Group 9 - Project Proposal

Team Members:

- Joseph Monahan
- Paul Smith

Abstract:

This project intends to analyze the possible relationship between the federal interest rate and total consumer debt levels. This analysis will utilize data from 1991 to 2020, and will primarily pull data from the New York Federal Reserve (newyorkfed.org), and an open source, online community of data scientists known as Kaggle (kaggle.com). This data will be compiled into Excel spreadsheets, processed with Pandas, and finally loaded into a PostgreSQL database. This data, once loaded into the PostgreSQL database, will then be joined together based on their associated date.

Processing through Excel will take advantage of the ease with which data can be pulled from our source websites, with each website providing a way to directly down their data as an Excel spreadsheet. These spreadsheets will then be further processed via Pandas, where our data can be directly translated into a dataframe format. This initial data will give the federal interest rate in a monthly format, and the household debt data will in turn have a quarterly format. Additional data processing will be necessary at this point to get both data sets into the same format and translating our interest rate data into a quarterly format can be accomplished with Python scripting. Once this data processing is complete, we can then move onto loading our data into a PostgreSQL database.

The final data entry into our PostgreSQL database will be handled by Python coding, utilizing Pandas, inside of a Jupyter notebook. Data integrity will not be easy to validate, but we believe that a comparable number of entries for each data set will provide an initial indicator of our successful data transfer. Our final notes from this data transfer will then be compiled into a final report, for presentation at the conclusion of this project.