STOCK ANALYSIS USING PYTHON

Module: MIS41110 – Programming for Analytics

Assignment: Business Analytics Assignment: Individual Project

Lecturer: Darren Redmond & Dr. Miguel Nicolau

Course: MSc Business Analytics

Date Submitted: 6th December 2020

Submitted By: Vaibhav Sharma - 20200516

USER MANUAL – STOCK ANALYSIS USING PYTHON

Overview

This is a user guide for using the stock analysis using the python program. As a user one can find detailed information about executing the program efficiently and in a hassle-free manner. The users should go through this user guide to effectively use the program for stock analysis. This user manual covers a step-by-step guide to run the program and obtain the desired analysis of a stock. All high-level to low-level aspects of the program are mentioned in this document.

TABLE OF CONTENTS

1. INTRODUCTION

In 1987, the era of digitalization of the stock market began in America when the SEC declared that electronic communication will enhance the security, accuracy, and speed of stock market trading. Since then there have been tremendous advancements in terms of involvement of technology for stock trading across almost all major global exchanges in the world. From floor to phone, the stock market is now accessible to common people and the only prerequisite is to own a decent digital communications device. In this era of utmost technological advancements and advanced analytics, many traders and investors still depend on guts to make a decision, which is neither based on the historical data or study of the desired stock. The use of analytics is not only highly recommended but also a tried and tested method of investment in the stock market. This program aims to provide the user with the most accurate trend analysis, the study of stock behavior, and calculated forecasting. It also enables users personalized selection, a wide range of visualization options, and real-time KPI indicators with the best possible predictions. Let's see how this program works, it's usability, and features.

2. PURPOSE

This program has been developed to simplify the analysis of a stock belonging to NYSE and NASDAQ. The development of this program has been done while keeping in mind the ease of use and clarity. The program serves the purpose of providing a unified platform to the users where all aspects of stock analysis can be studied. The program is having data updated at 12:00 AM New York, NY, USA (GMT-5) everyday, this enables the user to get real-time data till the last day of trading from the date of use.

3. AUDIENCE

Stock traders, stock investors, trading enthusiasts, business analysts, and students who want to study stocks are the target audience of this program. The ease of use, clear and presentable data, loaded functions, trend analysis, and prediction with low error margin make it suitable for all kind of audience pool. Also, this program keeps in mind the individuals who want to work in the field of analytics and finance as it provides a good reference for the use of python and its features and libraries extensively.

4. PREREQUISITES

1) Operating System - Windows 8 or higher,64-bit / MAC OS 10.11 or higher, 64-bit / RHEL 6/7, 64-bit

2) Processor - x86 64-bit CPU (Intel / AMD architecture)

3) RAM – 4 GB or more

4) Storage Space – At least 5 GB free disk space

5) Python 3.9.0 release date 5th October 2020 – Added to the system path

6) IDE – like PyCharm (preferred), Eclipse, Spyder, IDLE, ATOM, NetBeans, etc.

7) Monitor Resolution – 1024x768 (for graphs and visualization)

8) Launcher – add the IDE launchers dir to the path (if not added)

9) Required Python Libraries (also mentioned in the code as per each module use)

    9.1) import pandas

    9.2) import datetime

    9.3) import pandas_datareader

    9.4) import matplotlib.pyplot

    9.5) from pandas_datareader._utils import RemoteDataError

    9.6) import warnings

    9.7) from pandas.core.common import SettingWithCopyWarning

    9.8) import tkinter

    9.9) from tkinter import *

    9.10) import numpy

    9.11) from mplfinance.original_flavor import candlestick_ohlc

    9.11) import matplotlib.dates

    9.12) from tabulate import tabulate

    9.13) from tkinter import messagebox

    9.14) from sklearn import preprocessing

    9.15) from sklearn.tree import DecisionTreeRegressor

    9.16) from sklearn.model_selection import train_test_split

    9.17) from sklearn.linear_model import LinearRegression

    9.18) from sklearn import metrics

    9.19) import unittest

10) Internet Connection established and running.

Apart from this, it is expected that the user has a basic knowledge of how to run a python program and input the desired values as demanded by the program in the CLI as well as the GUI version.

5. PROGRAM MODULES

The program is divided into 6 modules which are described below: -

1) **CLI_Version: -** This module is created to run the program in the IDE's CLI. All the inputs will be taken in the console and the results will be printed in the console. The graphs will open in a new window created by the matplot library.

2) **GUI Version: -** This module is created to run the program in a user interface created by a python library called Tkinter. All the inputs will be taken in an easy-to-use user interface window and outputs will be generated in the user interface window as well.

3) **Descriptive_Analysis: -** This module contains all the functions responsible for the descriptive analysis of the stock. This module is used by the GUI_Version and CLI_version for descriptive analysis as per the user's selection.

4) **Predictive_Analysis: -** This module contains all the functions responsible for the predictive analysis of the stock. This module is used by the GUI_Version and CLI_version for descriptive analysis as per the user's selection.

5) **Error_Messagebox: -** This module is consumed by GUI_Version and contains error boxes that are displayed if the user enters any incorrect or unacceptable input value as per it's definition in the program.

6) **Unit_Test: -** This is an independent module and is not consumed by either GUI_Version or CLI_Version rather is made for the purpose of unit testing i.e. testing done by the developer to check the feasibility of a function.

6. HOW TO EXECUTE THE GUI VERSION

STEP 1 – RUN THE GUI_Version MODULE

The user should run the GUI_Version module and within some seconds a window will appear demanding inputs from the user to fetch the information about the desired stock, date range, etc. The user is requested to fill all inputs to get an output.

This is the user interface window which will open once the mentioned module is run. The next step is to input the values as demanded by the interface one by one.

STEP 2 – ENTER THE START DATE

The user is requested to enter the start date in YYYY-MM-DD format. For example, a valid date is '2020-08-20'. Just in case the user enters the date in the wrong format, after clicking the submit button after entering all information the program will prompt an error of wrong input.
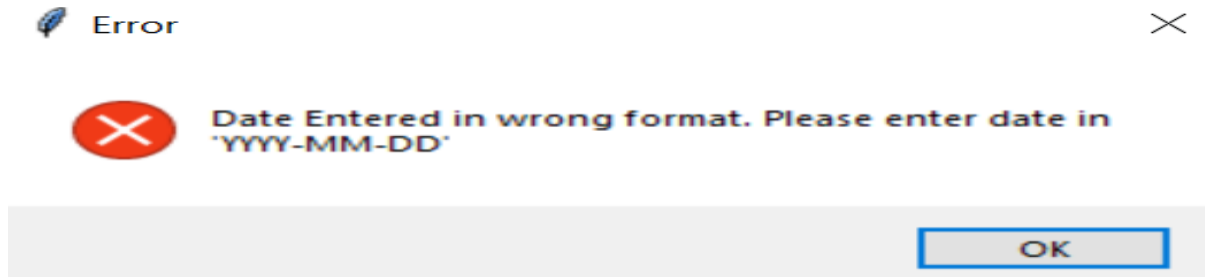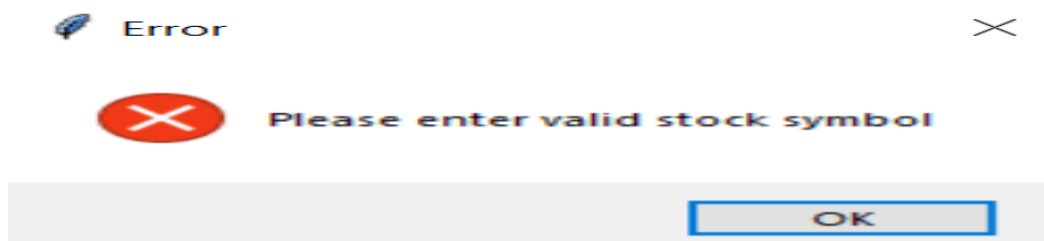

STEP 3 – ENTER THE END DATE

The user is requested to enter the end date in YYYY-MM-DD format. For example, a valid date is '2020-11-20'. Just in case the user enters the date in the wrong format, after clicking the submit button after entering all information the program will prompt an error of wrong input.
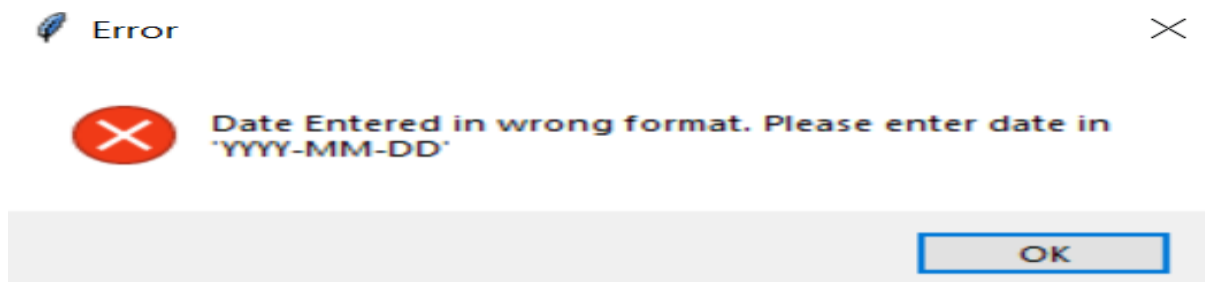
STEP 4 – ENTER THE STOCK SYMBOL

As the program fetches data using yahoo API of NYSE and NASDAQ stocks, the user is expected to enter a valid stock symbol belonging to NYSE and NASDAQ stock exchanges in either lower or upper case. Some examples of valid stock symbols are, 'AAPL', 'TSLA', 'GOOGL', 'MSFT', 'AMZN', 'AAL', 'NVDA', 'ABBV', 'BRK.A'. Just in case the user enters the wrong stock symbol, after clicking the submit button after entering all information the program will prompt an error of wrong input.

STEP 5 – ENTER THE NUMBER OF DAYS FOR PREDICTION

The user should input the number of days for which the prediction is required. This will provide the prediction graph for n (input by the user) days after the end date entered by the user. It will also provide the predicted price on the nth (input by the user) day after the end date provided by the user earlier. As the input should be an integer, if the user provides the wrong input, after clicking the submit button after entering all information the program will prompt an error of wrong input.

STEP 6 – ENTER THE NUMBER OF DAYS FOR CALCULATING MOVING AVERAGE

The user should input the number of days for moving average calculation. As the input should be an integer, if the user provides the wrong input, after clicking the submit button after entering all information the program will prompt an error of wrong input.

STEP 7 – CHOOSE THE PRICE FOR WHICH ANALYSIS IS NEEDED

As the stock has various prices like open, close, etc. The user has the option to choose on which he wants the analysis to be carried out.

STEP 8 – CHOOSE THE DISPLAY OPTION

Here the user should choose the display option. The details of the display will be provided after the completion of all steps.

STEP 9 – CHOOSE THE REGRESSION TYPE

This is where the user can choose the type of method to be used for prediction. There are two options either linear regression or decision tree regression. Based on these methods the predicted price will be shown a relevant graph will appear.

STEP 10 – CHOOSE THE FRACTION OF DATA TO BE USED FOR TRAINING THE REGRESSION MODEL

This enables the user to select what fraction of data should be trained to predict the price as per the user's selection.

OUTPUT

Based on the user's selection in step 8, the output will be displayed. Going choice by choice: -

1) **Descriptive Analysis** – This option will display the user the mean, median, mode, range, deviation, variance, quartiles, and IQR for the selected stock, date range, and price.

2) **Stock Information** – This option will display the user all price comparison time-series graphs, volume bar chart, Daily return chart, and linear trend chart for the selected stock, date range, and price.
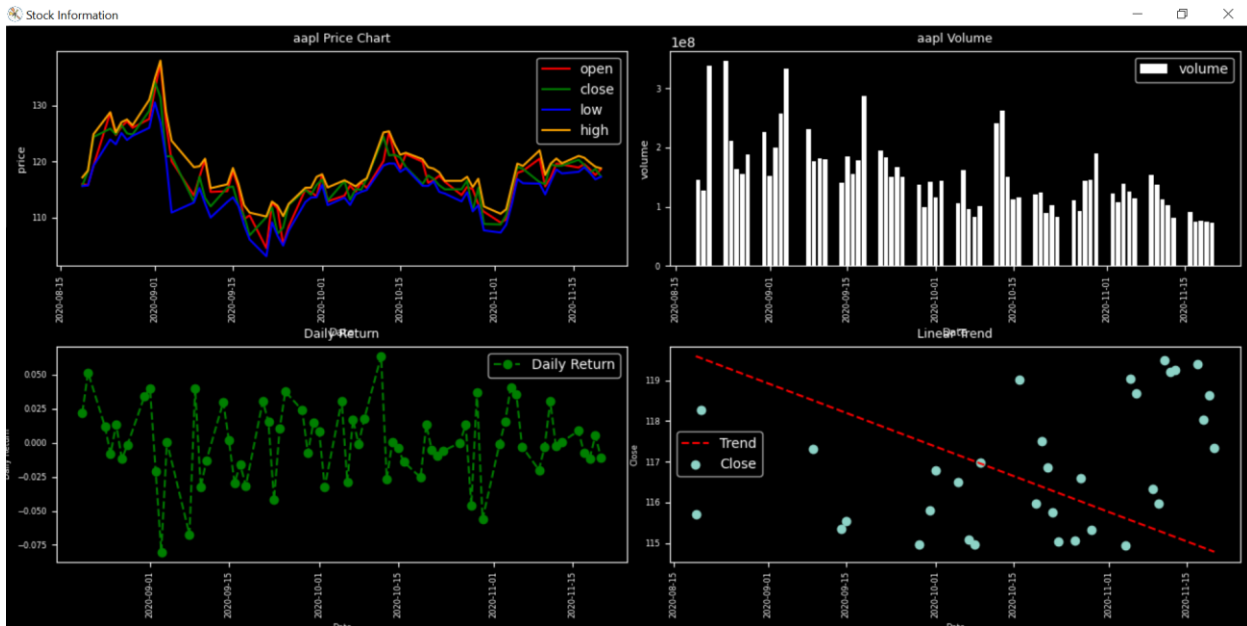
3) **Predictive Analysis** – This option will display the user firstly the predicted price after nth (input by the user in step 5), MAE, MSE, RMSE, R-square value for the selected stock, date range, price, regression method selected (9), m training days (input by the user in step 10) in a message box. After clicking "OK" in the message box user will see the graph for prediction using the selected method of regression.
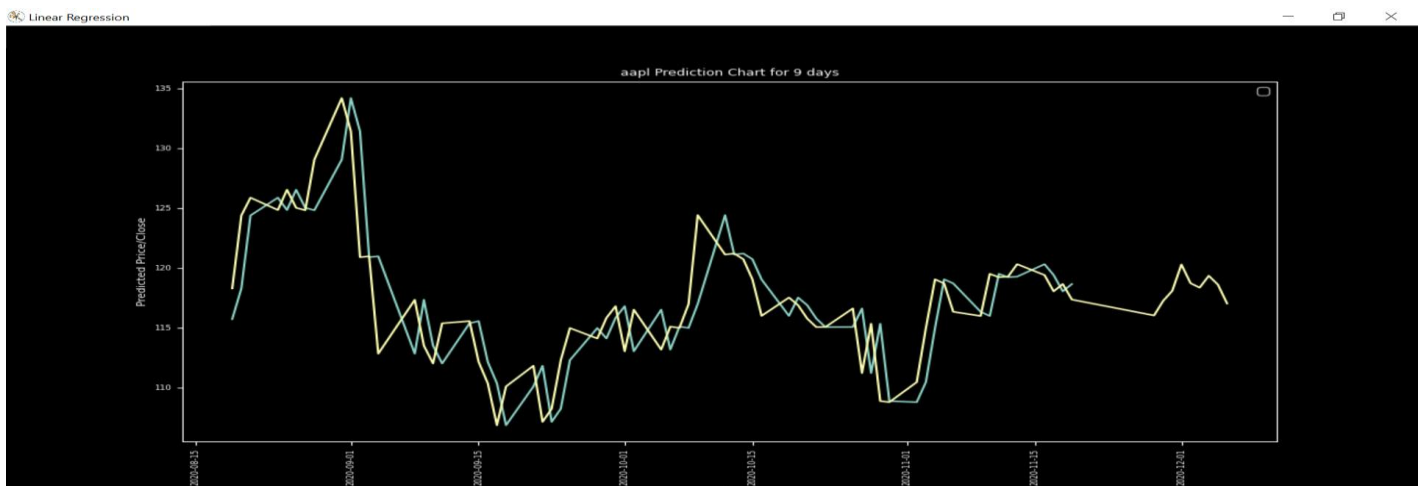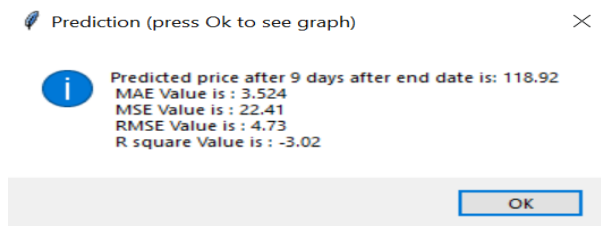
4) **CandleStick Chart** – This option will display the user a candlestick chart for the selected stock, date range, and price.
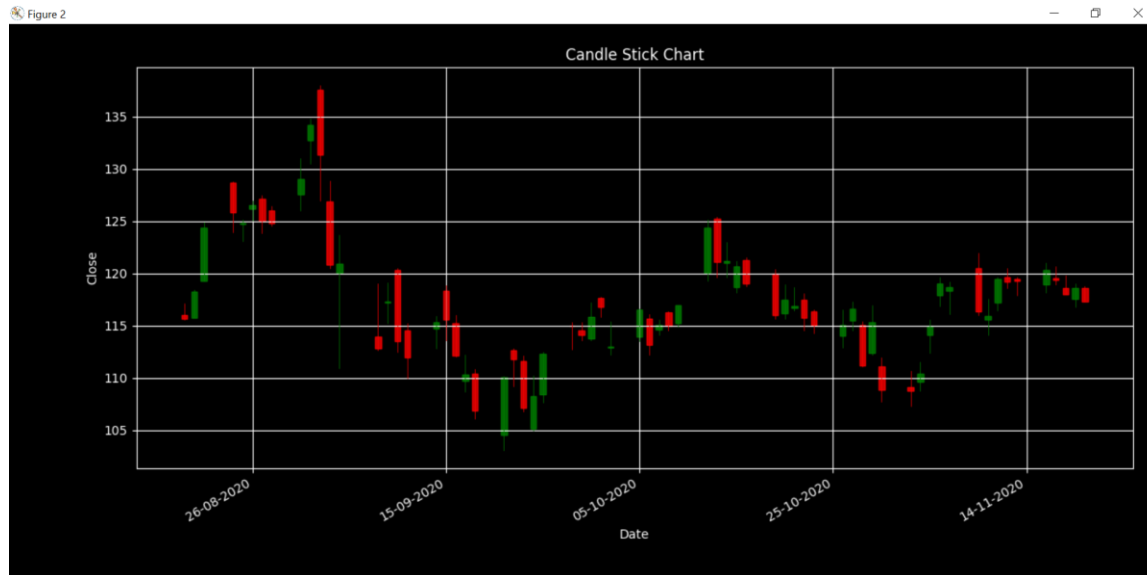
5) **Series Graphs** – This option will display user series graphs such as raw time series, moving average, weighted average, MACD for the selected stock, date range, price, and the number of days selected for moving average (input by the user in step 6).
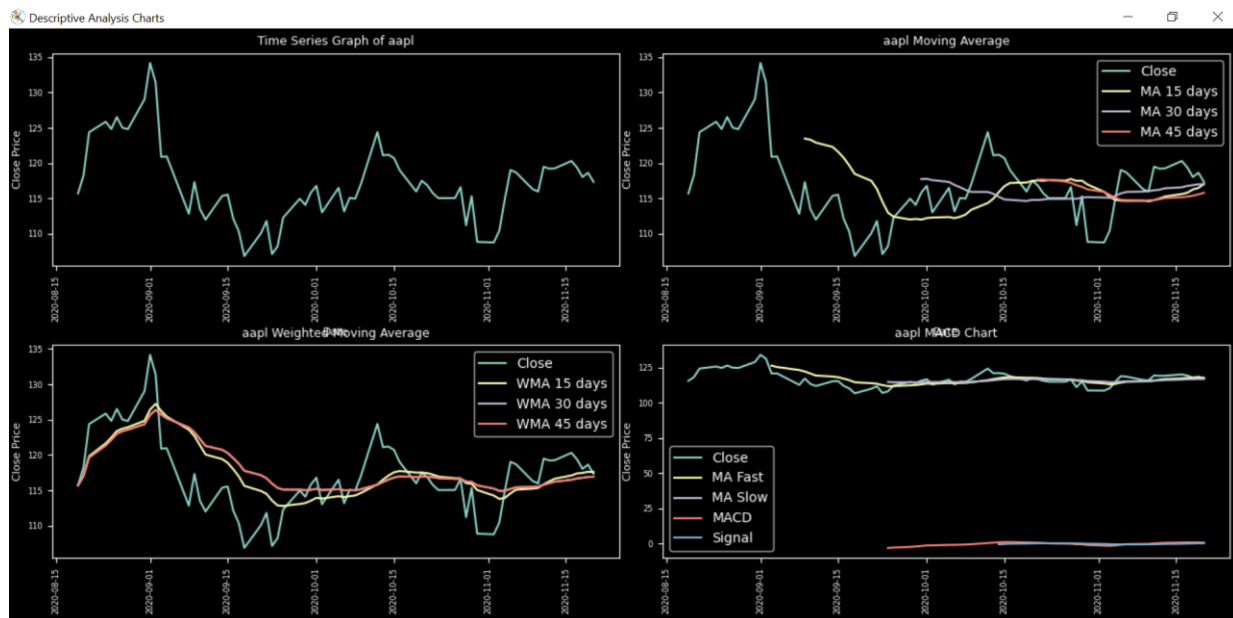
NOTE: The user can reuse the input window to select other display options and change input value after closing the pre obtained data/graph. This cannot be done in CLI_Version but GUI enables this option for the users.

7. HOW TO EXECUTE THE CLI VERSION

STEP 1 – RUN THE CLI_Version MODULE

The user should run the CLI_Version module and one by one in the console the program will demand inputs from the user to fetch the information about the desired stock, date range, etc. The user is requested to fill all inputs one by one to get an output.

STEP 2 – ENTER THE START DATE

The user is requested to enter the start date in YYYY-MM-DD format. For example, a valid date is '2020-08-20'.

```
Enter start date (YYYY-MM--DD):
```

Just in case the user enters the date in the wrong format, after pressing enter the program will print an error of wrong input and the user will be requested to fill the input again.

```
Entered date is wrong. Please re-enter
Enter start date (YYYY-MM--DD):  |
```

STEP 3 – ENTER THE END DATE

The user is requested to enter the end date in YYYY-MM-DD format. For example, a valid date is '2020-08-20'.

```
Enter end date (YYYY-MM--DD):   |
```

Just in case the user enters the date in the wrong format, after pressing enter the program will print an error of wrong input and the user will be requested to fill the input again.

```
Entered date is wrong. Please re-enter
Enter end date (YYYY-MM--DD):   |
```

STEP 4 – ENTER THE STOCK SYMBOL

As the program fetches data using the yahoo API of NYSE and NASDAQ stocks, the user is expected to enter a valid stock symbol belonging to NYSE and NASDAQ stock exchanges in either lower or upper case. Some examples of valid stock symbols are, 'AAPL', 'TSLA', 'GOOGL', 'MSFT', 'AMZN', 'AAL', 'NVDA', 'ABBV', 'BRK.A'.

```
Enter the Stock Symbol: |
```

Just in case the user enters the wrong stock symbol, after pressing enter the program will print an error of wrong input and the user will be requested to fill the input again.

```
Wrong Stock Symbol. Please re-enter
Enter the Stock Symbol: |
```

STEP 5 – ENTER THE NUMBER OF DAYS FOR PREDICTION

The user should input the number of days for which the prediction is required. This will provide the prediction graph for n (input by the user) days after the end date entered by the user. It will also provide the predicted price on the nth (input by the user) day after the end date provided by the user earlier.

```
Enter the number of days for prediction: |
```

As the input should be an integer, just in case the user enters the wrong input, after pressing enter the program will print an error of wrong input and the user will be requested to fill the input again.

```
 Prediction Days entered is wrong. Please enter again.
 Enter the number of days for prediction: |
```

STEP 6 – ENTER THE NUMBER OF DAYS FOR CALCULATING MOVING AVERAGE

The user should input the number of days for moving average calculation.

```
Enter the number of days for MA: |
```

As the input should be an integer, just in case the user enters the wrong input, after pressing enter the program will print an error of wrong input and the user will be requested to fill the input again.

```
 MA days entered is wrong. Please enter again.
 Enter the number of days for MA: |
```

STEP 7 – CHOOSE THE PRICE FOR WHICH ANALYSIS IS NEEDED

As the stock has various prices like open, close, etc. The user has the option to enter the choice as asked, on which he wants the analysis to be carried out.

```
Price selection; enter 1 for Close, 2 for High, 3 for Low, 4 for Open : |
```

STEP 8 – CHOOSE THE FRACTION OF DATA TO BE USED FOR TRAINING THE REGRESSION MODEL

This enables the user to enter the choice as asked to select what fraction of data should be trained to predict the price as per the user's selection.

```
fraction of training data for prediction; enter 1 for 0.25, 2 for 0.30, 3 for 0.50: |
```

STEP 9 – CHOOSE THE REGRESSION TYPE

This is where user input is asked to choose the type of method to be used for prediction. There are two options either linear regression or decision tree regression. Based on these methods the predicted price will be shown a relevant graph will appear

```
Regression Selection; enter 1 for Linear Regression, 2 for Decision Tree Regression: |
```

STEP 10 – CHOOSE THE DISPLAY OPTION

Here the user should input as asked to choose the display option. The details of the display will be provided after the completion of all steps.

```
Display selection; enter 1 for Descriptive Analysis,2 for Stock Information
 3 for Predictive Analysis, 4 for Series Graph
 5 for Candlestick Chart : |
```

OUTPUT

Based on the user's selection in step 10, the output will be displayed. Going choice by choice: -

1) **Descriptive Analysis** – This option will display the user the mean, median, mode, range, deviation, variance, quartiles, and IQR for the selected stock, date range, and price in the output console.

```
----- ---------
count   67
mean   117.149
std     5.54032
min    106.84
Q1     113.79
Q2     116.5
Q3     119.44
max    134.18
Range  27.34
IQR     5.65
----- ---------
```

2) **Stock Information** – This option will display the user all price comparison time-series graphs, volume bar chart, Daily return chart, and linear trend chart for the selected stock, date range, and price.
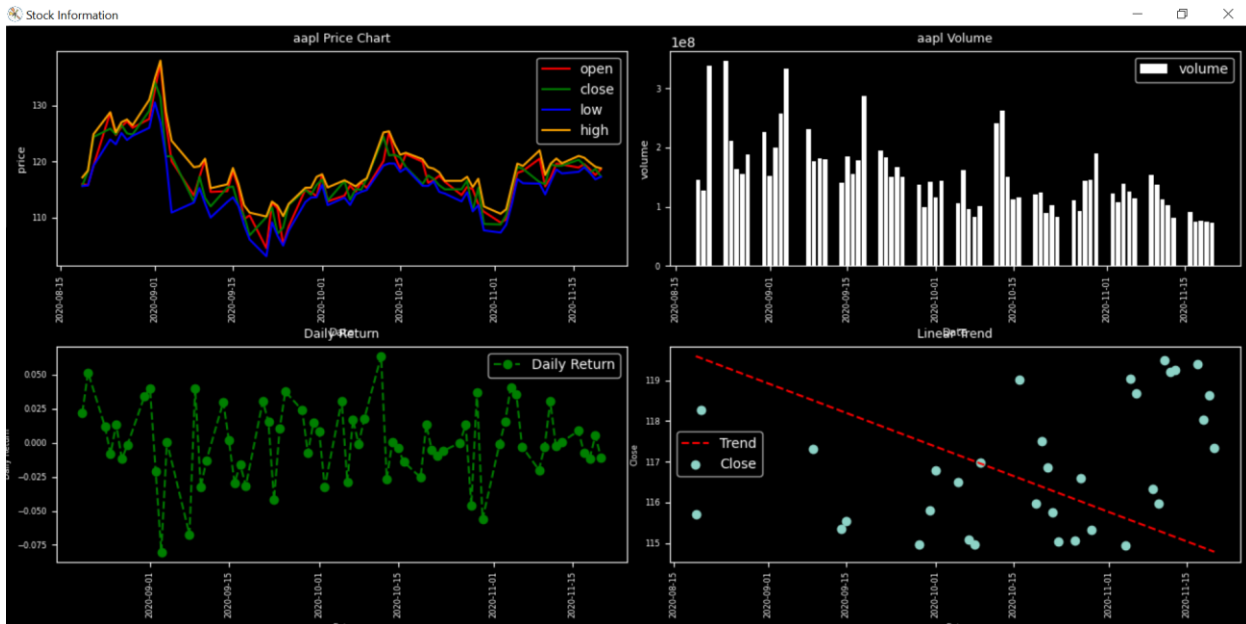
3) **Predictive Analysis** – This option will display the user firstly the predicted price after nth (input by the user in step 5), MAE, MSE, RMSE, R-square value for the selected stock, date range, price, regression method selected (9), m training days (input by the user in step 8) in output console and simultaneously graph for prediction using selected method of regression will appear in a window.

The predicted price after 15 days after the end date is: 118.13

MAE Value is : 2.933

MSE Value is : 17.211

RMSE Value is : 4.15

R square Value is : 0.03
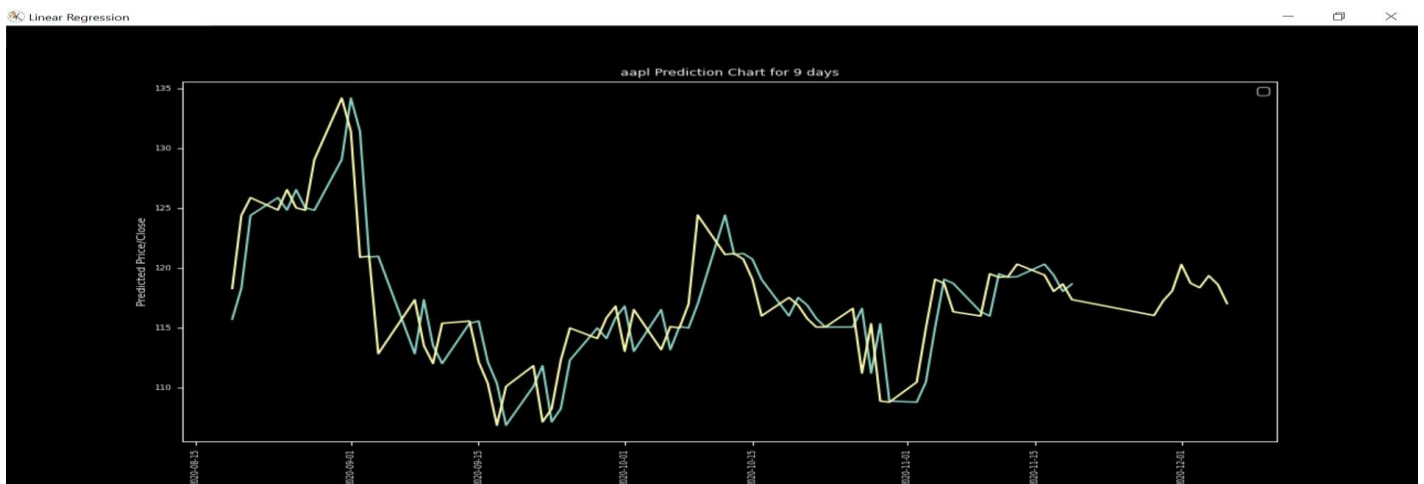
4) **CandleStick Chart** – This option will display the user a candlestick chart for the selected stock, date range, and price.
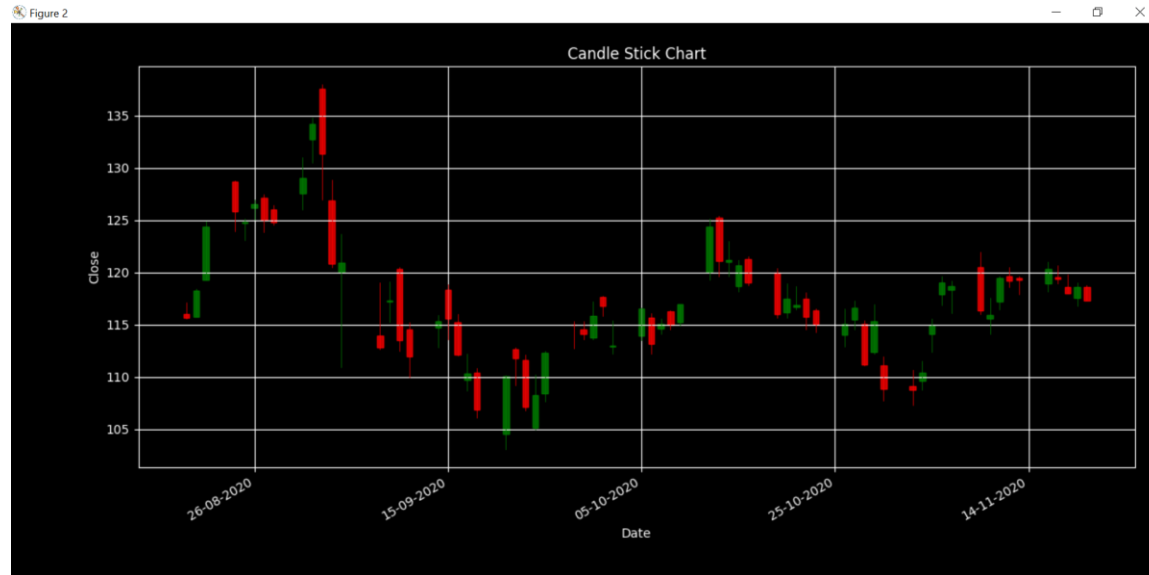
5) **Series Graphs** – This option will display user series graphs such as raw time series, moving average, weighted average, MACD for the selected stock, date range, price, and the number of days selected for moving average (input by the user in step 6).

8. HIGH-LEVEL UML

```mermaid
flowchart TD
    START((START))
    START --> A[ENTER START DATE]
    A --> B{TRY CATCH BLOCK}
    B -->|INVALID| A
    B -->|VALID| C[ENTER END DATE]
    C --> D{TRY CATCH BLOCK}
    D -->|INVALID| C
    D -->|VALID| E[ENTER STOCK SYMBOL]
    E --> F{TRY CATCH BLOCK}
    F -->|INVALID| E
    F -->|VALID| G[ENTER NUMBER OF DAYS FOR PREDICTION]
    G --> H{TRY CATCH BLOCK}
    H -->|INVALID| G
    H -->|VALID| I[ENTER DAYS FOR MOVING AVERAGE]
    I --> J{TRY CATCH BLOCK}
    J -->|INVALID| I
    J -->|VALID| K[CHOOSE PRICE SELECTION this passes the selection to defined funtions]
    K --> L[CHOOSE FRACTION OF DATA FOR DATA TRANING this passes the selected option to defined funtions]
    L --> M[SELECT THE REGRESSION TYPE OPTION passes the option to the defined funtion]
    M --> N{OPTION 1? OR 2?}
    M --> O[SELECT THE DISPLAY OPTION]
```

START

ENTER START DATE

TRY CATCH BLOCK — INVALID / VALID

ENTER END DATE

TRY CATCH BLOCK — INVALID / VALID

ENTER STOCK SYMBOL

TRY CATCH BLOCK — INVALID / VALID

ENTER NUMBER OF DAYS FOR PREDICTION

TRY CATCH BLOCK — INVALID / VALID

ENTER DAYS FOR MOVING AVERAGE

TRY CATCH BLOCK — INVALID / VALID

CHOOSE PRICE SELECTION (this passes the selction to defined funtions)

CHOOSE FRACTION OF DATA FOR DATA TRANING (this passes the selected option to defined funtions)

SELECT THE REGRESSION TYPE OPTION (passes the option to the defined funtion)

OPTION 1? OR 2?

SELECT THE DISPLAY OPTION

OPTION 1,2,3,4,5?

1 — CALLS Descriptive_Analysis_Terminal FUNCTION FROM Descriptive_Analysis MODULE

2 — CALLS stock_Info FUNCTION FROM Descriptive_Analysis MODULE

CALLS linear_Regression_Terminal FUNCTION FROM Predictive_Analysis MODULE

CALLS decisionTree_Regression_Terminal FUNCTION FROM Predictive_Analysis MODULE

4 — CALLS series_graphs FUNCTION FROM Descriptive_Analysis MODULE

5 — CALLS candlestick_Chart FUNCTION FROM Descriptive_Analysis MODULE

END