**ETL Project: S&P 500**

**Data required:**

* JSON : stock.json <https://datahub.io/core/s-and-p-500-companies>
* CSV: all\_stocks\_5yr.csv <https://www.kaggle.com/camnugent/sandp500>
* CSV: wiki.csv <https://en.wikipedia.org/wiki/S%26P_500_Index>

**Description:**

From the **URL** - url = 'https://pkgstore.datahub.io/core/s-and-p-500-companies/constituents\_json/data/64dd3e9582b936b0352fdd826ecd3c95/constituents\_json.json'

We directly extracted data and got the following information:

Company Names, Sectors and Ticker Symbols

From **Wikipedia**, we extracted the following information:

Ticker Symbols, Location of the company headquarters, Date first added and the Founded year.

We then merged the above 2 datasets into a third dataframe which has the following information:

Company Names, Ticker Symbols, Sectors, Headquarters, Date first added and the Year founded

This dataframe was then cleaned by replacing the empty cells with “Value Unavailable” so as to preserve the list of total stocks.

We then connected to postgres. Created a new database which was then used to convert the final dataframe into an SQL Table

The client or API may use this PostgreSQL data as they wish.