**ETL Project Report** : For our Data Deep Dive, we wanted to look at equity fundamental data vs. corresponding price action, on a yearly basis. The ultimate goal, if we were to embark on the statistical endeavor, would be to correlate fundamental company data (assets, liabilities, etc) vs. price action, to see how companies performed vs. the index.

**Extract**:

We used a Kaggle dataset for our fundamental data set, which had a list of 447 tickers, along with year fundamental data points derived from SEC filings over 2012-2017. Using pandas, we trimmed this dataset to include only the date range and data that we were interested in.

For the pricing data, we used a package called Quandl to pull publicly available pricing data using the tickers provided in the fundamental data set. We had to create an API key to do this.

**Data sources**:

Fundamentals CSV File: <https://www.kaggle.com/dgawlik/nyse#fundamentals.csv>

API Price Data : Quandl

**Transform**:

We extracted the data from both our sources into Pandas Dataframes.

For the fundamental’s data we only kept the columns that were relevant to our Project. And extracted all the Ticker’s into a List.

**Load:**

Using SQL engine, we then sent both data frames to SQL database (finance\_db) for manipulation and merging. Using groupby function, we created a table with tickers and avg price by year. Using year and ticker, we then merged this onto our table of fundamental data.

**Takeaways**:

-Pulling data was easier than expected. So much publicly available data

-Manipulating the data was harder than expected. SQL more intuitive than pandas when it comes to merging on double index