Krishna Pulipaka
Lupita Lopez
Corey McCulley

ETL Project: Group 12

Presentation of the Top 21 SP 500 Stocks

Purpose:

We decided to investigate the process of pulling the closing price on a certain day for the top 21 SP 500 stock prices and then presenting them in an API format. This project would give us experience in the ETL process and in creating a way for other users to rapidly access this data.

Data Sources:

The financial data for the SP 500 was pulled from the https://finance.yahoo.com/ using the yfinance module in python. This provided us the SP 500 data for an extended period of dates and with a variety of price and trade information. We narrowed the data until we had the top 21 stocks on the date of the pull (10-22-2020) as defined by their closing price.

We pulled API data from https://rapidapi.com/natkapral/api/currency-converter5 which provided an API containing the currency and conversion rate to USD.

These data sources were chosen as they are free, globally available and extensive.

Data Processing:

The yahoo financial data was pulled using yfinance and the Wikipedia table of SP500 stock symbols. The ticker symbols were collected from the Wiki table and turned into a dictionary which was passed to yfinance to pull the historical closing price data. The top 21 stocks were then selected for further analysis and their data pulled for the date the analysis was performed (10-22-20).

The currency and conversion data was pulled from the rapid API. Due to the huge number of countries listed, 5 were selected for use in this analysis. The US, EU, China, India and Philippines were picked as the first 3 present the largest economies and the final 2 represent the home currencies of group members. These 5 currencies and conversion information were passed into a csv for further use.

Data Transformation:

The two sources were combined to create a final database containing the stock symbol/ticker, and the close prices in the 5 currencies. The yahoo financial data was collected in USD, so the

conversion factors were multiplied to calculate the equivalent value in the other 4 currencies. The 6 columns were then passed into a pgAdmin database using the stock symbol as the primary key. This database was then passed into a flask server and presented online as an API.

Data Type (Final):

The final database was presented as an API. This was chosen as it provides many ways for the final user to access the various currency values for each stock depending on their specific need. It also allows this collected data to be used for further data analysis projects through a simple API call.

Hypothetical Use:

End users of this database could be anyone looking to pull the closing prices for top stocks in the SP 500 for further analysis in financial trading. By providing different currency values, we are providing information that traders across the world could access and find useful in their market analysis.

As our work is just a basic framework, it could be expanded to include historical data and include the entire SP500. It also could be designed to allow the end user to select the present currency from the world wide list. While being a very basic analysis, with some imagination this project could be expanded to provide an important database of financial information for people across the world.

Schema of Final Database:

```
CREATE TABLE sym_price (
    symbol varchar,
    close_usd float,
    close_cny float,
    close_eur float,
    close_inr float,
    close_php float,
    PRIMARY KEY (symbol)
);
```