ETL Project

By: Filipa and Nathan

Date: 01/29/20

Cryptocurrency enables permissionless access to money and financial tools. Unlike traditional markets, anyone with the adequate programming skills can create blockchains or even create exchanges. Given the nascency of this technology and the lack of cohesive international legislation, there is strong reason to believe that nefarious activity is commonplace. Therefore, an analysis of cryptocurrency exchanges may expose questionable activity. In this project, we wanted to compare the volume of exchanges (converted to USD) with the Alexa rank of their websites. We used this comparison as a first approach to evaluating the possibility that the reported traded volumes are inflated.

We determined which exchanges to analyze by calling the top 100 exchanges with the highest volume between the following trading pairs: Bitcoin/USD, and Bitcoin/USDT. Bitcoin was chosen because it is the most popular cryptocurrency with the largest marketcap. USD and USDT are the most common trading pairs with Bitcoin. USDT is a crypto-equivalent of the USD trading at a 1:1 ratio. It is used because there is less friction required to trade two cryptocurrencies than a cryptocurrency into a fiat currency.

**EXTRACT**

We sourced our data from two APIs: Crypto Compare (<https://min-api.cryptocompare.com/documentation>) and Alexa Rank (<https://awis.alexa.com/developer-guide>).

We called the top exchanges for each of the two pairs being analyzed. Naturally there was plenty of overlap between the list of these two exchanges. To combine the two lists, we used a union join:

exchanges = list(set(exchanges\_usdt) | set(echanges\_usd))

With a for loop, this list of exchanges was used to call an API to find the volume for each exchange, as well as a second API to find the website for each exchange.

With the Alexa Rank API, we extracted four variables performing two different requests, following the documentation. The first request returned the rank, page views per million and page views per user for each exchange, feeding their url. These were saved in lists, which were then put together in a dataframe with the list of urls.

The second request returned the visitors by country per exchange, also feeding their url. Each list of pair of countries and respective percentage of visitors for each respective exchange (url) was stored in a dataframe, resulting in 43 dataframes.

**TRANSFORM**

We combined the dataframes with information on traded volumes and with the rank, page views per million and page views per user for each exchange and cleaned it.

We kept the dataframes with visitors per country for a later stage of the project.

**LOAD**

Mongo was used for our database because we anticipate updating this table as we progress with this project. Having a flexible model such as MongoDB will allow us to increase the variables of our database in a lean way that does not require significant restructuring.