Data Collection

Except for Bloomberg data, all other data used in this research was free to download. Free alternative places to download Bloomberg Market Data do appear to be available https://datahub.io/. The data is contained in 3 csv files named: **Project_Set_BlockchainData.csv**, **Project_Set_MarketData.csv**, **Project_Set_OHLC.csv**. The csv files contain data which dates from 2014-02-16 to 2022-02-28. Some cleaning was applied to the data in excel for ease of use. **Project_Set_OHLC.csv** contains (3.2.1) Open High Low Close Volume Traded data for Bitcoin (BTC), and Close Traded data for Ripple (XRP) and Ethereum (ETH), for the Kraken exchange. As neither XRP nor ETH were trading on Kraken in 2014, much of their initial datetime values are 0. The data on Kraken's website only goes up to 2021-12-31 so to collect up to date data the KrakenexAPI was installed to python and utilised. **Project_Set_MarketData.csv** contains (3.2.5) macroeconomic market data. Whilst downloading this data from Bloomberg I realised the formatting was inadequate. Data downloaded contained missing weekend entries as traditional markets do not operate on weekends. When copying the data to a csv file the formatting from Bloomberg's database only allocated a single cell space in between weekends, some bank holiday days were also omitted. As different markets had different days missing data, the spacings for each market feature were not the same. I looked through each feature individually and realigned the cell spacing so that the csv contains spacing which represents 7-day weeks, with blank data left on weekend days, blank data from weekdays was filled with the previous days value. **Project_Set_BlockchainData.csv** contains (3.2.3) Blockchain data, (3.2.2) Kraken Bitcoinity data, and (3.2.4) Google Trends data. Google trends data was not difficult to obtain but costly in terms of time. This was because downloading Google trends data for our entire 8-year period automatically formatted the data into monthly periods. To obtain daily data, I had to download each individual month for each of the 8 years for all 3 keyword searches. The values of this feature represent the volume of interest on google search engines for the key words specified. Each month I downloaded had weighted scores that were local to that month. This meant rescaling each daily data point for each month by the overall monthly weights from the whole time-period. It should also be noted that words of interest are not position specific, i.e., 'bitcoin buy' is the same as 'buy bitcoin'. Blockchain data was collected directly from the Blockchain.com charts API. Even though their API contains functionality that lets you specify what period and frequency to collect, certain features could not be obtained as the APIs same functionality had not been implemented for those features at that time. The graphs I am referring to are listed under the 'Market Signals' heading for Blockchain.com charts. Bitcoinity contains bitcoin data for many different exchanges, I chose to only take Kraken specific data as this was the exchanges price I was looking to forecast. (3.2.6) Technical indicators data is not contained within these csv files as it was created in Python. It might be important to note for anyone who wishes to download and use this data in conjunction with my code that Microsoft Excel can sometimes reformat the date column. To ensure the date column is formatted properly right click on the date column and click format cells. Once there, navigate to date, select the bottom configuration i.e., yyyy-mm-dd.

Note: The numbers in parenthesis before data categories are a reference to the project write-up, section 3.2 – Data. There you can find links to the websites that the data was collected from as well as a list of all features and their descriptions