**ETL Project: Aggregating Data for Analyzing Stock Patterns in NASDAQ**

**Data Sources:**

**NASDAQ Datasets**

<https://datahub.io/core/nasdaq-listings#resource-nasdaq-listed>

**STOCK Datasets**

<https://www.kaggle.com/borismarjanovic/price-volume-data-for-all-us-stocks-etfs>

**Summary:**

I found a CSV file for NASDAQ with their listings for corporations, and field information, and NASDAQ listed symbols, and field information. I think it would be interesting to link the data and come up with the number of times the market was crashed. It would also be interesting to plot the earnings per share, and quarter ends to see the forecast vs the actual earnings of the company. Next, would be to see the relationship between the corporation and the number of times their stocks went down. I would be able to find a CSV file for specific corporations and the number of times its stocks were low. Linear regression could be used to analyze the relationship between corporation’s stock market information and the number of times the stock fluctuated overall. The goal of this project is to create a SQL table with clean data ready for above mentioned analysis.

**Technical Steps**

**Extract:**

Download the CSV files from Kaggle and load them into one Python notebook using pandas. Then browse the file to check column names and data types.

**Transform:**

Clean up the company’s name, earnings per share, quarter end, and net income column in the NASDAQ dataset to be the same format as the company’s name column in the Stock dataset. To do this, I would strip the stock string in the stock dataset’s company name column. I then would drop unrelated columns in the NASDAQ and Stock datasets, and rename the rest of the columns to ensure the naming is concise and clear. Then move on to the NASDAQ dataset, where I can extract the time component (date column) of the stock dataset. Then merge the NASDAQ dataset with the cleaned, merged stock dataset using inner join.

**Load:**

After cleaning the data and merging the datasets, I will create a database in MySQL titled Stock\_Market\_DB. Also create a table titled Stock\_Market\_Trends in Python notebook. Then, create an engine in the Python notebook to establish a connection between the SQL database and the notebook. Next step is to push the merged dataset into the Stock dataset in the SQL database. After that in MySQL alter the table to add a primary key and re-ordered the columns to have the columns with most useful data in the front.

**Note:**

Earnings per Share represents the portion of a company's profit allocated to each outstanding share of common stock. The net income (reported or estimated) for a period divided by the total number of shares outstanding during that period. The data that I would like to analyze would be for Apple Inc.

**Below graph can be created based on the above data sets:**

