***Certificate in Data Analytics for Finance – Assessment***

***Cian Tighe***

***GitHub URL –***

***Abstract***

With the report, I am aiming to analyse the historical pricing information of Bitcoin and other cryptocurrency prices, to see if the movement in the prices in these coins have an effect on the other. If so, why? Is there a leader? Were there periods where this correlation was lost? Throughout this project I plan to amalgamate the closing prices of key cryptocurrencies over a period of almost 5 years into a single data frame to compare and contrast. I also plan to compare these datasets to the S&P500, a key index in indicating the overall health of the world economy – particularly in the USA.

Using key lessons that I have obtained during the Data Analytics for Finance course I will use the Python programming language in the JupyterLab environment using Anaconda Navigator to take datasets of each cryptocurrency and index used. This will allow me to take data from various sources, modify this data, make it ‘cleaner’ and visualise it to give us a deeper intuition into each element of the data and develop our own insights.

***Introduction***

In the past 5 years, cryptocurrencies have exploded. From being a niche interest of a handful of people to a part of the general public’s vocabulary, the world of cryptocurrencies is something that excites, confuses, scares and divides people with over 20,000 cryptocurrencies currently traded publicly (Smith and Rosen, 2022). Simply put, the mission statement of any crypto currency is to stem away from the control of money by a central authority and to give the general public control over their money – without discrimination. The scary thing about this decentralized world is the uncertainty, the lack of a guarantee, we have all heard of these ‘crypto winters’ where a cryptocurrency’s value can completely wipe out and lose someone their entire investment – it is extremely volatile.

Bitcoin was created in 2009 and is the most well-known coin and with good reason. Bitcoin has the highest market cap, near double of its nearest competitor Ethereum and a single Bitcoin at one point cost nearly $70k, an eye-watering amount in comparison to the $0.30 it was worth nearly 10 years ago. It’s almost so influential that crypto and Bitcoin are nearly synonymous with each other. People often equate the two, even though it’s one of over twenty thousand. Using techniques and applications learned in this course, I have analysed information from these multiple coins to see how linked these coins are to each.

***Dataset***

The historical pricing data I used was sourced from Yahoo Finance. For each coin, the data stemmed from the 9th of November 2017 to the 7th September 2022, the flat files were all in .csv formats and all had 1764 rows with 8 columns (for the purpose of this project we are only concerned with the ‘Date’ and ‘Close’ columns).

As I alluded earlier, there were many coins to choose from. Ultimately, I chose eight of the most popular cryptocurrencies – Bitcoin (BTC), Ethereum (ETH), Cardano (ADA), Ripple (XRP), Bitcoin Cash (BCH), Dogecoin (DOGE) and Stellar (XLM), all tracking its price relative to the USD over the same timeframe.

***Implementation***

* Starting off my Python notebook I imported the pandas and numpy packages as I knew straight away I would be making use of these. Then, I left the rest of that cell free so I could keep all my imported packages in one easy to find spot.
* Before I began to load in any data, I created a blank data frame using pandas that I could load the information I needed from each file into the same data frame to analyse. I called this data frame *‘dfcom1’***.**
* I created a bespoke function called *csvtodataframe* which allowed me to:
  + Read in a .csv file and turned it into a data frame.
  + Renamed each ‘Close’ column to ‘Close *\*name of coin\*’* so I could differentiate between each coin’s closing price when I merged the datasets together.
  + Added in a ‘Year’ column to each dataframe with the year the closing price was taken from. This allowed me to split to the data frame in future to multiple data frame – one from each year. For this I created another function called *year* which entailed of splitting up each date in the Date column in the format DD/MM/YYYY using *str.split* and splitting by each “/”.
  + The result of this function would return a new data frame with a year column and the close column renamed from the inputs of the file and the name of the coin.
* Starting with the Bitcoin csv I used this function to transform the file into a data frame, assigning the name *dfybit* to it, I would later repeat this process with each other coin used.
* I created two more blank data frames, one for simply the historical pricing of the coins during 2021 and one for the historical pricing of the coins throughout the year to date – calling them *dfcom2021* and *dfcom2022* respectively.
* After this, I used a tailored function called *splitbyyear* to split our data frame into a data frame for each unique year, using the *groupby* function on the Year column we created to help us achieve this. I applied this function to the Bitcoin csv.
* This returned a list of data frames for each year and added a Closing column to each of their corresponding dfcom data frame.
* By importing the pyplot module from matplotlib and the seaborn library I was able to move onto creating plots to visualize this data:
  + I used a pandas .plot() function to represent my dataframe using a bar chart. Using the *groupby* function again to group the dataframe to each year and took the mean closing price from each to graph the mean closing price of each year. I added a title and label for each axis. I put this project into a function called *plotmeanbyyear* so I could use it again for each other cryptocurrency I analysed.
  + Using similar methodology, I formed a function *plot2021plot* to create a plot displaying the closing price average for a given coin during 2021. I created a new function *month* in the same guise as the *year* function mentioned before.
  + Following this, I formulated a one-line function *scatterplot* to implement across all coins. Using the seaborn function *.scatterplot()* I was able to represent the day-by-day pricing figures of each unique currency visually. The hue argument in the scatterplot allowed me to distinctively show how the price transitioned year by year. I called this function *scatterplot* and like all my plot functions thus far, considers two arguments, the data frame to visualise and the coin it represents.
  + I applied these three functions to create a new encompassing function called *plots*, which I planned to use to transform a .csv file into a data frame and create the three plots above for any piece of historical data I use, each time adding in their own column to the different combined data frames.
* Testing these functions on the Bitcoin file with success, I applied the *plots* functionto the other six .csv files for the other six coins – this allowed me to also join and merge all the data I needed for analysis together by adding each coin’s closing price column to the original data frames.
* After checking the now completed *dfcom1* data frame for any nulls using the *.isnull().sum()* function, I set the index to be the ‘Date’ column using *.set\_index()*. This allowed the data frame to look cleaner.
* For each data frame *dfcom1* (2017-2022), *dfcom2021* (2021) and *dfcom2022* (2022 YTD), I created a correlation matrix using *.corr()*. To visualise this better I represented these correlation matrices using seaborn’s *.heatmap()*.
* The last bit of analysis on the cryptos was finding the Coefficient of Variation of each. I created a formula for this *coeffvar* using the numpy library to get the mean and standard deviation of a given column. I signified this data in a data frame.
* When thinking of perhaps predicting future movements in a particular crypto’s price, regression first came to mind. So, I used seaborn’s *lmplot* to showcase the line of best fit on the Bitcoin v. Ethereum graph. How machine learning and linear regression models could advance this analysis definitely entered my mind when looking at this graph.

***Results***

For each coin analysed, I have the visual representation of their closing price throughout the time period in the data sets used. I only included the Closing Price Average by Year and Closing Price Average by Month in 2021 for Bitcoin to make the write-up a little cleaner but these are all available in the .ipynb file but these allowed us to deeper dive into certain time periods inside our greater data set. Using scatterplots for the larger data frame allowed me to easily showcase my data even though it had a lot of data points. It also neatly each data point, including our extremes and illustrates very clearly trends in the data in a specific time period or also over a longer period of time. Scatterplot would also be the most suitable if we wanted to kick on and go deeper with regression analysis.

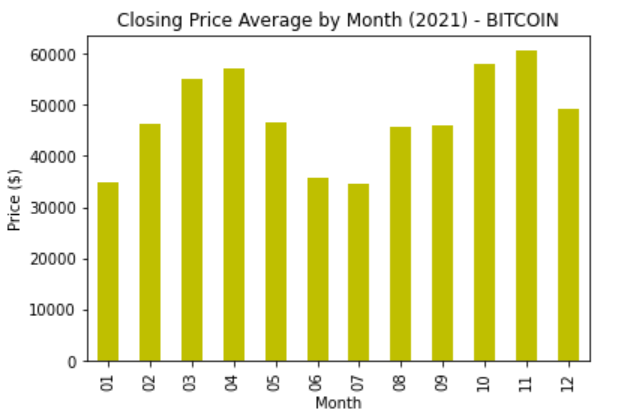
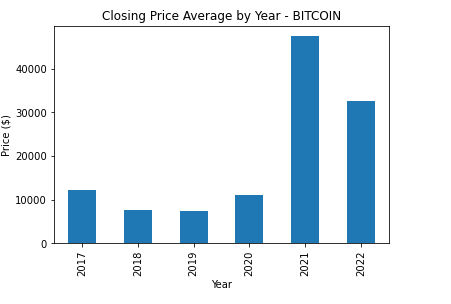
Each graph I used was titled clearly, as were each x-axis and y-axis. Each scatterplot made use of a legend to denote what year the data point belonged to via colour differentiation.

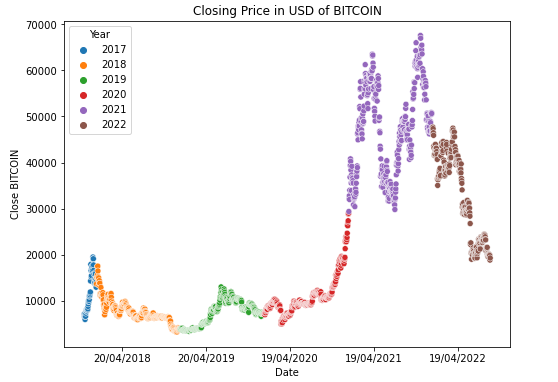
Correlation Plots are also formulated to see the correlation between each pair of coins in the full time period, 2021 and 2022 YTD. Seaborn’s Heatmap through the use of colour saturation clearly exemplifies the strength of each correlation and makes analysis that much easier.

Included also is a data frame of each coin and their corresponding Coefficient of Variation ().

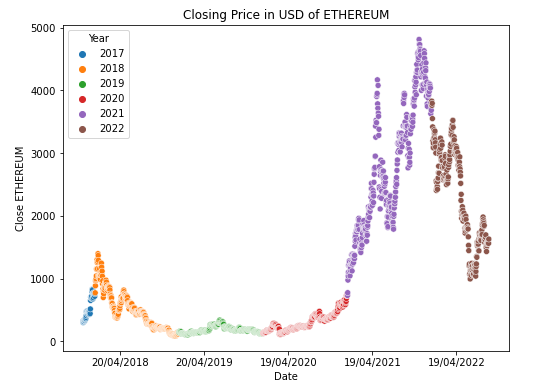
Also included is the closing price of Bitcoin throughout 2022 YTD plotted versus its corresponding Ethereum with a line of best fit added, this gives us a visual of how linear regression can enhance our understanding.

***Bitcoin***

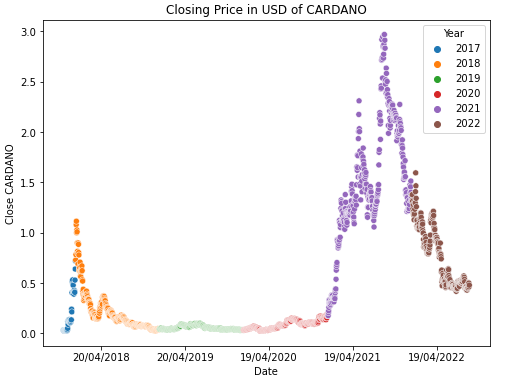




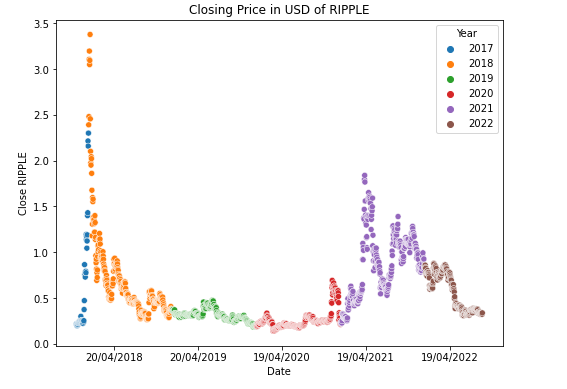
***Ethereum***



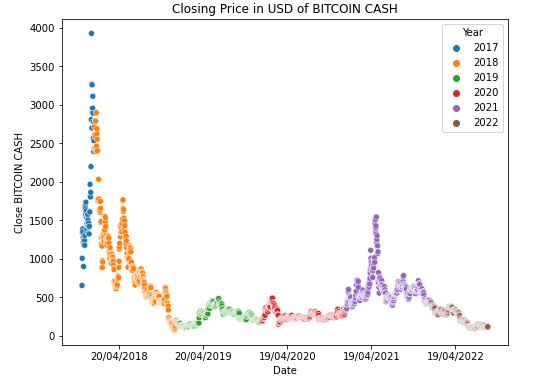
***Cardano***



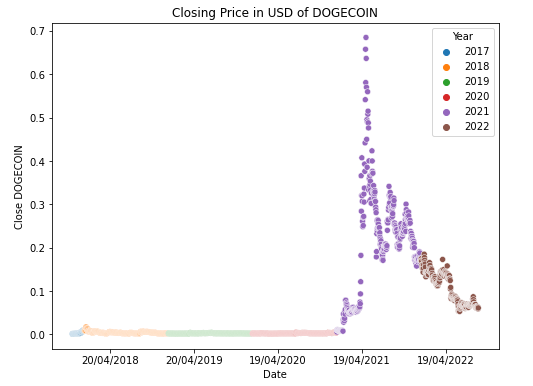
***Ripple***



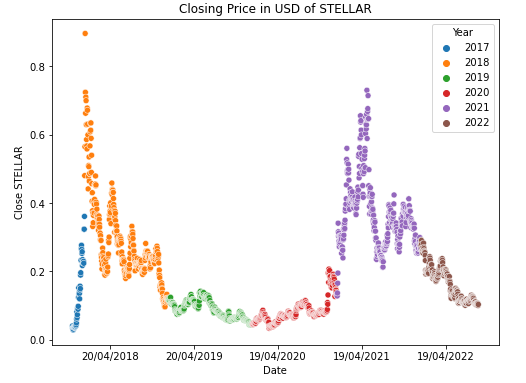
***Bitcoin Cash***



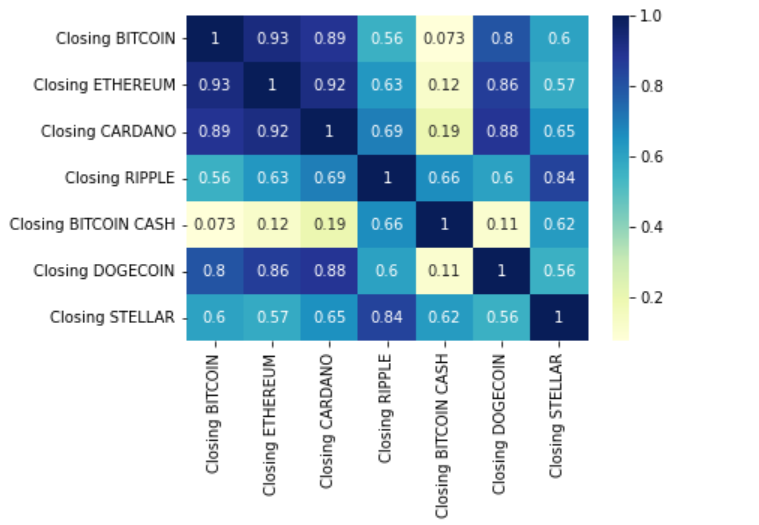
***Dogecoin***



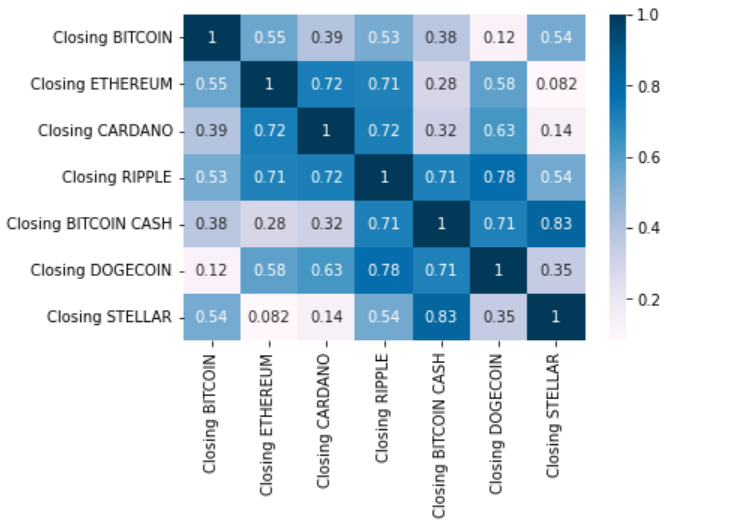
***Stellar***



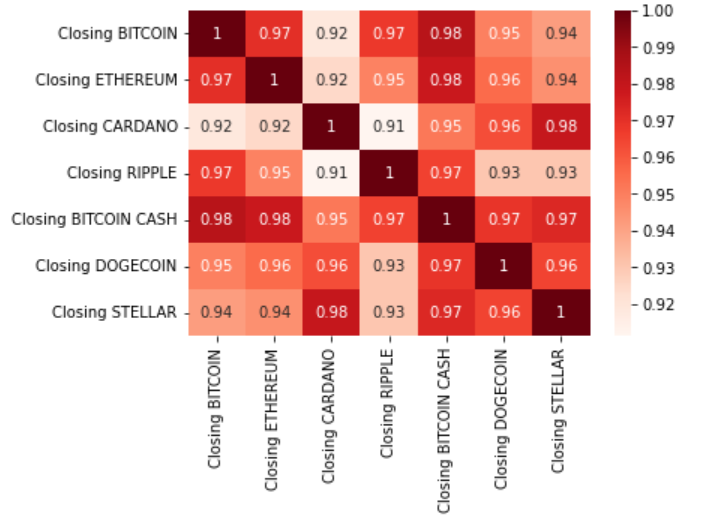
***Correlation Plots***



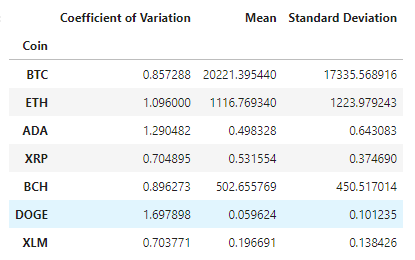
**1** - Correlation Plot for entire time frame (9/11/17 - 7/9/22)



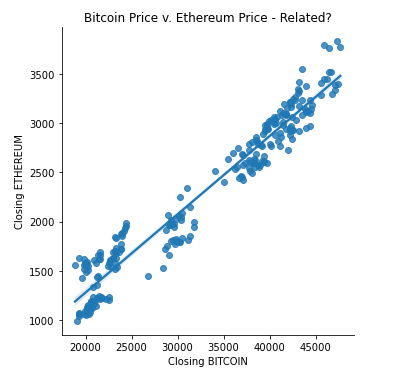
**2** - Correlation Plot from 2021



**3** - Correlation Plot from 2022



4 - Coefficient of Variation, Mean and Standard Deviation



5 - Bitcoin Closing Price on a Given Day Plotted v. the Etherum Price (2022 YTD) with Line of Best Fit added

***Insights***

* First looking at all the scatterplots we can see all the coins find a similar pattern of price movement, some of the graphs look pretty identical at time, but with the various peaks and valleys sometimes a touch more exaggerated for the more volatile coins.
* Four of the coins reached there all time high in 2021, Bitcoin, Ethereum, Dogecoin and Cardano. Bitcoin Cash crashed in 2018 due largely to it splitting into two separate coins, Bitcoin Cash and Bitcoin SV, the price of Bitcoin almost halved in a week after this (Kim, C. 2018) and hasn’t recovered to previous all time high levels. Stellar and Ripple also both hit all time highs in 2017 it hasn’t been able to replicate.
* Over the entire time period covered in the csv files, each coin has a positive correlation with each other. The crypto market is largely linked to each other and can move in tandem. We can see in the first heatmap that Bitcoin and Ethereum are the closest to perfect positive linear correlation with a correlation coefficient of 0.93.
* We see this correlation to dip across the board in 2021, especially for Bitcoin. Bitcoin steadily increased throughout the year, gaining 60% over the year. But some of the more minor coins experienced an explosion, with some gaining over 5,000% on the year (The Motley Fool, 2022). This dampened the correlation a little bit here as the ‘altcoins’ all behaved differently throughout 2021 and didn’t follow the market trends as much.
* However, this trend comes back in 2022, the heatmap lights up and all the 7 coins have a near linear positive relationship with each other, indicating that the price of the crypto market in general have a decent effect now on each coin.
* We can visualise perfectly how linked Ethereum and Bitcoin are to each other in the *lmplot.* When the line of best fit is added in, the data fits it almost perfectly.
* Dogecoin has the highest coefficient of variation, so it experienced the biggest swings in prices, given its mean price. This is no surprise when you look at the scatterplot and see the explosion and regression of Dogecoin’s price during 2021. The lack of certainty surrounding the general crypto market makes all these coins volatile relative to some steadier investment options.

***References***

* NerdWallet. (Voigt, K. and Rosen A.). *What Is Cryptocurrency? Here’s What You Should Know*. [online] Available at: https://www.nerdwallet.com/article/investing/cryptocurrency.
* Kim, C. (2018). *Bitcoin Cash Just Split Into Two Blockchains*. [online] www.coindesk.com. Available at: https://www.coindesk.com/markets/2018/11/15/bitcoin-cash-just-split-into-two-blockchains/ [Accessed 19 Sep. 2022].
* The Motley Fool. (2022). *5 Altcoins That Left Bitcoin in the Dust in 2021*. [online] Available at: https://www.fool.com/the-ascent/cryptocurrency/articles/5-altcoins-that-left-bitcoin-in-the-dust-in-2021/ [Accessed 19 Sep. 2022].