**ETL PROJECT REPORT 3.10.20**

**EXTRACT:**

Data for our ETL project came from the following sources:

* Data from Trumptwitterarchive.com extracted as a JSON file
* Stock data from Yahoofinace.com extracted with "yahoo\_fin.stock\_info import get\_data".

**TRANSFORM:**

Before joining the two datasets:

* **Trumptwitterarchive Data:**
  + Export the data as a JSON file
  + Removed unnecessary columns
  + Created data frame using Pandas.
  + Removed the time stamp from the date by creating a new column to reflect date only
  + Removed the column with date & time stamp.
  + Reordered column to match the stock data.
* **Yahoo Stock Data:**
* Stock data was read into Jupyter notebook.
* Reset index which was reflected as dates.
* Renamed index “date”
* Removed the time stamp from the date by creating a new column to reflect date only
* Removed the column with date & time stamp.
* Reordered column to match the stock data.

**LOAD:**

It was decided the data would be loaded into an SQL database with “date” as Primary Key.

* Both datasets were loaded into SQL using SQLAlchemy
* Data tables were created for each of the different types of Stock Data (NASDAQ, Dow Jones, S&P and Russell) and a data table for the Trump Tweets data.