Introduction

How do the cryptocurrencies behave? What are the causes of the fluctuation of cryptocurrency value? How can we predict the future value of cryptocurrencies?

There are a huge number of articles on the cryptocurrencies, are widespread with speculation these dats, with thousands of guide and forum advocating for the trends that experts expect to emerge.

Therefore, in the Data Mining assignment, we are going to explore and analyze the trend of cryptocurrency value by using the data mining technique.

Web scraping and BeautifulSoup

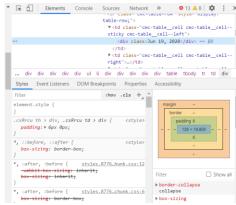
Normally, the cryptocurrency data we are looking for is on a web page, which is neither ready to be downloaded nor available with API. Meanwhile, we can perform web scraping technique by scraping the web pages CoinMarketCap with Python using BeautifulSoup and requests.

In CoinMarketCap, we have selected the top 20 cryptocurrencies from the list and identified the URL structure as the page we want to scrape for 20 cryptocurrencies.

The base url for web scraping is shown as above, we are required to change the highlighted part to the name of cryptocurrencies that would like to scrape, the historical data including closing, opening, highest, lowest, volume and market capacity, will being scrape.

Then, inspect on any of the value from the browser. The HTML line highlighted in gray corresponds to what we have seen on the web page.



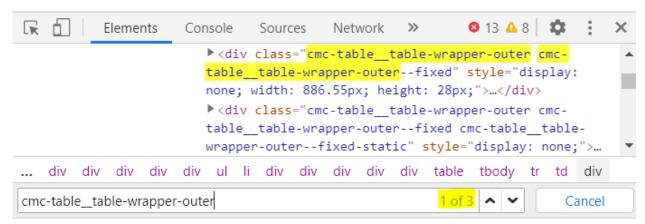


In Python, we were using BeautifulSoup to parse the HTML content. We parsed the response.text by creating a BeautifulSoup object, and assign this object to 'soup'.

```
url = base_url.format(cc)
response = requests.get(url, headers = headers)
soup = BeautifulSoup(response.content, 'html.parser')
```

There are many *div* containers on the webpages, we need to figure out what distinguishes the cryptocurrencies data from other *div* elements on that page. Often, the distinctive mark resides in the *class* attribute.

However, we noticed that there are three *div* containers were using the same *class* id, which is 'cmc-table_table-wrapper-outer',



We found out that the data value for cryptocurrency that we would like to scrape is with the last *div* container. Inside the container, we search for all rows with tag *tr* and all columns *td* within each *tr*. Then we parse all the content of columns from row to row, and create a dataframe for the dataset. Finally, we cleaned and changed the data types for certain columns and saved it as csv file.

Date	Open	High	Low	Close	Volume	Market Capacity
2020-05-25	8786.11	8951.01	8719.67	8906.93	31288157264.0	163760453116.0
2020-05-24	9212.28	9288.4	8787.25	8790.37	32518803300.0	161610414643.0
2020-05-23	9185.06	9302.5	9118.11	9209.29	27727866812.0	169305492440.0
2020-05-22	9080.33	9232.94	9008.64	9182.58	29810773699.0	168807619957.0
2020-05-21	9522.74	9555.24	8869.93	9081.76	39326160532.0	166947987864.0
2020-05-20	9725.33	9804.79	9447.2	9522.98	36546239703.0	175050963475.0
2020-05-19	9727.06	9836.05	9539.62	9729.04	39254288955.0	178831635026.0
2020-05-18	9675.69	9906.03	9570.36	9726.57	41827139896.0	178779483464.0
2020-05-17	9374.93	9823.0	9349.55	9670.74	40084250663.0	177745404470.0
2020-05-16	9333.24	9564.2	9260.69	9377.01	36164766408.0	172340956579.0