Clayton Nagle

Abigail Carreon

UCSD Data Science & Visualization Bootcamp

01 May 2021

ETL Project – Final Report

**Topic**: Comparing the daily Dow Jones Industrial Average (DJIA), Bitcoin, and the dollar value of gold, starting from the year 2013. This was changed from our original proposal from starting in the year 2009 when Bitcoin was released, as the only data available starts in 2013.

**Extract**: Historical data for the DJIA was pulled from the Wall Street Journal; Bitcoin data was pulled from CoinDesk.com, and gold price data was pulled from Gold.org. Data was either already in csv format, or was converted into csv from an Excel workbook.

**Transform**: The data for DJIA and Bitcoin pulled from the sources included the date, Opening Price/Score, 24-hour high, 24-hour low, and Closing Price/Score. Data was cleaned to use the date and Closing Price/Score. Data pulled for gold prices included annual, quarterly, and monthly averages, annual, quarterly, and monthly end of period, and daily prices in USD, EUR, JPY, GBP, CAD, and CHF. Data was cleaned to only use the dates, and daily price information in USD. We each worked on pulling and cleaning up the data sets for this project. Clayton was responsible for the data for DJIA and Bitcoin, and Abigail the gold prices. Each worked in their own Jupyter Notebook and pgAdmin databases, which were later merged into a single Jupyter Notebook and database.

**Load**: The data sets, final Jupyter Notebook, and SQL file are loaded into a shared Github repository. This topic was chosen because we were looking through Kaggle for ideas of what topic to use, and we found a data set on historical Bitcoin and gold prices. We thought it would be interesting to compare the prices of the two, along with the DJIA as a base index.