**Project Report – Alan, Eileen, Kevin, Ritika**

Project Proposal

* **E**xtract:
  1. Get historical stock prices and associated industry and sector information.
     + historical\_stock\_prices.csv
     + historical\_stocks.csv
  2. Scrape current market price for a small set of tickers from [https://www.morningstar.com/stocks/xnas/{ticker name}/quote](https://www.morningstar.com/stocks/xnas/%7bticker%20name%7d/quote)
* **T**ransform:
  1. Drop “name” column from historical\_stocks.csv.
  2. Sort the data in historical\_stock\_prices.csv by ticker.
  3. Only use ticker, close and date columns from historical\_stock\_prices.csv
  4. Filter the data for the year 2018.
  5. Create a list of tickers for web scrapping current market values.
* **L**oad:
  1. Create 3 tables in pgAdmin for historical stocks, historical stock prices and current market value.
  2. Create queries for some stock analysis.

Project Results

* **E**xtract:
  1. Obtained datasets from Kaggle - <https://www.kaggle.com/ehallmar/daily-historical-stock-prices-1970-2018>
  2. Found a reliable and repeatable source for daily price quotes the Motley Fool - <https://www.fool.com/quote/>
* **T**ransform:
  1. The name was changed to firm\_name to make it clearer and eliminate ambiguity
  2. We sorted the data as we wanted to obtain a year’s worth of data for historical volatility analysis as compared to the current market price.
  3. We used Beautiful Soup to easily obtain the scraped current market price. Motley Fool was chosen as other providers, e.g. Yahoo/Finance and Morningstar has inconsistent results with different browsers.
* **L**oad:
  1. Created 3 tables in pgAdmin for historical stocks, historical stock prices and current market value.
  2. Created 2 queries for sample stock analysis. We wanted to compare volatile periods that occurred historically in the stock market vs current conditions.
  3. Created the flat file Resources/StockPriceSubset.csv to store the subset data for 2018 from the historical stock price file.