

# HPQ Stock Analysis

By Richard Twitty

# HP Inc

- A leading global provider of personal computing and printing solutions.
- Founded in 1939 by Bill Hewlett and Dave Packard.
- Known for pioneering products like desktop calculators and the LaserJet printer.
- Innovates in areas such as 3D printing and premium laptops.

# The Problem

Stock market in general is difficult to anticipate. The price flow, whether it buy or sell, or a dead market.

The average stock market return is 10%. We want greater returns

What are ways to increase profit returns?

Can we predict price flow of the market?

# Who might be interested?

- Companies who invested in HPQ stocks
- Day trader
- Positional traders
- The company itself

# What condition might affect the market price?

- Employee layoff
- Major events
- News trend
- Volume

# Data Information

- Data acquired for the period: 1/2/1970 to 7/2/2020
- Data: trading data for HPQ
- Number of records: 12741
- Number of fields:
  - 7(initially)
  - 9(final analysis)

# Steps for HPQ Stocks

Data Analysis of HPQ stocks

- Sorting data
- Replacing or dropping null data
- Checking for duplicates
- Remove outliers

Install TA-Lib for candle stick analysis

Generate a signal features for buying, selling, and holding

Generate a difference, Log, square root, and boxcox column for open, close, high, low features

Generate a algorithm for technical indicators such as RSI, MFI, etc.

Run adfuller check on each features

Generate Time series for each features

Generate PACF and ACF for close, open, high, low and corresponding differences

Generate ARIMA model with/out cross validation

# Feature Stationary Check (p value test)

Volume: 4.96e-9

Close\_diff: 0

Open\_diff: 0

High\_diff: 0

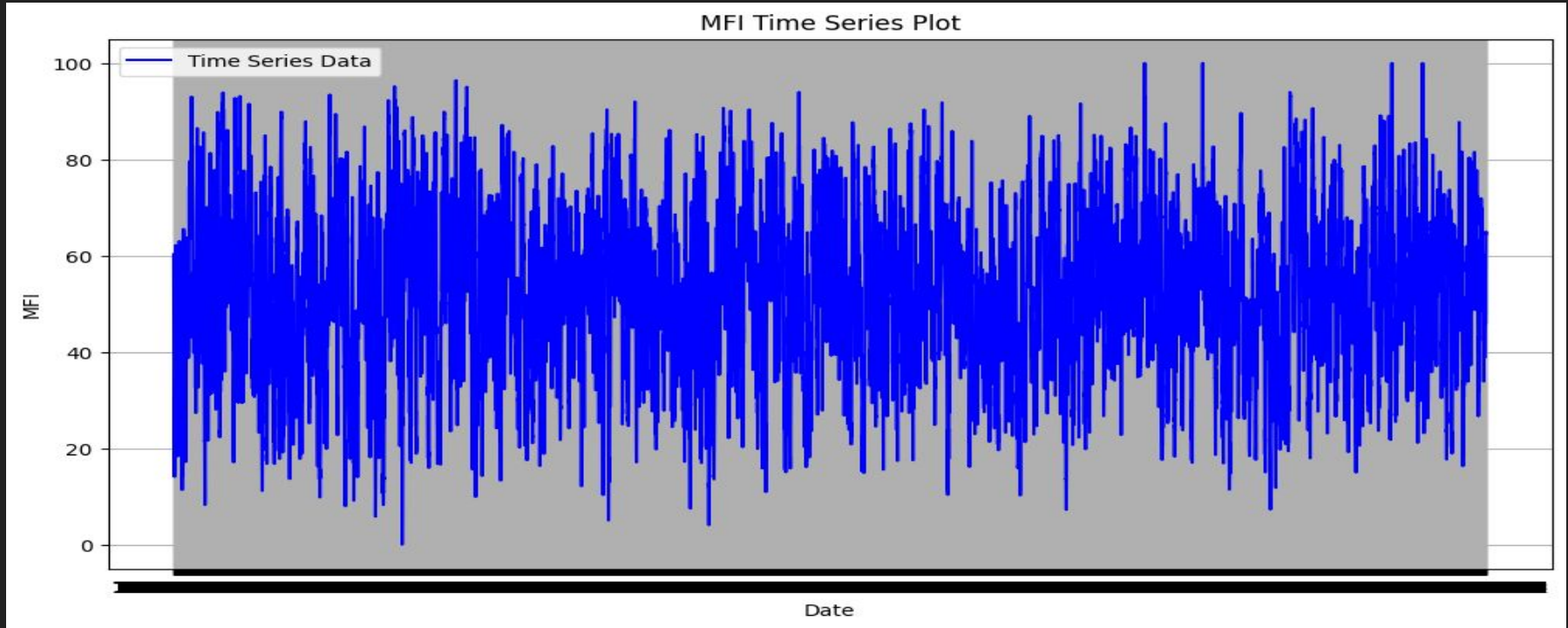
Low\_diff: 0

MFI: 0

RSI\_14: 0



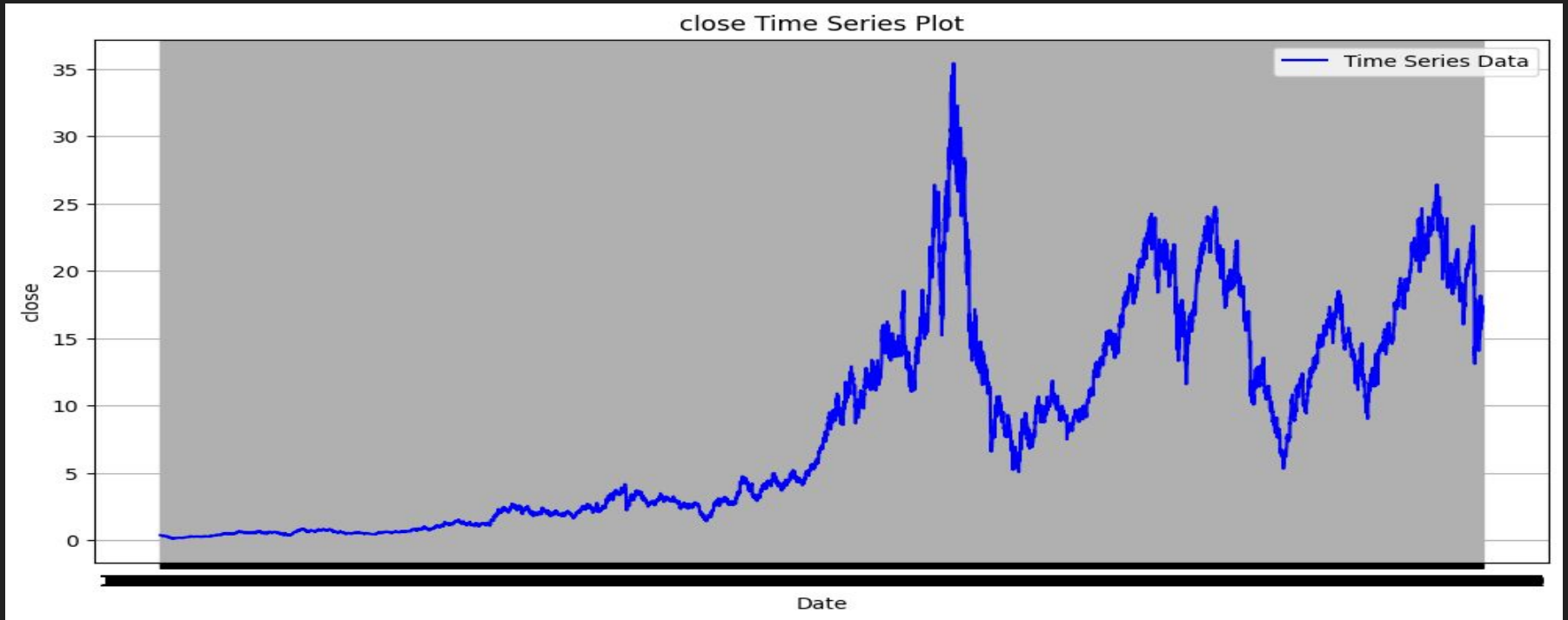
# Data Exploration



# Data Exploration

The MFI time is stationary because it consolidating between 0 and 100 through the time series. It good for ARIMA model.

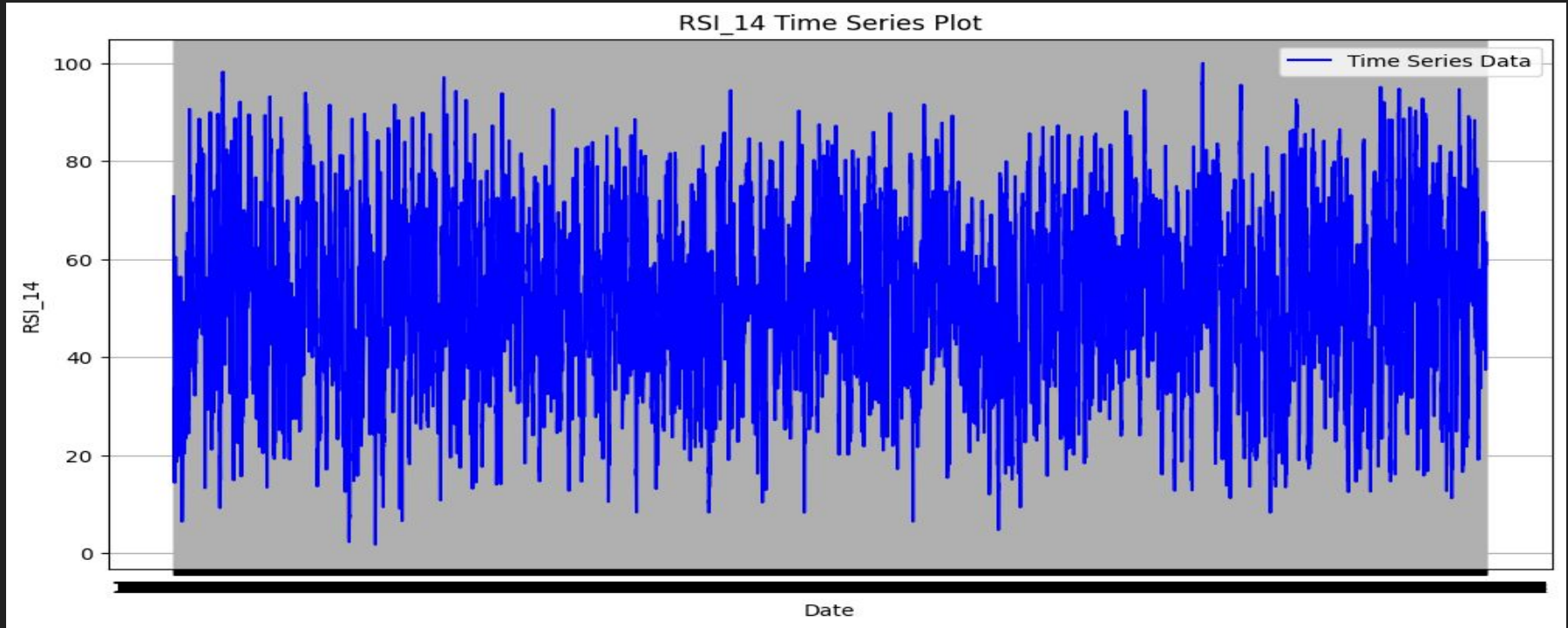
# Data Exploration



# Data Exploration

We have a time series that not stationary, which is the “close” feature. This means that the variance and means are not constant through out the time sere. You can see different part of the charts that have more variance than the other.

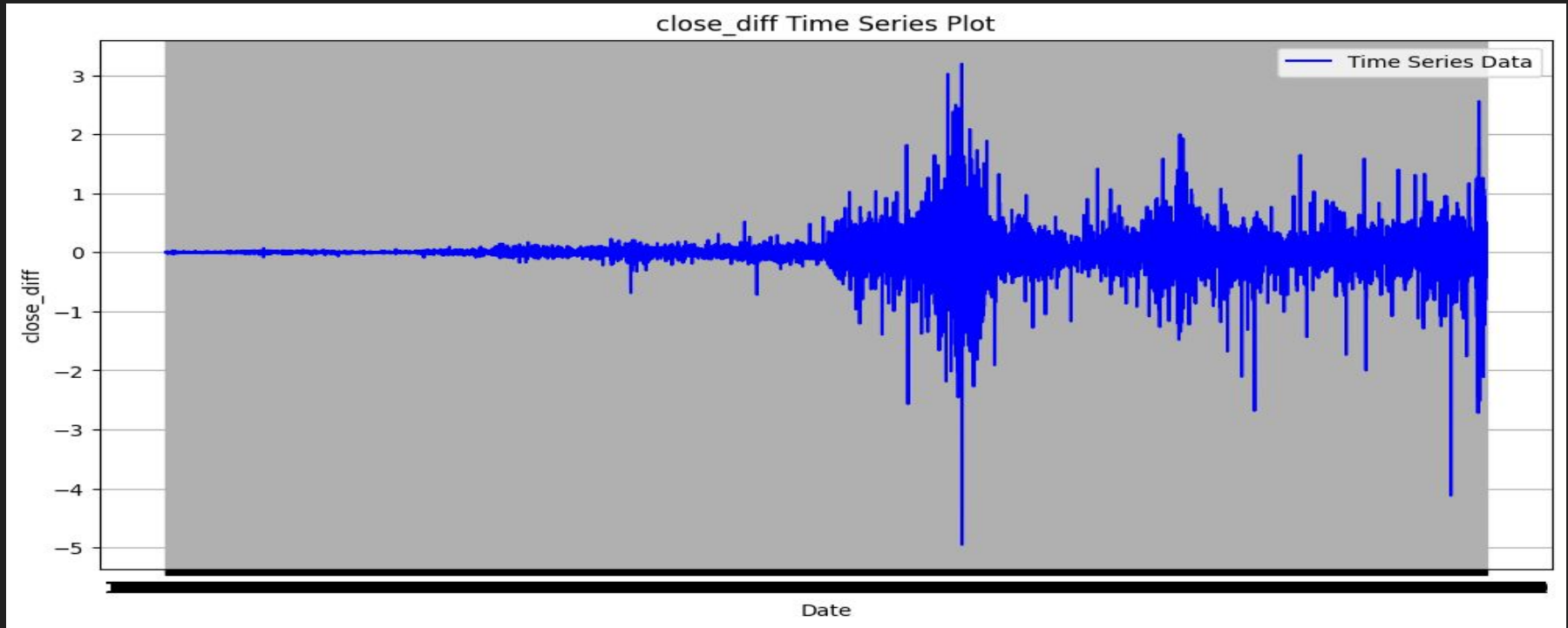
# Data Exploration



# Data Exploration

The RSI\_14 time is stationary because it consolidating between 0 and 100 through the time series. It good for ARIMA model.

# Data Exploration

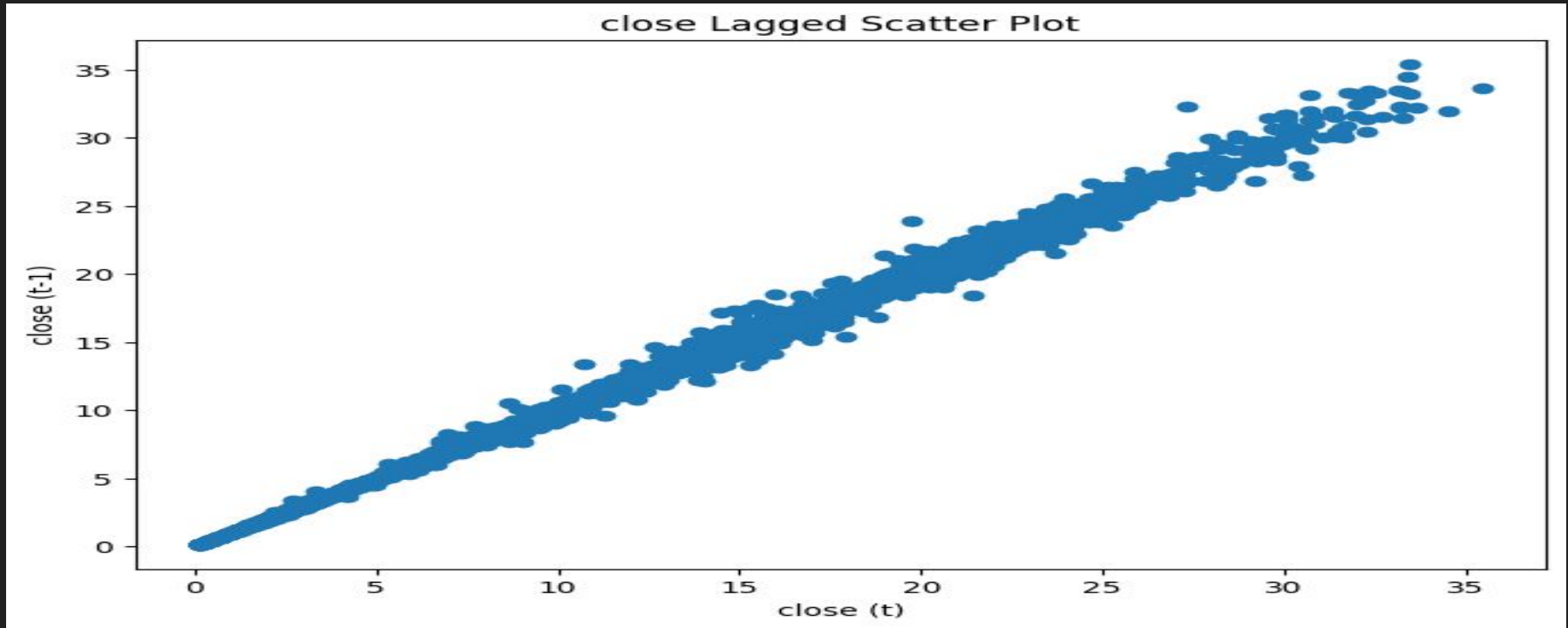


# Data Exploration

This is an interesting case because it show change in varaince, but it still stationary. It have no define trend. It consolidating with several different ranges. This is the close\_diff feature. This feature also pass the adfuller stationary test.



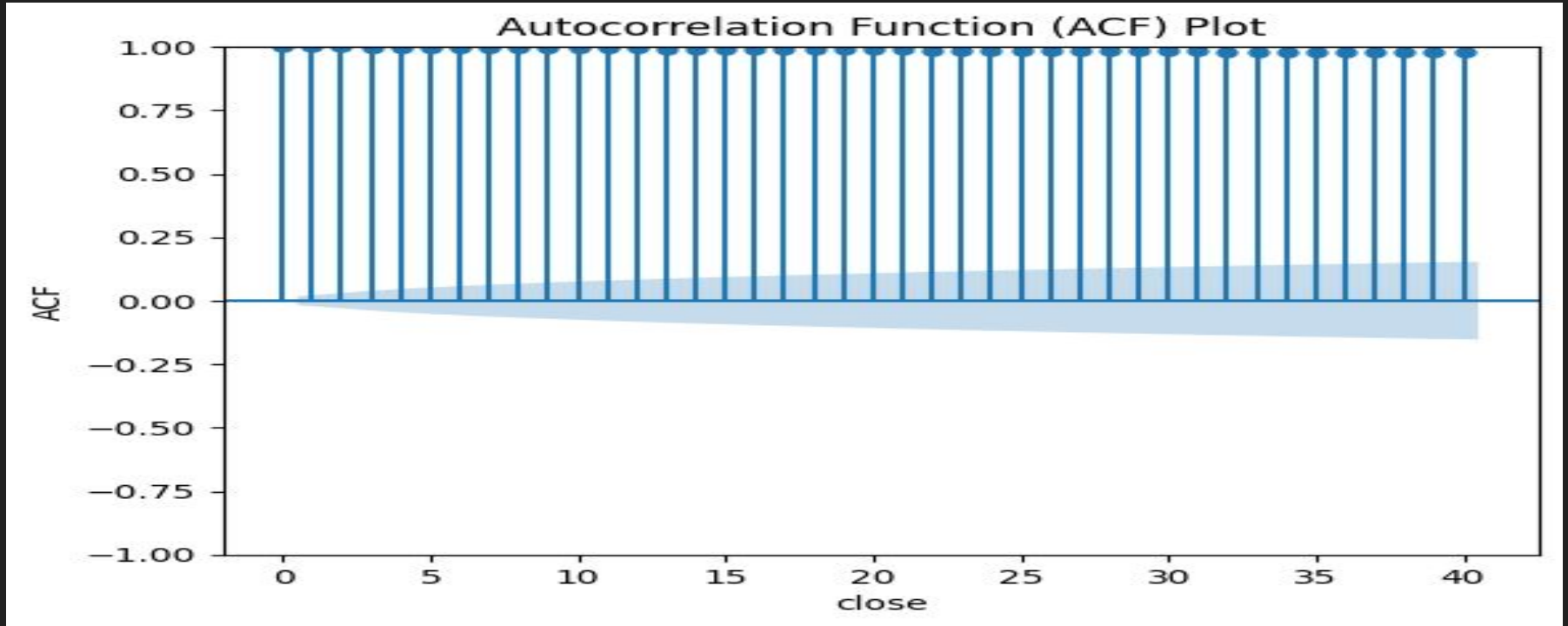
# Data Exploration



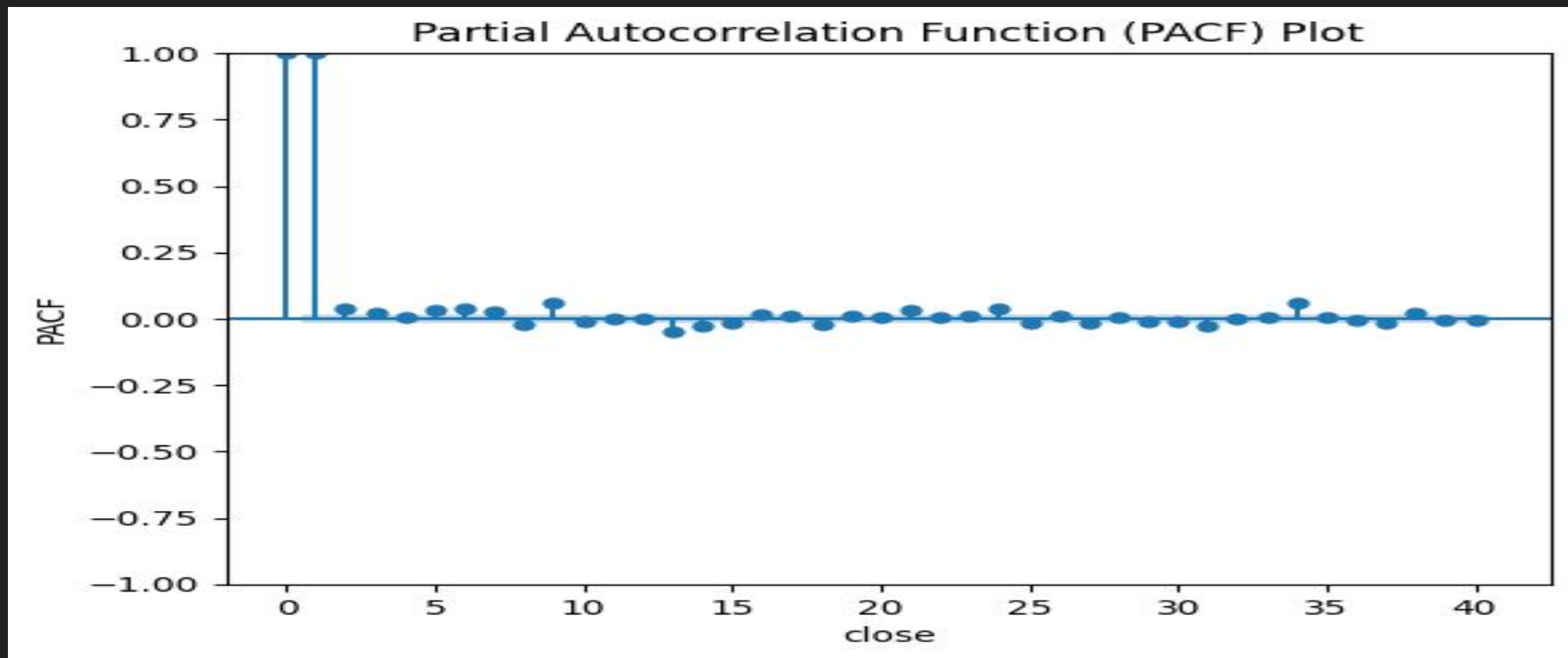
# Data Exploration

This is a sign of Autocorrelation because the current “close” value has positive correlation with it lagged 1 value. This is a sign of a feature variable being non-stationary.

# Data Exploration



# Data Exploration

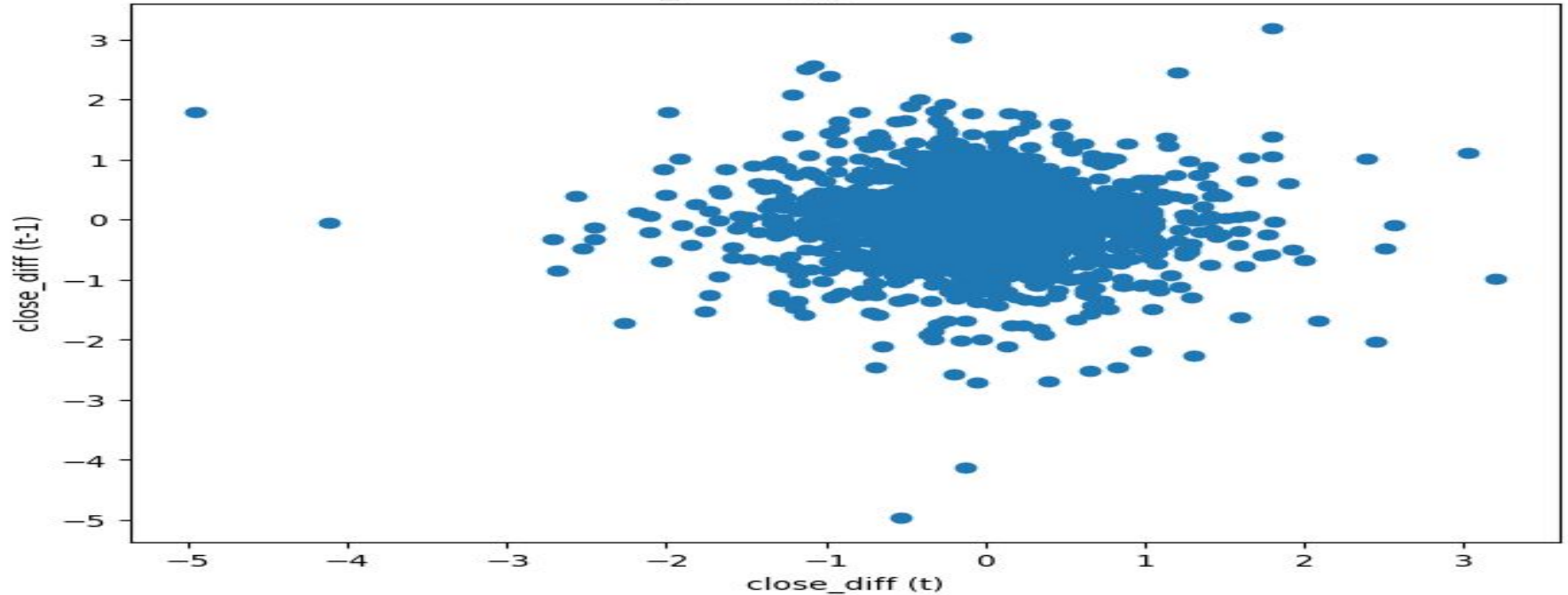


# Data Exploration

Both ACF and PACF show that the close variables is non-stationary. The ACF shows a series of significant spike without decay. This indicates the series has some trend going on. The PACF has a significant spike at lagged 1 indicating that this lagged has potential to be AR(1)

# Data Exploration

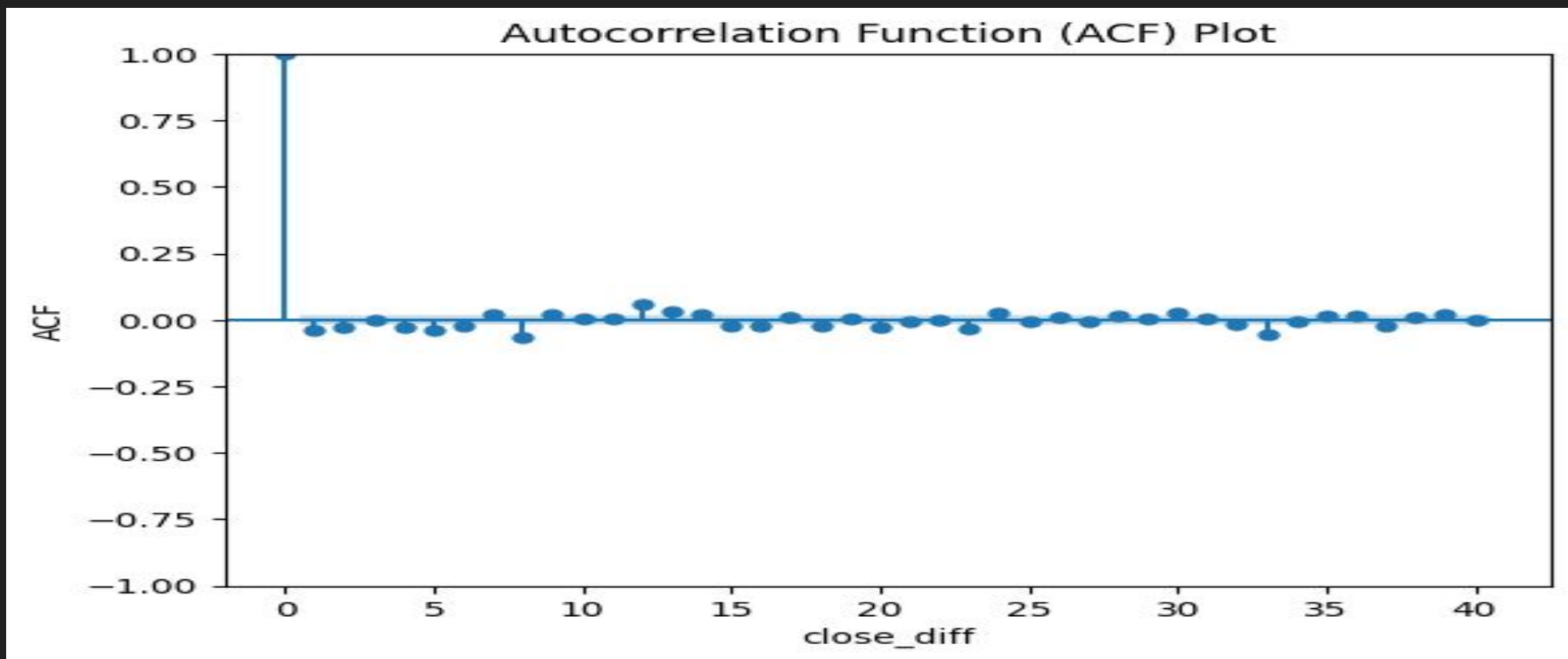
close\_diff Lagged Scatter Plot



# Data Exploration

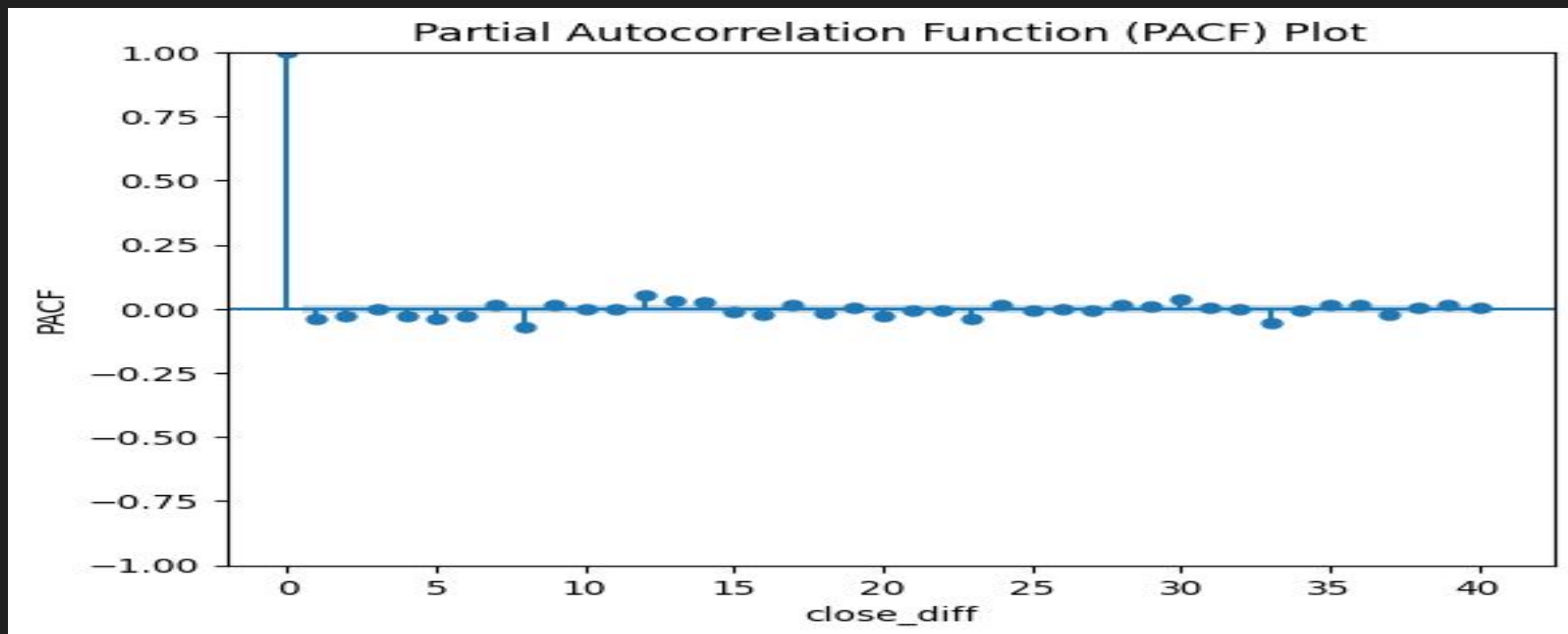
We have a lagged scatter chart for `close_diff` where most of its points are at the center and it forms an oval. This is a sign of no trend and the variable is stationary. This is good for the ARIMA model. This also determines my parameters for `d` or order of difference, which is 1.

# Data Exploration





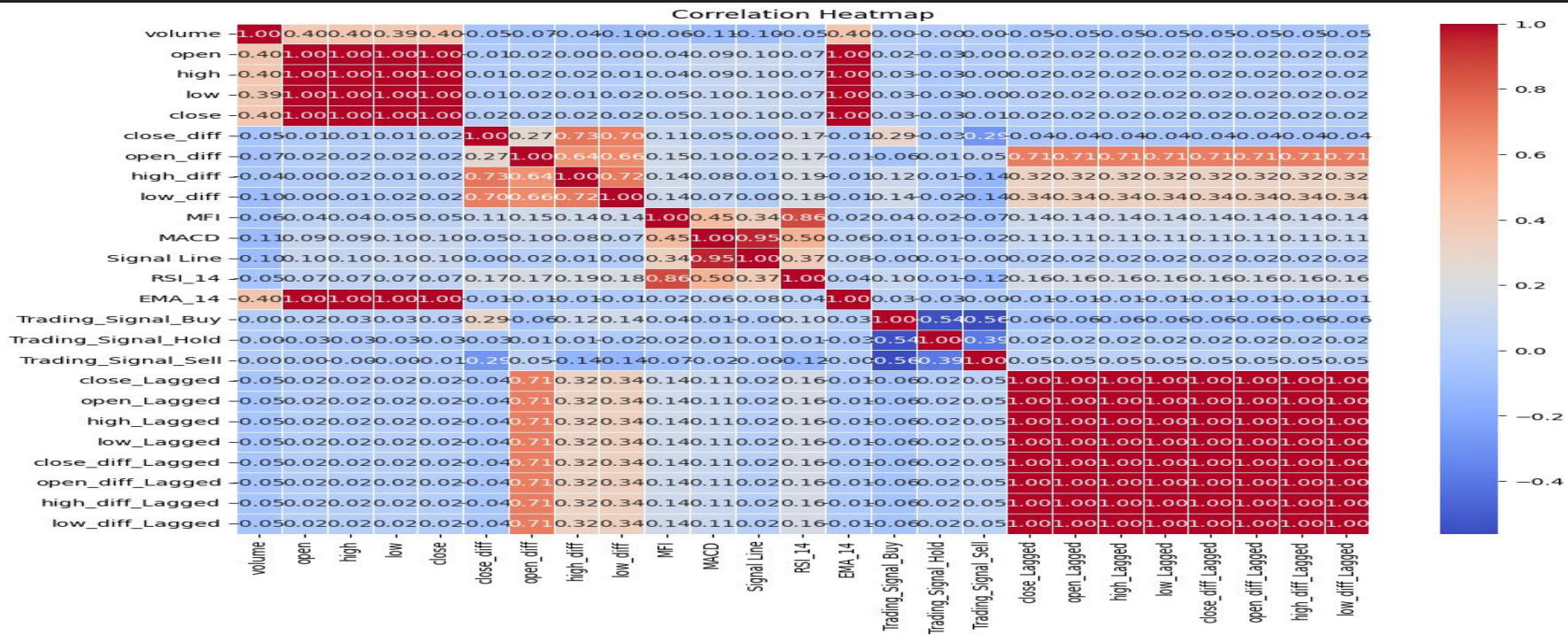
# Data Exploration



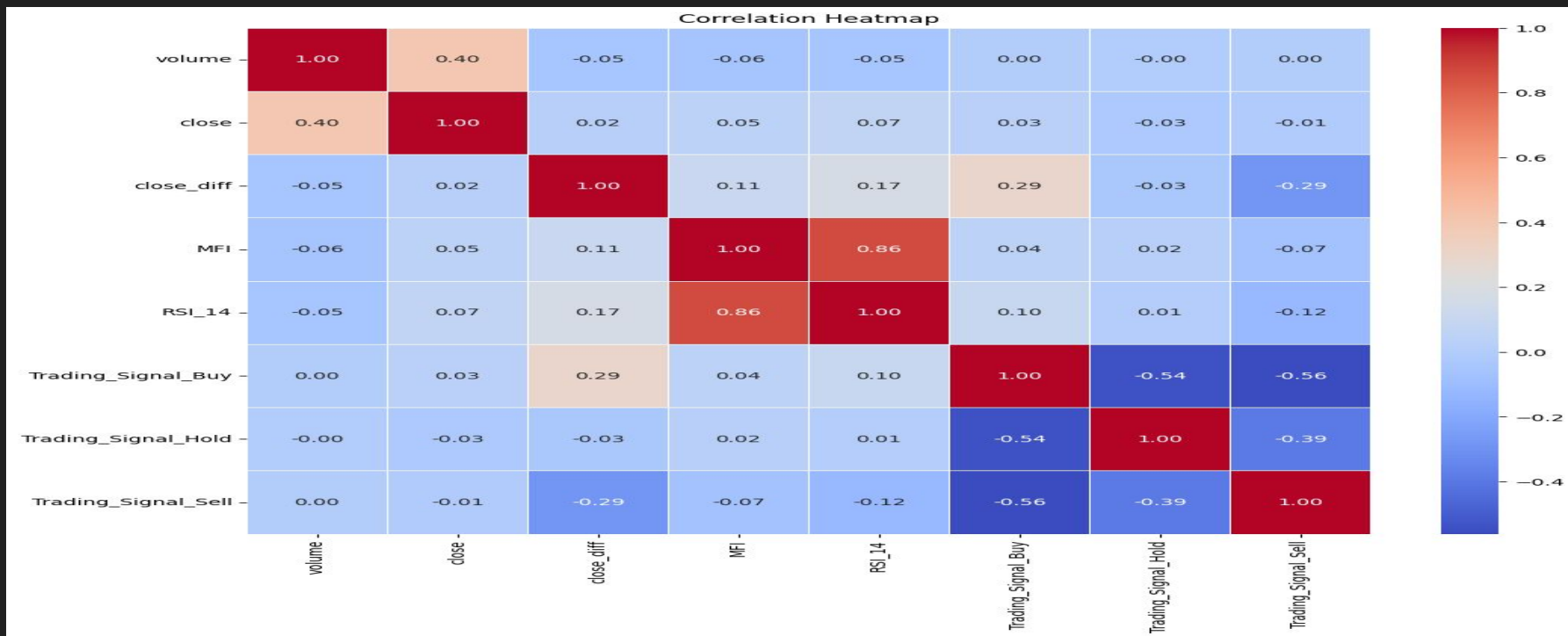
# Data Exploration

We have a small negative significant spike at lagged 1 in both ACF and PACF models. This is a sign of stationary feature. This also indicates we can utilize either AR(1) or MA(1) model.

# Data Exploration



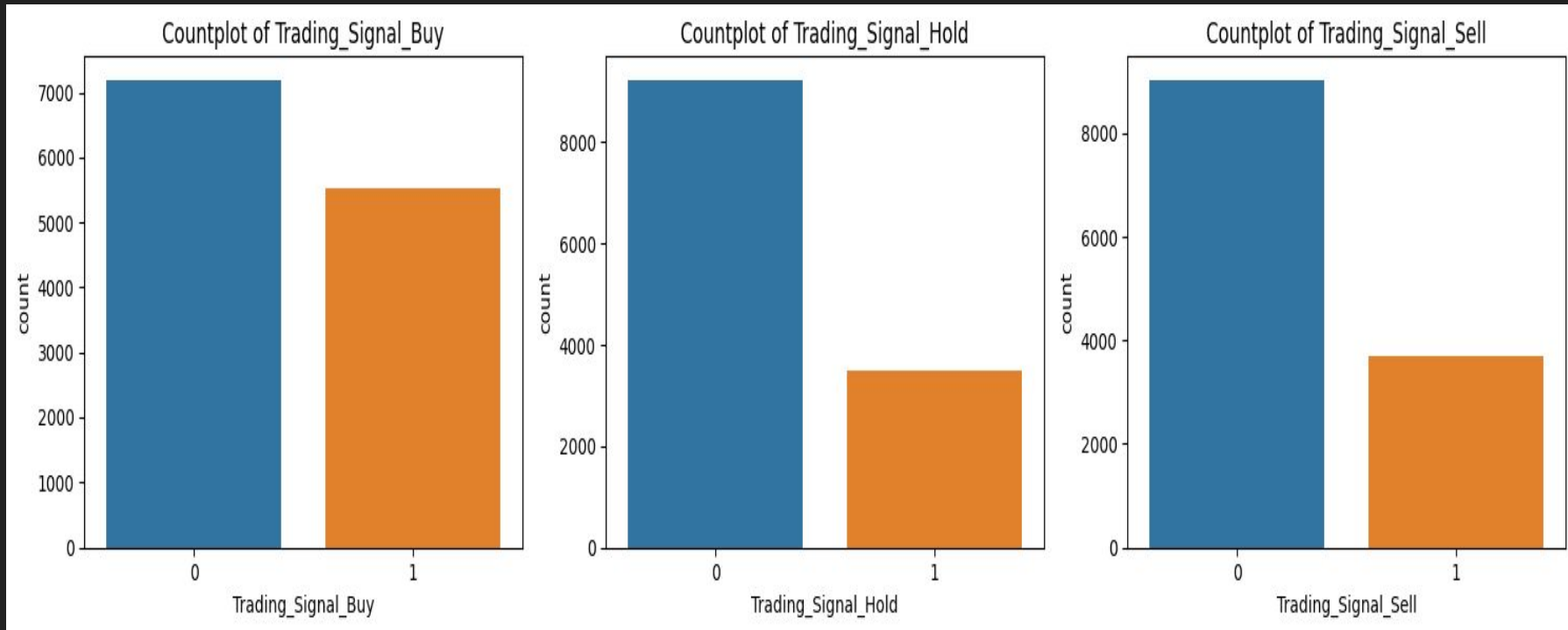
# Data Exploration



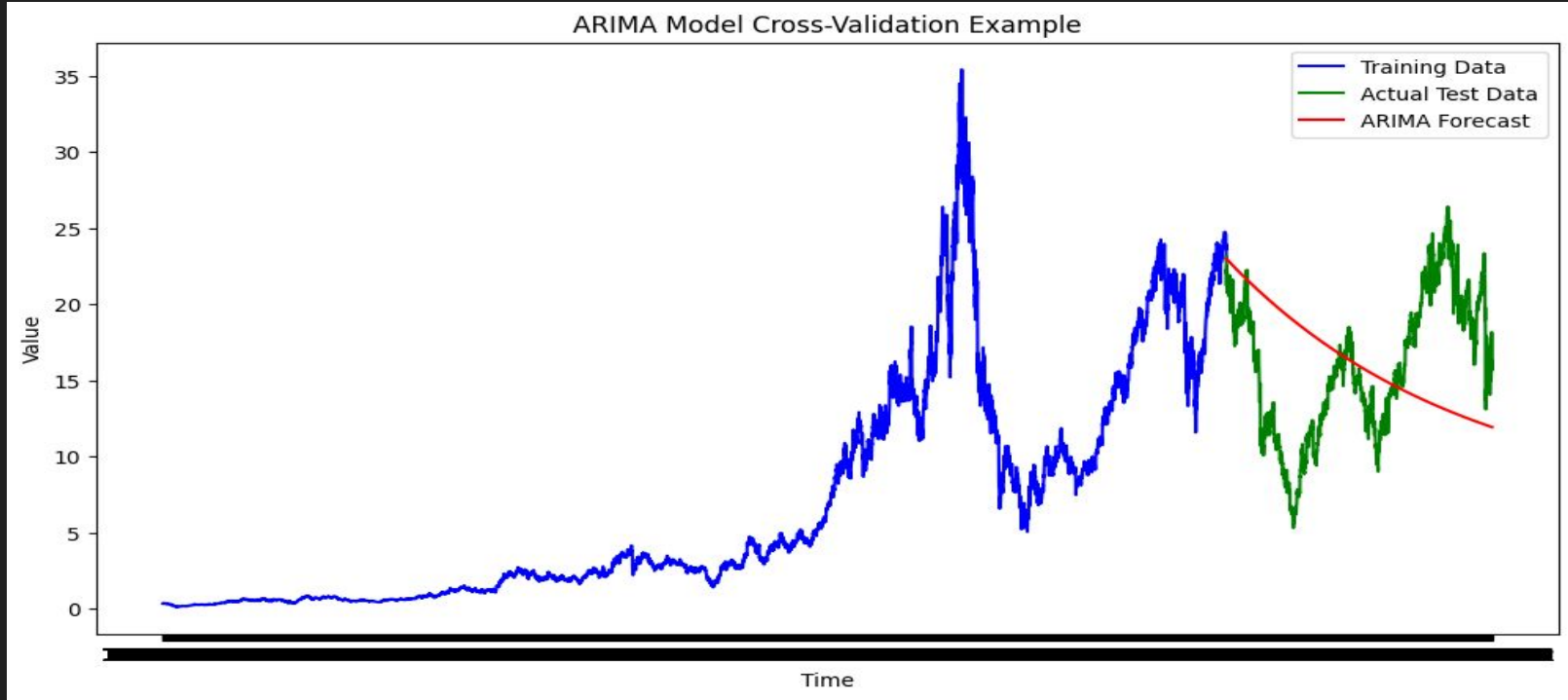
# Data Exploration

I planned on using VARMA model later on. I done correlation analysis and remove some correlated variable to avoid multicollinearity.

# Data Exploration



# Arima Model Analysis Without Cross Validation

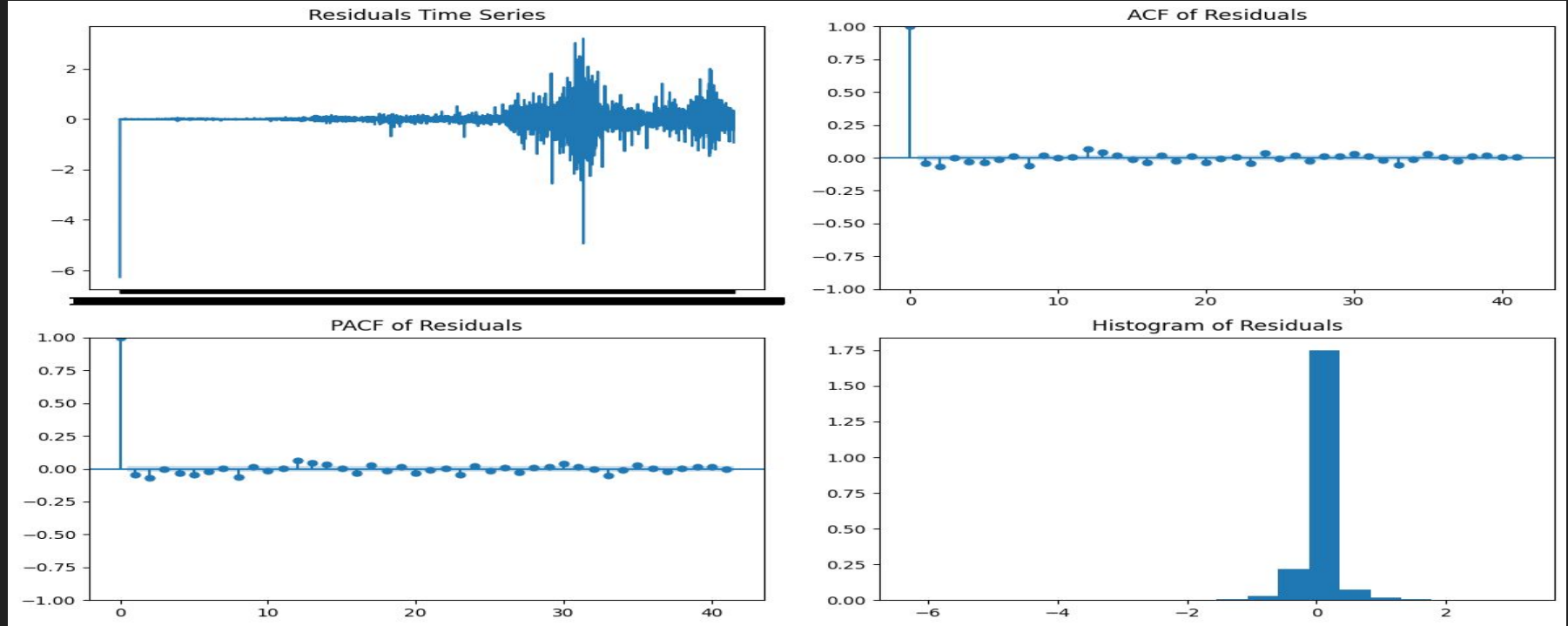


# Arima Model Analysis Without Cross Validation

We have an ARIMA(1,0 0) model. As you can it not doing too well in predicting.  
We lack cross validation.



# Arima Model Analysis Without Cross Validation



# Arima Model Analysis Without Cross Validation

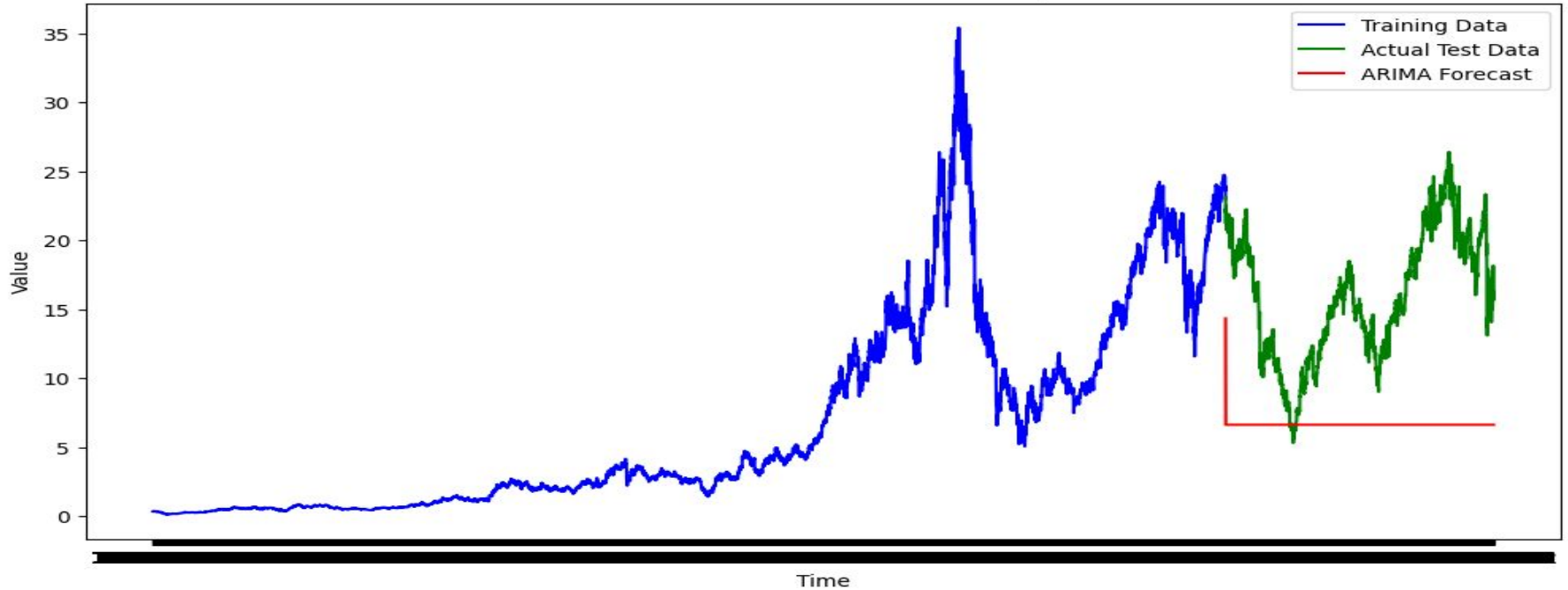
Time series chart has no pattern or trend. It consolidating. It look almost like white noise. This is sign that there maybe external factor or pattern that I missing in my model, but really minimum.

Both ACF and PACF shows a series of several significant spike, but really close to the significant threshold. So these are signs of weak autocorrelation among those lags, but some of these may improve the model.

The histogram shows a normal distribution, which is a good thing. My residual has an equally likely to be negative or positive.

# Arima Model Analysis Without Cross Validation

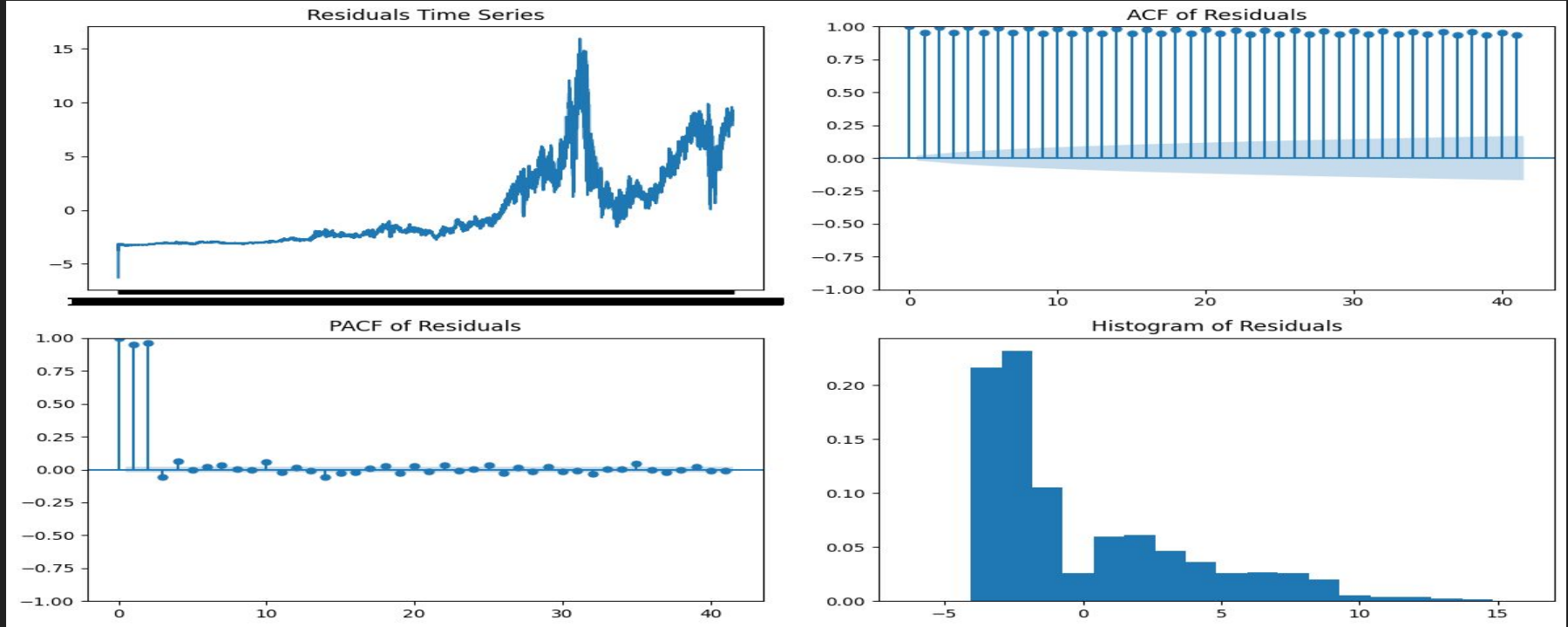
ARIMA Model Cross-Validation Example



# Arima Model Analysis Without Cross Validation

We have an ARIMA(0,0, 1) model. As you can it not doing too well in predicting.  
We lack cross validation.

# Arima Model Analysis Without Cross Validation



# Arima Model Analysis Without Cross Validation

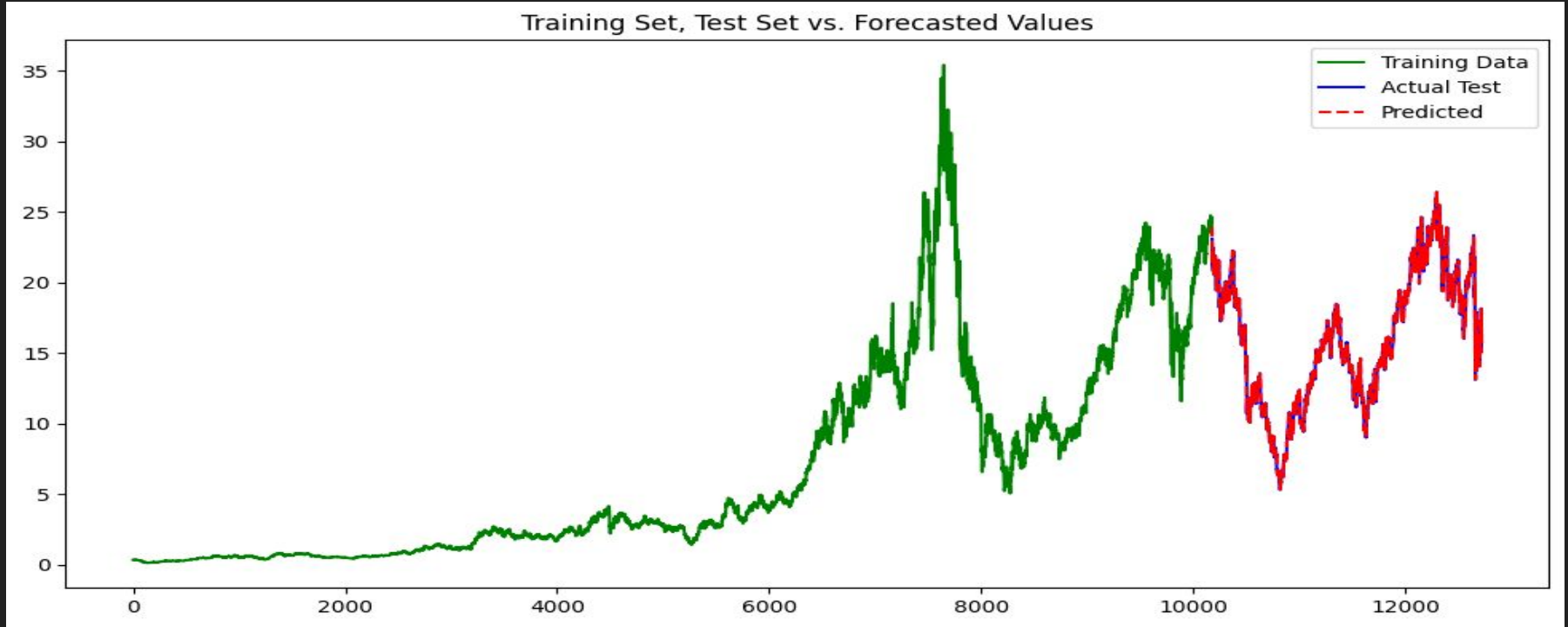
The residuals time series clearly shows a pattern, which is sign that our model is missing pattern.

Our ACF shows a series of significant spike with no decay, this shows that the model are missing information about the time series

Our PACF also have several significant spikes, implying we missing some information. There may be some potential AR terms

Our histogram is not normally distribute, so that means our residuals is not equally likely to be positive and negative.

# Arima Model Analysis With Cross Validation

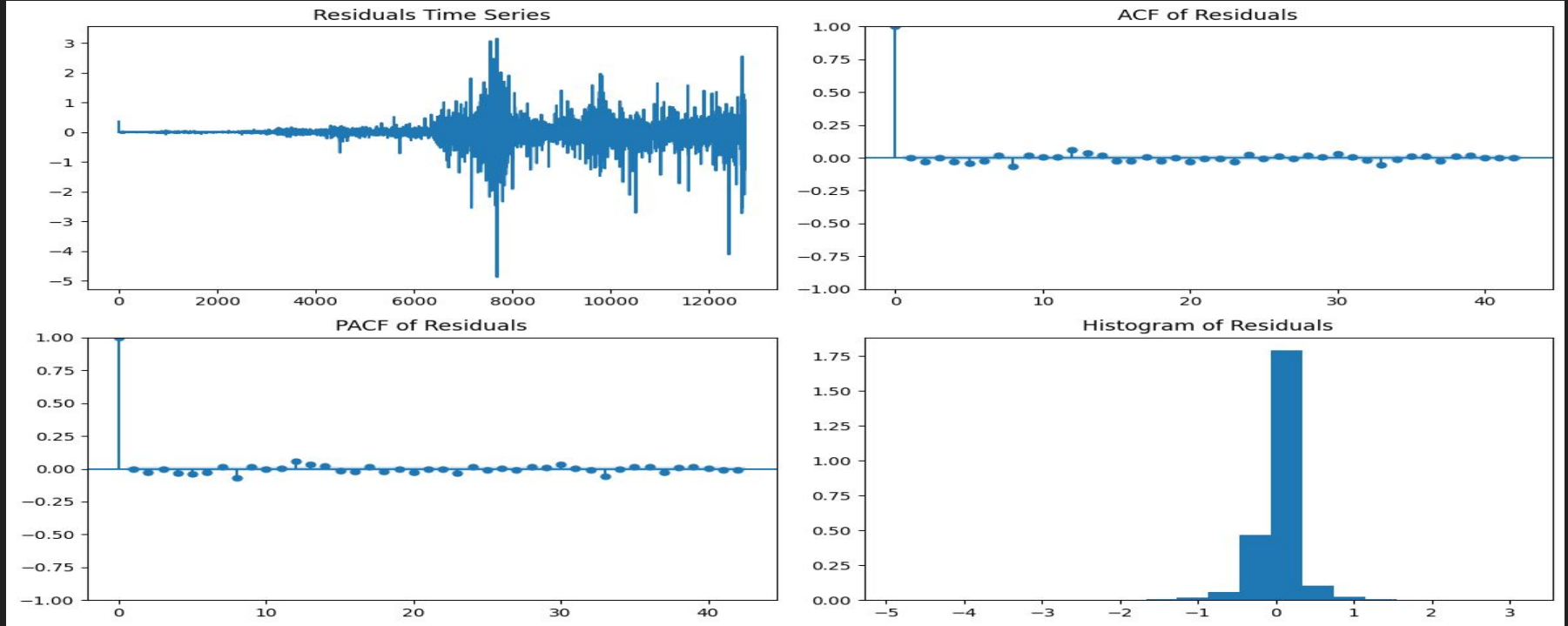


# Arima Model Analysis With Cross Validation

We have a great prediction. We barely see any blue, which is the test values. This is a ARIMA(1,1,0) model. We use model with one step cross validation also known as walk-forward validation with 1 step. This model is slow, it takes about 30 to 40 minutes to train. We only have about 13000 rows. You can imagine if it was 100000 rows. The accuracy is great though.



# Arima Model Analysis With Cross Validation



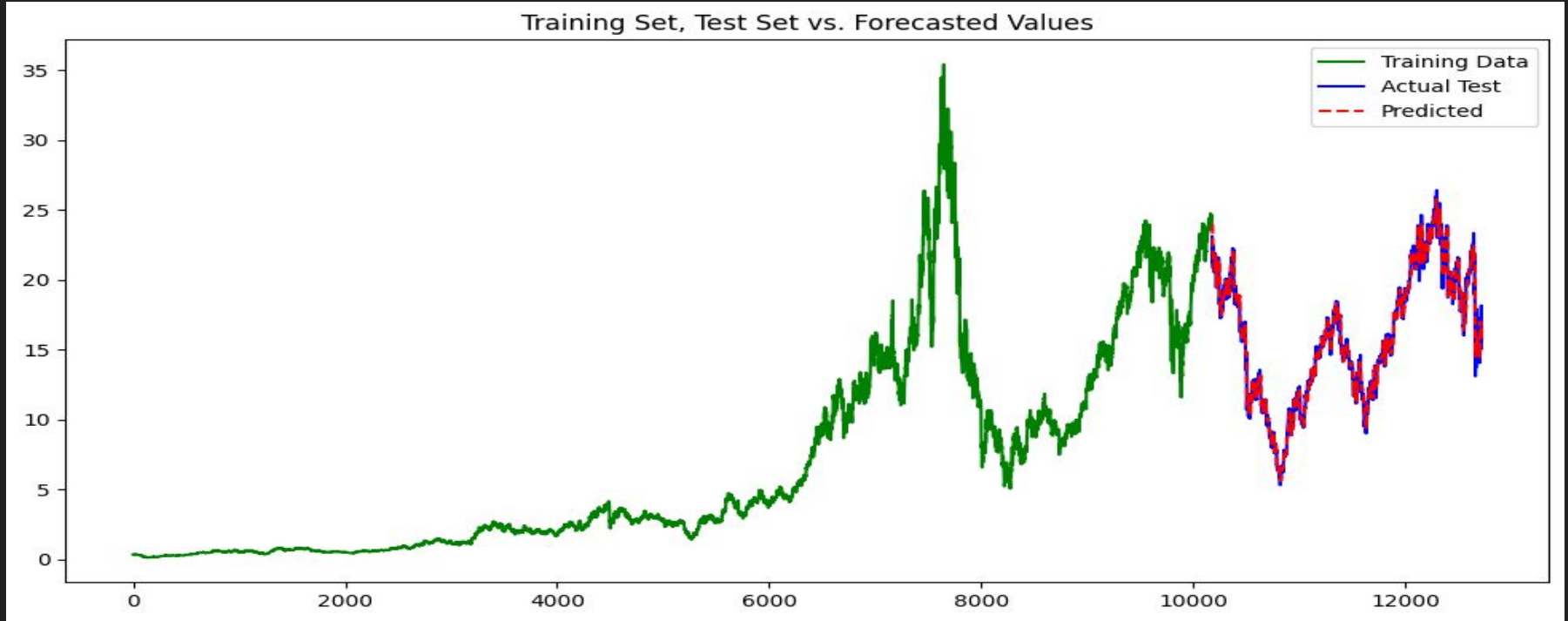
# Arima Model Analysis Without Cross Validation

Time series chart has no pattern or trend. It consolidating. It look almost like white noise. This is sign that there maybe external factor or pattern that I missing in my model, but really minimum.

Both ACF and PACF shows a series of several significant spike, but really close to the significant threshold. They even smaller than before. So these are signs of weak autocorrelation among those lags, but some of these may improve the model.

The histogram shows a normal distribution, which is a good thing. My residual has an equally likely to be negative or positive.

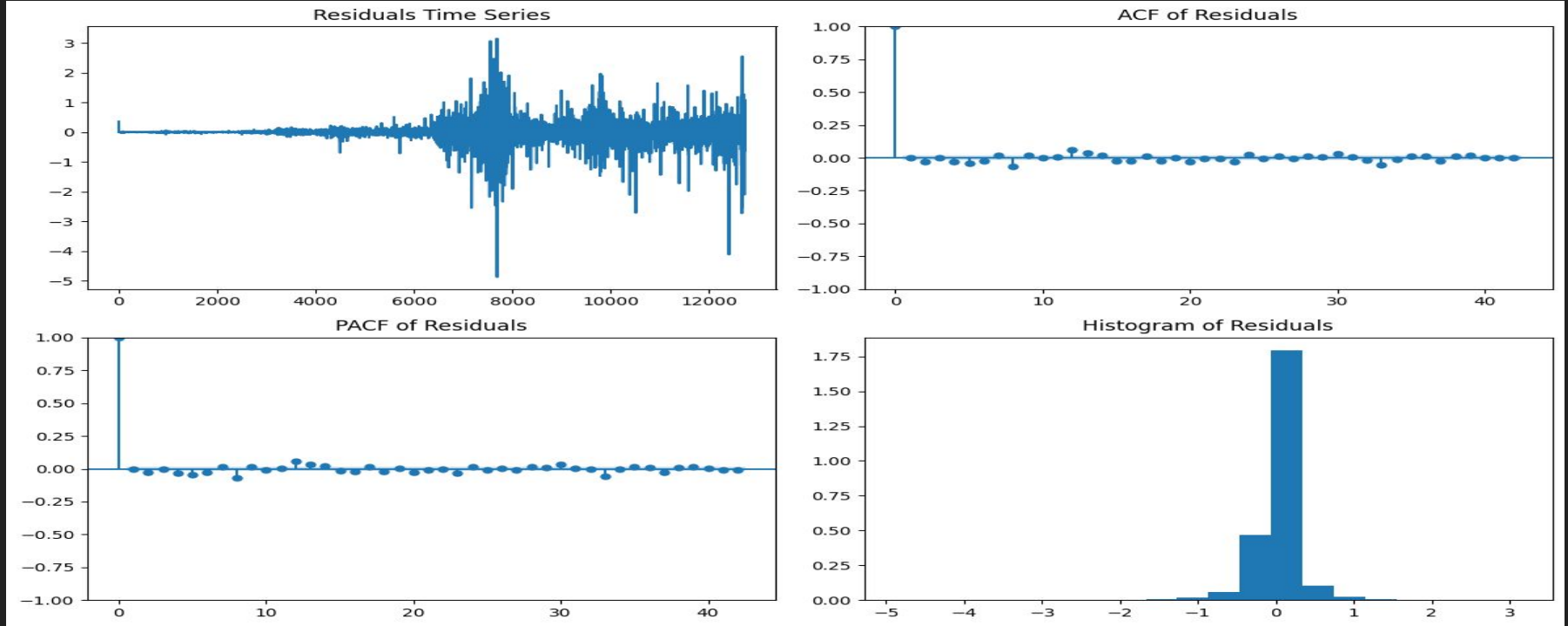
# Arima Model Analysis With Cross Validation



# Arima Model Analysis With Cross Validation

We have a great prediction. We see a little more blue, which is the test values. This is a ARIMA(1,1,0) model. We use model with walk-forward validation with 6 step. This model is faster, it takes about 7 to 15 minutes to train. The accuracy is lower, but still great though.

# Arima Model Analysis With Cross Validation



# Arima Model Analysis Without Cross Validation

Time series chart has no pattern or trend. It consolidating. It look almost like white noise. This is sign that there maybe external factor or pattern that I missing in my model, but really minimum.

Both ACF and PACF shows a series of several significant spike, but really close to the significant threshold. They even smaller than before. So these are signs of weak autocorrelation among those lags, but some of these may improve the model.

The histogram shows a normal distribution, which is a good thing. My residual has an equally likely to be negative or positive.

# Conclusion

Cross Validation with Arima is a must. Arima can't predict too far in the future because it a linear model. The farther we go to the future, the more uncertainty we have our prediction.

# Future Research

News sentiment analysis

Social media analysis

Allocate more data

Do a hourly analysis

Utilizing a multi variate model like VARMA

Analyzing technical indicators with ARIMA

Incorporating fundamental analysis into machine  
and more.



# Reference

<https://www.kaggle.com/datasets/qks1lver/amex-nyse-nasdaq-stock-histories>