**Group 5 Project Proposal- Tariq Attarwala, Sai Reddy, and Veronica Valencia**

**Summary:**

We seek to organize Standard & Poor(S&P) 500 and macro-economic data. The organized data source will provide us the opportunity to understand the impact of 1.Inflation(CPI), 2.Employment(U.R), 3.Housing(HPI), 4.Consumer Spending (CS), and 5.Consumer Confidence (CC) on S&P 500 between 2019-03 to 2020-03.

**Scope:**

The macro-economic data will be extracted from FRED (Federal Reserve Economic Data) and the S&P 500 from yahoo finance. An ERD diagram that depicts the data lineage along with the primary and foreign keys for each table will be provided. A data dictionary that provides variable description and data type will also be included as part of the project. No analysis will be done on the data extracted.

**Technologies and resource contributions:**

Extract - Tariq

* Extract the Macro-economic variables from FRED
* Extract the S&P data from Yahoo finance

Transform- Veronica

* Read in and transform the extracted data into the required format using python.
* Create an ERD diagram that shows table relationships, data types and Keys using database diagrams.

Load- Sai

* Load the transformed data in postgres DB using python.
* Check the data quality, accuracy and integrity of the tables uploaded using postgresql.

**Source:**

* CPI : <https://fred.stlouisfed.org/series/CPIAUCSL>
* UR: <https://fred.stlouisfed.org/series/UNRATE>
* HPI: <https://fred.stlouisfed.org/series/CSUSHPISA>
* CS: <https://fred.stlouisfed.org/series/PCE>
* CC: <https://fred.stlouisfed.org/series/UMCSENT>
* S&P: <https://finance.yahoo.com/quote/%5EGSPC/history?period1=1553558400&period2=1590451200&interval=1d&filter=history&frequency=1d>