Stocks Data Analysis and Visualization Solution Using Python

Problem Statement:

For Investors to properly manage their portfolios, they need to visualize datasets, find useful patterns and gain valuable insights such as stock daily returns and risks.

In this project we will use the power of python to perform stock data visualization and stock return calculation.

Dataset:

Link: https://drive.google.com/file/d/1VX2IO8FVaFgv7dCPzH-7WxSLHhxwkbK8/view?usp=sharing

The Dataset contains stock prices of certain companies listed in S&P 500. S&P 500 is a stock market index that measures the stock performance of 500 large companies listed on U.S. stock exchange.

Check the list of S&P 500 companies here:
https://en.wikipedia.org/wiki/List_of_S%26P_500_companies

Below is the list of stocks considered:

AAPL = Apple Stock

BA = Boeing

T = AT&T

MGM = MGM Resorts International (Hotel Industry)

AMZN = Amazon

IBM = IBM

TSLA = Tesla Motors

GOOG = Google

	Date	AAPL	BA	T	MGM	AMZN	IBM	TSLA	GOOG	sp500
0	2012-01-12	60.198570	75.510002	30.120001	12.130000	175.929993	180.550003	28.250000	313.644379	1295.500000
1	2012-01-13	59.972858	74.599998	30.070000	12.350000	178.419998	179.160004	22.790001	311.328064	1289.089966
2	2012-01-17	60.671429	75.239998	30.250000	12.250000	181.660004	180.000000	26.600000	313.116364	1293.670044
3	2012-01-18	61.301430	75.059998	30.330000	12.730000	189.440002	181.070007	26.809999	315.273285	1308.040039
4	2012-01-19	61.107143	75.559998	30.420000	12.800000	194.449997	180.520004	26.760000	318.590851	1314.500000
...	...	...	...	...	...	...	...	...	...	...
2154	2020-08-05	440.250000	174.279999	29.850000	16.719999	3205.030029	125.449997	1485.020020	1473.609985	3327.770020
2155	2020-08-06	455.609985	172.199997	29.840000	18.459999	3225.000000	126.120003	1489.579956	1500.099976	3349.159912
2156	2020-08-07	444.450012	170.020004	30.020000	19.030001	3167.459961	124.959999	1452.709961	1494.489990	3351.280029
2157	2020-08-10	450.910004	179.410004	30.200001	21.650000	3148.159912	127.110001	1418.569946	1496.099976	3360.469971
2158	2020-08-11	437.500000	180.130005	30.200001	21.500000	3080.669922	126.750000	1374.390015	1480.319946	3333.689941

2159 rows × 10 columns

Questions:

1) **How do you read the data from stock.csv file using pandas module.**
2) **Print out the number of stocks.**
3) **Print the name of stocks.(Try using for loop)**
4) **What is the average return of the S&P500?**
5) **Which stock or index has the minimum dispertion from the mean in dollar value?**
6) **What is the maximum price for AMZN stock over the specified time period?**
7) **Check if data contains any null values.**
8) **Get dataframe information.**
9) **Define a function to plot RAW STOCK PRICES (WITHOUT NORMALIZATION/WITHOUT SCALING).**

   # The function takes in a Dataframe df as an input argument and does not return anything back!

   # The function performs data visualization.

   # Pandas works great with matplotlib, you can simply plot data directly from a Pandas DataFrame using plot() method(Take Date on X – axis).
10) **Plot Normalized/Scaled stock prices.**

    # To plot Normalized/Sclaed stock prices means to start all of the stock prices at one point so that they can be easily compared.This will help you to see the amount of gains you are going to get from the listed companies.

    # Create a Function that takes in Dataframe and returns Normalized Dataframe.

    # For Normalization take the stock prices and divide them by the value in the first row of dataframe.(Notice the massive gains in Tesla Stock)
11) **Perform Interactive Data visualization using Plotly.**

    Ref: https://plotly.com/python/

    Ref: https://plotly.com/python/line-charts/

Write a Function to perform an interactive data plotting of RAW STOCK PRICES (WITHOUT NORMALIZATION/WITHOUT SCALING) using plotly express.
Plotly.express module which is imported as px includes functions that can plot interactive plots easily and effectively.
Loop through each stock (while ignoring time columns with index 0)
add a new Scatter trace.(hint:fig.add_scatter)

12) **SimilarlyPlot normalized stock data in an interactive way.**

13) **It seems that most stocks experienced massive drops in 2020, let's assume that you own 100 shares of the S&P500 and you bought them on Feb 19th, 2020. How much did you lose (in $) by March 23rd, 2020? Hint: S&P500 dropped from $3386.15 on Feb 19th, 2020 to $2237.4 by March 23rd, 2020.**

14) **Calculate the daily return for** S&P 500**.**

get sp500 data from the dataframe.

#l loop through every element in the data frame.

Calculate the percentage of change from the previous day.

Hint: df_daily_return[j] = ((df[j]- df[j-1])/df[j-1]) * 100

put zero in the first line item

15) **Calculate the daily return for Amazon stock.**

16) **Define a function to calculate stocks daily returns (for all stocks**)

With in the function Loop through each stock (while ignoring time columns with index 0)
Loop through each row belonging to the stock
Calculate the percentage of change from the previous day
Hint: df_daily_return[i][j] = ((df[i][j]- df[i][j-1])/df[i][j-1]) * 100
Set the value of first row to zero since the previous value is not available
Use the function to get Daily returns of all the stocks.

17) **Plot the Daily returns vs. time graph using both static and interactive plots.**

Hint: use function obtianed in question 10 and question 11.

18) Calculate correlation between Daily returns of all stocks.

Drop Date column.
Use .corr() function
Use heat map to show correlation.

19) From the heatmap obtained ans the following questions.

What are the top 2 stocks that are positively correlated with the S&P500?

#What is the correlation between Amazon and Boeing? Comment on your answer

What is the correlation between MGM and Boeing? Comment on your answer

20) Plot Histogram for Daily returns.

21) Based on the histogram, which of the following stocks are more risky? T or TSLA