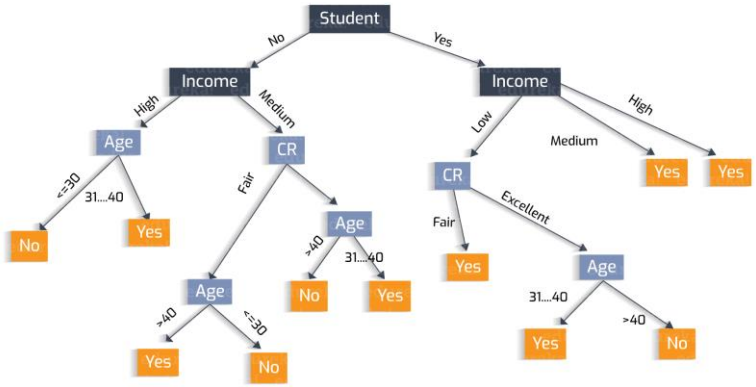
# Lasso Regression & Elastic Net on Trading Days and Stocks Price



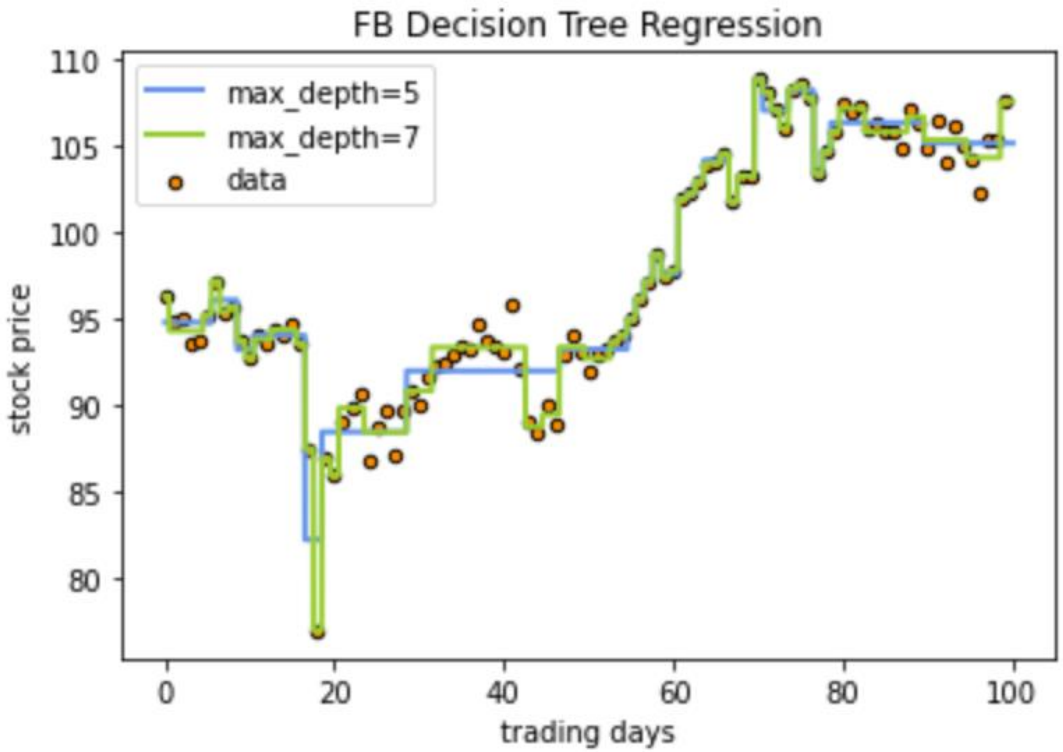
# Introduction to Decision Tree

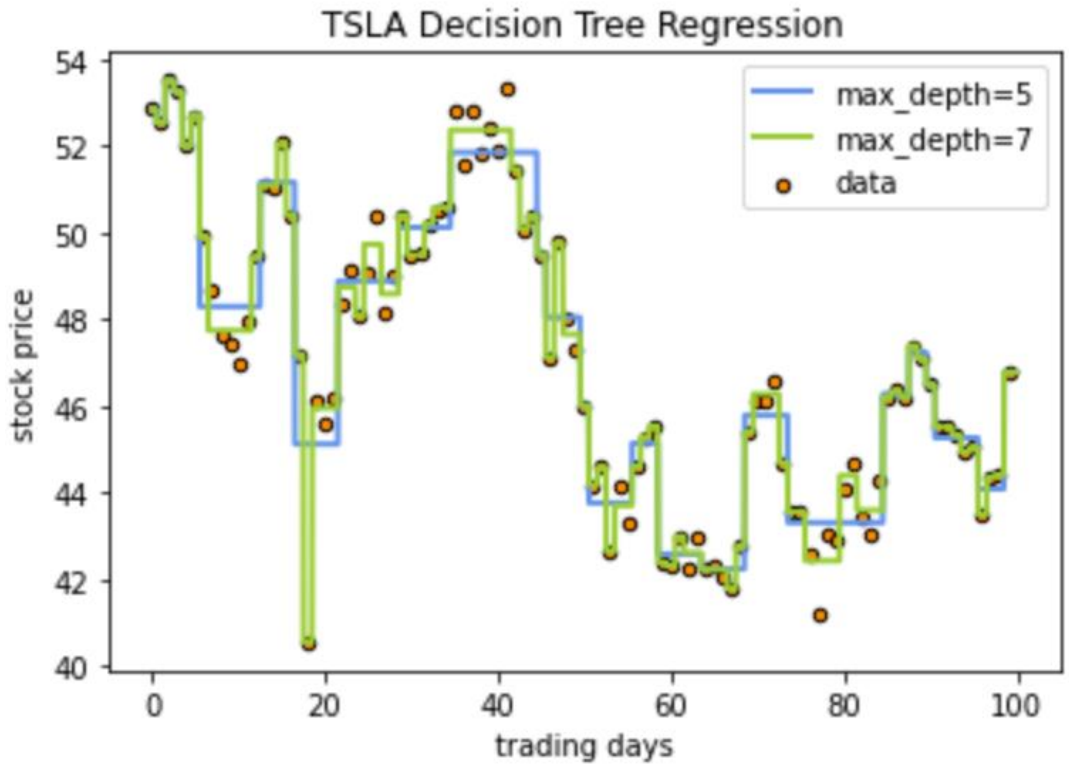
- A decision tree is a decision support tool that uses a tree-like model of decisions and their possible consequences, including chance event outcomes, resource costs, and utility. It is one way to display an algorithm that only contains conditional control statements.
- Decision trees are commonly used in operations research, specifically in decision analysis, to help identify a strategy most likely to reach a goal, but are also a popular tol in machine learning.
- Reference: https://en.wikipedia.org/wiki/Decision_tree




Reference: https://heartbeat.fritz.ai/understanding-the-mathematics-behind-decision-trees-22d86d55906

# Decision Trees Regression on Trading Days and Stocks Price

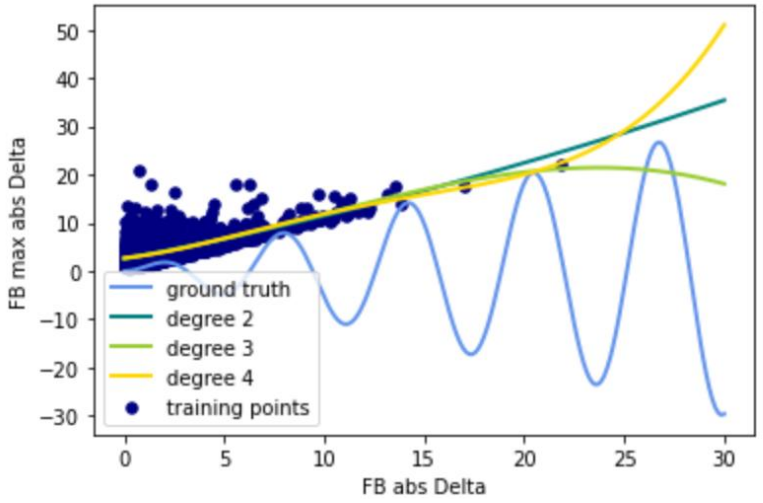


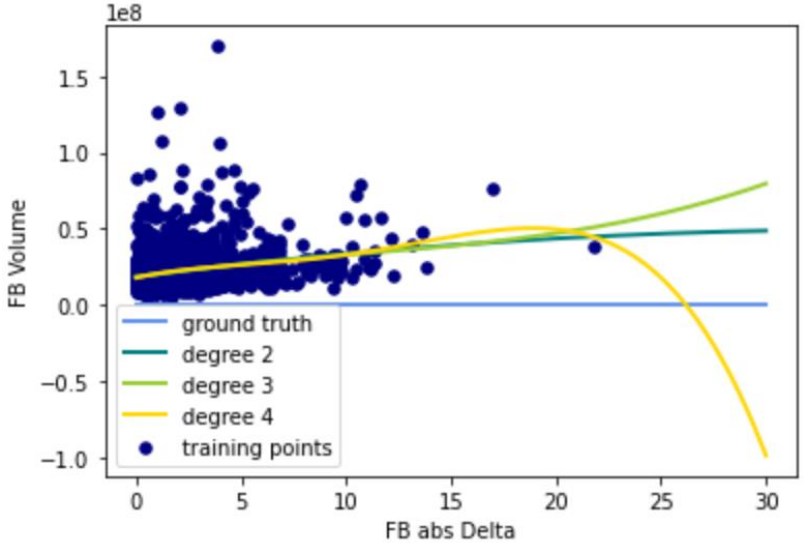
# Introduction to Polynomial Regression

- In statistics, **polynomial regression** is a form of regression analysis in which the relationship between the independent variable x and the dependent variable y is modelled as an nth degree polynomial in x. Polynomial regression fits a nonlinear relationship between the value of x and the corresponding conditional mean of y, denoted E(y |x). Although polynomial regression fits a nonlinear model to the data, as a statistical estimation problem it is linear, in the sense that the regression function E(y | x) is linear in the unknown parameters that are estimated from the data. For this reason, polynomial regression is considered to be a special case of multiple linear regression.
- Reference: https://en.wikipedia.org/wiki/Polynomial_regression

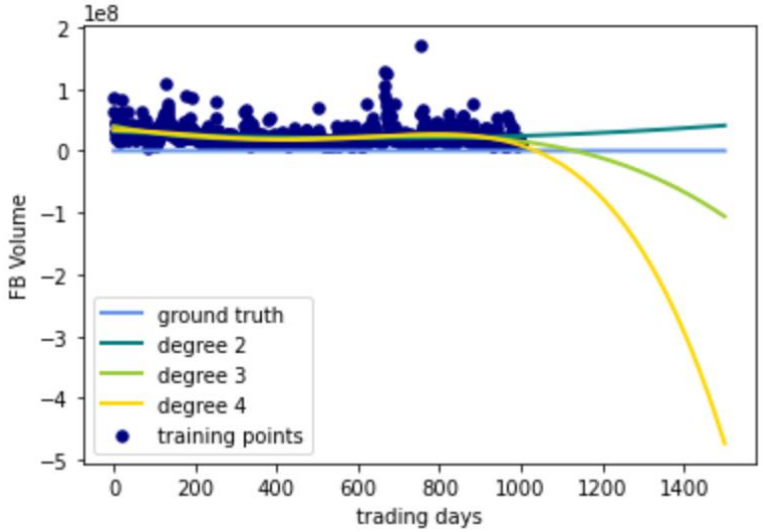
# Polynomial Regression




Polynomial Regression on abs Delta Price and Max Delta Price of Facebook




Polynomial Regression on abs Delta Price and Volume of Facebook

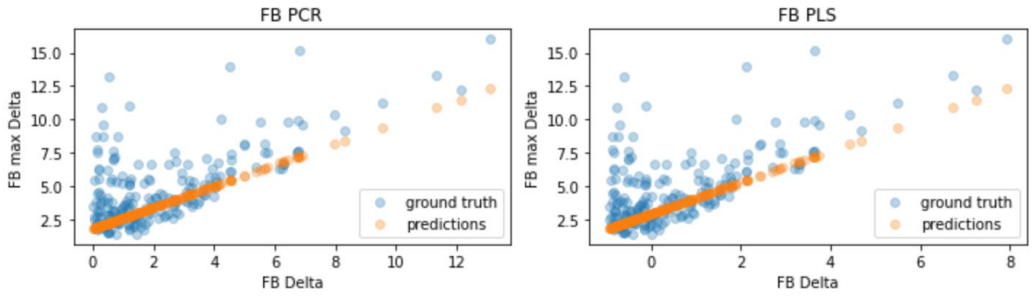



Polynomial Regression on Trading Days and Stocks Price

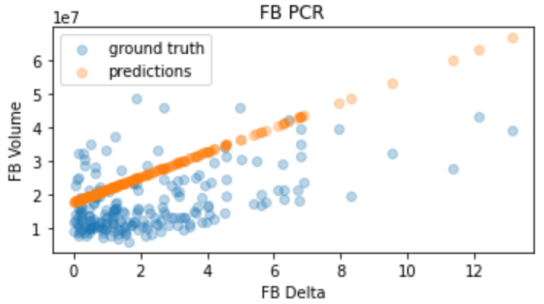
# Introduction to PCR &PLS

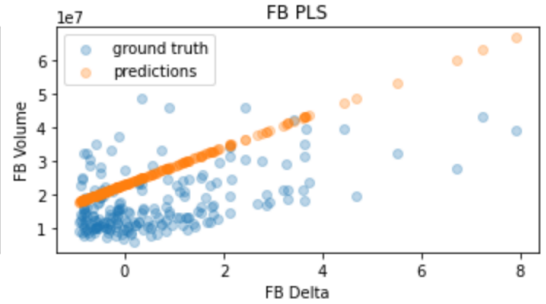
- In statistics, **principal component regression** (**PCR**) is a regression analysis technique that is based on principal component analysis (PCA). More specifically, PCR is used for estimating the unknown regression coefficients in a standard linear regression model. In PCR, instead of regressing the dependent variable on the explanatory variables directly, the principal components of the explanatory variables are used as regressors. One typically uses only a subset of all the principal components for regression, making PCR a kind of regularized procedure and also a type of shrinkage estimator.
- Reference: https://en.wikipedia.org/wiki/Principal_component_regression
- **Partial least squares regression** (**PLS regression**) is a statistical method that bears some relation to principal components regression; instead of finding hyperplanes of maximum variance between the response and independent variables, it finds a linear regression model by projecting the predicted variables and the observable variables to a new space. Because both the X and Y data are projected to new spaces, the PLS family of methods are known as bilinear factor models. Partial least squares discriminant analysis (PLS-DA) is a variant used when the Y is categorical.
- Reference: https://en.wikipedia.org/wiki/Partial_least_squares_regression

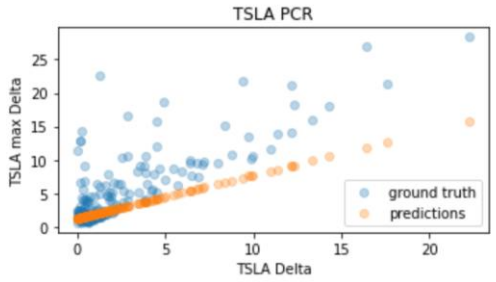
# PCR & PLS

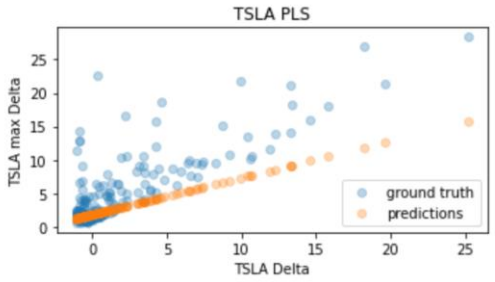

FB PCR
FB max Delta
FB Delta
ground truth
predictions

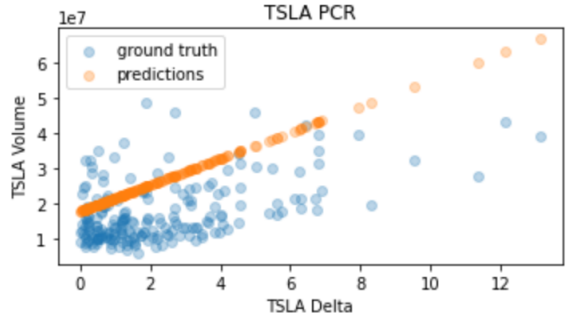

FB PLS
FB max Delta
FB Delta
ground truth
predictions

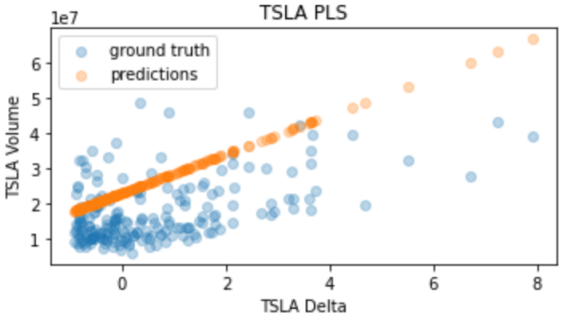

1e7
FB PCR
FB Volume
FB Delta
ground truth
predictions


1e7
FB PLS
FB Volume
FB Delta
ground truth
predictions

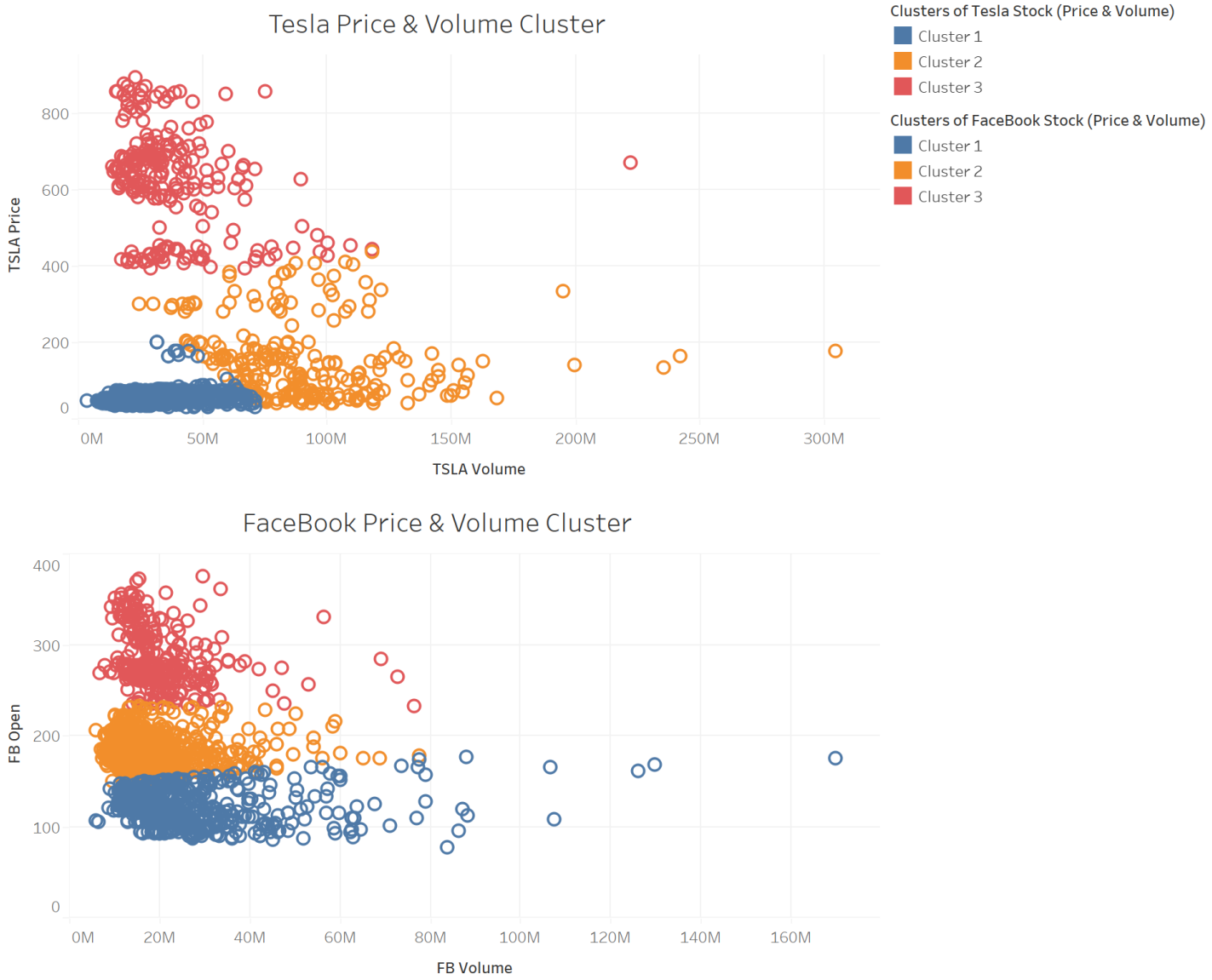

TSLA PCR
TSLA max Delta
TSLA Delta
ground truth
predictions


TSLA PLS
TSLA max Delta
TSLA Delta
ground truth
predictions

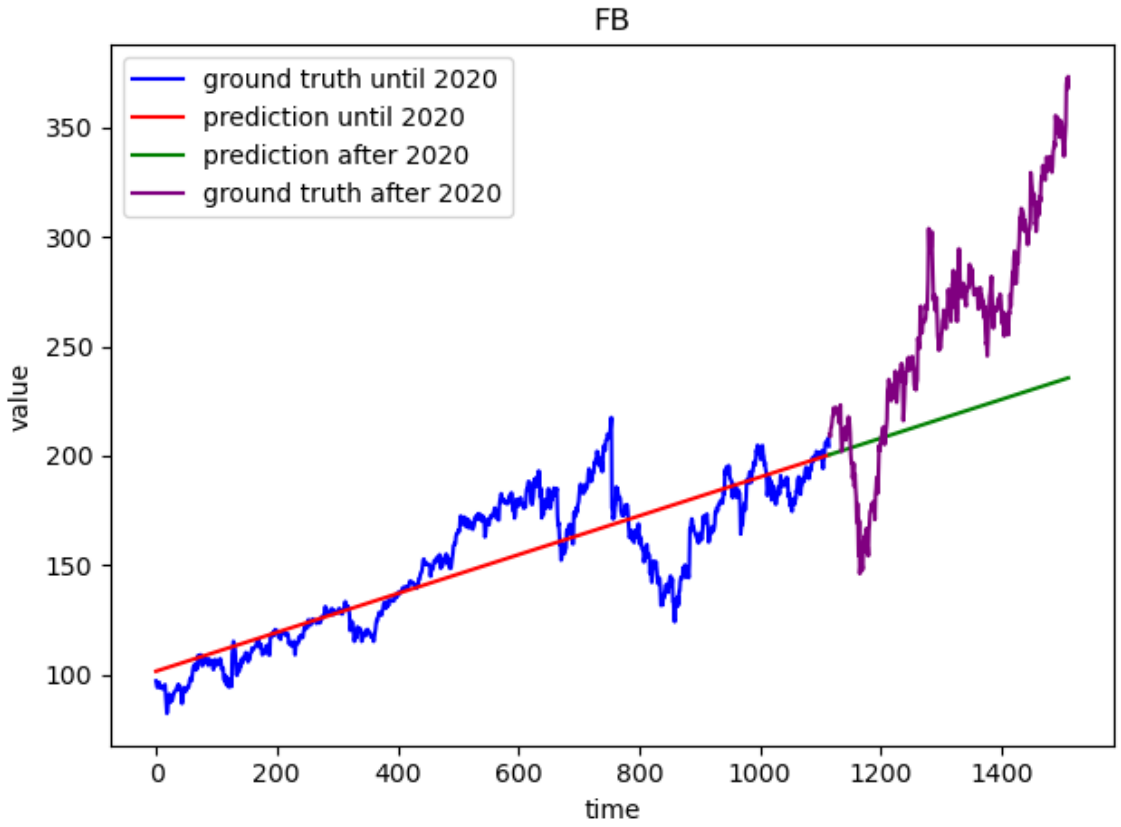

1e7
TSLA PCR
TSLA Volume
TSLA Delta
ground truth
predictions

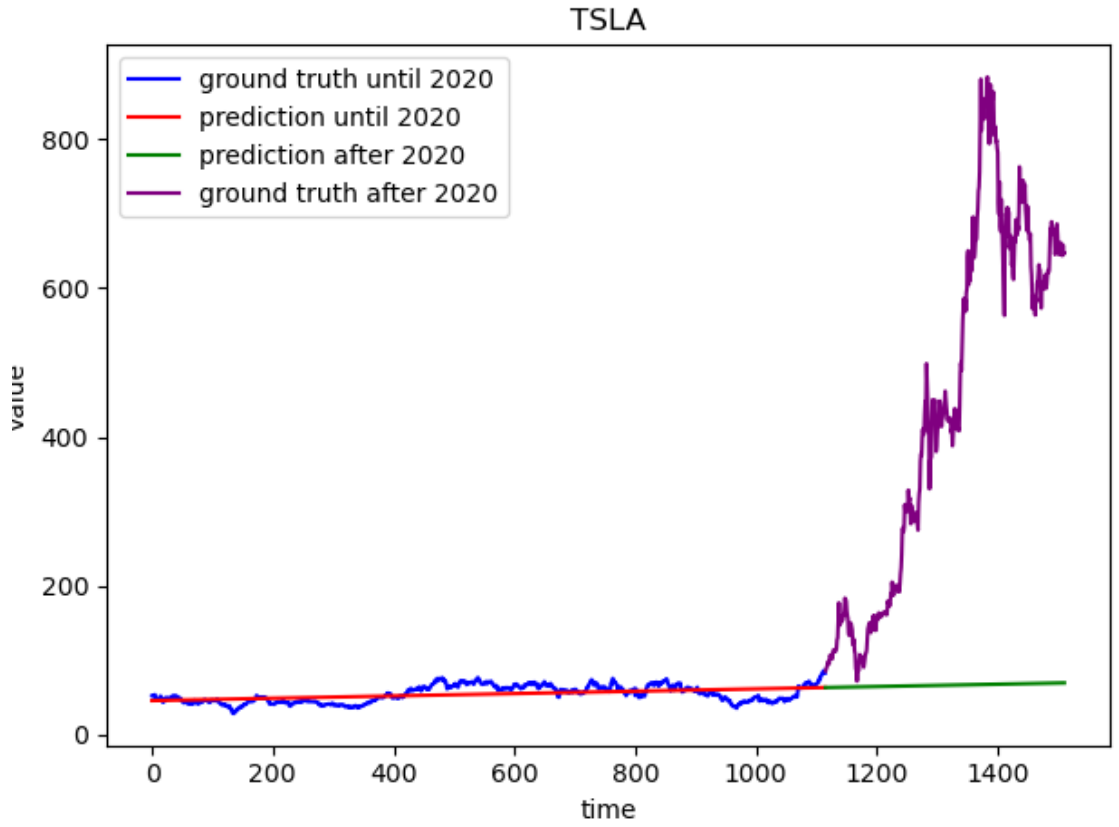

1e7
TSLA PLS
TSLA Volume
TSLA Delta
ground truth
predictions

# Clusters of Price and Volume of Two Stocks



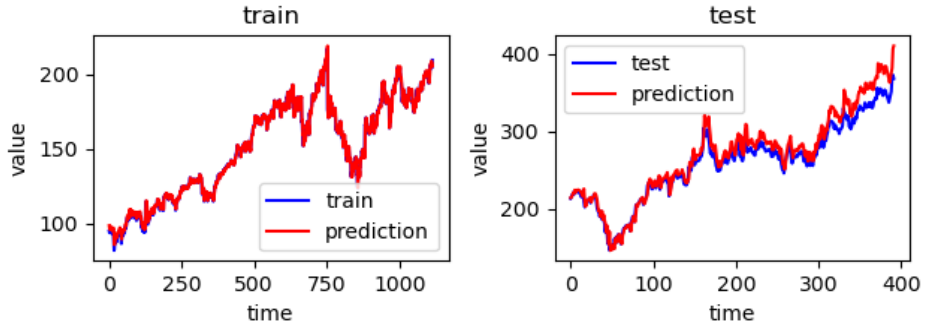
# Linear Regression on Trading Days and Stocks Price



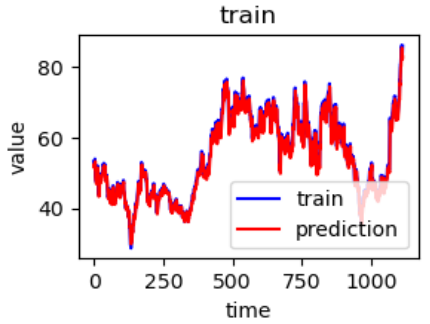
# LSTM & Fully Connected Network

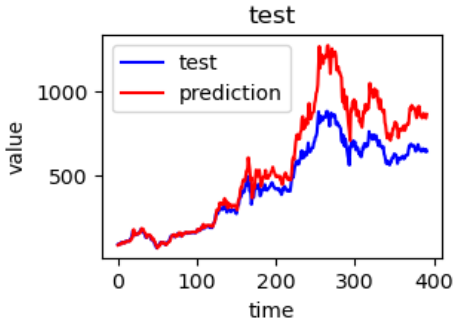
Long Short-Term Memory is an advanced version of recurrent neural network (RNN) architecture that was designed to model chronological sequences and their long-range dependencies more precisely than conventional RNNs.
Fully connected layer is the most common and widely used layer. It has a simple structure that the output stems from a singular formula *output = activation (dot (input, kernel) + bias*).

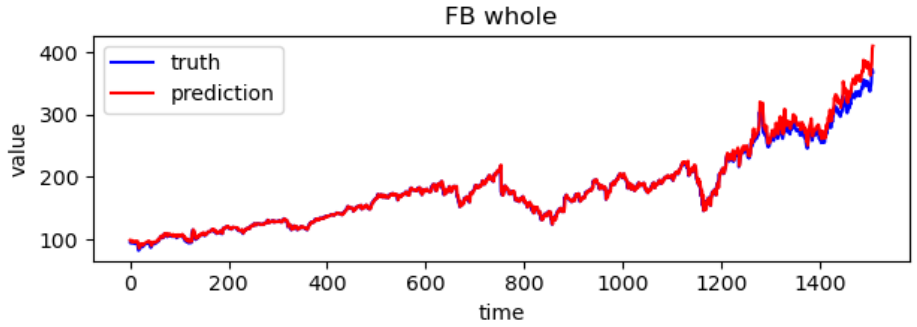


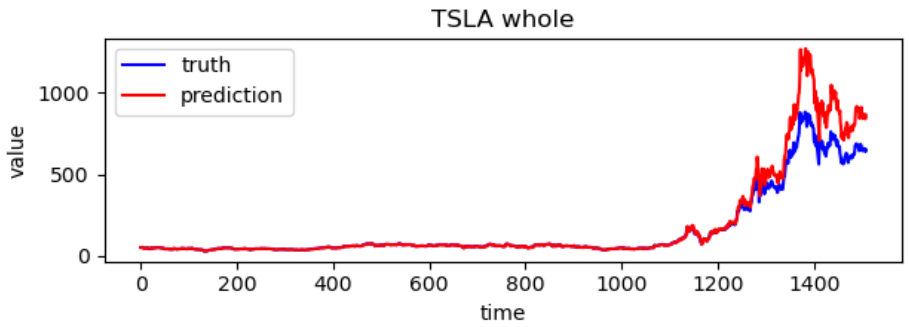






# Thank you