**Project Name:** Tesla Stock Analysis-Insights from Historical Data Analysis

**Course:** PROG38063 Data Analysis & Visualization 2

**Team Members:** Ruttansh Bhatelia, Xiaohong Deng

## Section1: Data Exploration:

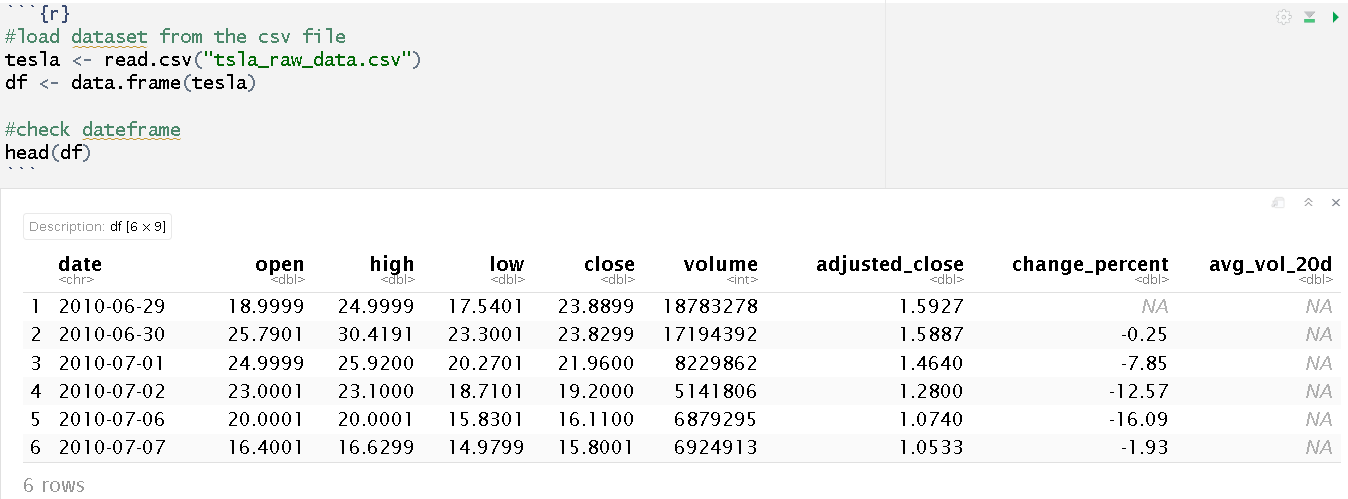
### 1- Introduction

Tesla, Inc., founded by Elon Musk and his partners, has experienced remarkable growth in market value, making its stock a focal point in investment circles. As stockholders, we are keen on delving into the intricacies of Tesla's stock performance and understanding the factors shaping its market behavior.

The dataset, sourced from [kaggle](https://www.kaggle.com/code/kkhandekar/tesla-stock-forecast-sklearn-skforecast/input?select=tsla_raw_data.csv), encompasses Tesla's stock prices from its initial public offering (IPO) to August 15, 2023. It comprises essential parameters such as date, open, high, low, close, volume, adjusted\_close, change\_percent, and avg\_vol\_20d, providing a comprehensive perspective on Tesla's stock performance over time. This project will analyze the Tesla stock dataset across five phases: Data Exploration, Hypothesis Testing, Deriving Insights, Discussion and Conclusion, and Presentation.

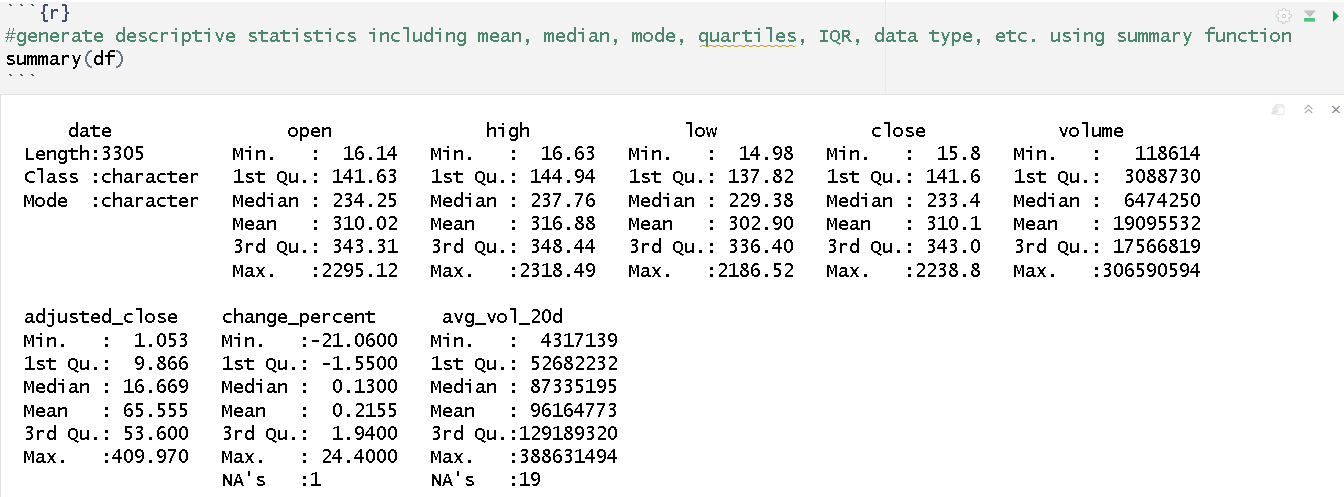
### 2- Data Summarization

Provide descriptive statistics of your data like mean, median, mode, quartiles, IQR, data types, etc., and interpret the output.

**Solution:**

**Observation:**

1. The dataset contains several columns including 'open', 'high', 'low', 'close', 'volume', 'adjusted\_close', 'change\_percent', and 'avg\_vol\_20d'. These columns represent different aspects of Tesla stock data such as opening price, closing price, trading volume, etc.
2. The table reveals that the values across these columns fluctuate over time, as indicated by the 'date' column.



**Observation:**

1. The dataset contains 3305 entries, each representing a specific date, with a total of 9 columns. The "date" column is of character type, indicating dates, while the other columns, including "open," "high," "low," "close," "volume," "adjusted\_close," "change\_percent," and "avg\_vol\_20d," are numerical.
2. For numerical columns such as "open," "high," "low," "close," "volume," "adjusted\_close," "change\_percent," and "avg\_vol\_20d," descriptive statistics like minimum, maximum, quartiles (1st, 2nd, and 3rd), and mean are provided. However, for the character column "date," descriptive statistics include length, class, and mode.
3. There are some missing values in the "change\_percent" and "avg\_vol\_20d" columns, as indicated by the NA values. These missing values will need to be handled appropriately in subsequent analyses.

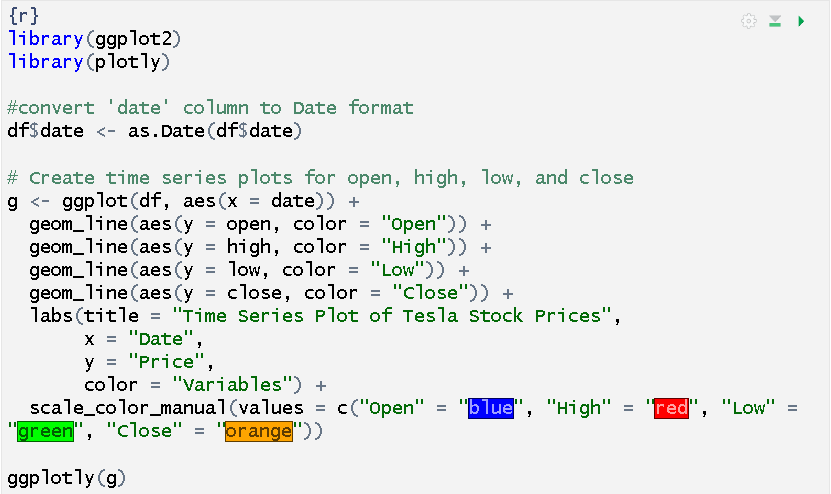
### 3- Data Visualization

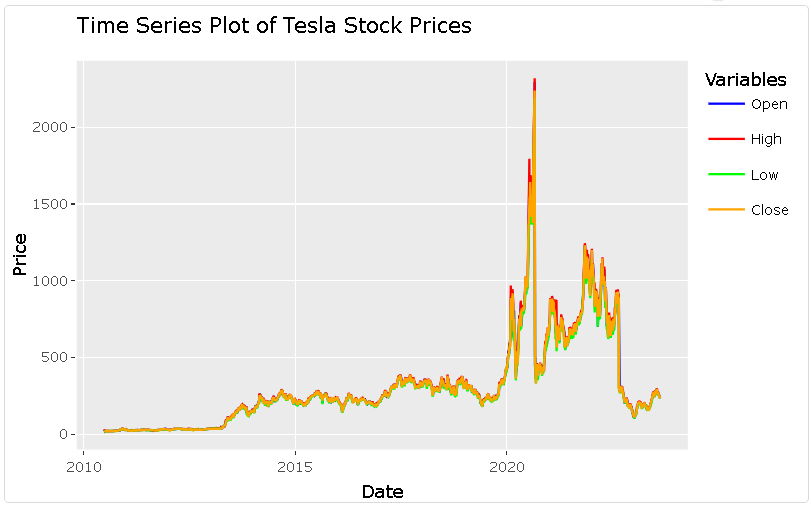
Use tables and graphs to organize, visualize, and explore your dataset.

**Solution:**

**3.1) using time series plots**

to visualize how variables change over time. This can help identify trends, seasonality, and anomalies in the dataset.





**Observation:**

* Tesla Prices dropped suddenly on 2020/08/31 and 2022/08/25 because of stock splits. These splits made each existing share into multiple shares, causing the price per share to decrease. For example, if you had one TSLA share before August 31st, 2020, you'd now have 5 shares.
* To keep the historical stock data accurate and avoid distortions from stock splits, the original dataset "tsla\_raw\_data" was adjusted as follows

A computer screen shot of a program

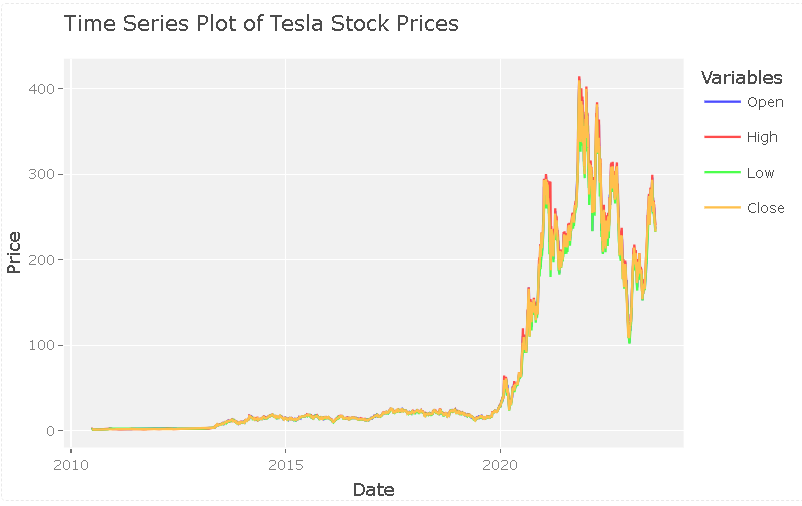
Description automatically generated

A screenshot of a computer

Description automatically generated

A computer screen shot of a program

Description automatically generated



**Observation:**

* The time series plots above provide a clear visualization of how the variables "open," "high," "low," and "close" change over time. These plots are valuable for identifying trends, patterns, seasonality, and any unusual fluctuations in Tesla's stock prices dataset.

A screen shot of a graph

Description automatically generated

**Observation:**

* The plots above depict fluctuations in trading volume over time, with peaks often aligning with notable events or news releases that influence investor behavior and market activity.

A screen shot of a graph

Description automatically generated

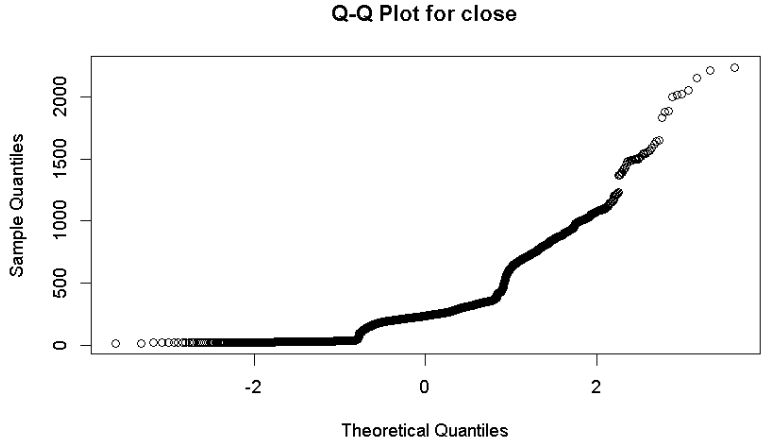
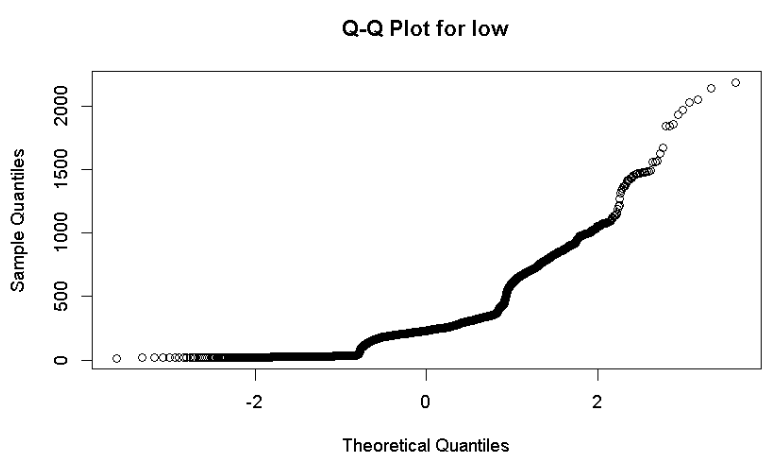
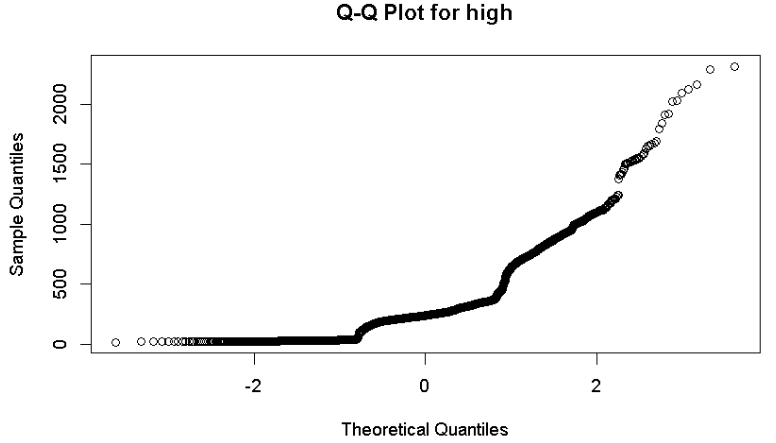
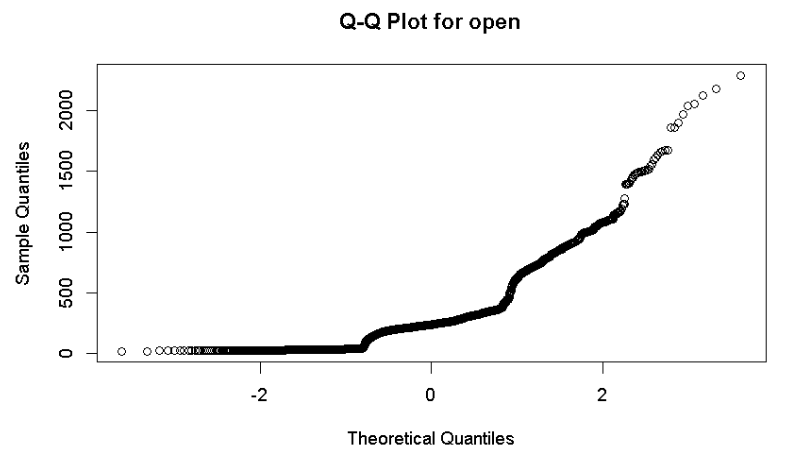
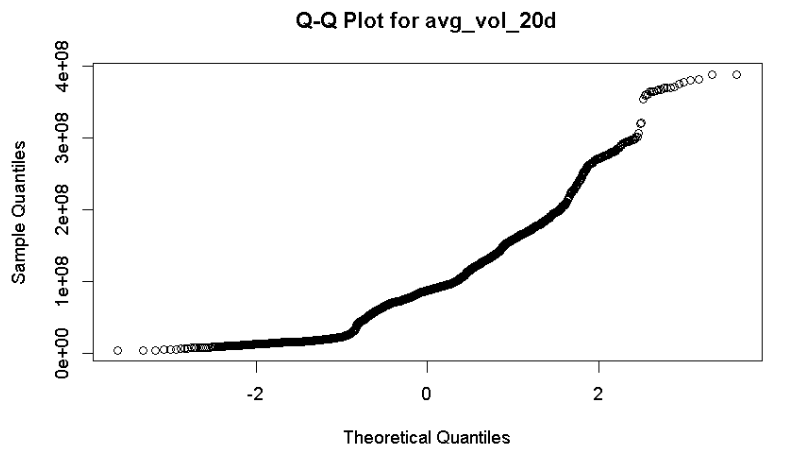
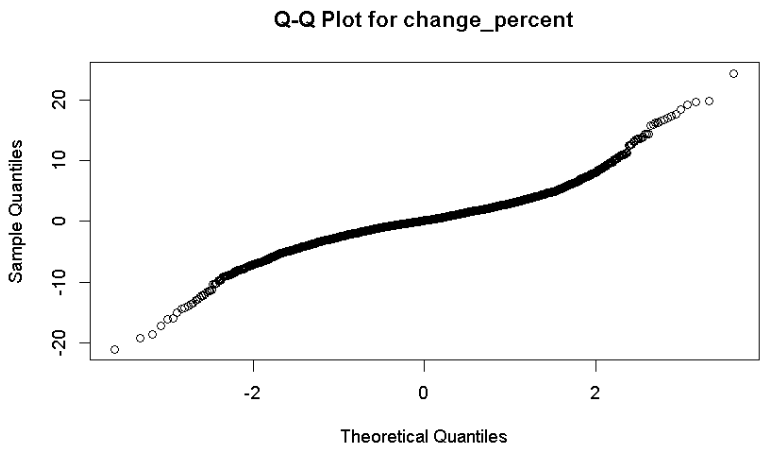
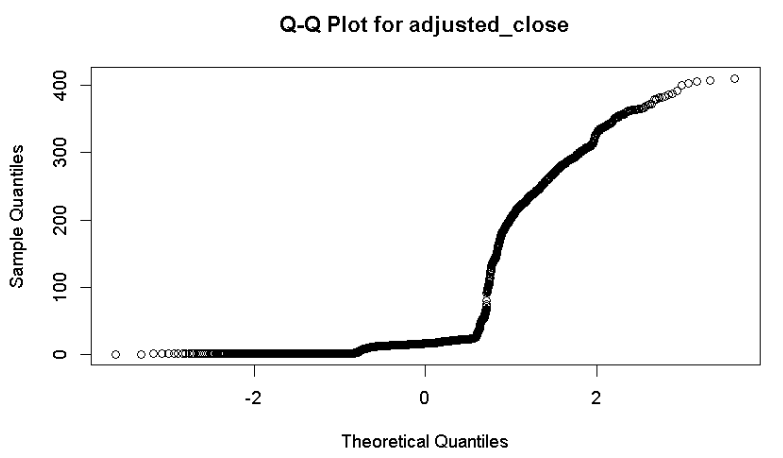
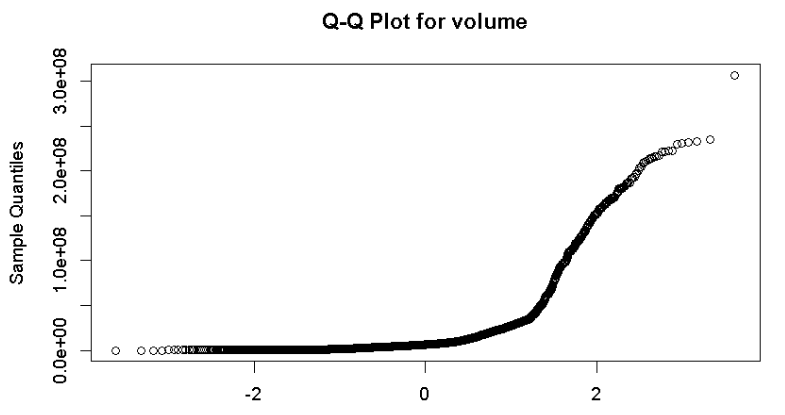
**Observation:**

* The plot displays the daily percentage change of Tesla's stock price, highlighting the volatility over the observed period. Sharp spikes and dips depict significant price movements, possibly influenced by news events, earnings reports, or broader economic conditions.
* The warning message indicates the presence of null values in the 'change\_percent' column, which should be addressed during data preparation before proceeding with further analysis.

**3.2) using a quantile-quantile plot**

to verify normality for each column. If the qqplot is linear, then we can assume normality.

A screenshot of a computer

Description automatically generated****

**Observation:**

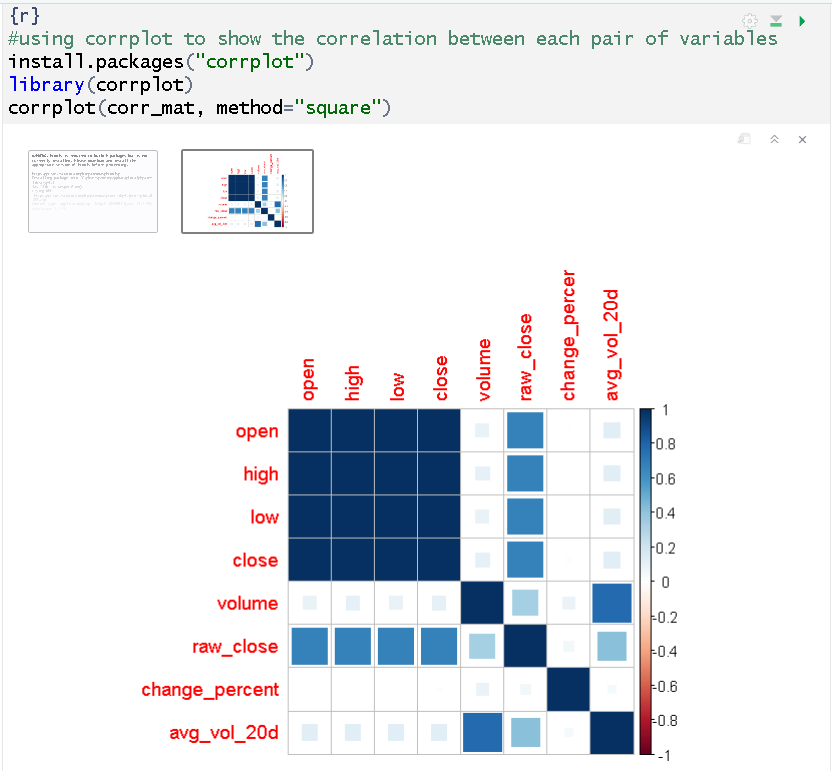
* The plot of the Q-Q plot for the 'change\_percent' column appears to follow a linear pattern, suggesting that the data distribution may approximate normality.

**3.3) using corrplot**

to explore the linear relationships between each pair of variables in the tesla dataset. This can help identify patterns, correlations, or trends within the datasets.

A screenshot of a computer

Description automatically generated



**Observation:**

* The splot indicates that there is strong positive linear relationship between each pairs of "open," "high," "low," and "close".
* A moderate positive linear relationship is observed between 'volume' and 'avg\_vol\_20d'.
* There exists a moderate positive linear relationship between 'raw\_close' and each price variable ("open," "high," "low," or "close").
* For other variable pairs, the linear relationship is weak or nearly zero.

### 4- Data Quality

Data quality: based on the values and visualizations, discuss the quality of your dataset and discuss how issues with data quality might impact your analysis (for example, if one of your graphs in (3) shows many extreme values, how does that impact the values you calculated in (2)). The provided metadata can help you with evaluating the quality of your data. In addition, please take the necessary steps to clean your data and prepare it for analysis (steps 2 and 3 can help you identify the steps needed to clean the dataset).

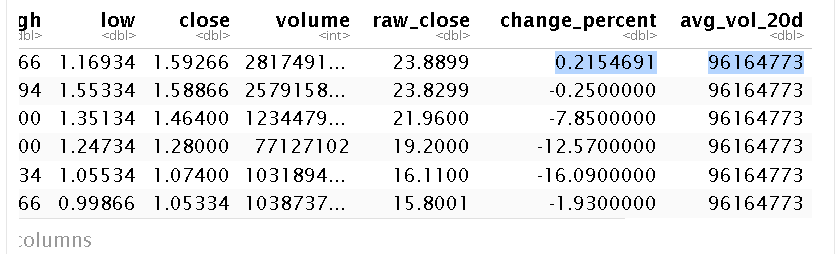
**Solution:**

**4.1) Missing data**

Based on the result of head and summary functions, we can find that there are some missing values in the "change\_percent" and "avg\_vol\_20d" columns, as indicated by the NA values. These missing values can impact the accuracy of our analysis, especially when calculating summary statistics or visualizing trends over time. To address this issue, we will fill in the missing values with the mean of each respective column.

A screenshot of a computer code

Description automatically generated



#### 4.2) Inconsistency data

During data visualization, inconsistencies were observed before/after 2020/08/31 and 2022/08/25 due to stock splits. These splits caused Tesla's prices to drop suddenly on those dates. To ensure the historical accuracy of the stock data and avoid distortions from stock splits, the original dataset "tsla\_raw\_data" was adjusted to provide a clearer and more accurate depiction of Tesla's stock performance over time.A screenshot of a computer

Description automatically generated

Reference

Jorgenson, D. W., & Weitzman, M. L. (2024). Tesla stock forecast: NASDAQ:TSLA. AC Investment Research, v.220. Retrieved from <https://www.ademcetinkaya.com/2023/05/tesla-stock-forecast-nasdaqtsla.html>

Khandekar, K. (n.d.). Tesla Stock Forecast - sklearn + skforecast [Data set]. Kaggle. Retrieved from<https://www.kaggle.com/code/kkhandekar/tesla-stock-forecast-sklearn-skforecast/input>

Raza, S. H. (2022, July 31). Creating and Visualizing Correlation Matrices Using Heatmap. Medium. https://medium.com/@ai.mlresearcher/creating-and-visualizing-correlation-matrices-using-heatmap