# Project 3 – The Use of Data Engineering to Analyze Stock Prices

# Project members:

Manal Bayoumi, Rocio Cantu, and Melissa Mosby.

# Project Overview:

We decided to do the data engineering track. We did this track because we aspire to become Data Engineers after this program.

Data engineers are the backbone of data analysis. Data Engineering is becoming in great demand due to the growing of data analytics and big data and has grown by 30% in recent years. Data engineers make a significant impact on the world.

Here is a high-level outline of our project:

- Obtain stock price information,

- Clean and modify the data,

- Load the data into a database (SQLite3),

- Generate an ERD,

- Create HTML.

# ETL - Extract, Load and Transform

We created a jupyter notebook that uses pandas to pull historical stock price data from Yahoo Finance. This is a jupyter notebook file, with this file, user can download stock data from Yahoo finance API and then save to SQLite3 database.

We transformed the data by changing the dollar decimal points from 6 to 2 and removing a redundant floating point. We added these columns: stock ticker; normalization, which compares the gains and losses of stock prices; percentage change, which shows the percentage change in stock price from the day before; accumulative return, shows the returns from the market.

We added normalization because

We added percent change because

We added accumulative return because

We loaded our data into SQLite. We chose SQLite because our data is structured. Also we chose SQLite because it is good for lightweight applications, stand-alone applictions and situations that require portability. Our situation does call for portability in that we can upload the SQLite database file to GitHub for ease of use with our group members.

# Part Three: Creating HMTL with Streamlit to Present Our Project

For better user experience, our group build a interactive webpage with a streaming package not covered in the lecture called streamlit, this package use javascript plotly package build web.

## There are several reasons why my group decided to use SQLite:

\* SQLite is a SQL based database that can create a single database file that can easily be uploaded to GitHub and shared with other group members.

\* SQLite facilities the work when using multiple databases.

\* We can DB Browser read SQLite file and check if the file performs well.

\* SQLite is small like a paper; however, the other SQL is a big library.

\* Download DB Browser at this link https://sqlitebrowser.org.

# Project Ethical Considerations

Our project used public historical stock data. When considering ethics, we should handle the data with integrity. We should be transparent about any factors that could influence their analysis, interpretation, or reporting of results.

# References

“Alphabet Inc. (GOOG) Stock Historical Prices & Data.” Yahoo! Finance, Yahoo!, 16 Jan. 2024, https://finance.yahoo.com/quote/GOOG/history?p=GOOG

“Apple Inc. (AAPL) Stock Historical Prices & Data.” Yahoo! Finance, Yahoo!, 16 Jan. 2024, https://finance.yahoo.com/quote/AAPL/history?p=AAPL

“Intel Corporation Stock. (INTC) Stock Historical Prices & Data.” Yahoo! Finance, Yahoo!, 16 Jan. 2024, https://finance.yahoo.com/quote/INTC/history?p=INTC

“Microsoft Corporation Stock. (MSFT) Stock Historical Prices & Data.” Yahoo! Finance, Yahoo!, 16 Jan. 2024, https://finance.yahoo.com/quote/MSFT/history?p=MSFT

“Oracle Corporation Stock. (ORCL) Stock Historical Prices & Data.” Yahoo! Finance, Yahoo!, 16 Jan. 2024, https://finance.yahoo.com/quote/ORCL/history?p=ORCL

Smallcombe, Mark. “SQL VS NOSQL: 5 Critical Differences.” Integrate.Io, 9 Nov. 2023, www.integrate.io/blog/the-sql-vs-nosql-difference/#two.