Required libraries and packages:

pip install pmdarima

import os  
import warnings  
warnings.filterwarnings('ignore')  
import numpy as np  
import pandas as pd  
import matplotlib.pyplot as plt  
from statsmodels.tsa.stattools import adfuller  
from statsmodels.tsa.seasonal import seasonal\_decompose  
from statsmodels.tsa.arima\_model import ARIMA  
from pmdarima.arima import auto\_arima  
from sklearn.metrics import mean\_squared\_error, mean\_absolute\_error  
import math

Time series analysis only works with stationary data, we must first determine whether a series is stationary.

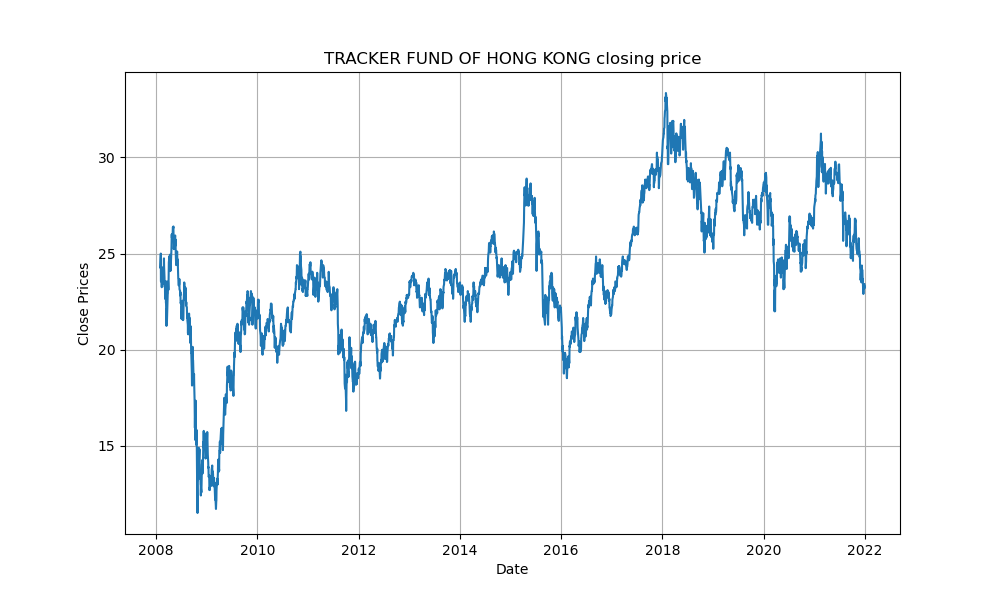


Fig 1 Closing price of 2800 against time

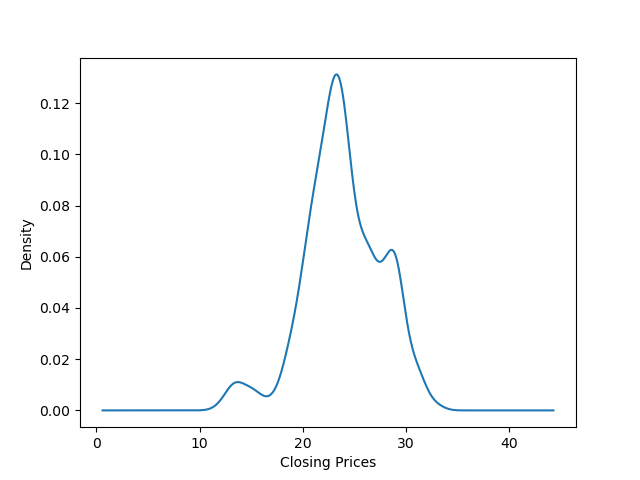


Fig 2 Distribution of Closing price

One of the most widely used statistical tests is the Dickey-Fuller test. It can be used to determine whether or not a series has a unit root, and thus whether or not the series is stationary. This test’s null and alternate hypotheses are:

Null Hypothesis: The series has a unit root (value of a =1)

Alternate Hypothesis: The series has no unit root.

If the null hypothesis is not rejected, the series is said to be non-stationary. The series can be linear or difference stationary as a result of this.

Results of dickey fuller test

Test Statistics -2.438352

p-value 0.131203

No. of lags used 2.000000

Number of observations used 3429.000000

critical value (1%) -3.432258

critical value (5%) -2.862383

critical value (10%) -2.567219

dtype: float64

We can’t rule out the Null hypothesis because the p-value is bigger than 0.05. Additionally, the test statistics exceed the critical values. As a result, the data is nonlinear. The series is stationary if both the mean and standard deviation are flat lines (constant mean and constant variance).

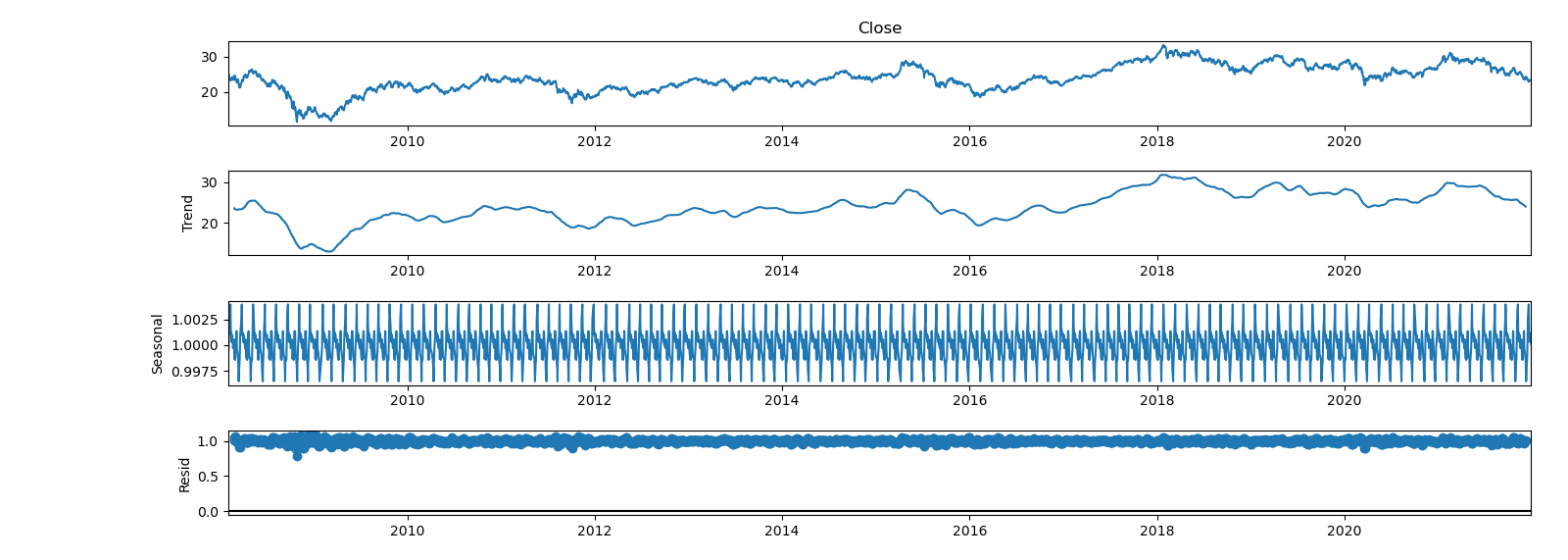


Fig 3 The pattern repeats in the third graph, which implies the existence of seasonality in the 2800 closing price data.

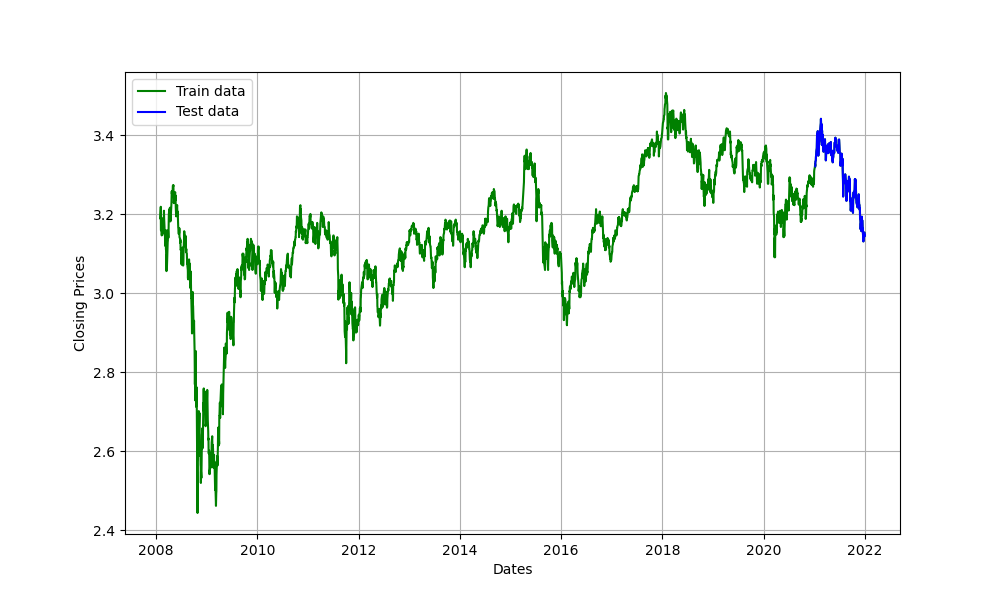
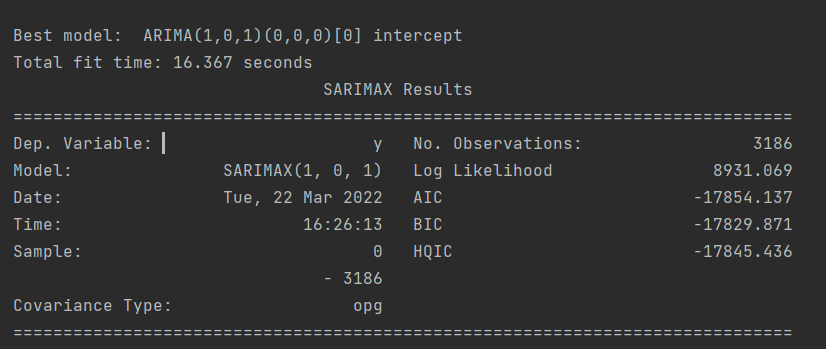
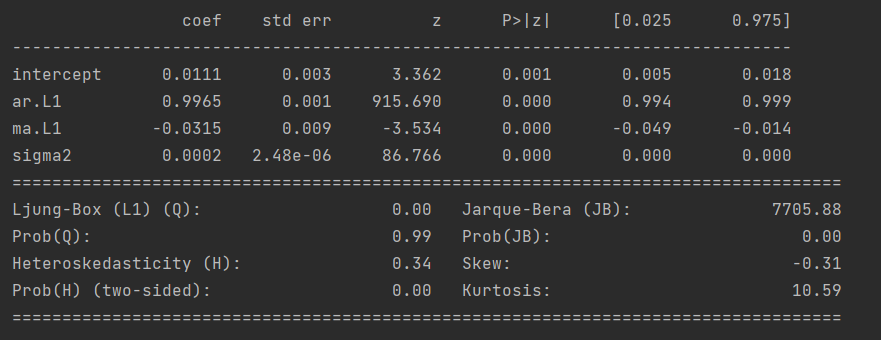
To reduce the magnitude of the values and the growing trend in the series, we first take a log of the series. We then calculate the rolling average of the series after obtaining the log of the series. A rolling average is computed by taking data from the previous 12 months and calculating a mean consumption value at each subsequent point in the series. 

Fig 4. Plot of log(closing\_price) against time

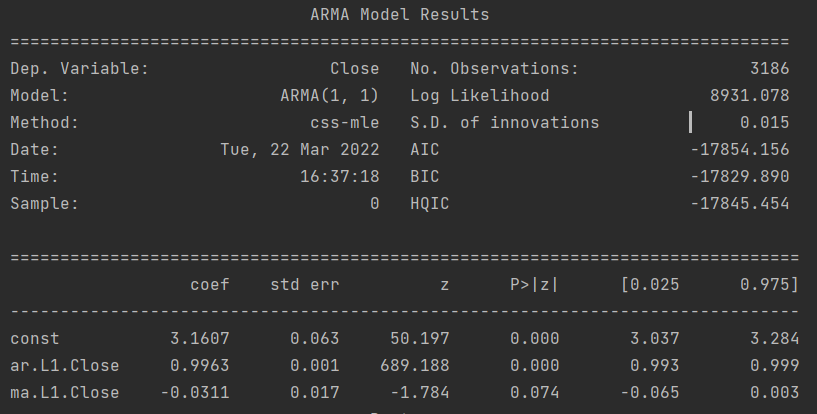
We’ll utilize Auto ARIMA to find the best parameters without looking at the ACF and PACF graphs.

The best model is found to be ARIMA(1,0,1).





Now we train the model using the training data and the model parameters are as follows:



Finally we test the model on the testing data and obtain the following result:

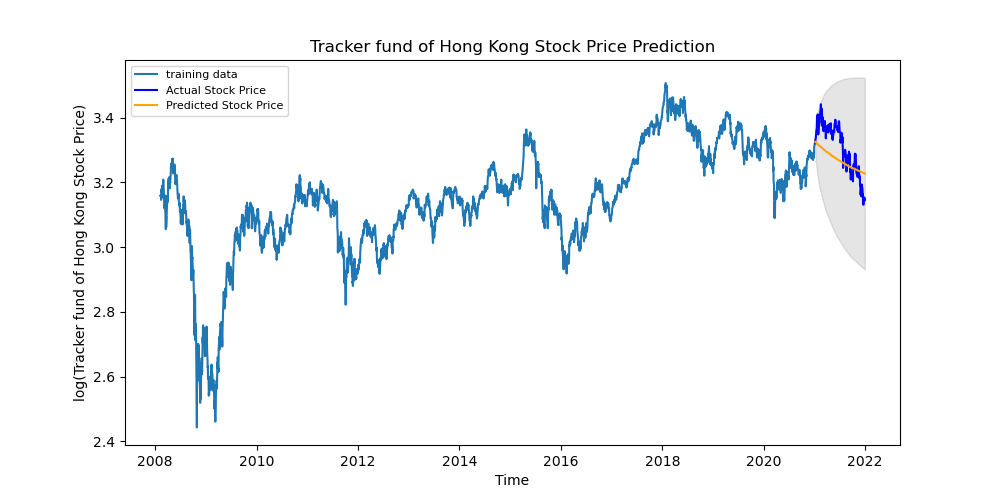


Fig 5. Plot of log(closing\_price) against time, the orange line indicates the model prediction

The prediction is evaluated using the testing data. MSE is found to be 0.0043867 which is acceptable.

