Kevin Black

Gninet Dagnogo

Stephanie Nickerson

Jourdan Norman

Pete Turnham

With increasing hype around cryptocurrencies, we decided to dig into the prices of bitcoin and gold. We understood that these prices would contrast but finding a data set proved more difficult than expected. On Kaggle, we discovered that many of the formatted, available datasets could make our data transformation more difficult or, in some cases, impossible. Some data sets had in-depth exploratory data analysis geared more towards machine learning. Other data sets had sparse information, several null values, or did not provide a side-by-side comparison. In addition to the exploratory analysis, many of the Bitcoin datasets included minute intervals.  These minute intervals also made comparisons to the Gold datasets almost impossible.

Eventually, we came across a couple of workable data sets. The first and most difficult data set to find was for Bitcoin. This data set was cumbersome due to the limited historical information provided coupled with the daily dates. Considering that the bitcoin is valued using the USD, we had to compare it to another USD valued commodity. The commodity we decided on was gold. Fortunately, the data set for gold was easier to find, but the information did not align well with our bitcoin data set. Our first issue was finding corresponding time frames for the data. We were able to find datasets for Gold dating back to 1980 but we were only able to find more recent data for Bitcoin.  As a result, we focused on timeframes that were common between both Gold and Bitcoin, 2012 through 2020.

Once we settled on the relevant years, we needed to merge these data sets, but the values provided for gold were monthly averages. These various data formats required converting the bitcoin dates into months and averaging the close price to obtain *the price per month*. We accomplished this by changing the date format from YEAR/MONTH/DAY to YEAR/MONTH. In addition to changing the data format, we also had to change both files from Object to Float for the price and datetime for the date. From here, we were able to group by the year and month. Once the group-by was complete, we took the close prices and averaged them by month. The result was monthly bitcoin data that we compared and merged with our gold data set. The primary change required in SQL was using varchar for the date since our newly formatted date functions wouldn’t work in SQL as a normal date function.

Overall, we found our project and the resulting analysis to be very interesting, considering the increased interest and investment in Bitcoin.  It was beneficial to find and explore data that would allow us to compare Bitcoin against a US currency standard, Gold.  The  most important benefit of our project is that it allows anyone to perform exploratory data analysis (EDA) on whatever you can think of.  This is possible because the data is accessible via CSV and a SQL database.  As a team, we learned how to work together to pull together the information.