Stock Analysis - A Data Engineering Perspective

By: Ardavan and Nithya

Microsoft

ORACLE

intel

Project Outline

The project uses fundamental data engineering and analytical principles to perform a volume and pricing analysis of a small sample of five prominent technology stocks (Microsoft, Apple, Oracle, Intel and Google). It uses a combination of ETL workflows to ingest data from Yahoo Finance (a popular stock market reference) using Python libraries and SQLite, Pandas Dataframe to perform first level analysis of the above stocks.

Data Source: Yahoo Finance

Pandas Dataframe

SQLite DB

Data Analysis

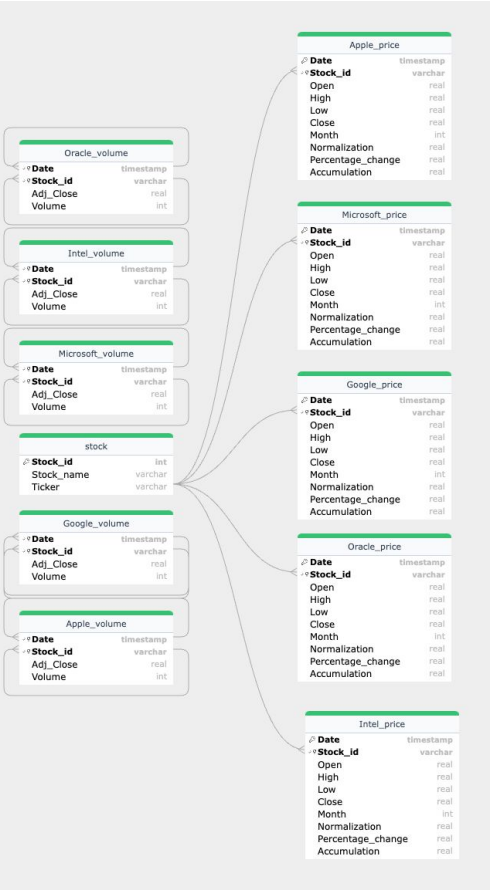
Our data analysis provides the following insights based on each of the stocks in our portfolio:

- What are the broader trends in pricing changes across each of the different stocks in 2023?
- Does change in trading volume reflect a related change in stock price?
- How does normalization of a particular stock's price affect its overall investability?
- Are there broader market trends in play that need to be inferred based on this portfolio analysis?
- The following key concepts are used in the analysis:

  - ***Percent Change:*** Increase or decrease in the price of a stock over time. This provides a standardized way to compare investments with different starting price values.

  - ***Normalization:*** Based on the first day's closing price, the analysis normalizes the closing price of each stock throughout 2023. This provides an insight into if the stock is overpriced or underpriced relative to its fundamental value.

ER Diagram and SQL Flow



Primary Table:

- *Stock* (contains all the relevant data for each of the stocks in the portfolio)

Derived Tables:

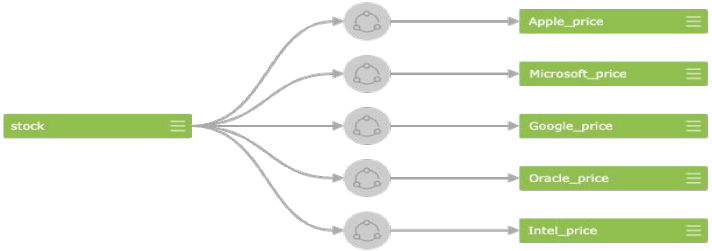
- *Volume* (contains trading volume data for each of the stocks in the portfolio)
- *Price* (contains pricing data including normalized metrics for each stock in portfolio)

Database:

- SQLite is used for the related schema and all relevant SQL query transformations.

SQL Flow:

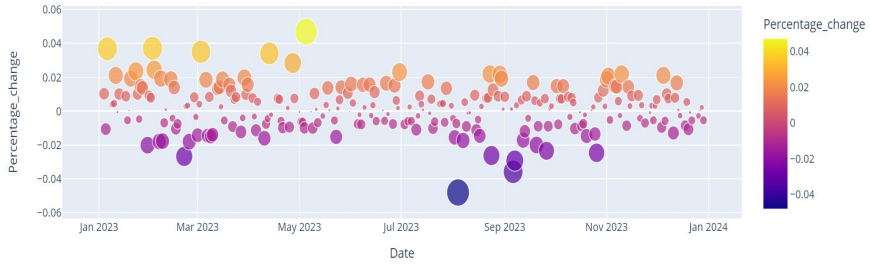
-



Candle Chart and Scatter Bubble - AAPL Analysis

```
candle_plot(Apple_df,"Apple Stock").show()
```


```
scatter_bubble(Apple_df).show()
```

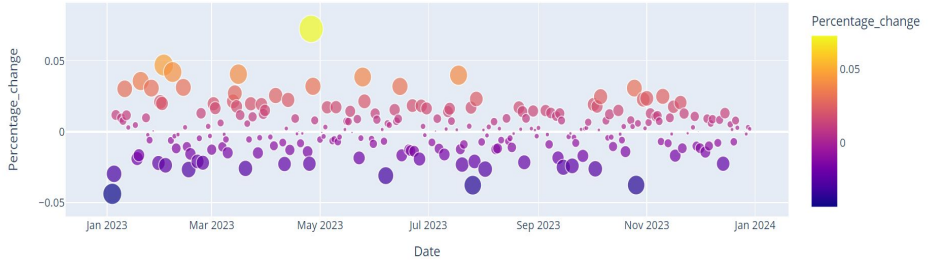


- Apple's stock price is largely range bound within +/- 2% based on 2023 volumes.
- The cumulative increase in stock price for Apple (AAPL) for 2023 is almost 60%.
- Negative (decrease) in pricing coincide with prominent price changes (around Aug 2023).
- May 2023 marks the only instance where AAPL price change recorded an increase in excess of 4%.
- Apple is largely a safe stock to invest in with limited risk profile but also steady returns.

Candle Chart and Scatter Bubble - MSFT Analysis

```
candle_plot(Microsoft_df,"Microsoft Stock").show()
```


```
scatter_bubble(Microsoft_df).show()
```



- Microsoft's stock price is largely range bound within +/- 5% based on 2023 volumes.
- The cumulative increase in stock price for Microsoft (MSFT) for 2023 is almost 50%.
- MSFT had multiple instances of negative price swings throughout 2023.
- May 2023 marks the only instance where MSFT price change recorded an increase in excess of 7%.
- Microsoft's price trends indicate that it is a slightly higher risk-reward profile stock than AAPL.

Candle Chart and Scatter Bubble - ORCL Analysis

```
candle_plot(Oracle_df,"Oracle Stock").show()
```


```
scatter_bubble(Oracle_df).show()
```

- Oracle's stock price is largely range bound within +/- 5% based on 2023 volumes.
- The cumulative increase in stock price for Oracle (ORCL) for 2023 is only around 25%.
- ORCL had very few, if any, negative price drops and that too only in the 1% range.
- Oracle is largely a neutral risk-reward stock with limited upside, based on the above analysis.
- ORCL may suit investors who seek to preserve capital and grow their investment via dividends.

Candle Chart and Scatter Bubble - INTC Analysis

```
candle_plot(Intel_df,"Intel Stock").show()
```


```
scatter_bubble(Intel_df).show()
```

- Intel's stock price has a high degree of variability from upswings of 1% to downswings in excess of 5% based on 2023 volumes.
- Despite a largely lower stock price, Intel has cumulatively increased in value by almost 65% in 2023.
- There are several price swings in it's daily stock price trends, perhaps reflective of the broader semiconductor market trends.
- Intel may largely suit niche investors in the semiconductor space with limited upside.
- Majority of the price increase with INTC is relatively recent (past six months) - potential investors may need to be wary of this.

Candle Chart and Scatter Bubble - GOOG Analysis

```
candle_plot(Google_df,"Google Stock").show()
```


```
scatter_bubble(Google_df).show()
```

- Google's stock price is largely range bound within +/- 5% based on 2023 volumes.
- The cumulative increase in stock price for Google (GOOG) for 2023 is almost 1.8x.
- GOOG had very few, if any, negative price drops but these have been large (in excess of 5%).
- Google is largely a higher risk-reward stock with more upside, based on the above analysis.
- GOOG may suit investors who prefer higher returns on their investment but also have a lower risk aversion.

Conclusion

- ETL framework using a combination of Pandas and SQLite provides a solid foundation to analyze broader market trends in the stock market.
- There is potential for a deeper analysis around correlation of stock price changes to broader market indices movement (DOW, NASDAQ) based on this simple framework.
- Investor sentiment around a broader portfolio of stocks can also be inferred based on such datasets, including the *fair market value* of a company relative to its peers.
- Augmenting the data from Yahoo Finance with other datasets like Kaggle around sector specific stocks (Technology, Industrial, Semiconductor to name a few) can provide segment-level analysis.
- Pandas (Dataframe) is a powerful way to transform data from multiple sources into a predictable SQL flow and be customized with different levels of varying visualization elements.


thank you!