**DESCRIPTIVE AND PREDICTIVE ANALYSIS ON SHIBA INU (Cryptocurrency)**

**MAJOR PROJECT REPORT**

**Submitted in partial fulfilment of the requirement for the degree of**

**B.TECH (CSE)**

**BY**

**Ritwik Bahukhandi (2018-310-116)**

Under the Supervision of

**Mr. Piyush Kumar Gupta**

****

**JAMIA HAMDARD UNIVERSITY**

**HAMDARD NAGAR , NEW DELHI - 110062**

**(Deemed to be University)**

**MAY , 2022**

**CANDIDATE’S DECLARATION**

It is hereby certified that the work which is being presented in the B. Tech Minor Project Report entitled "DESCRIPTIVE AND PREDICTIVE ANALYSIS ON SHIBA INU (Cryptocurrency)" in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology and submitted in the Department of Computer Science & Engineering of JAMIA HAMDARD UNIVERSITY, New Delhi (AFFILIATED TO AICTE) is an authentic record of our own work carried out during a period from March 2022 to May 2022 under the guidance of Mr. Piyush Kumar Gupta. The matter presented in the B.Tech Major Project Report has not been submitted by us for the award of any other degree of this or any other Institute.

**Ritwik Bahukhandi (2018-310-116)**

**CERTIFICATE**

This is to certify that the project entitled **Fake News Detection using NLP** is a bonafide work carried out by **Mr. Ritwik Bahukhandi** under my guidance and supervision and submitted in partial fulfillment of B. Tech. degree in CSE of JAMIA HAMDARD UNIVERSITY affiliated by AICTE, New Delhi. The work embodied in this project has not been submitted for any other degree or diploma.

**MR. PIYUSH KUMAR GUPTA**

**(Assistant Professor)**

**Department of Computer Science & Engineering**

**DR. FARHEEN SIDDIQUI**

**Professor & HOD,**

**Computer Science & Engineering,**

**Jamia Hamdard University (SEST)**

**Acknowledgements**

I express my deep gratitude to **Mr. Piyush Kumar Gupta**, Assistant Professor, Department of Computer Science and Engineering for his valuable guidance and suggestion throughout my project work.

I would like to extend our sincere thanks to **Dr. Farheen Siddiqui, Professor & HOD,**

**Computer Science & Engineering,** for his time to suggestions to complete our project work.

# **TABLE OF CONTENTS**

**Topic Page no.**

Candidate’s Declaration……………………………….………………….…….II

Certificate…………………………………………………………...................III

Acknowledgement……………………………………………………….….…IV

Table of Contents………………………………………….……………..V – VI

List of Figures…………………………………………………………VII – VIII

Abstract……………………………………………………...………………..X

Chapter-1. Introduction……………………………………………………1 – 6

Chapter-2. Literature Review…………………………………..…………7– 11

Chapter-3. Proposed Solution……………………………………..............12-28

3.1Problem Statement…………………………………...…………12

3.2 Methodology…………………………………………...………12

3.2.1 Dataset…………………………………………............12

3.2.2 Describing Dataset………………………...................13 – 14

3.2.3 Descriptive Analysis…………………………............15 – 18

3.2.4 Moving Averages…………………………………….19 – 22

3.2.5 Correlation Analysis…………………………………21 - 25

3.2.6 Seasonal Analysis……………………………………26

3.2.7 Predictive Analysis…………………………………..27-28

Chapter-4. Results………………………………………………………..29 – 30

4.1 Forecasting Graph…………………………………………….29

5.2 Evaluation…………………………………………………….30

Chapter-5 Observations…………………………………………………31

Chapter-6. Conclusion………………………………………………….32

Reference...………………………………………………….…………….33 - 34

**List of Figures**

|  |  |
| --- | --- |
| **S NO**. | **FIGURES** |
| 1.1 | Shiba Inu Logo |
| 1.2 | Compoment of Time Series |
| 1.3 | Applications of Time Series Analysis |
| 2.1 | Additive Seasonality |
| 2.2 | Multiplicative Seasonality |
| 3.1 | Workflow of the Project |
| 3.2. | .describe() (Python) |
| 3.3 | Volatility of Asset |
| 3.4 | Probability Distribution of Close and Volume |
| 3.5 | Line Graph (Close and Adj Close) |
| 3.6 | Augmented Dickey Fuller Test Result |
| 3.7 | Simple Moving Average (SMA 21) |
| 3.8 | Exponential Moving Average (EMA 21) |
| 3.9 | Cummulative Moving Average (CMA) |
| 3.10 | Pearson Correlation |
| 3.11 | Spearman Correlation |
| 3.12 | Kendall Correlation |
| 3.13 | Seasonal Decompose of Closing Price1 |
| 3.14 | Residual Plot (ARIMA Model) |
| 4.1 | Forecasting on Test Data (ARIMA Model) |

**ABSTRACT**

Cryptocurrency and Blockchain is one of the most trending technologies out there , Just like Stock Market which is actually Centralised , these Cryptocurrency market (Decentralised) has also gained popularity and plays an important role in Economy. This project is proposed to give insights and visualizations on how **Shiba-Inu** has performed in past and how will it going to perform in future by analysing it’s time series components .

**AIM**

This project aims to give **Historical (Descriptive)** and **Future (Predictive)** using Statistics , Time Series Analysis , computational models of machine learning models using Python. The extraction of insights from the Shiba-Inu dataset is done by Statistical Analysis Various algorithms like ARIMA , AUTO – ARIMA , EXPONENTIAL SMOOTHENING , AD FULLER TEST etc. is used to visualize the predictive and Descriptive insight

**Chapter-1**

**INTRODUCTION**

**1.1BLOCKCHAIN and CRYTOCURRENCY**

Blockchain is a revolutionary technology that allows people to record transactions on a digital, **decentralised,** distributed ledger, without any central authority. Some consider this technology as “the trust machine” and cryptocurrency is without doubt the most notable by-product of the blockchain revolution. Initial Coin Offering (ICO) is a new way to raise entrepreneurial finance, newly created cryptocurrencies are being sold to the public by startups in exchange of capital. This book chapter intends to clarify this phenomenon by explaining the concepts of blockchain technology, cryptocurrency and ICO, in order to provide valuable insights into this new trend of entrepreneurial finance.

A **cryptocurrency** is a digital or virtual currency that is secured by cryptography, which makes it nearly impossible to counterfeit or double-spend. Many cryptocurrencies are decentralized networks based on blockchain technology—a distributed ledger enforced by a disparate network of computers. A defining feature of cryptocurrencies is that they are generally not issued by any central authority, rendering them theoretically immune to government interference or manipulation.

A cryptocurrency is a form of digital asset based on a network that is distributed across a large number of computers. This decentralized structure allows them to exist outside the control of governments and central authorities.

Experts believe that blockchain and related technology will disrupt many industries, including finance and law.

The advantages of cryptocurrencies include cheaper and faster money transfers and decentralized systems that do not collapse at a single point of failure.

The disadvantages of cryptocurrencies include their price **volatility, high energy consumption for mining activities, and use in criminal activities**.

**SHIBA INU (MEME COIN)**

The Shiba Inu coin launched on August 1, 2020 and at the time was priced at less than $0.00000001. With 1 quadrillion tokens minted it remained at that level for some time, but more recently thanks to the increasing interest in Dogecoin, Shiba Inu has also come to life. Created by an anonymous developer going by the name Ryoshi, the Woofpaper for Shiba Inu says the project was begun from the simple question of:

“What would happen is a cryptocurrency project was 100% run by its community?”



Fig 1.1 SHIBA-INU (A meme Coin)

While there are certainly proponents of the Shiba Inu coin, there are just as many who call the coin a joke. Which might seem cruel or unjust, but as a direct competitor of Dogecoin, which was openly called a joke by its creator, it’s actually very accurate. And yet even as a joke it seems to have massive potential for gains as people have poured massive amounts of capital into the token, hoping for it to continue building on the more than 1 million percent gains already delivered. The Shiba Inu coin uses the ticker symbol SHIB and is an ERC-20 token that runs on the Ethereum blockchain. Interestingly, the creators of the Shiba Inu coin sent half the total coin supply to Ethereum’s founder Vitalik Buterin, while the other half was locked on Uniswap, with the key thrown away. Or so the Shiba Inu woofpaper claims.

**TIME SERIES ANALYSIS (SHIBA-INU)**

Time series analysis is a specific way of analyzing a sequence of data points collected over an interval of time. In time series analysis, analysts record data points at consistent intervals over a set period of time rather than just recording the data points intermittently or randomly. However, this type of analysis is not merely the act of collecting data over time.

What sets time series data apart from other data is that the analysis can show how variables change over time. In other words, time is a crucial variable because it shows how the data adjusts over the course of the data points as well as the final results. It provides an additional source of information and a set order of dependencies between the data.

Time series analysis typically requires a large number of data points to ensure consistency and reliability. An extensive data set ensures you have a representative sample size and that analysis can cut through noisy data. It also ensures that any trends or patterns discovered are not outliers and can account for seasonal variance. Additionally, time series data can be used for forecasting—predicting future data based on historical data.

**Why organizations use time series data analysis ?**

Time series analysis helps organizations understand the underlying causes of trends or systemic patterns over time. Using data visualizations, business users can see seasonal trends and dig deeper into why these trends occur. With modern analytics platforms, these visualizations can go far beyond line graphs.

When organizations analyze data over consistent intervals, they can also use time series forecasting to predict the likelihood of future events. Time series forecasting is part of predictive analytics. It can show likely changes in the data, like seasonality or cyclic behavior, which provides a better understanding of data variables and helps forecast better..

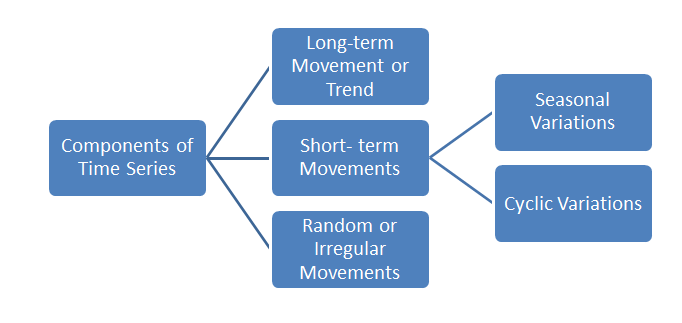


Fig 1.2 Component of Time Series Analysis

**Time Series Analysis Types**

Because time series analysis includes many categories or variations of data, analysts sometimes must make complex models. However, analysts can’t account for all variances, and they can’t generalize a specific model to every sample. Models that are too complex or that try to do too many things can lead to a lack of fit. Lack of fit or overfitting models lead to those models not distinguishing between random error and true relationships, leaving analysis skewed and forecasts

**Models of time series analysis include:**

* **Classification:** Identifies and assigns categories to the data.
* **Curve fitting:** Plots the data along a curve to study the relationships of variables within the data.
* **Descriptive analysis:** Identifies patterns in time series data, like trends, cycles, or seasonal variation.
* **Explanative analysis:**Attempts to understand the data and the relationships within it, as well as cause and effect.
* **Exploratory analysis:**Highlights the main characteristics of the time series data, usually in a visual format.
* **Forecasting:** Predicts future data. This type is based on historical trends. It uses the historical data as a model for future data, predicting scenarios that could happen along future plot points.
* **Intervention analysis:**Studies how an event can change the data.

**Segmentation:** Splits the data into segments to show the underlying properties of the source information.

### **Data classification**

Further, time series data can be classified into two main categories:

* **Stock time series data**means measuring attributes at a certain point in time, like a static snapshot of the information as it was.
* **Flow time series data** means measuring the activity of the attributes over a certain period, which is generally part of the total whole and makes up a portion of the results.

### **Data variations**

In time series data, variations can occur sporadically throughout the data:

* **Functional analysis**can pick out the patterns and relationships within the data to identify notable events.
* **Trend analysis** means determining consistent movement in a certain direction. There are two types of trends: deterministic, where we can find the underlying cause, and stochastic, which is random and unexplainable.
* **Seasonal variation**describes events that occur at specific and regular intervals during the course of a year. Serial dependence occurs when data points close together in time tend to be related.

Time series analysis and forecasting models must define the types of data relevant to answering the business question. Once analysts have chosen the relevant data they want to analyze, they choose what types of analysis and techniques are the best fit.

### **Important Considerations for Time Series Analysis**

While time series data is data collected over time, there are different types of data that describe how and when that time data was recorded. For example:

* **Time series data** is data that is recorded over consistent intervals of time.
* **Cross-sectional data** consists of several variables recorded at the same time.
* **Pooled data** is a combination of both time series data and cross-sectional data.

**Applications of Time Series Analysis**



Fig 1.3 Applications of Time Series Analysis

There are many applications of Time Series Analysis, among which Crypto/Stock Analysis is one of them, which is used in this project to the Future forecast of a particular Cryptocurrency.

I’ve used the machine learning library PANDAS and STATSMODEL Library in Python since it has built-in methods that implement different statistical and Data oriented methodologies.

**NOTE : Time Series Analysis of Shiba Inu can be done through various Mathematical and Statistical Analysis but there are many more things which can effect the prediction so we can’t directly say its 100% accurate to perform time series analysis on cryptocurrency**

**Chapter-2**

**LITERATURE REVIEW**

According to Trautman (2014), cryptocurrencies are a subset of digital currencies that may have either centralized institutions or are based on a decentralized network. In simple terms, cryptocurrencies are a new type of currency (Duque, 2020; Hudson & Urquhart, 2019) that is digital and produced from cryptographic algorithms, exchanged across the Internet using protocols such as peer-to-peer networking (Nakamoto, 2008). Another way to define cryptocurrencies is the fact that they are based on the use of complex cryptographic techniques to provide users with a secure and safe medium of exchange (Bulut, 2018). The creation of value (or money) and the triggering of transactions are governed by the mining process, which is a set of mathematical algorithms that are implemented within the underlying protocol (Adhami et al., 2018; Cennamo et al., 2020). Most cryptocurrencies are created to introduce new units of currency with a limited total amount (Baur et al., 2015). Unlike state-issued currencies, cryptocurrencies are not governed by established laws, but by technology (Dodd, 2018). Accordingly, this makes cryptocurrencies a new invention that is different from traditional currencies. In the field of finance, the advent of cryptocurrencies represents a new area that requires additional public and academic attention

In paper [1] In an open-ended analysis where there are many options and avenues for the investigator to explore, we have found a method that helps to keep a common thread throughout. Multiple iterations of creating a story that accurately portrays a metaphor for a bitcoin transaction involve a deep learning of the details by the investigator. We discovered new meanings and details in the process, perhaps more than our traditional research would have provided. We presented our exploratory analysis of bitcoin data to an audience of 8, many of who were unfamiliar with technical bitcoin terminology. We received positive feedback from the audience that our method of telling a story, and revisiting that story throughout the presentation of data helped the overall understanding of the new topics. We received great feedback, suggestions and additional questions about the technical aspects of the analysis. Without a control group its unclear whether we succeeded at improving communication of our analysis.

In paper [2] The stock market is very unpredictable, any geopolitical change can impact the share trend of stocks in the share market, recently we have seen how covid-19 has impacted the stock prices, which is why on financial data doing a  reliable trend analysis is very difficult. The most efficient way to solve this kind of issue is with the help of Machine learning and Deep learning. A popular and widely used statistical method for time series forecasting is the ARIMA model. It is one of the most popular models to predict linear time series data. This model has been used extensively in the field of finance and economics as it is known to be robust, efficient, and has a strong potential for short-term share market prediction. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting and provide complementary approaches to the problem. While exponential smoothing models are based on a description of the trend and seasonality in the data, ARIMA models aim to describe the auto-correlation(Autocorrelation is the degree of similarity between a given time series and a lagged version of itself over successive time intervals) in the data.

In paper [3] The Author approach was

2. Dataset Quick Overview & Pre-Processing

3. Features generation

4. Distribution of Features

5. Correlation Analysis

6. Candle stick chart on Bitcoin historical data

7. Time series analysis and prediction using prophet

8. Regression analysis using pycaret

In Paper [4] , The author proposed the idea of unit root or AD-Fuller testA unit root test tests whether a time series is not stationary and consists of a unit root in time series analysis. The presence of a unit root in time series defines the null hypothesis, and the alternative hypothesis defines time series as stationary.

There are various tests which include unit root tests.

* Augmented Dickey-Fuller test.
* Phillips-perron test.
* KPSS test.
* ADF-GLS test
* Breusch-godfrey test.
* Ljung-Box test.
* Durbin-watson test.

The augmented dickey- fuller test is an extension of the dickey-fuller test, which removes autocorrelation from the series and then tests similar to the procedure of the dickey-fuller test. The augmented dickey fuller test works on the statistic, which gives a negative number and rejection of the hypothesis depends on that negative number; the more negative magnitude of the number represents the confidence of presence of unit root at some level in the time series.

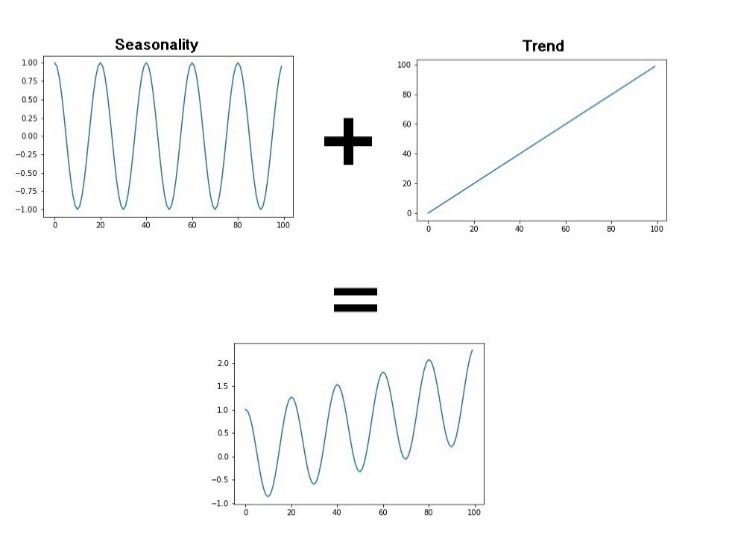


In Paper[5] **Seasonality** in time-series data refers to a pattern that occurs at a regular interval. This is different from regular cyclic trends, such as the rise and fall of stock prices, that re-occur regularly but don’t have a fixed period. There’s a lot of insight to be gained from understanding seasonality patterns in your data and you can even use it as a baseline to compare your time-series machine learning models.

**Additive Seasonality**

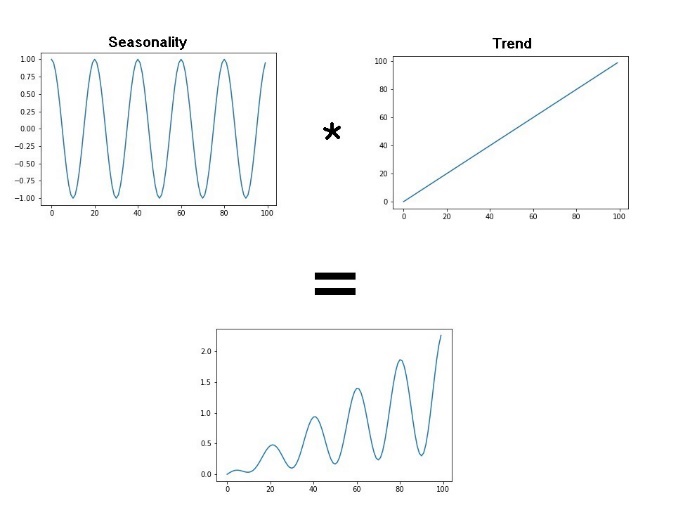
It’s pretty rare for actual time series to have constant crest and trough values and instead, we typically see some kind of general trend like an increase or a decrease over time. In our sales price plot, for example, the median price tends to go up over time.

If the amplitude of our seasonality tends to remain the same, then we have what’s called an additive seasonality. Below is an example of an additive seasonality.



**Multiplicative Seasonality**

The other type of seasonality that you may encounter in your time-series data is multiplicative. In this type, the amplitude of our seasonality becomes larger or smaller based on the trend. An example of multiplicative seasonality is given below.



[5]Exponential smoothing was proposed in the late 1950s ([Brown, 1959](https://otexts.com/fpp2/expsmooth.html#ref-Brown59); [Holt, 1957](https://otexts.com/fpp2/expsmooth.html#ref-Holt57); [Winters, 1960](https://otexts.com/fpp2/expsmooth.html#ref-Winters60)), and has motivated some of the most successful forecasting methods. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight. This framework generates reliable forecasts quickly and for a wide range of time series, which is a great advantage and of major importance to applications in industry.

**Chapter – 3**

**PROPOSED SOLUTION**

**3.1 PROBLEM STATEMENT**

To perform Time Series Analysis on Shiba Inu historical data to gain descriptive and Predictive insights on Closing Price of Shiba Inu using Statistical and Machine Learning Models

**3.2. METHODOLOGY**

**3.2.1 DATASET**

The dataset was taken from Yahoo Finance (https://finance.yahoo.com/quote/SHIB-INR/history/) It’s often considered to be the ﬁrst step towards analysis . The Data is decribed in a such way that it has record of more that 600 days of daily shiba inu price since August 2020.

**ATTRIBUTES:**

1. **Date**
2. **Open**
3. **High**
4. **Low**
5. **Close**
6. **Adjacent Close**
7. **Volume**

NOTE : The most important Attribute is Closing Price cause we have to predict the future values of the Closing Price.

**3.2.2 DESCRIBING THE DATA SET**

is a mandatory step when we are working with Time Series Data.  Changing the Data Types of Attribute for e.g. changing the Date attribute to Date Time Signature , Summarising the complete Data Set for further Descrictive analysis comes under this part

****

Fig. 3.1 Workflow of Project

**A) .info()**

The info() method prints information about the DataFrame. The information contains the number of columns, column labels, column data types, memory usage, range index, and the number of cells in each column (non-null values).

**B) .desctibe()**

The describe() method is used for calculating some statistical data like percentile, mean and std of the numerical values of the Series or DataFrame. It analyzes both numeric and object series and also the DataFrame column sets of mixed data types.

include: It is also an optional parameter that includes the list of the data types while describing the DataFrame. Its default value is None.

exclude: It is also an optional parameter that exclude the list of data types while describing DataFrame. Its default value is None.

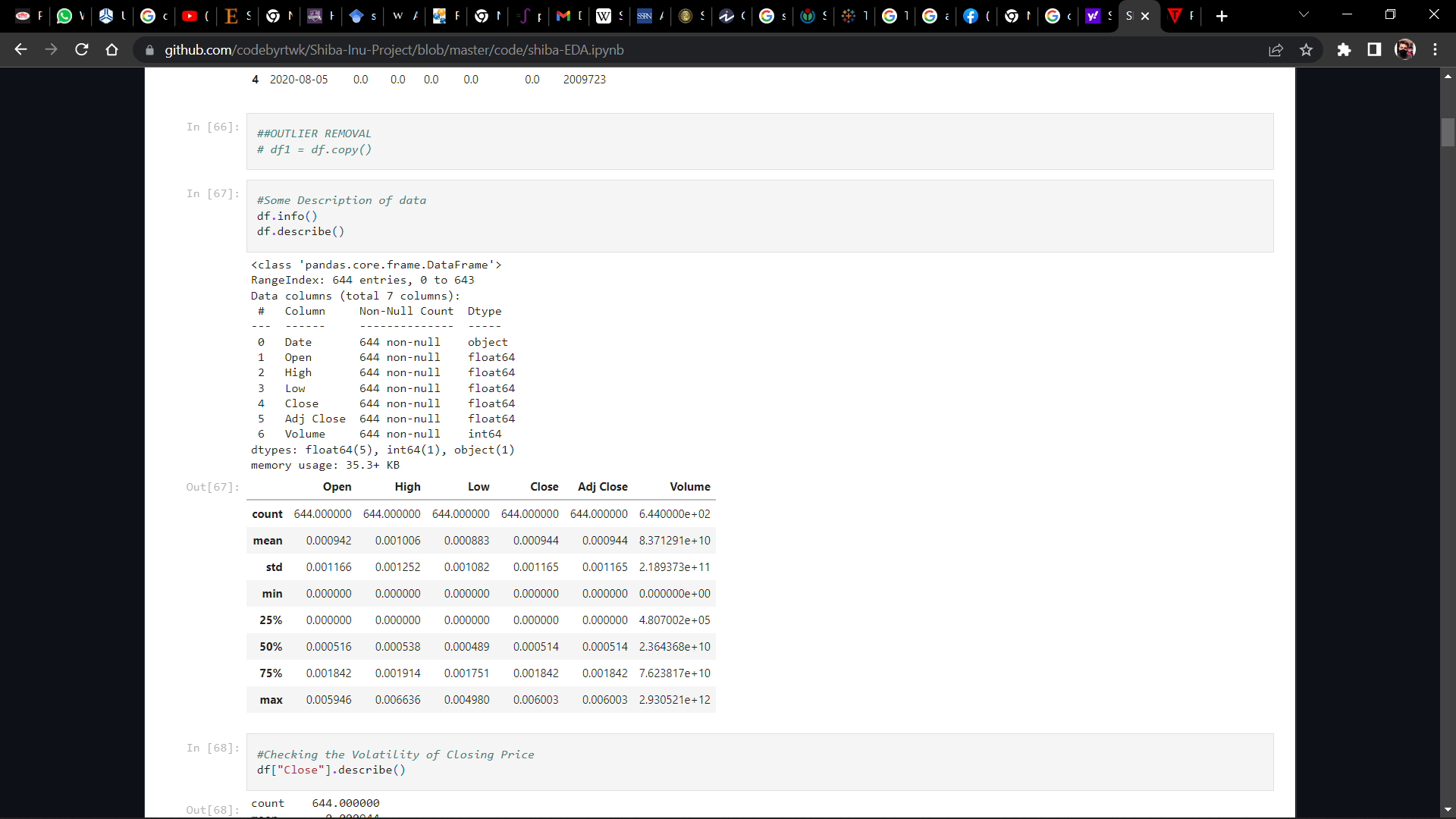


Fig. 3.2 .describe() (python3)

**C) Volatility of Closing Price (.describe())**

Volatility is a statistical measure of the dispersion of returns for a given security or market index. In most cases, the higher the volatility, the riskier the security. Volatility is often measured as either the standard deviation or variance between returns from that same security or market index.

In the securities markets, volatility is often associated with big swings in either direction. For example, when the stock market rises and falls more than one percent over a sustained period of time, it is called a "volatile" market. An asset's volatility is a key factor when pricing options contracts.

**After Analysis it turned out to be Quite a Volatile Asset but not much (Standard Deviation > 0)**

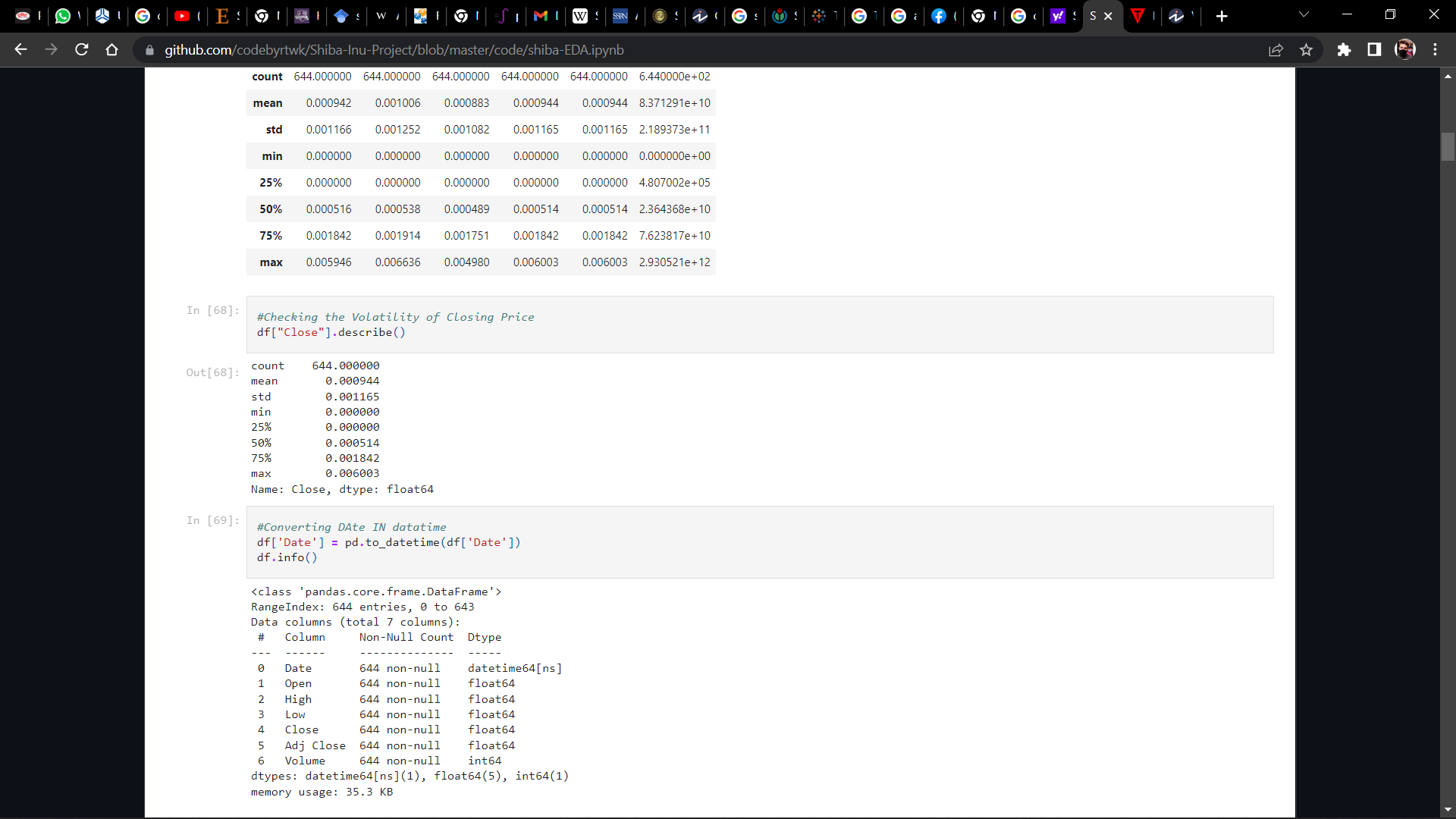
****

Fig. 3.3 Volatility

**3.2.3 DESCRIPTIVE VISUALIZATION / ANALYSIS**

­­­­­it summarizes and organizes all of the collected data into something manageable and simple to understand. The descriptions can include the entire data set or just a part of the data set.

One of the most important things to know about descriptive data analysis is that it focuses on the data instead of on the implication that can be far reaching and go beyond the represented data.

This is the main difference between inferential statistics and descriptive statistics. Inferential statistics uses complicated calculations to make predictions while descriptive statistics does not.

High degree of objectivity and neutrality of the researchers are one of the main advantages of Descriptive Analysis. The reason why researchers need to be extra vigilant is because descriptive analysis shows different characteristics of the data extracted and if the data doesn’t match with the trends then it will lead to major dumping of data.

Descriptive analysis is considered to be more vast than other quantitative methods and provide a broader picture of an event or phenomenon. It can use any number of variables or even a single number of variables to conduct a descriptive research.

**A) Probability Distribution**

A probability distribution is a statistical function that describes all the possible values and likelihoods that a [random variable](https://www.investopedia.com/terms/r/random-variable.asp) can take within a given range. This range will be bounded between the minimum and maximum possible values, but precisely where the possible value is likely to be plotted on the probability distribution depends on a number of factors. These factors include the distribution's mean (average), [standard deviation](https://www.investopedia.com/terms/s/standarddeviation.asp)

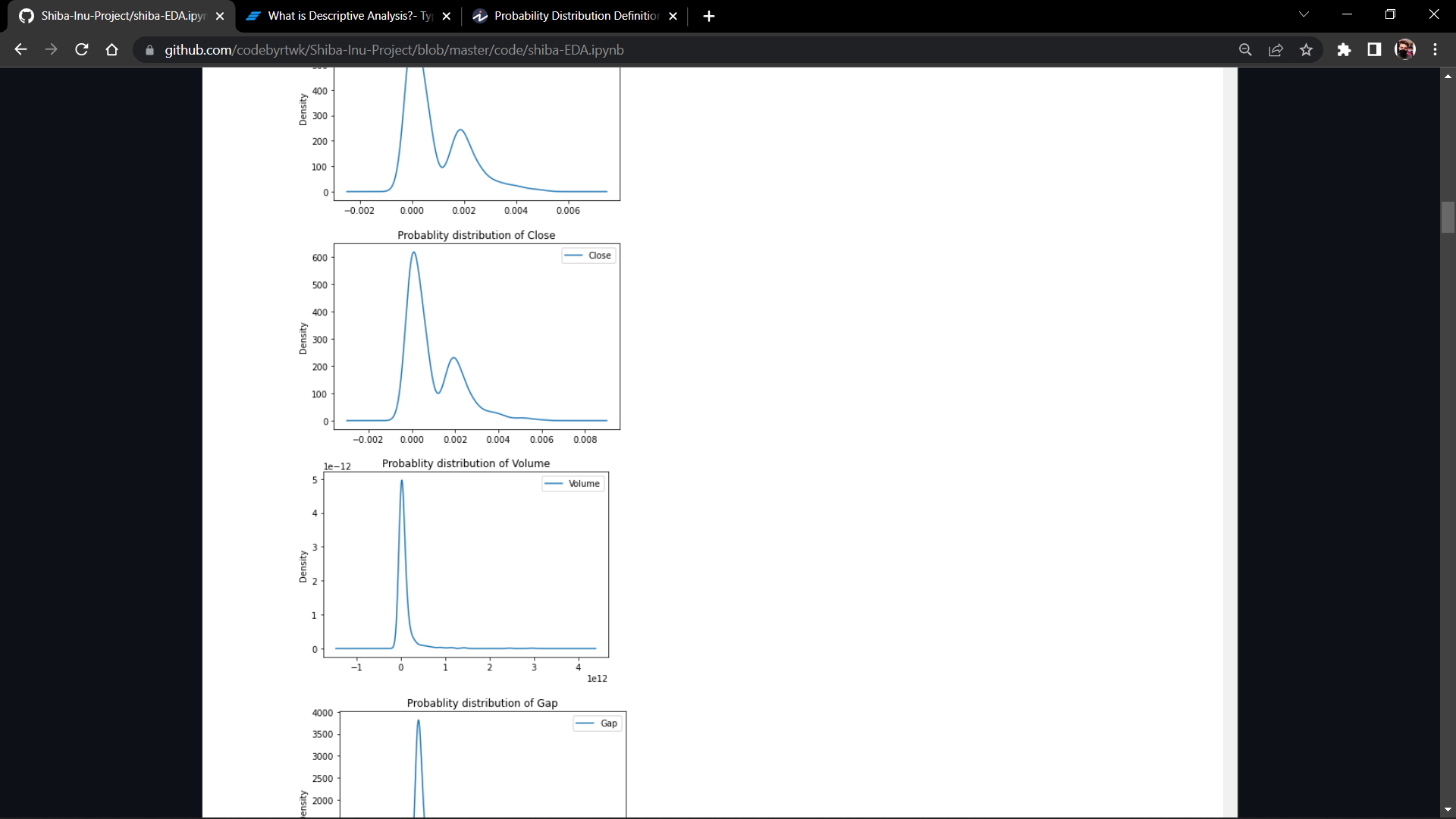


Fig. 3.4 Probability Distribution of Close and Volume

**B) Line Graph vs Time**

The closing price is the raw price or cash value of the last transacted price in a security before the market officially closes for normal trading. It is often the reference point used by investors to compare a stock's performance since the previous day—and closing prices are frequently used to construct line graphs depicting historical price changes over time.The adjusted closing price factors in anything that might affect the stock price after the market closes, such as dividends or splits. Most stocks and other financial instruments are traded after-hours, although in far smaller volumes. Therefore, the closing price of any security is often different from its after-hours trading price.

NOTE : LINE GRAPH IS USED .

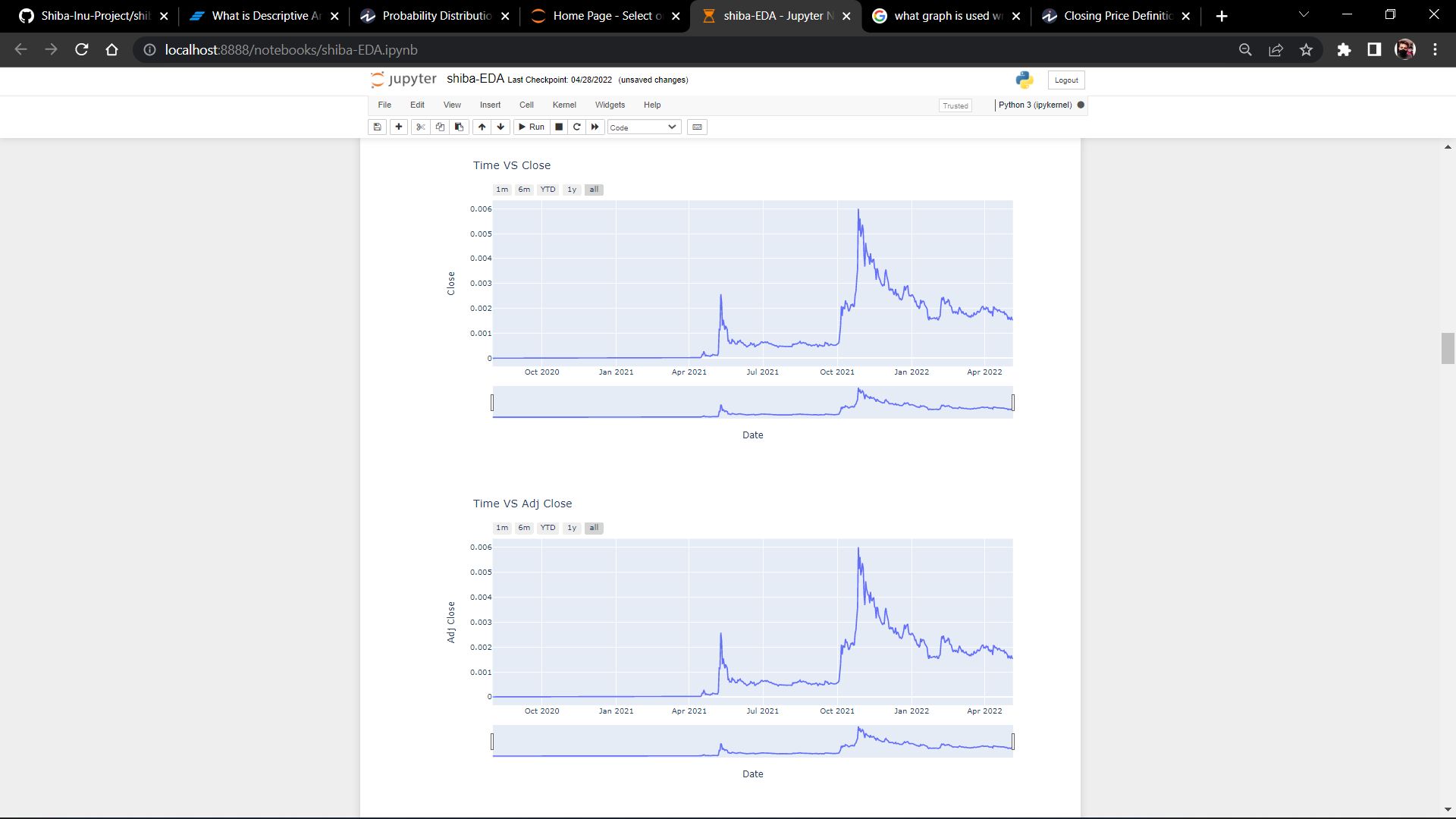


Fig. 3.5 Line Graph (Close and Adj Close)

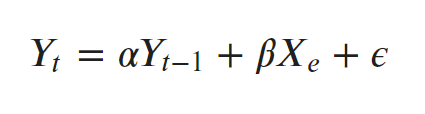
**C) Hypothesis Testing (AD-Fuller Test)**

Augmented Dickey Fuller test (ADF Test) is a common statistical test used to test whether a given Time series is stationary or not. It is one of the most commonly used statistical test when it comes to analyzing the stationary of a series

Since testing the stationarity of a time series is a frequently performed activity in autoregressive models, the ADF test along with KPSS test is something that you need to be fluent in when performing time series analysis.

This approach is used in ARIMA

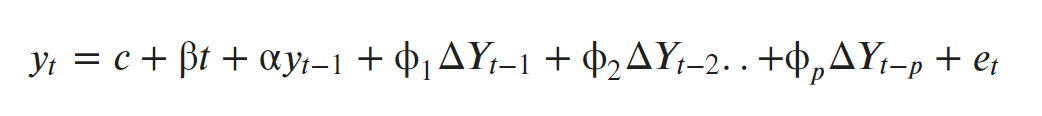
The ADF test belongs to a category of tests called ‘Unit Root Test’, which is the proper method for testing the stationarity of a time series. Unit root is a characteristic of a time series that makes it non-stationary. Technically speaking, a unit root is said to exist in a time series of the value of alpha = 1 in the below equation.



where, Yt is the value of the time series at time ‘t’ and Xe is an exogenous variable (a separate explanatory variable, which is also a time series).

As the name suggest, the ADF test is an ‘augmented’ version of the Dickey Fuller test.

The ADF test expands the Dickey-Fuller test equation to include high order regressive process in the model.



If you notice, we have only added more differencing terms, while the rest of the equation remains the same. This adds more thoroughness to the test.

The null hypothesis however is still the same as the Dickey Fuller test.

A key point to remember here is: Since the null hypothesis assumes the presence of unit root, that is α=1, the p-value obtained should be less than the significance level (say 0.05) in order to reject the null hypothesis. Thereby, inferring that the series is stationary.

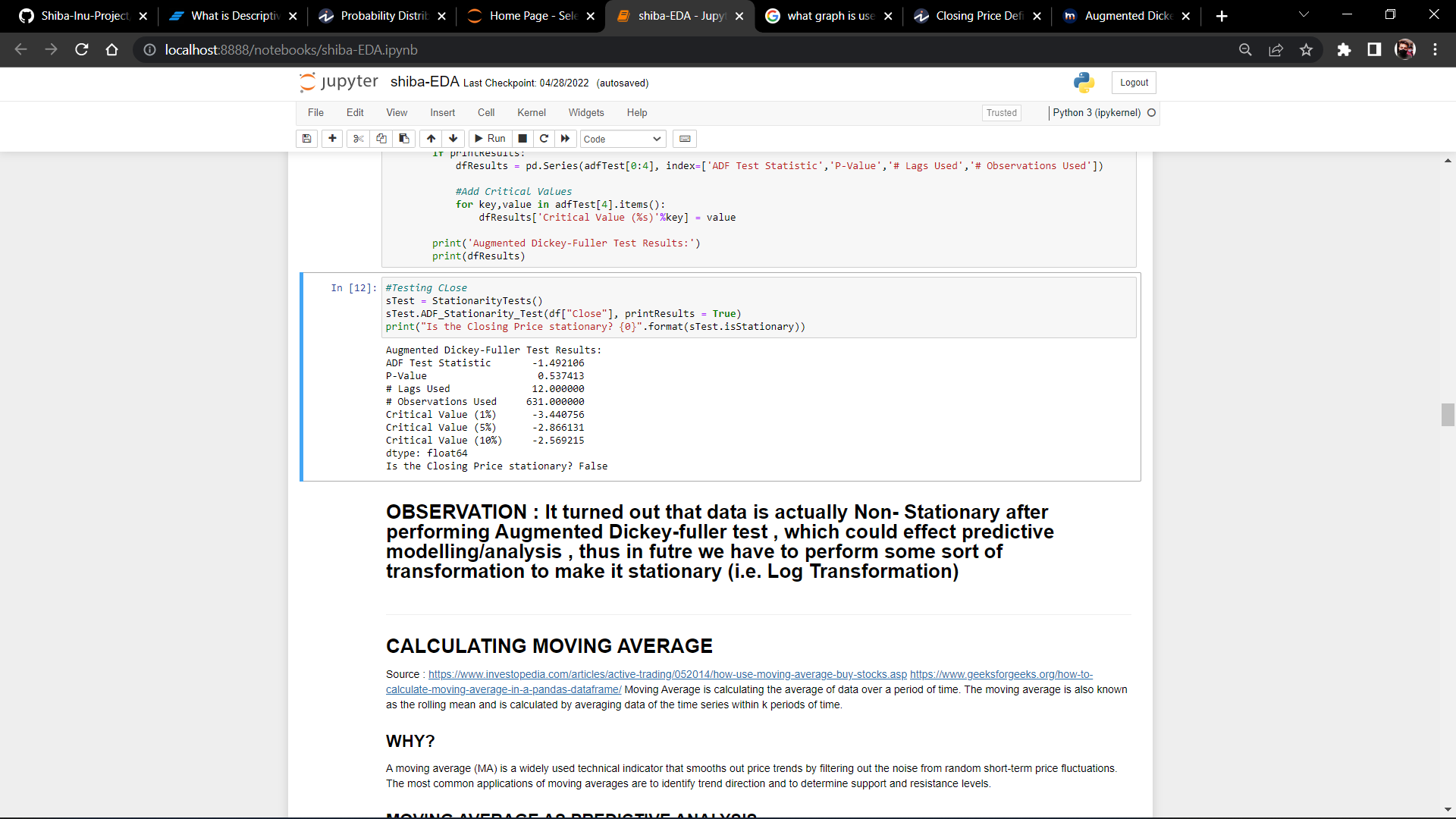


Fig. 3.6 Augmented Dickey Fuller Test Result (Close Price)

**3.2.4 MOVING AVERAGE (Trend Analysis)**

Moving Average is calculating the average of data over a period of time. The moving average is also known as the rolling mean and is calculated by averaging data of the time series within k periods of time

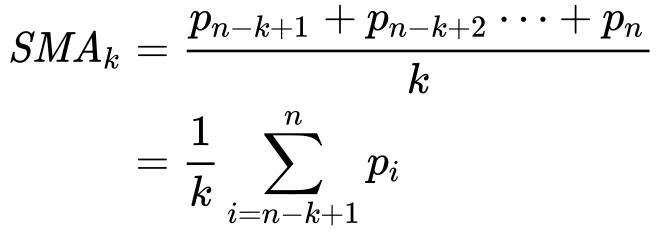
A moving average (MA) is a widely used technical indicator that smooths out price trends by filtering out the noise from random short-term price fluctuations. The most common applications of moving averages are to identify trend direction and to determine support and resistance levels

The concept of a moving average is straightforward. You generally apply a moving average to a series of observations taken over time. The average moves forward with the data series. So, the first moving average might include Days 1 through 3, the second moving average might include Days 2 through 4, and so on. This chapter discusses some of the reasons why you might want to make those calculations.

It is the first step towards the predictive/Inferential Analysis

**A) Simple Moving Average (SMA)**

A simple moving average tells us the unweighted mean of the previous K data points. The more the value of K the more smooth is the curve, but increasing K decreases accuracy. If the data points are p1, p2, . . . , pn then we calculate the simple moving average.



In Python, we can calculate the moving average using .rolling() method. This method provides rolling windows over the data, and we can use the mean function over these windows to calculate moving averages. The size of the window is passed as a parameter in the function .rolling(window)

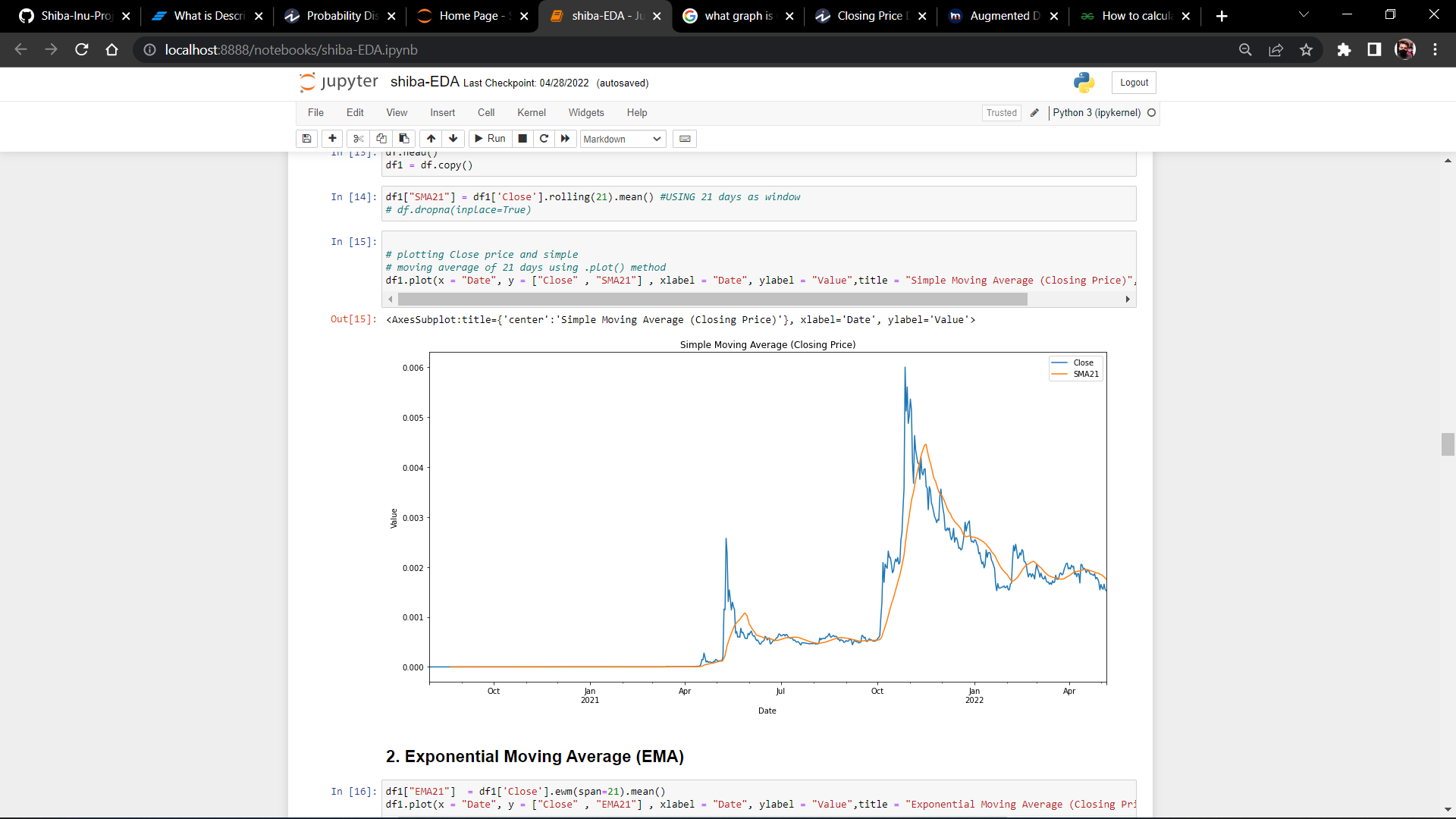
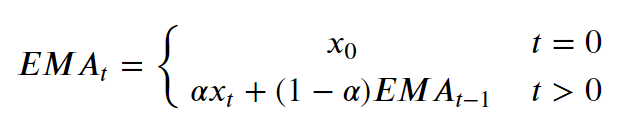


Fig. 3.7 Simple Moving Average ( SMA21)

**B) Exponential Moving Average (EMA)**

Exponential moving average (EMA) tells us the weighted mean of the previous K data points. EMA places a greater weight and significance on the most recent data points. The formula to calculate EMA at the time period t is:



where xt is the value of observation at time t & α is the smoothing factor. In Python, EMA is calculated using .ewm() method. We can pass span or window as a parameter to .ewm(span = ) method.

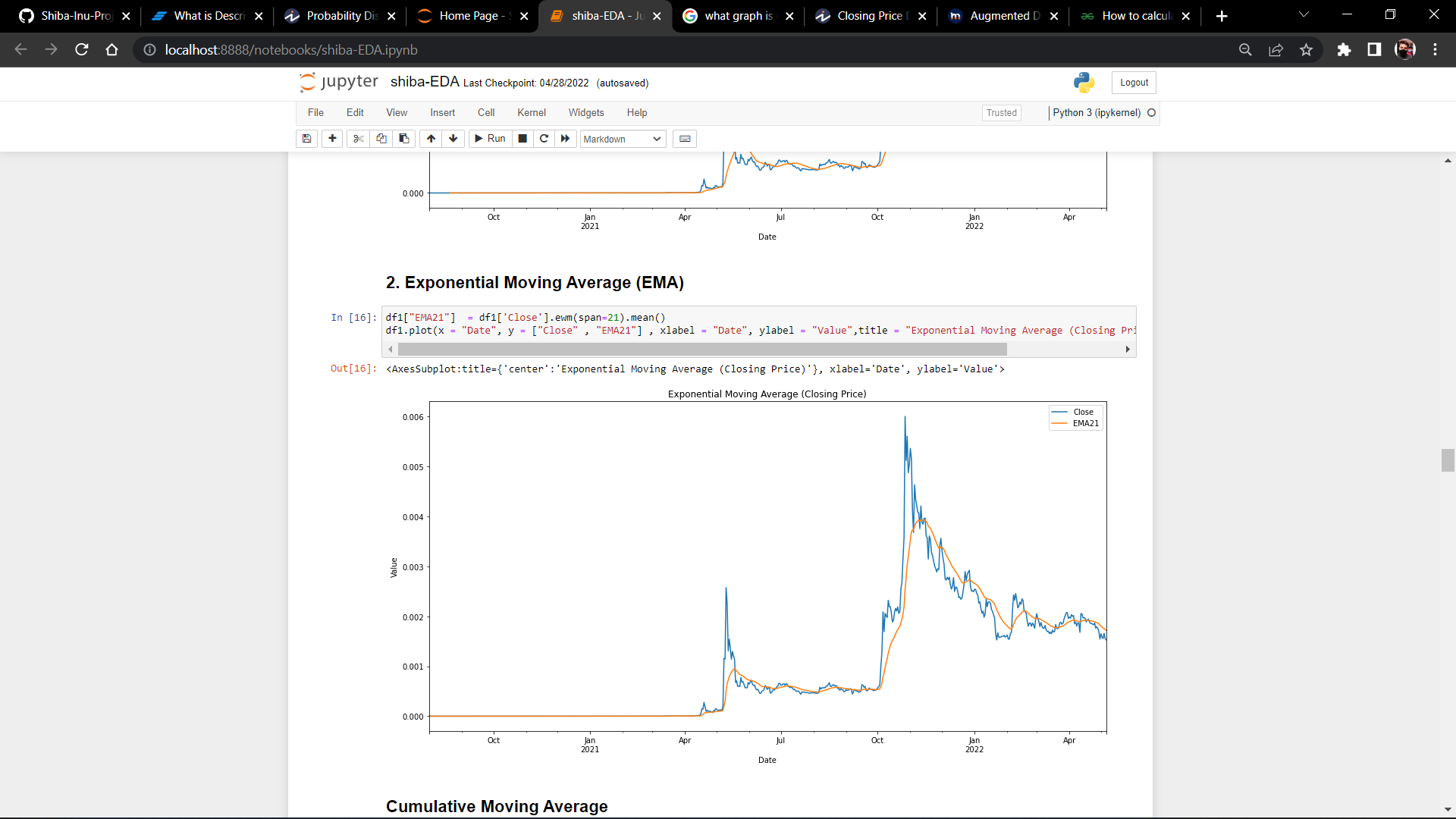
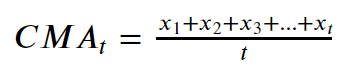


Fig 3.8 Exponential Moving Average (EMA21)

## C) Cummulative Moving Average

The Cumulative Moving Average is the mean of all the previous values up to the current value. CMA of dataPoints x1, x2 ….. at time t can be calculated as,



While calculating CMA we don’t have any fixed size of the window. The size of the window keeps on increasing as time passes. In Python, we can calculate CMA using .expanding() method.

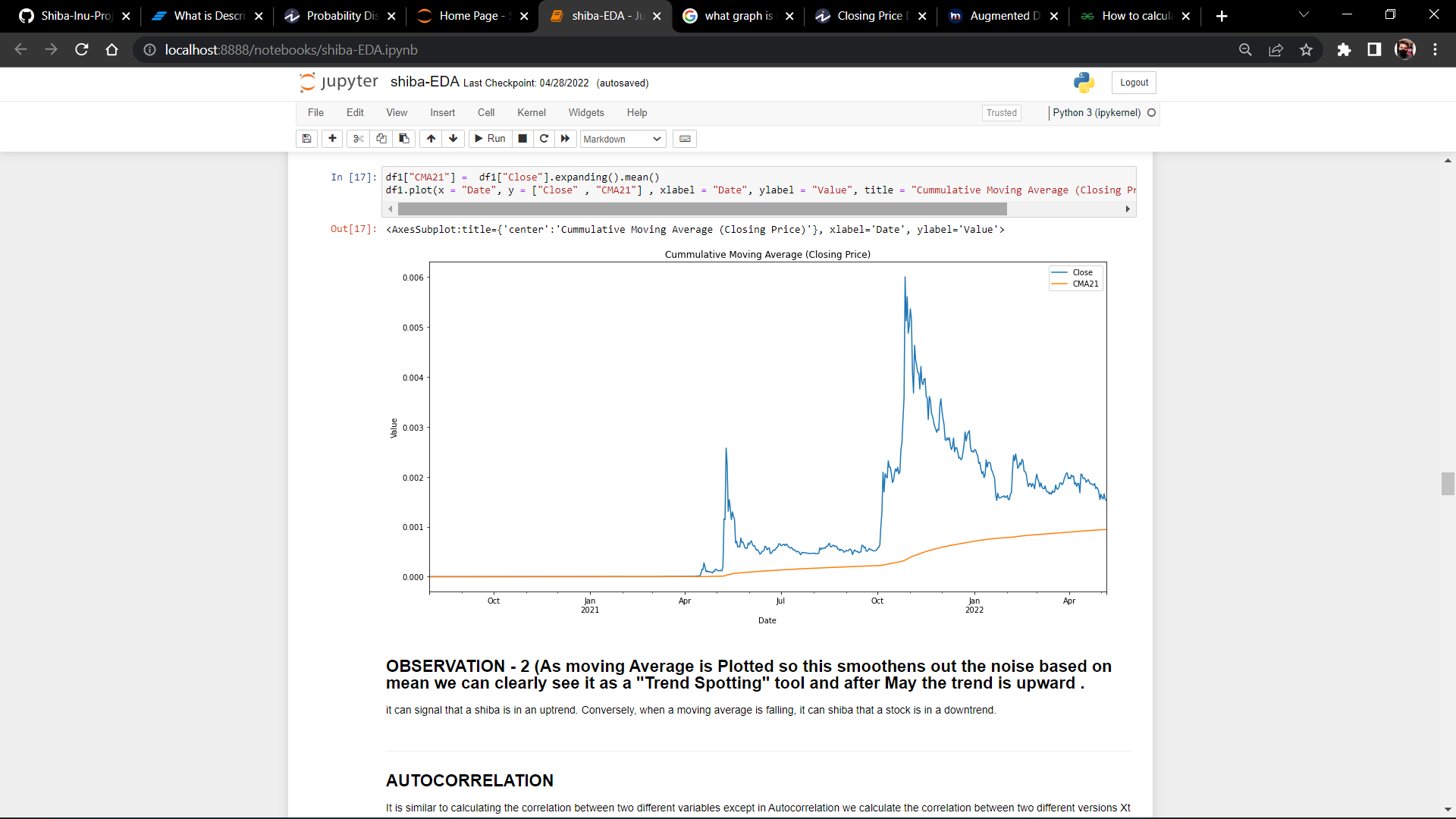
.

Fig. 3.9 Cummulative Moving Average (CMA)

**3.2.4 CORRELATION ANALYSIS**

In statistics, correlation or dependence is any statistical relationship, whether causal or not, between two random variables or bivariate data. In the broadest sense correlation is any statistical association, though it commonly refers to the degree to which a pair of variables are linearly related.

**The Correlation Coefficient**

One of the statistical concepts that is most related to this type of analysis is the correlation coefficient.The correlation coefficient is the unit of measurement used to calculate the intensity in the linear relationship between the variables involved in a correlation analysis, this is easily identifiable since it is represented with the symbol r and is usually a value without units which is located between 1 and -1.

**Advantages of correlation analysis**

In statistics, correlation refers to the fact that there is a link between various events. One of the tools to infer whether such a link exists is correlation analysis. Practical simplicity is undoubtedly one of its main advantages.

To perform reliable correlation analysis, it is essential to make in-depth observations of two variables, which gives us an advantage in obtaining results. Some of the most notorious benefits of correlation analysis are:

Awareness of the behavior between two variables: A correlation helps to identify the absence or presence of a relationship between two variables. It tends to be more relevant to everyday life.

A good starting point for research: It proves to be a good starting point when a researcher starts investigating relationships for the first time.

Uses for further studies: Researchers can identify the direction and strength of the relationship between two variables and later narrow the findings down in later studies.

Simple metrics: Research findings are simple to classify. The findings can range from -1.00 to 1.00.

1. **Pearson Correlation**

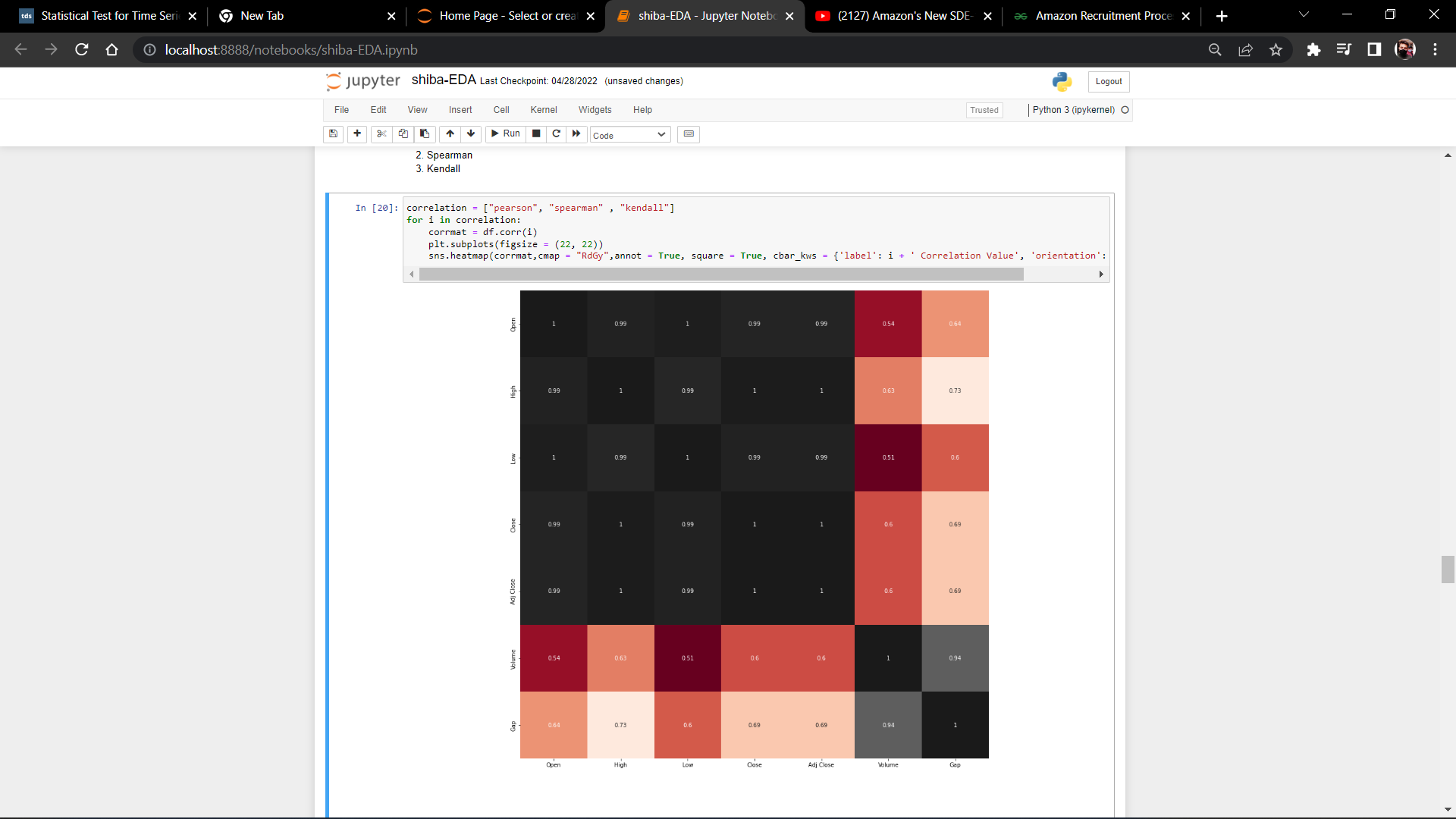


Fig. 3.10 Pearson Correlation

B**) Spearman Correlation**

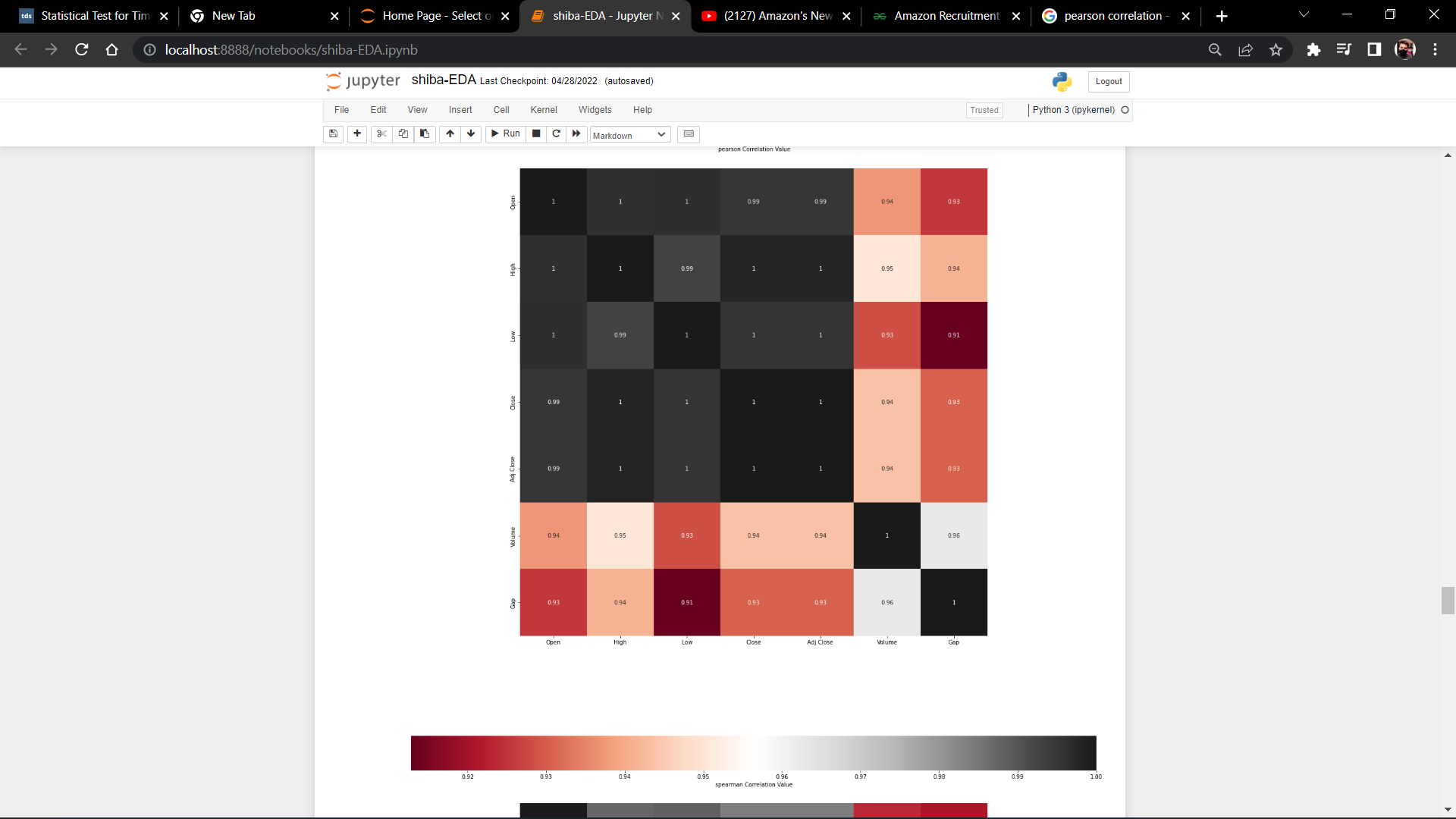


Fig. 3.11 Spearman Correlation

C**) Kendall Correlation**

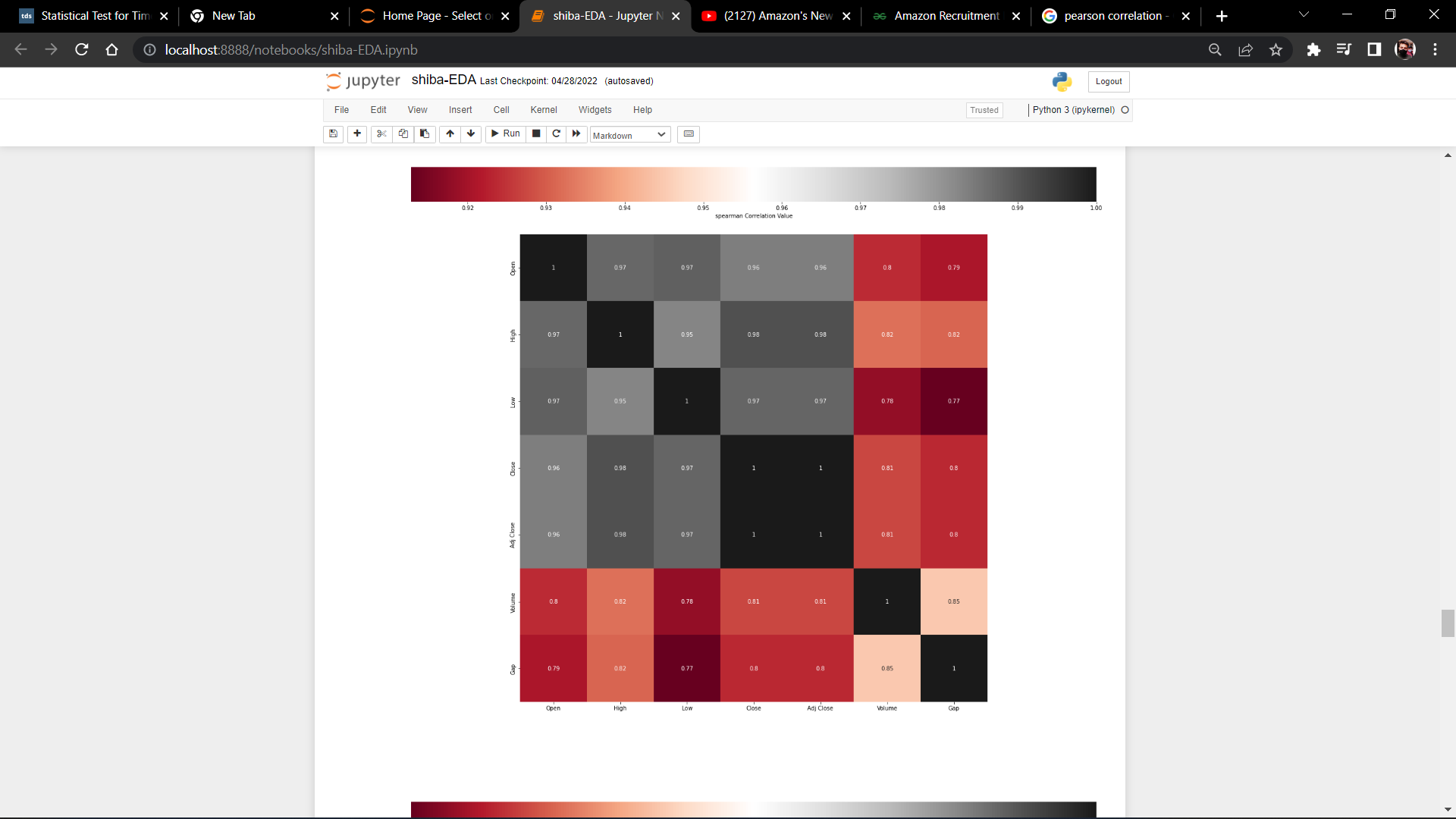


Fig. 3.12 Kendall Correlation

We can see pearson and spearman are roughly the same, but kendall is very much different. That’s because Kendall is a test of strength of dependece (one could be written as a linear function of the other), whereas Pearson and Spearman are nearly equivalent in the way they correlate normally distributed data. All of these correlations are correct in their result, it’s just that Pearson/Spearman are looking at the data in one way, and Kendall in another.

**3.2.5 Seasonal Analysis**

Time series decomposition involves thinking of a series as a combination of level, trend, seasonality, and noise components.

Decomposition provides a useful abstract model for thinking about time series generally and for better understanding problems during time series analysis and forecasting.

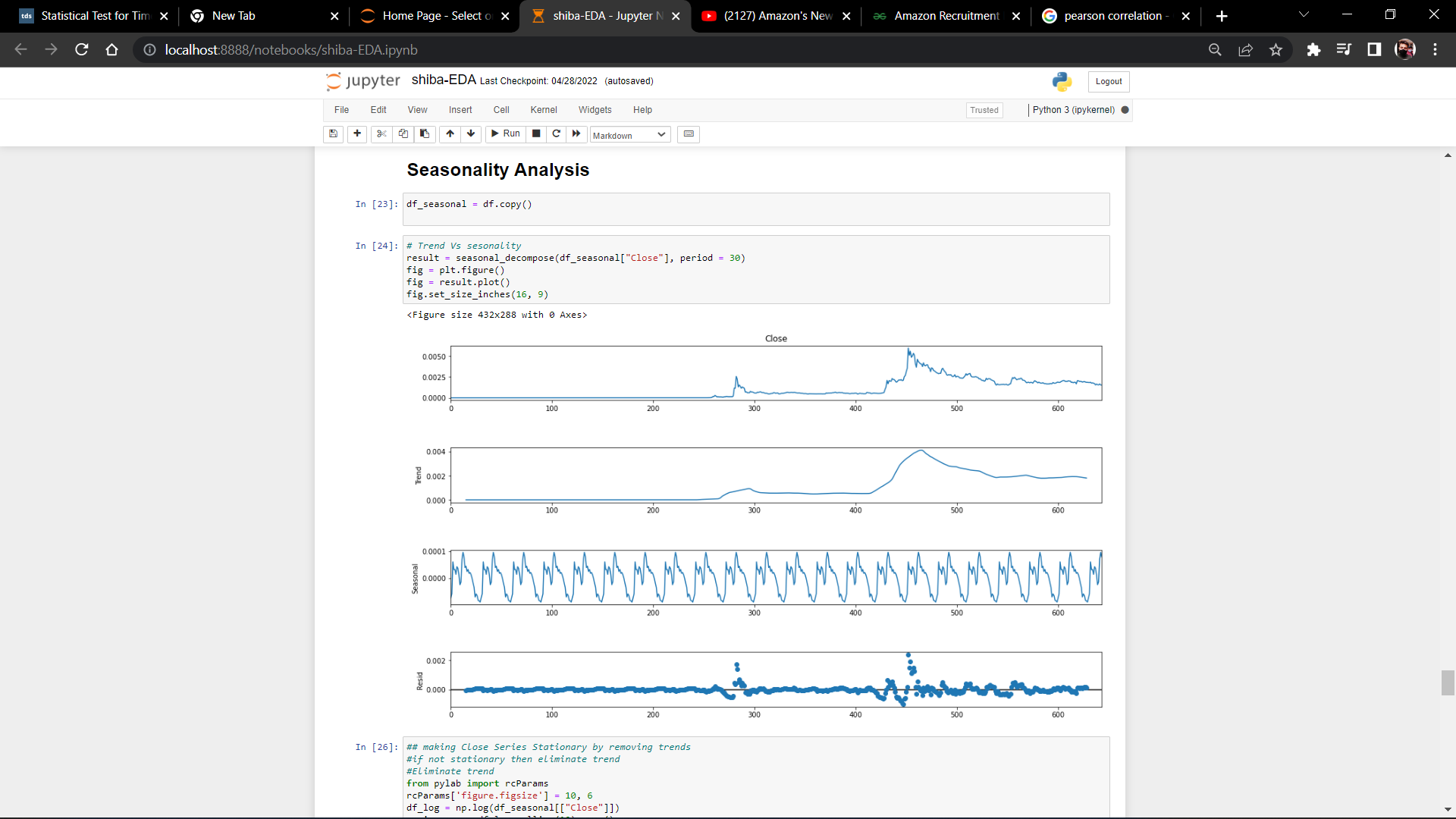


Fig 3.13 Seasonal Decompose of Closing Price

**3.2.5 Predictive Analysis**

Predictive Analysis / modeling is a commonly used statistical technique to predict future behavior. Predictive modeling solutions are a form of data-mining technology that works by anayzing historical and current data and generating a model to help predict future outcomes .

**ARIMA MODEL**

An autoregressive integrated moving average, or ARIMA, is a statistical analysis model that uses [time series data](https://www.investopedia.com/terms/t/timeseries.asp) to either better understand the data set or to predict future trends.

A statistical model is autoregressive if it predicts future values based on past values. For example, an ARIMA model might seek to predict a stock's future prices based on its past performance or forecast a company's earnings based on past periods.

An ARIMA model can be understood by outlining each of its components as follows:

* [***Autoregression (AR****)*](https://www.investopedia.com/terms/a/autoregressive.asp): refers to a model that shows a changing variable that regresses on its own lagged, or prior, values.
* ***Integrated (I):***represents the differencing of raw observations to allow for the time series to become stationary (i.e., data values are replaced by the difference between the data values and the previous values).
* [***Moving average (MA)***](https://www.investopedia.com/terms/m/movingaverage.asp): incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations.

Each component in ARIMA functions as a parameter with a standard notation. For ARIMA models, a standard notation would be ARIMA with p, d, and q, where integer values substitute for the parameters to indicate the type of ARIMA model used. The parameters can be defined as:

* p: the number of lag observations in the model; also known as the lag order.
* d: the number of times that the raw observations are differenced; also known as the degree of differencing.
* q: the size of the moving average window; also known as the order of the moving average.

[**Seasonality**](https://www.investopedia.com/terms/s/seasonality.asp), or when data show regular and predictable patterns that repeat over a calendar year, could negatively affect the regression model. If a trend appears and stationarity is not evident, many of the computations throughout the process cannot be made with great efficacy.

****

**Residual Plot :**

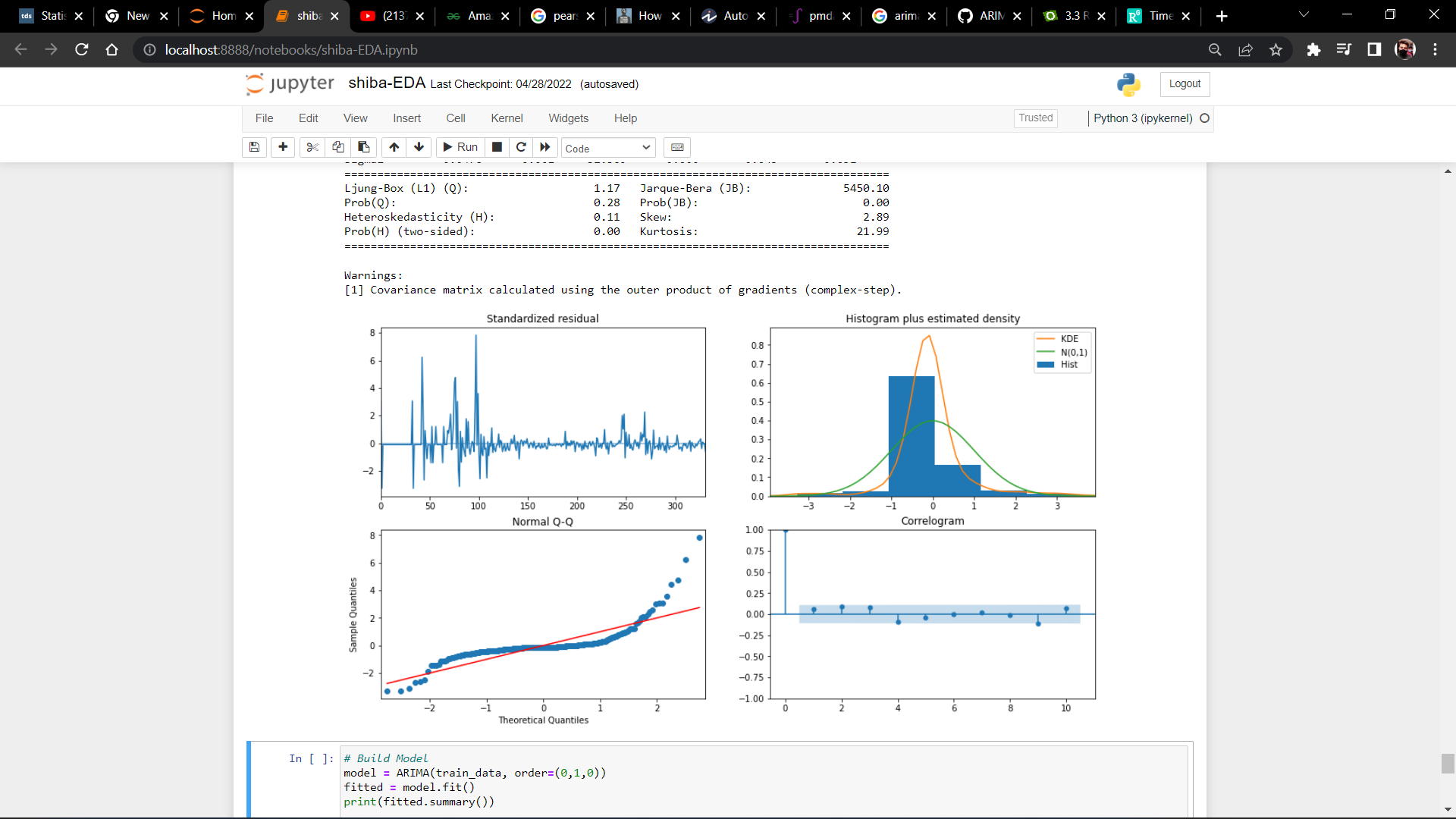


Fig 3.14 Residual Plot (Arima)

**Chapter – 4**

**RESULT**

After Performing bunch of predictive Analysis The Predictions are as follow (**ARIMA MODEL**)

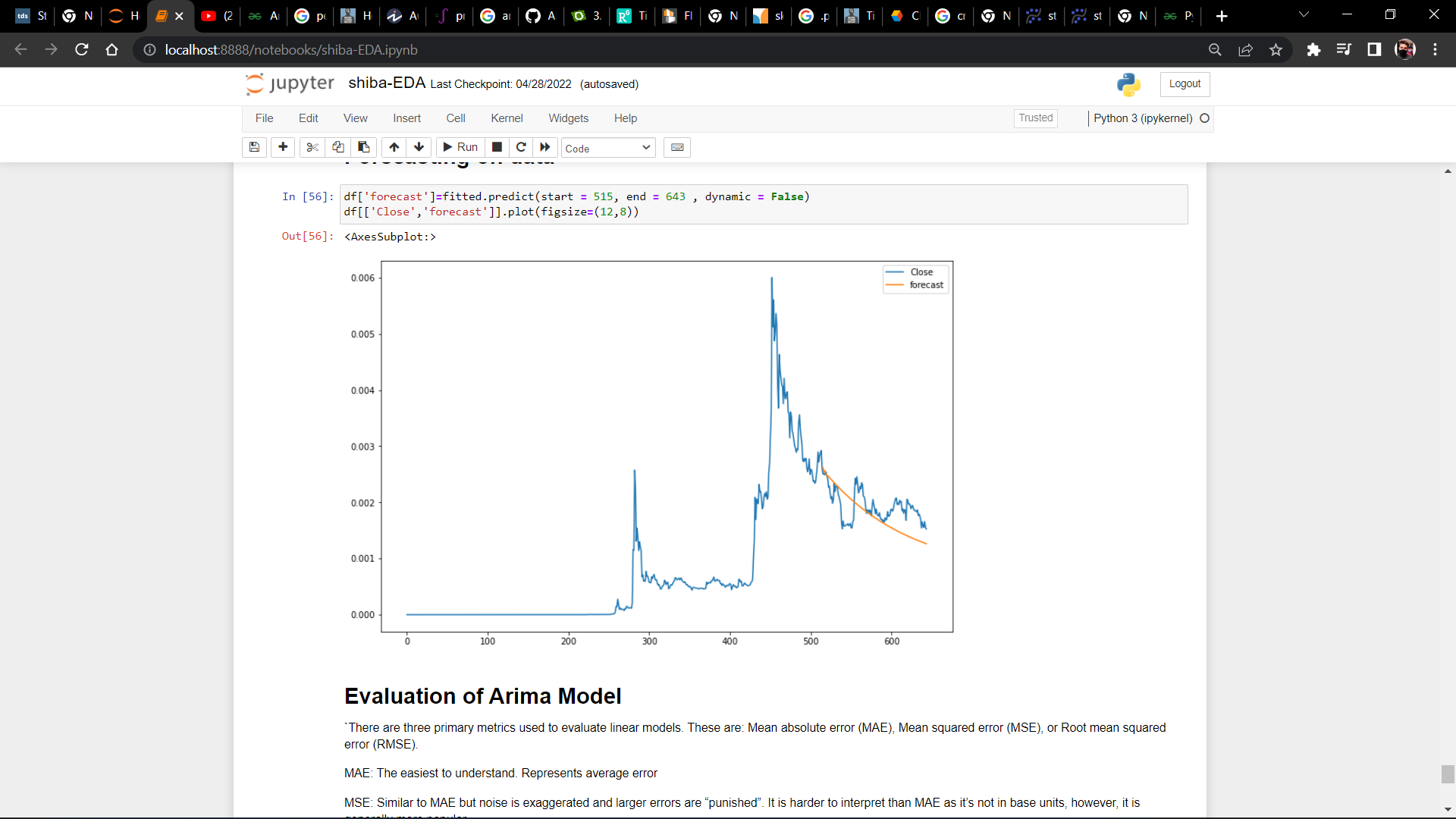


Fig 4.1 Forecast on Test Data (ARIMA)

**NOTE : The optimal p,d,q = (8,0,0)**

**MODEL EVALUATION**

## TABLE : COMPARISON BETWEEN MAE , MSE AND RMSE

|  |  |  |  |
| --- | --- | --- | --- |
|  | MAE | MSE | RMSE |
| ARIMA | 0.000308 | 0.000000 | 0.000364 |

## With the ARIMA , I was able to minimize the Error and MSE has showed 0 (Zero)

**Chapter – 5**

**OBSERVATIONS**

**OBSERVATION1 - The historical trend of Shiba inu before April2021 is kind of constant and quite dead in case of growth , Since this series contain Lots of zero value it might affect the predictive analysis hence we have to perform some Preprocessing further**

**OBSERVATION 2 : It turned out that data is actually Non- Stationary after performing Augmented Dickey-fuller test , which could effect predictive modelling/analysis , thus in future we have to perform some sort of transformation to make it stationary (i.e. Log Transformation)**

**OBSERVATION - 3 (As moving Average is Plotted so this smoothens out the noise based on mean we can clearly see it as a "Trend Spotting" tool and after May the trend is upward .**

**OBSERVATION - 4 : Here, we can see that Durbin-Watson statistics are closer to 0 (0.05). Hence, there is some positive autocorrelation to the linear model of Closing Values**

**OBSERVATION - 5 - We can see pearson and spearman are roughly the same, but kendall is very much different. That’s because Kendall is a test of strength of dependece (one could be written as a linear function of the other), whereas Pearson and Spearman are nearly equivalent in the way they correlate normally distributed data. All of these correlations are correct in their result, it’s just that Pearson/Spearman are looking at the data in one way, and Kendall in another.**

**Chapter – 6**

**CONCLUSION**

In this report we evaluated the insights gained from the Shiba Inu Historical Data, we come to an understanding that how the it is mathematically and statistically possible to make sense out of Time Series Data. We came up with lots of Observations to justify whether it’s a good asset or not, In particular the ARIMA and Auto-Arima has given good results on test data. However, comparing these mathematical approach to more general approach we can conclude that Shiba Inu Is not just about getting good prediction using machine learning , It’s a quite Volatile market and investing is quite risky because there are many more technical factor that effects the price of Shiba Inu . Considering it as a meme-coin we can also conclude that the Community and Social Media has also effect this market .

**References**

[1] Aslan & Sensoy, 2020; Baumöhl, 2019; Cerqueti et al., 2020; Corbet et al., 2019; Platanakis et al., 2018; Vidal-Tomás et al., 2019

[2] *Bitcoin prices : Eda and prediction (R2~0.99) | kaggle*. Bitcoin Analysis . (n.d.). Retrieved May 19, 2022, from https://www.kaggle.com/kaushiksuresh147/bitcoin-prices-eda-and-prediction-r2-0-99

[3] Sveneschlbeck. (2021, October 28). *Time Series visuals*. Kaggle., from https://www.kaggle.com/code/sveneschlbeck/time-series-visuals

[4] Hayes, S. (2021, June 8). *Finding seasonal trends in time-series data with python*. Medium., from https://towardsdatascience.com/finding-seasonal-trends-in-time-series-data-with-python-ce10c37aa861

[5] *Forecasting: Principles and practice (2nd ed)*. 2.8 Autocorrelation. (n.d.). Retrieved from https://otexts.com/fpp2/autocorrelation.html

[6] *How to calculate autocorrelation in python?* GeeksforGeeks. (2022, January 17). Retrieved from https://www.geeksforgeeks.org/how-to-calculate-autocorrelation-in-python/#:~:text=A%20plot%20of%20the%20autocorrelation,can%20be%20considered%20for%20modeling.

[7] *Augmented dickey-fuller test in python*. HackDeploy. from https://www.hackdeploy.com/augmented-dickey-fuller-test-in-python/

[8] Majumder, S. (2020, May 3). *Time series analysis - stationarity check using statistical test*. Medium. Retrieved from https://medium.com/analytics-vidhya/time-series-analysis-stationarity-check-using-statistical-test-f106e9045370#:~:text=them%2C%20if%20any.-,Augmented%20Dickey%20Fuller%20Test,in%20a%20time%20series%20sample.

[9] Portilla, J. M. (2018, June 10). *Using python and auto arima to forecast seasonal time series*. Medium. Retrieved from https://medium.com/@josemarcialportilla/using-python-and-auto-arima-to-forecast-seasonal-time-series-90877adff03c

[10] Khalid, I. A. (2020, May 27). *Statistical test for time series*. Medium. Retrieved from https://towardsdatascience.com/statistical-test-for-time-series-a57d9155d09b

[11] Bajaj, A. A., Aayush Bajaj Data Scientist at NatWest GroupA Data head who loves solving data problems whether it’s engineering or science. You’ll find a lot of conceptual blogs explained in the most illustrative way and framework related tutorials here. St, Bajaj, A., Data Scientist at NatWest GroupA Data head who loves solving data problems whether it’s engineering or science. You’ll find a lot of conceptual blogs explained in the most illustrative way and framework related tutorials here. Stay tuned if you want to fi, & on, F. me. (2022, May 6). *Arima & Sarima: Real-world time series forecasting*. neptune.ai. Retrieved from https://neptune.ai/blog/arima-sarima-real-world-time-series-forecasting-guide

[12] *Recurrent neural networks models¶*. Recurrent Neural Networks Models - darts documentation. (n.d.). Retrieved from <https://unit8co.github.io/darts/examples/04-RNN-examples.html>

[13] **Citation** I Djakaria and S E Saleh 2021 *J. Phys.: Conf. Ser.* **1882** 012033