Andrew Otte

ECON 4980

Clemson University

Problem Set 1

For my assignment I originally decided to scrape BoardGameGeek.com, but I was struggling with retrieving the data that I wanted so I switched to CoinMarketCap.com.

After testing that my scraping program was working I scraped CoinMarketCap.com for three days, requesting every 7200 seconds, or 2 hours. Afterwards I wrote the parsing program to take the data I wanted from the html files and put them into a .csv file. I parsed the name of each currency, the symbol, price, market cap, circulating supply, volume and scrapping time. When recording the numerical values such as price or volume, I made sure to not just parse the text from the tag as this would give me unusable formatting with dollar signs and commas. I then reordered the csv after saving it for better organization and changed all the ‘None’s in the file to ‘0’. At this point I tried applying all the techniques shown off in class, such as KMeans and KNeighborsClassifier, but I could not get these to work with my data. I did however manage to run an OLS regression for the Market Cap, with Price, Volume and Circulating Supply being the x variables. This helped me determine the relationship between these variables and the Market Cap, with an R squared of .524. After this I ran another linear regression to predict the Market Cap of a few currencies when presented with Price, Volume and Circulating Supply data. The program gave me accurate predictions with the data provided. I spent quite a bit of time trying to apply modules such as Kmeans and KNeighborsClassifier to my data but it always ended in errors.