Proposal by Thomas A Duffy

Capstone 1: Due 9/5/2019

The Real Worth of Cryptocurrency

Introduction

Cryptocurrency is currently a phenomenon that has taken social media and wall street by storm. It came from being an unknown digital token based off of an ingenious math equation to becoming a way for the average person to make millions of dollars without even knowing the functionality of it. The purpose of this capstone is finding out the reason on why this wealth and dramatic increase of price happened in the cryptocurrency market by analyzing relevant data. First we must look at the what creates the value of a currency; normally the value of a currency is relative to the health of the country/s which it comes from or used(i.e the USD($) value is based of the economic health of the U.S.A). For cryptocurrency though the value can not be based off of a country due to its decentralized nature, so therefor what makes this value? Though I do believe there is inherent value in cryptocurrency, my hypothesis is the value recently has been socially constructed and does not represent its true value. In the data I have collected I will be looking for a correlation between social media(focusing on twitter data) mentions/metrics and the price/volume of two cryptocurrencies. I will be focusing on Bitcoin and Ethereum as those are the two coins I have collected sufficient data for proper analysis.

The Data so Far

So far I have found raw data from a public data set which contains individual .json files for all the cryptocurrencies. These separate json files contain four key headers: market_cap_by_available_supply, price_btc, price_usd, and volume_usd. Though there is a clean version of this data, in order to get more practice I am going to structure this data from the raw source files. For each key there is 1733 rows of data, each row containing a list of two values which I believe are a UNIX time stamp and the relevant metric beside it. I am also using a twitter data set of relevant cryptocurrency tweets that is relatively clean, I will have to do analysis on this data though to find the count per day and possibly a sentiment if possible. Location and also time can also play a factor as cryptocurrency markets do not close so usually when it is late in the western side of the world, the east is just waking up. The majority of my EDA is going to be surrounding these .json files to figure out what these values actually mean and how to organize my data properly in order to be able to do my anaylsis properly. Next I will have to figure out a way to vector my twitter data in order to graph it and also produce a correlation statistic.

Minimum Viable Product

I am aiming to at least be able to graph the numeric count of tweets of the cryptocurrency by the price and volume metrics and create a correlation metric between the two. This probably will be utilizing numpy/pandas arrays and data frames to do so. If I am able to complete that, I would also like to look for more data to do this with more coins than just Bitcoin and Ethereum;In addition, graphing big events (like Facebook announcing Libra) and other big data points which I would like to research.