DATA ANALYTICS BOOT CAMP - PROJECT 2

**Cryptocurrency Values over Time**

Margaret Oluwatoyin sippiotitiloyeola, Dylan Bruce, and Mirella Mardakis

April 2019

**Purpose:**

The purpose of this project is to display the values per day for the past six year of some popular cryptocurrencies. This information is presented alongside the same historical data for the S&P 500 stock for comparison of the behavior of cryptocurrencies versus traditional stocks.

**Extract:**

We extracted CSV files from Kaggle.com for the cryptocurrency data, and one CSV file from Yahoo Finance for the stock market data.

**Transform:**

* We converted the CSV files of the cryptocurrencies into Pandas dataframes, deciding to use all the existing columns, as the information is relevant to analyzing value fluctuations over time.
* When creating table columns in SQL, some column names had to be used as strings, as their names were also existing data types (such as “date”).
* The date columns in the cryptocurrency dataframes had to be converted to the SQL format of YYYY-MM-DD. To do this, we used datetime, strftime and strptime to convert from Mon DD, YYY to YYYY-MM-DD. This function was looped over the “date” column in all dataframes.
* The null values in the Volume column in the original data were specified by a dash (“-”); we had to replace that with a blank character in order to import to SQL as an integer.
* The data for the S&P 500 has a column called “Adj Close” which we removed when cleaning the data, as it reflected the same values from the “Close” column.

**Load:**

* The data is loaded into a SQL relational database. This type of database works best for displaying financial trends because it makes joining tables and manipulating numerical data very easy, and this type of information is more efficiently displayed in tables.
* In order to more easily share our work, we decided to use SQLite. A SQLite file was saved in our repo.

**Final Tables/Collections:**

* Five tables were created, one for each cryptocurrency and one for the S&P 500.
* The primary key in all tables is the date.

**Data Sources:**

* Cryptocurrency historical data: <https://www.kaggle.com/sudalairajkumar/cryptocurrencypricehistory>
* S&P 500 historical data: <https://finance.yahoo.com/quote/%5EGSPC/history?p=%5EGSPC>