

# **Capstone Project 1: Data Wrangling**

## **Description of initial dataset**

The initial data set consisted 785020 entries of 1643 unique crypto currencies. Each entry has 13 variables, the slug, symbol and name of the crypto currency are the first three variables. The next two columns recorded the date that each observation was recorded, the rank the coin held during that date respectively. The next few columns are as follows, the opening price in dollars, the highest price in dollars the crypto currency peaked at during that day, the lowest price in dollars the crypto currency dropped to during that day and the closing price in dollars. Then the total volume of coins traded that day and the market cap are recorded and finally the last two column record the close ratio and the spread.

## **Preparing the data**

For this analysis only the top three ranked crypto currencies will be utilized, Bitcoin(BTC), Ethereum(ETH) and Ripple(RPX). The observations for these crypto currencies have been sliced out of the original data frame and put into their own dataframes respectively. The date column is then used as the index for the data frames and it is sorted in an ascending order.