ETL Project- Data Analytics and Visualization

Members: Chris, Victor, Shannon

The Ask:

• The sources of data that you will extract from.

* We are extracting data from:
  + <https://www.kaggle.com/jessevent/all-crypto-currencies>
  + <https://www.kaggle.com/philmohun/cryptocurrency-financial-data>

• The type of transformation needed for this data (cleaning, joining, filtering, aggregating, etc).

* We found a csv file from Kaggle, which was formatted to correct US numbers.
  + Date column was inverted to date/time type
  + Data was filtered and cleaned
* We converted a csv file into a .json using Python in Jupyter notebook.
  + We parsed the data from the .json file and renamed the columns we need to match up with the columns of our csv file.

• The type of final production database to load the data into (relational or non-relational).

* We used mySQL (relational database) to load the csv and .json files.

• The final tables or collections that will be used in the production database.

* Final collection to include "Currency", "Date\_open", "Open\_price", "High", "Low", "Close\_price", "Volume", "Market\_Cap" as combined from our two datasets.

You will be required to submit a final technical report with the above information and steps required to reproduce your ETL process.

Our ETL Process:

Our first step was to find two datasets that showed a history of cryptocurrency. We decided to use two files from Kaggle which showed consolidated financial information.

\*\* Datasets Used:

1. A total history of Cryptocurrency prices, including opening and closing numbers, highs/lows, and market values.

<https://www.kaggle.com/jessevent/all-crypto-currencies>

2. This data was pulled from coinmarketcap.com to collect information on the trends of cryptocurrency with the intention of predicting future trends.

<https://www.kaggle.com/philmohun/cryptocurrency-financial-data>

ETL Project Report

Logic: We decided to use historical price information pulled from CoinMarketCap.com using an API request. By joining this dataset with the set showing the rank and spread, we expect to do exploratory analysis to answer the following questions for investors:

- What are the trends for the top 10 currencies?

- Which day of the week is the best time to buy/sell?

- Which currencies are the least volatile?

- How does the shape of the trendline open prices correlate with that of Bitcoin?

- Which models can we use for forecasting highs/lows?