ISGB 79AA – Advanced Python for Financial Programming
Assignment 5 – Stock Price Statistics

I) Summary

For this assignment, you will download data from Yahoo Finance on a stock's daily trading measures (prices, volume) as well as that of the S&P500 index, and visualize and analyze this data in Python with these packages: pandas, MatPlotLib, and StatsModels or NumPy.

II) Notes

You will analyze stock trading data for a particular company of interest that you select.

In finance.yahoo.com, each company's stock has a "ticker" which is a short unique identifier. For example, Apple's stock's ticker is AAPL. In the descriptions below, XYZ should be replaced with the ticker of your company. (Your company should be other than AAPL or MSFT.) Tickers are also defined for funds and indexes, for example, the S&P500 index's ticker is ^GSPC.

III) Assignment Tasks

A) Data Preparation
A.1) From Yahoo Finance's Historical Data feature, download .csv files of daily data for the past 5 years – one .csv for your stock and 1 .csv for the S&P500. The measures include Date, Open, High, Low, Close, Adj Close, and Volume. See appendix A for information on data access using Yahoo Finance.

B) Price Time Series – Stock
Background: https://www.investopedia.com/terms/a/adjusted_closing_price.asp
B.1) Create a line chart for the adjusted closing prices of your stock.

C) Price Time Series – Compare
C.1) Create a single chart with both time series for the adjusted closing prices of your stock and the closing prices of the S&P500 (using left and right Y axes).
C.2) Over the period of the chart, did your stock outperform or underperform the S&P500?

D) Analysis - Stock Beta
Background: https://www.investopedia.com/terms/b/beta.asp
D.1) Create a Scatter Plot for the daily returns of your stock vs. the daily returns of the S&P500. See Appendix B for the definition of daily return. The S&P500 returns should be the X axis. The Scatter Plot should include daily returns from 2018 and 2019 (filter out data from before 2018 and from 2020).
D.2) For the daily returns in part D.1 (2018-2019), perform a linear regression between the S&P500 daily returns (independent variable) and your stock's daily returns (dependent variable). What is the equation of the regression line, and the R-squared? What is the slope of the line (this is the beta of the stock against the S&P500)?

E) Analysis – Outliers
E.1) For your company, identify the two dates of the largest (most positive) and smallest (most negative) daily price returns. For these two dates, report the date, the date's adjusted price, and the daily return. (Not required: was there a company-specific event that occurred on that date?)

F) Analysis – Price Move Distribution
background: https://www.investopedia.com/articles/investing/102014/lognormal-and-normal-distribution.asp
For this section, use one year of data (from 2019).
F.1) Create a Histogram chart showing the distribution of daily price returns for XYZ.
F.2) Report the mean and standard deviation of the daily price returns for your stock.
F.3) Determine whether the daily returns for XYZ follow a normal distribution (appendix C).

IV) Submission

Your submission to this assignment should be to Blackboard, with these attachments:

1) LastnameFirstnameAsn5.ipynb – your Python analyses, include code, output, and plots. You can also include your answers to the specific questions in your Notebook.
2) LastnameFirstnameAsn5.docx – If you want to put your answers to questions in a Word document, please attach that as well.
3) .csv(s) – attach the .csv's that your Notebook loads.

Appendix A – Downloading Data from Yahoo Finance

1) In a web browser, go to finance.yahoo.com

2) Search for a company of interest, in the search box at top.

3) choose the Historical Data tab, in the middle of the page

4) select the time period of 5Y, then Done

5) select frequency of Daily

6) select Apply

7) Download data via Download Data

Appendix B – Definition of Daily Return

The daily return is given by:

$$R_t = (P_t - P_{t-1}) / P_t$$

where P_t is the adjusted closing price for a given day t, and P_{t-1} is the adjusted closing price for the prior business day.

Appendix C – Checking for Normality

For this assignment you can check if the daily returns distribution is likely Normal if skewness is in the range (-.3, .3) and excess kurtosis is in the range (-.6, .6). (Note these values assume 250 data points, which is 1 year of data.)

Skewness – is the data sample symmetrical?
A normal distribution has a skewness of 0.

Kurtosis – are the sample distribution tails as expected, or too heavy or too light?
A normal distribution has a kurtosis of 3 and an "excess kurtosis" of 0. Excess Kurtosis is defined as Kurtosis-3; Excel and DataFrame.kurt() both report "excess kurtosis".
https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.DataFrame.kurt.html