INTRODUCTION **(MANAL)**

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.

WHY & HOW HTML FOR PRESENTION (**MANAL)**

We built an interactive webpage with a streaming package called streamlit. Streamlit is a user-friendly interactive HTML for data engineers. This is also a unique way to present our project and data. Our HTML is dynamic. We will show this throughout our presentation.

**<< CLICK ON REQUIRED PACKAGE BUTTON >> (MANAL)**

Here is the code and instructions for the HTML << Show them>>

INTERACTION WITH HTML (**MANAL)**

**<< CLICK ON THE DISPLAY THE DATA BUTTON >>** (**MANAL)**

Show the different companies data.

**MANAL’S SPEAKING PART ENDS**

EXTRACT LOAD AND TRANSFORM **(MELISSA)**

**<< CLICK ON THE ORIGINAL AND ETL DATA BUTTON >>** (**MANAL)**

I will discuss our processes for the ELT workflows. For all of the workflows, we used:

* Jupyter notebook as the interface to write our code.
* Pandas for reading, manipulating, cleaning our datasets.

EXTRACTION (**MELISSA)**

We used jupyter notebook to write our code, pandas to capture stock data from Yahoo Finance. We used the python library for yahoo finance, yfinance.

Our datasets are the historical stock data for Google, Apple, Intel, Microsoft and Oracle for the date range of January 1, 2023 through December 31, 2023.

TRANSFORMATION **(MELISSA)**

Original Data Includes: Date, Open, Close, High, Low, Adj Close & Volume

We transformed the data by:

* Changing the dollar decimal points from 6 to 2
* Removing a redundant floating point.
* Adding columns:
* Stock Ticker
* Index Number
* Normalization – This compares the gains and losses.
* Percentage Change – Shows the % change from the day before.
* Accumulative Return – Shows returns from the market.

LOADING (**MELISSA)**

**<< CLICK ON THE SQLITE BUTTON >>** (**MANAL)**

We chose SQL because: Our data is structured and the original data was in table form; can perform analytical queries, such as filters, joins, merges, and aggression on the data.

We chose SQLite because it allows for

* Lightweight applications
* Stand-alone applications
* Situations requiring portability – EXAMPLE: SQL can create a single database file that can easily be uploaded to GitHub and shared with other group members.

ERD (ENTITY RELATIONSHIP DIAGRAM) **(MELISSA)**

**<< CLICK ON THE ERD BUTTON >>** (**MANAL)**

This is the visualization for the relationships in our database.

We have 1 table to rule them all. The primary key is the “STOCK ID”.

Stock\_ID joins each company’s price table with their volume table.

The “date timestamp” joins the price and volume tables.

**MELISSA’S SPEAKING PART ENDS**

PLOTTING DATA **(ROCIO)**

**<< CLICK ON THE PLOTTING THE DATA BUTTON >>** (**MANAL)**

Used plotly in pandas/python to create our charts

This bubble chart shows the percentage change.

Talk more about the chart. Mention:

* What the size of the bubble means
* What the colors mean

Notice if we click on a company, the data changes:

* For the first company, pick a point & read it
* For the second company, pick a point & read it
* Bigger the circle the greater the percent change & vice versa
* Also, look at the top right, look at all the options you can use to interact with the data.

CANDLE DATA **(ROCIO)**

Also, notice if we click on a company, the data changes:

* Green means the close price for stock on that date is high.
* Red means the close price for stock on that date is low.
* The bottom is the range slider. It is automatically created when you plot the data.

ETHICAL CONSIDERATIONS **(ROCIO)**

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.

CONCLUSION **(ROCIO)**

**<< CLICK ON THE CONCLUSION BUTTON >>** (**MANAL)**

After completing this project, here are our thoughts:

1. Data engineering is not simple. We had a few stumbling blocks such as: picking the data, which SQL or NoSQL would be best for us, and how to present our project.
2. We enjoyed using jupyter notebook and pandas in our project.
3. Even though the process of data engineering was challenging, we think it can be useful for our data. For example:

* Data analysts can quickly query for these potential questions:
  + How do investors pick their stocks?
  + When buying and selling stocks, do investors depend on the opening or the closing prices of stocks?
  + What factors make stock prices rise or fall for a company like Apple, Google, Intel, Oracle or Microsoft?
  + What makes an investor pick one company’s stock over another?

Q&A Possible

Do we have any insights into future buying & selling.

No. This data looks at the past, not the future.

Can we tell what is affecting the rise & fall of each stock

Since this is a historical snapshot of companies, further research to pin point the reasons why the stocks rose and fell. In general, stocks rise and fall to due some of the follow:

* Supply and demand of stocks
* Market Trends
* Liquidity of a Company
* Company health
* Company’s Quarterly & End of Year reports
* Inflation
* Economic