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

The two primary\* data sources used were:

* Federal Reserve Economic Data
  + <https://fred.stlouisfed.org/>
* Yahoo Finance
  + <https://finance.yahoo.com>

\*NASDAQ was also used to obtain a list of stock tickers: <https://www.nasdaq.com/screening/company-list.aspx>

#### Transformation

Datasets:

1. List of all tickers from NASDAQ showing sector (industry)
   1. downloaded manually as a .csv file
      1. Technology
      2. Health Care
      3. Finance
      4. Energy
2. Stock prices for 5 years for all publicly traded Technology companies on NASDAQ as listed in Yahoo Finance.
   1. This data is presented as daily listings and was obtained using API
3. GDP for 5 years
   1. This data is presented as monthly listings and was obtained as flat files (.csv)
4. CPI for 5 years
   1. This data is presented as monthly listings and was obtained as flat files (.csv)
5. Unemployment rates for 5 years
   1. This data is presented as monthly listings and was obtained as flat files (.csv)

#### Normalizing:

* Specific industries were extracted by sector column to create tables based on key industries
* Stock prices, GDP CPI and unemployment datasets were grouped by quarter
* Normalization of the format of quarter to "YYYY-MM-DD"

*\*The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).*

*\*The type of final production database to load the data into (relational or non-relational).*

*\*The final tables or collections that will be used in the production database.*

#### Load

All datasets were loaded in Mongo DB as the following tables

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

## Final Report contents:

\* Extract: your original data sources and how the data was formatted (CSV, JSON, pgAdmin 4, etc).

\* Transform: what data cleaning or transformation was required.

\* Load: the final database, tables/collections, and why this was chosen.

#### 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/