In this project, we tried to track the stock price change starting from January 1, 2019 to August 5, 2019, and further group the changes to evaluate performance by sector. To perform the ETL, we chose data in two types of formats from data sources, transformed and merged in Python environment and eventually manipulated and output in pgAdmin 4.

Extract: We extracted data from two different sources and format. CSV file listed all stock tickers, August 5, 2019 last price, sectors and other information in New York Exchange (source link see Appendix) and Json file through API retrieved from Financial Modeling Prep (source link see Appendix).

Transform: We transformed data in Python environment. For CSV. file cleaning, we performed dropna function in important columns such as Sectors and Last price, drop extraneous columns such as industry, IPO year and output the cleaned data in CSV. format for further processing. The uniqueness and standardization of stock symbols made them a good choice for a primary key in a relational database. On the other hand, we filtered only stocks last price on December 31, 2018 in Json file through API and joined cleaned CSV. data. After that we did calculations to show the stock price percentage change during time period.

Load: We loaded transformed data in pgAdmin 4 using Python psycopg2 and sqlalchemy packages. In pgAdmin 4, we wrote queries to group the average percentage change in sectors to show our conclusion.

pgAdmin 4 is a SQL database, there are several reasons we chose SQL database. First, it is quick and productive to run queries in SQL database. Second, our data are standardized and are good in relational database. Lastly, SQL database helps in removing redundancy and organized data in a better way.

After ETL process, we find that the best three performed sectors from January 1, 2019 to August 5, 2019 are Transportation (11.46%), Technology (7.54%), Finance (6.20%). For other sectors’ performance please refer to Graph 1.

A screenshot of a social media post

Description automatically generatedGraph 1. Performance record grouped by Sectors

Appendix:

New York Exchange stock data, source link <https://www.nasdaq.com/screening/company-list.aspx>

Financial Modeling Prep, source link: https://financialmodelingprep.com/api/v3/historical-price-full/