Group ETL Project

### RAAW\_data\_group

Collaborators: Vien Vo, Wilson Aliaga, Angel Alejandro, Amy Reynolds

Question: How are stock prices affected by market conditions such as gross domestic product (GDP), consumer price index (CPI) and unemployment rates?

#### Extraction

1. Stock Information:
   1. Source: <https://finance.yahoo.com>
   2. Method: API call using module fix\_yahoo\_finance
   3. File type: API call.
   4. Tickers list for API call extracted from Nasdaq Exchange. <https://www.nasdaq.com/screening/company-list.aspx>
   5. Range: daily value for 15 years from 05/03/2004 – 05/03/2019.
   6. Data include: Ticker, Date, Close, and Volume.
   7. Four industries were selected: Healthcare, Energy, Finance, and Technology.
2. Economics Stats:
   1. Source: Federal Reserve Economic Data <https://fred.stlouisfed.org/>
   2. Method: direct download.
   3. File type: flat file: CSV.
   4. Range: quarterly value for 15 years from 05/03/2004 – 05/03/2019.
   5. Data include: CPI-Monthly, GDP, GDP\_Change, and Unemployment\_rate.
   6. All values are for USA only.

#### Transformation

1. Stock Information:
   1. Because economics stats were presented at quarterly value, we calculate mean of each ticker by quarter instead of by day. This process allows direct comparison between stock price and Economics stats.
   2. Normalization of the format of quarter to "YYYY-MM-DD"
   3. Convert downloaded data to Dictionary for ease of transfer to MongoDB.
   4. Specific industries were extracted by sector column to create tables based on key industries
2. Economics Stats:
   1. GDP CPI and unemployment datasets were grouped by quarter
   2. Convert them to dictionary to added to MongoDB.

#### Load

We decided to use MongoDB for ease of deployment. All datasets were loaded as the following tables

* Stock - All
* Stock - Finance
* Stock - Healthcare
* Stock - Energy
* Stock – Technology
* CPI data
* GDP

#### Definitions:

\* CPI -The Consumer Price Index (CPI) is a measure that examines the weighted average of prices of a basket of consumer goods and services, such as transportation, food and medical care. It is calculated by taking price changes for each item in the predetermined basket of goods and averaging them. Changes in the CPI are used to assess price changes associated with the cost of living; the CPI is one of the most frequently used statistics for identifying periods of inflation or deflation.

\* GDP - Gross Domestic Product (GDP) is a broad measurement of a nation’s overall economic activity. GDP is the monetary value of all the finished goods and services produced within a country's borders in a specific time period.

\_Definitions obtained from <https://www.investopedia.com/>