



DePauw University
School of Business and Leadership
BUSA 310 Business Analytics III: Predictive and Prescriptive Business Analytics

Assignment 2: Exploratory Data Analysis (EDA) for Asset Price Analysis

Instruction:

- Use Google Colab Notebooks for this Assignment.
- Be organized. In your Google Drive, create a folder named “BUS310”. In this folder, create a folder named “Assignment2”. In “Assignment2” folder, create two folders, one named “Colab Codes” (save your Notebooks here), the other one named “Results” (save all your charts here).
- Put your results (charts, tables and answer to each question) in this Word file. Highlight your answers in red.
- Upload this file and with a share link to your Colab Codes to Moodle after you completed.
 - o **Link:**
https://drive.google.com/drive/folders/1-_PIJBaXM6E5eq7vAwR27MCnDkfW4E6w?usp=sharing
- **AI policy:** You may use ChatGPT to help you with coding. Don’t use ChatGPT for summaries and answers. These must be your own words and reflecting your own thoughts.

1. Data Collection and Organizing

In this section, you will collect stock price data and organize the data.

Steps (show your work and code in Colab Notebooks):

- First, you are going to collect stock data on SPY, TSLA, and GM (these are the tickers for companies) and another company of your interests or you can even try Bitcoin. You

can directly load the data to Colab using libraries called “pandas_datareader” and “yfinance”.

- Write a loop to load these four stock price data.
- Now, you should have four series of stock price data. Append the data as one long dataset (use “pd.concat”) and call it “df”. Print the first 15 datapoints of “df”.
- Sort the “df” data by company name and the date from the newest to the oldest. Call this new data “df_sorted” and print the first 15 datapoints of “df_sorted” to see if you successfully did the sorting.
- Make a new dataset named “df_close” that only keep the close price for each stock. Transform “df” from its current long format into a short, wide format. The first column should be the date. The next columns (from the second to the fifth) should be the closing prices for each stock.

Questions:

- o How many datapoint do we have for each company a year? Why don’t we have 365 datapoints for a year?
 - o Answers: There are 253 datapoints for each stock. This is because the stock market is not open on non-trading days such as weekends and holidays.
- o What is the data type of stock prices? Cross-section, time series, or panel data?
 - o Answers: The data type of stock prices in this dataset is panel data because it combines cross-sectional and time-series, involving multiple entities over time. The data record daily stock prices for four companies (Apple, General Motors, S&P, Tesla) over several years.

2. Statistics

In this section, you will calculate basic statistics for each stock and interpret your findings. You must complete Section 1 before working on this section.

Steps:

- Calculate the minimum close price, maximum close price, the average close price, standard deviation of each stock. Keep two decimal places. Put these statistics in only one table and your table should only show the asked statistics.
- Calculate the annual return for each stock. Use the following formula: $(\text{Final Close} - \text{Initial Close}) / \text{Initial Close}$. This is based on the first date and the last date. Returns should be in percentage format with two decimal places.
- Calculate the correlation coefficient for each pair of stocks. You would get 6 correlation coefficient in total. Keep two decimal places.

Questions:

- Which of the stock had the highest annual return and why is this interesting? If you invested \$100 last year in this asset, how much you will have in total now?

Answers: Tesla with 1284.84%. This means Tesla was having a strong financial performance and indicating exponential growth in the market

- Which of the stock had the highest standard deviation and why is this interesting?

Answers: Tesla had the highest standard deviation with 89.97. This means that Tesla's stock experienced more significant fluctuations compared to other stocks

- Which of the stock had the lowest standard deviation and why is this interesting?

Answers: General Motors had the lowest standard deviation with 10.20 implies the stock is the stabliest with fewest flunctuations compared to other stocks

- Which of the stock pairs had the highest correlation and why is this interesting?

Answers: Apple and S&P at 0.919377. A high correlation indicates that Apple's stock price movements are closely aligned with the broader market index. Therefore, the market can easily affect Apple's stock and reverse

- Which of the stock pairs had the lowest correlation and why is this interesting?

Answers: Apple and General Motors at 0.305613 which means these two stock are less related to each other. If one of the stock performance go good/ bad, it does not really influence the other.

3. Data Visualization

A. Stock Price Index

In this section, you will visualize the data series by creating a simple stock price index and see which stock outperform the market (typically, we use SPY to represent the market fluctuation).

Steps:

- Create a line chart below that shows the four data series of the close prices. Give the chart a title and a legend at the bottom of the chart to label the series. Make sure the x axis provides a clear labeling of the date range.
- This is not a good chart to compare stock price. In this step, you are going to adjust the data. Create a new dataset (call it "normalized_df_close") that contains four new variables that calculate the normalized price. To do this, dividing each price by the first price and then times 100. This will convert the value on the first date to 100. Print the new dataset to see your new variables. These are the stock price indexes.
- Create a second line chart using the new variables you created. Give the chart a title and a legend at the bottom of the chart to label the series. Make sure the x axis provides a clear labeling of the date range.
- Save your line charts to your Google Drive "Results" folder.

Questions:

- According to your "Close 100" chart, which stock outperform the market?

Answers: S&P stock with yellow line is outperform the market compared to others. S&P represents the performance of the 500 largest companies in the US and could be seen as a safe choice with consistent earnings, lower volatility and more steady

B. Daily Return

Now, we've done the baseline analysis. Let's dive a little deeper to check the daily returns for stocks.

Steps:

- Calculate the daily return for each stock (use "pct_change").
- To visualize the daily returns for each stock using subplots within a single large plot, you can use the matplotlib library. Make a 2x2 subplots. Each subplot will show the daily returns for one stock. (use "plt.subplots").
- Save your chart to "Results" folder.

C. Distributions of Daily Return

Steps:

- Make similar charts that shows the distributions of the daily return for each stock. Show 30 bins for each subplot.
- Save your chart to "Results" folder.

D. Pair Plots for Correlation

To visualize the correlation between stocks, you can use pair plots (also known as pairwise plots or scatterplot matrices).

Steps:

- Use "seaborn" library to create the pair plot of daily return each pair of stocks. (use "sns.pairplot")
- Save your chart to "Results" folder.

4. Summary

Imagine that you are the financial manager trying to help your client select a stock or a portfolio that can outperform the market while maintaining reasonable risk. Summarize your analysis here and insert all graphs you made put them in this section. Explain the following:

- **Dataset Overview:** Describe the dataset used in this analysis.
- **Graph Interpretation:** Explain what each graph reveals about the data.
- **Investment Strategy:** Outline the investment strategy you'd suggest based on your analysis.

Dataset Overview

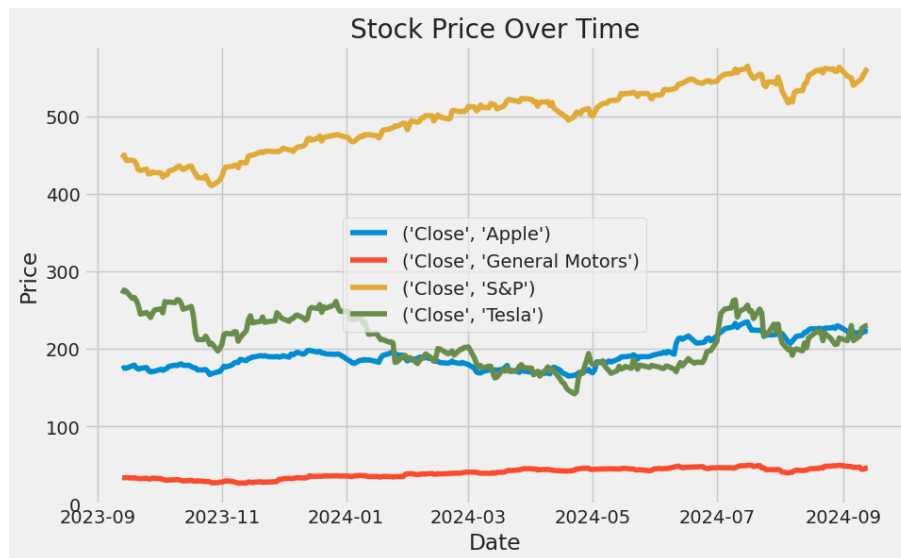
In this assignment, I choose 04 stocks: S&P (SPY), Tesla (TSLA), General Motors (GM) and Apple (APPL). The dataframe includes this information:

- Date: The date for each stock price observation
- Open: The stock price at market open for that day
- High: The highest stock price reached during the day
- Low: The lowest stock price reached during the day
- Close: The stock price at market close for that day
- Adj Close: The adjusted closing price
- Volume: The number of shares traded during the day
- Company name

Later, I made a new dataset that only keep the close price for each stock in a short, wide format. In order to portray more information about this dataset, I found both the annual and daily returns for each stock; and the correlations between stocks.

Graph Interpretation

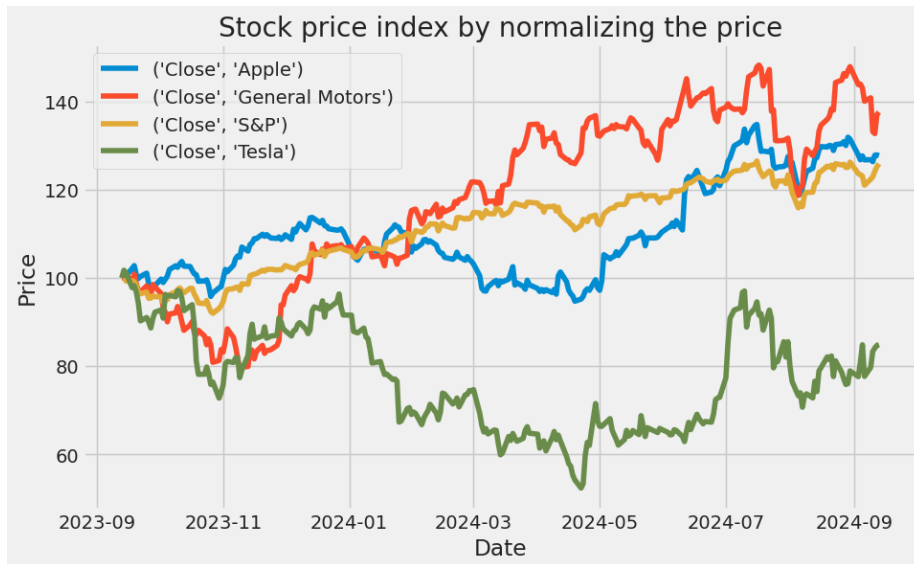
Graph 1



Graph 1 shows the trend of 4 stock prices at closing time during one year, starting from Sep 2023 to Sep 2024. SPY with the yellow is having the positive trend with overperform in the market at over 500USD. In contrast, GM has the lowest value which have not reach 50USD.

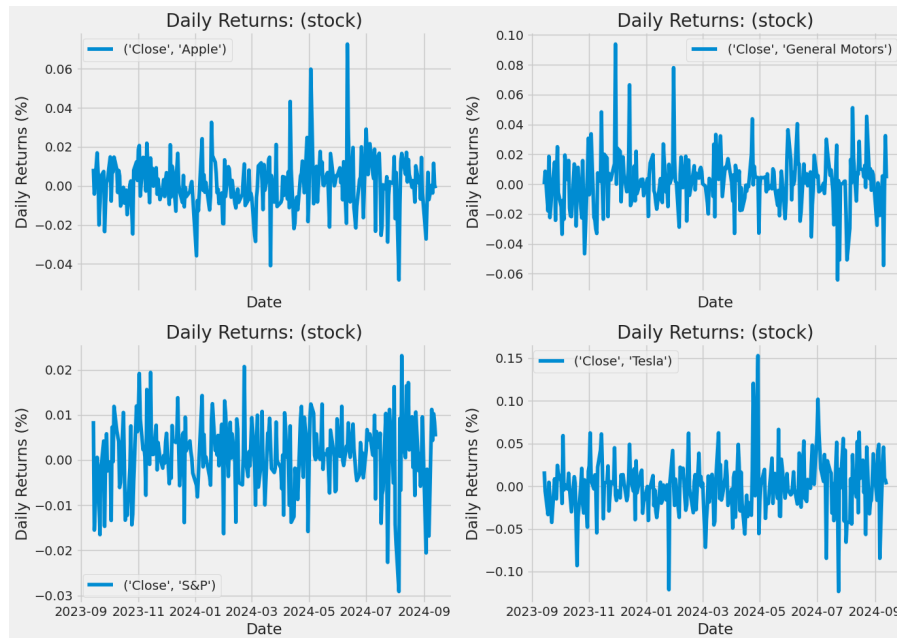
APPL and TSLA has a steady growth and fluctuation perpsectively; however, they still keep in the middle range from nearly 200 to 300USD.

Graph 2



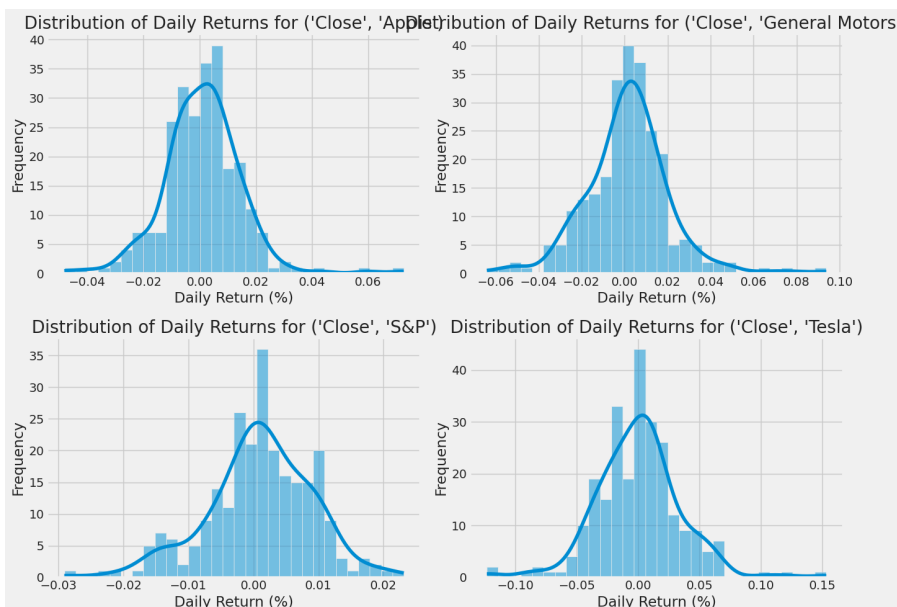
Each stock's price in the Graph 2 is being scaled to a common baseline. This graph helps to make comparison of performance across 4 stocks, regardless of their starting price. Most of the stocks are not stable. At the price of 100, they rise and fall and by Sep 12, 2024, GM is leading the market while TSLA is being left far behind. Following after the first place is APPL and SPY.

Graph 3



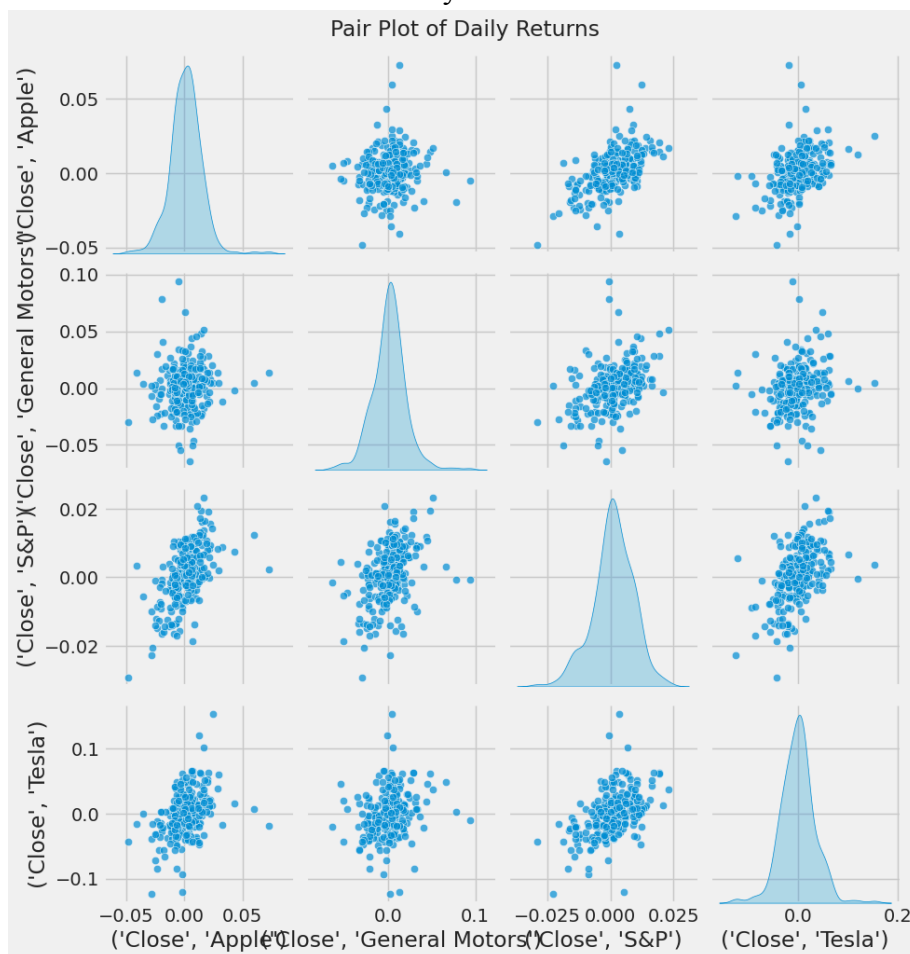
The daily returns for each stock, which analyze day-to-day performance, is display in Graph 3. I calculated the percentage change from one day to the next for each stock. Apple shows a relatively stable pattern (mostly within the range of -2% to +4%) with occasional peaks (above 6%). SPY has relatively small daily return changes, just within -2% to +2%, means less volaitile compared other individual stocks. On the other hand, Tesla has the highest volatility swinging from -10% to +10% and GM fluctuate between -6% to +8%. However, GM appearing less stable than APPL since the fluctuation is sharper.

Graph 4



This graph 4 demonstrate four histograms that representing the distribution of daily returns for 4 stocks. APPL shows a somewhat symmetrical shape, suggesting a roughly normal distribution and at the highest frequency of returns (slightly above 0%). GM's distribution on the top right is a little skewed to the left, suggesting that neutral or small positive returns are more frequent. Next, SPY appear fairly symmetrical and normal with small wider peak (less volatility). The distribution for TSLA is less symmetrical and has heavier tails.

In this
stocks are
diagonal
(the



graph, 4
present in
histograms

distribution of daily returns for a single asset) and scatterplots (the relationship between daily returns two different assets).

- Apple and S&P: they have a slight positive correlation
- Tesla and General Motors: a weak correlation or possibly no significant correlation
- S&P and General Motors: some positive correlation
- Tesla and Apple: little direct correlation
- Tesla and S&P: a mild correlation

In conclusion, graph 5 suggests that some assets (like Apple and the S&P 500) tend to move together, while others (like Tesla and General Motors) show less correlation in their daily returns.

Investment Strategy

As the financial manager, the investment strategy I would propose for the client would involve balancing both the performance trends and risk factors based on the analysis of the stocks across the different graphs I created above.

There is stable growth in these two stocks:

- SPY (S&P 500): Due to its positive trend, low volatility, and symmetrical risk distribution, SPY can act as a foundation for the portfolio, ensuring exposure to the broader market with lower risk.
- AAPL (Apple): Its steady growth and lower volatility make it a great complement to SPY. The positive correlation between AAPL and SPY suggests that these two stocks can work well together in maintaining market outperformance with reduced risk as APPL is a part of SPY

In contrast, GM is riskier than AAPL and SPY. However, its recent positive trend and upward shift in performance suggest potential short-term gains. Including GM as part of the portfolio provides a moderate risk/reward opportunity.

Lastly, taking high risk in TSLA could also be the high reward. For the client willing to take on high risk, TSLA offers significant growth potential despite its high volatility. However, this stock should constitute a smaller portion of the portfolio due to its unpredictable swings.