Is Tesla's Growth Trajectory Truly Unique Among Tech Giants?

Abstract

Our assignment aims to investigate whether Tesla's growth trajectory is genuinely unique compared to other leading tech giants. We are comparing Tesla with 4 other global giants Amazon, Apple, IBM and Microsoft. We aim to analyze this by comparing the monthly opening prices, closing prices, volumes of the mentioned stocks. By analyzing the number of trades (volume transactions) of Tesla stocks, the month-by-month/year-by-year closing stock price and % increase/decrease in stock price performance we aim to determine the distinct factors contributing to Tesla's astounding growth trajectory. Preliminary analysis highlights Tesla's efficiency against Amazon (the best performing stock among the remaining 4). We normalized the closing prices which helped us in comparing the relative performances of the companies over time on the same scale, independent of their initial price levels. Through the normalization we could see the sharp growth of Tesla stock through the period from 2013 to 2021.

1. Datasets

Data for this study was sourced from a publicly available website - https://www.alphavantage.co/documentation/ covering the period from Tesla's global boom in 2013 to 2024. The information that came through was the stock information – stock trade date, stock opening price, stock closing price, stock high point, stock low point, stock volume. We pulled this information through the API and the data came in JSON format. We stored the JSON data in RDBMS through ETL and the data gets updated every working day with incremental load.

The two important aspects of big data present in our dataset are **volume** and **velocity**. Firstly, we are pulling data from a stock website through API so that we get new data every day. This introduces velocity to the dataset.

The historical data for these stocks has been available since the 1990s, so there is a considerable amount of data accumulated. Simplifying the volume of constantly growing data for seamless integration and minimal server strain is essential to effectively manage such a large dataset. To tackle this, we use an ETL process that loads only the new data into our database, incrementally. The advantage of this approach is that it not only optimizes resource usage but also reduces server load, which in turn allows faster data processing and better system performance. The ETL also helps us keep our data always updated and avoid overwhelming the infrastructure by handling the data incrementally.

2. Data Exploration, Processing, Cleaning and/or Integration

Our investigation is trying to answer whether Tesla growth is truly unique amongst other tech giants. To showcase this, we decided to analyze stock performances from 2010 i.e. when Tesla stock was offered to the public. How do we compare the growth you may ask? We decided to compare the Tesla upswing via stock price and the volume, but the price didn't give us the correct picture. To overcome this and to compare the relative performance of the companies over time in a fair way, independent of their original price levels, we normalized the closing prices to the same scale.

(Normalization: In the context of stock prices, normalized closing price is the way to adjust closing prices of stocks for each company over time, so the first data point (first year) is set to 100%. This makes it possible to compare the performance of different companies over time on the same scale, without regard to their initial price level.)

Data Cleaning and Processing

Since our data came in JSON format, we couldn't process it directly so we needed to store that data in RDBMS along with its metadata. In RDBMS, we created 3 tables for ETL process – source_audit, task_audit and source_information. For every company stock, we created a separate table and each table gets updated every working day.

To get the API response and store it into the RDBMS we use **request** and **sqlaclhemy** libraries. After performing ETL on that data we again load that JSON response data into the python environment for further cleaning and processing.

For converting the JSON data into dataframe we have used the **ast** and **json** libraries which gave us the required dataframe. After that all the operations are done using **pandas.**

We did the necessary operations like checking for null values, checking datatypes of columns and checking whether there is unwanted information.

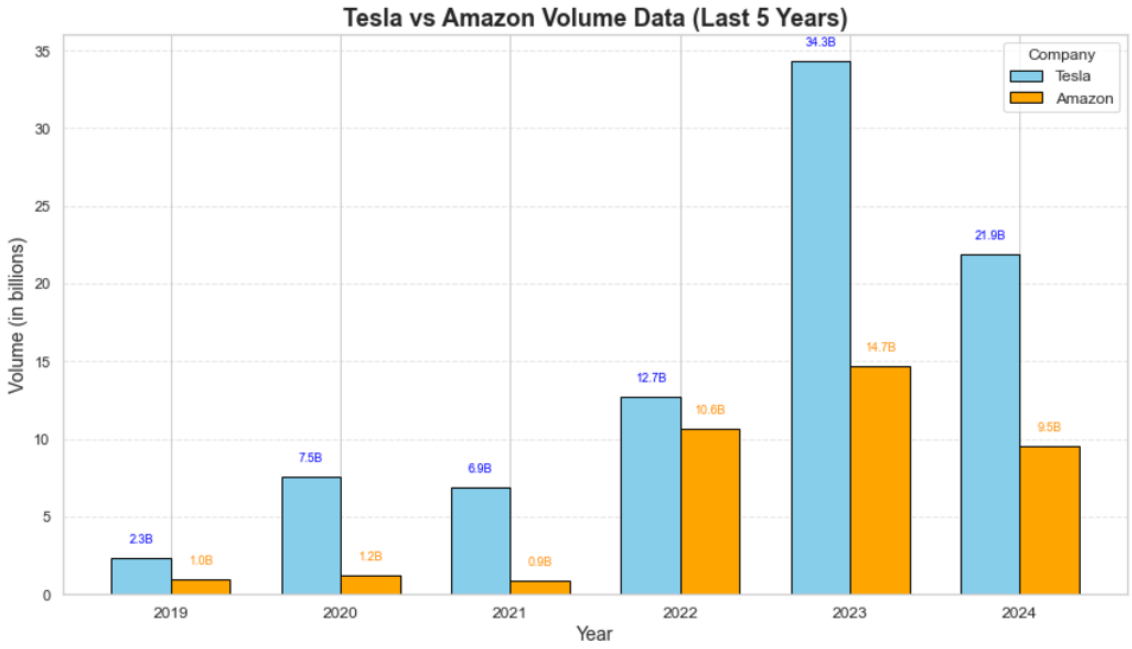
We found some null values in stock close price which we handled by replacing the null value with its nearest date closing price. We changed the data types according to the columns and removed unnecessary/unwanted characters from the column headers (for eg: 1. Low to LOW).

Calculations were then made so that we could get the necessary columns required for visualizations-

- month_id
- year_id
- % change closing price

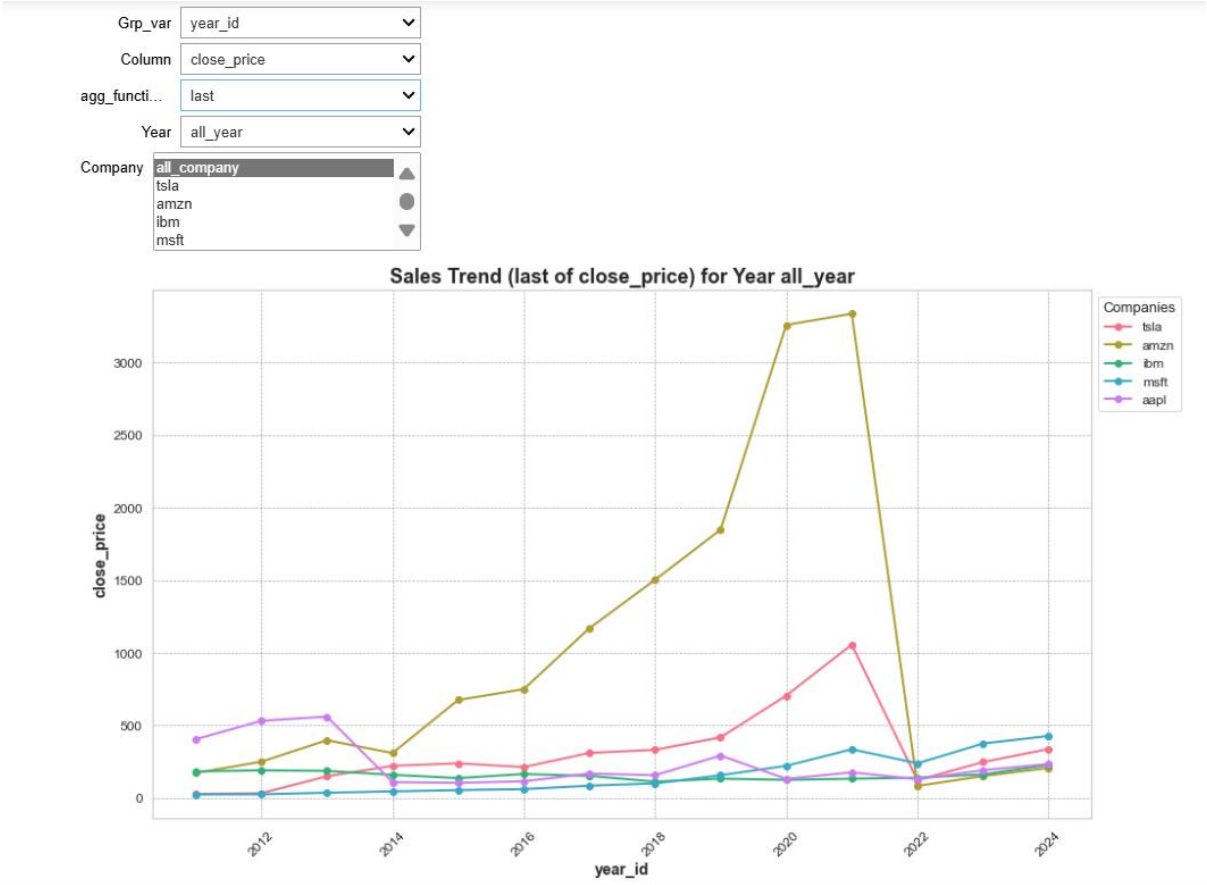
Data Exploration

For exploration, we have multiple key figures: opening price, closing price, high, low and the number of trades(volume) to check the growth of a stock. So first we decided to go with the volume parameter since we found out Tesla had the highest volume and Amazon had the second highest volume compared to the other 3 companies in question.



But we can't conclude Tesla had a significant growth through the volume parameter because having more volume doesn't necessarily mean that the stock price is also increasing.

So we moved forward to the closing price parameter -



Here we can see that amazon has the highest closing price, however, that is not indicative of its growth since amazon already had a high stock price. Due to this, this is not a correct way to see the closing price parameter for the performance of any stock. To overcome this, we normalized the closing price parameter. This made it possible to compare the performance of different companies over time on the same scale, without regard to their initial price level.

3. Visualisation

Choice of Chart Type

To answer our question on whether tesla's growth trajectory is truly unique among tech giants we would have to see the percentage increase of normalized closing price of each company at the end of years. We wanted to show that the Tesla stock was performing well year on year while the other companies were showing steady growth. We could also see that the normalized closing price rapidly increased from 2013 to 2021.

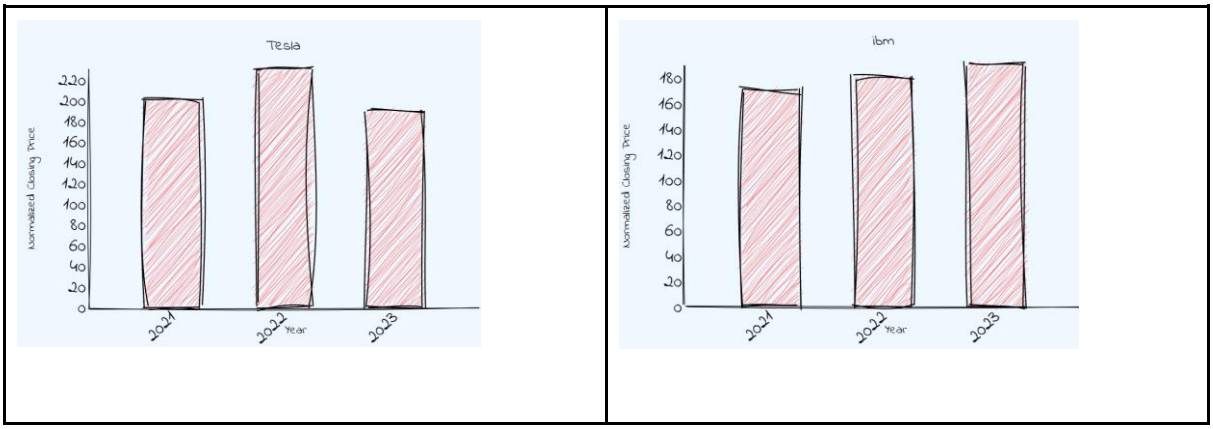
We are using a line chart because -

i) We have time-series data from 2011-2024.

ii) We want to show the comparison between 5 different companies.

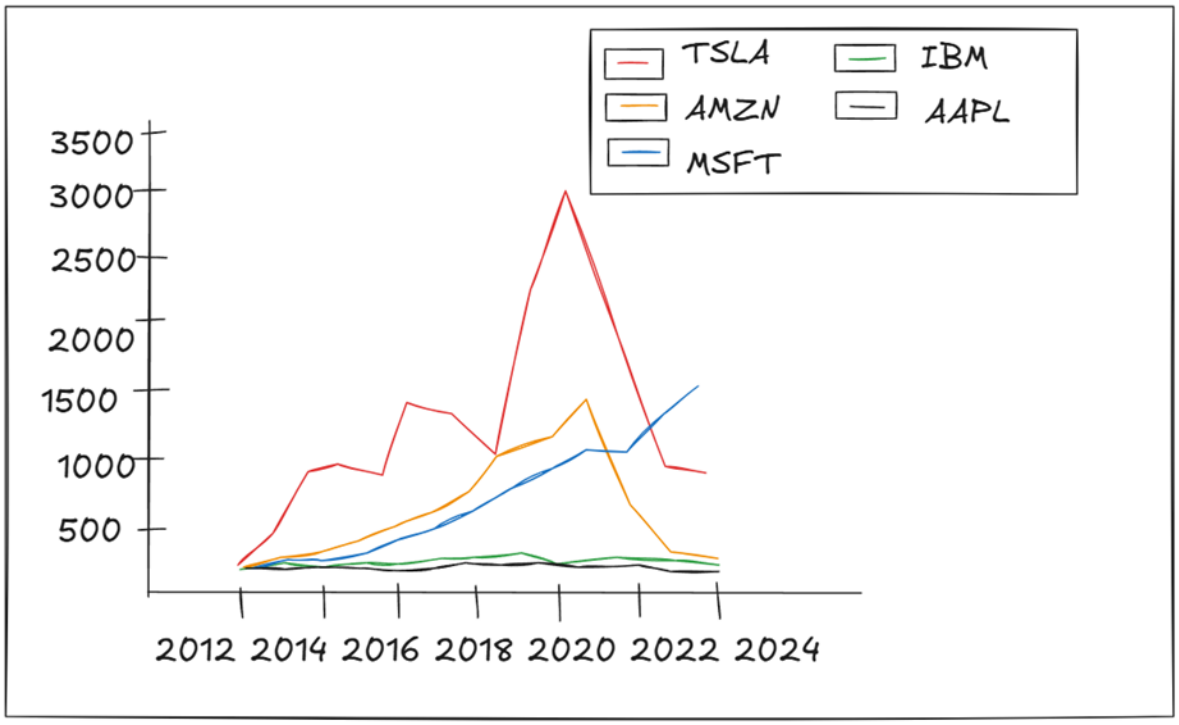
iii) We want to show the growth trajectory of Tesla stock in the given period.

Sketches



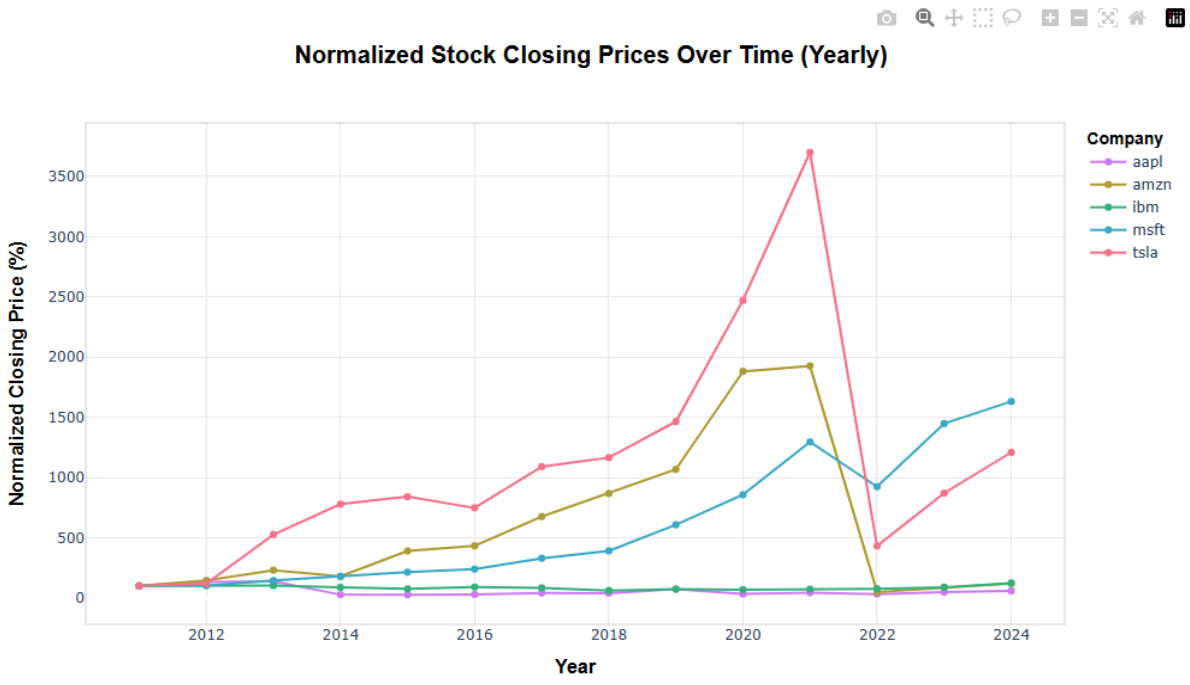
If we only wanted to see the Tesla stock performance then a bar chart would have been the best option but here we want to compare Tesla's performance with other companies.

Using a bar chart would require creating separate charts for each company, making it necessary to manually compare the closing prices. This approach would lack self-explanatory clarity.



In the above example, it is much easier to compare how different companies performed over time using a line chart. For this reason, we decided to use this chart.

Our Chart



Design Choices

We built upon the design of our sketch by making the following changes as per the checklist by Stephanie Evergreen & Ann K. Emery:

- We have used line chart to establish an appropriate relationship between years and the normalized closing price for the company.
- We have used title, x-label, y-label and legend so that it makes the graph self-explainable.
- We have used muted colors like lavender, olive green, slate blue, mustard for the remaining four companies apple, ibm, microsoft, amazon respectively so that they do not stand out more in the visualization. We have used pink as an action color for Tesla so that it stands out more amongst other tech giants.
- Our gridlines are subtly designed with a faint grey color, ensuring they don't stand out prominently while enhancing the graph's readability. This design allows the price labels on the y-axis to remain clear and facilitates effortless comparison between companies.
- To reduce the clutter on the x-axis we have plotted points every two years from 2012 to 2014 i.e. 2012, 2014, 2016 etc. This helped us in increasing the readability of our graph as well as reducing the congestion on x-axis.
- We also added interactivity such as highlighting a stock's percentage change in closing price as compared to the first year i.e normalized closing price for all the five stocks.

4. Conclusion

Tools and libraries used

For efficient data storage we used a Relational Database Management System (RDBMS) **MySQL** and for our development environment we used **Jupyter Notebook**. We used **json** and **requests** to handle API requests and JSON data. We used **pandas** for manipulation and **ast** for handling abstract syntax trees for data processing. We used **matplotlib** to create different plots and **seaborn** to have aesthetically pleasing and informative graphical representation which makes it easier to work from data storage to visualization.

Analysis of the the outcome of visualisation

The biggest challenge we faced was to illustrate growth performance on the visualization. We tried using the volume parameter but realized that having more volume doesn't necessarily mean that the company is also performing better. When we tried closing prices, we realized that every company has a different starting point so we cannot measure growth for a particular year based on its previous year. Therefore, we normalized the closing prices, enabling us to compare the performance of different companies over time on a consistent scale, regardless of their initial price levels.

Stock prices aren't the only thing that matters in growth. Big news events, worldwide situations like COVID, changes in leadership (a new CEO) or winning major awards are only some of the other things that can affect it. These are big factors in growth but it's hard to get them and add them to the data that we already have. It's tricky, but including them would really make the analysis much more meaningful and complete.

During the data cleaning and processing phase, the incoming data was in JSON format. The challenge was to convert this JSON data into appropriate dataframe. After this we did null value processing, data-type conversions and we also created some new columns like month_id, year_id, % change in closing price, normalized closing price to get better visualization.

Our chart tells us that Tesla has a truly unique growth trajectory among tech giants. We can see a sharp growth in Tesla stock from the period 2013-2021. Amazon and Microsoft are following Tesla's performance path but IBM and Apple show a steady growth.

Most of the work was evenly shared between both of us(Harsh Tyagi and Nitish Surve).
Harsh Tyagi did the data gathering and stored the data in RDBMS.
Nitish Surve did the data cleaning and processing.
We both contributed equally for visualization sharing important insights with each other about data-readability.
The report was collaboratively authored, with both members reviewing the content to guarantee accuracy and clarity.

References

1. https://www.alphavantage.co/documentation/

2. https://stephanieevergreen.com/wp-content/uploads/2016/10/DataVizChecklist_May2016.pdf

3. https://chatgpt.com/ OpenAI. *CHATGPT* (October 2023 version), Used for grammar and rephrasing support.