Introduction & Business Problem description - Comparing Various Cryptocurrencies & see which is more stable than other

In this project, the datasets of interest is the Meme Cryptocurrency
Historical datasets. Added a few other datasets related to the Currencies to help with the analysis.

Using datasets analyzing and getting more insights about various cryptocurrencies available in the market like Bitcoin, Ethereum, Litecoin & wrapped bitcoin. Perform comparative analysis between meme cryptocurrencies with famous cryptocurrencies

The features like Date, Volume, Market Cap & Price History played an important role in calculating Cryptocurrency comparison .

Features which are important will be used for analysis work and drop the ones which are not as critical. Used pandas_datareader to access and explore data.

This dataset of some top meme cryptocurrencies will help me compare it with some famous cryptocurrency like bitcoin and Ethereum.

Background & History-

Initial appearance of Cryptocurrency/Bitcoin/altcoin happened in 2009. Now we know it’s going to stay. Question being addressed is what the right percentage of crypto should be in one’s investment portfolio.

Unlike investing in traditional currencies, bitcoin is not issued by a central bank or backed by a government; therefore, the monetary policy, inflation rates, and economic growth measurements that typically influence the value of currency do not apply to bitcoin.

Dogecoin is an "inflationary coin," while cryptocurrencies like Bitcoin are deflationary because there is a ceiling on the number of coins that will be created. Every four years the amount of Bitcoin released into circulation via mining rewards is halved and its inflation rate is halved along with it until all coins are released.

DataSets & Sources-

Kaggle https://www.kaggle.com/datasets/deepshah16/meme-cryptocurrency-historical-data
https://www.kaggle.com/code/ayushggarg/cryptocurrency-meme-vs-famous-eda/notebook

The dataset consists of the historical price information of some of the crypto currencies by market capitalization. The dataset has some of the cryptocurrency from the following categories.Meme Cryptocurrency: Coins like Doge, Shibu inu, elongate ,etc which are derived from a meme or just for fun.

Penny Cryptocurrency: Coins whose price is below a 1 penny usd are known as penny coins.

Dead Cryptocurrency: Coins whose Market cap is 0 are known as dead coins.

Famous and largest cryptocurrency: Coins like Bitcoin, Ethereum etc. with a very high price and large market cap are listed below this category.

Overall, I like to clean the data as needed by dropping off columns which may not be required in further data analysis / visualizations, adding new columns / updating the data elements in selected columns to maintain consistent relationships between various data sources and then create the planned visualizations.

I analyzed the factors that impact the currency. We want to make a few visualizations to see which one has the most impact.

Dataset has 2983 unique values. Here is the glance at data:

Valid	2983	100%
Mismatched	0	0%
Missing	0	0%
Unique	2983	
Most Common	27-06-2021	0%

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2983 entries, 0 to 2982
Data columns (total 7 columns):
 #   Column      Non-Null Count  Dtype
---  ------      --------------  -----
 0   Date        2983 non-null   object
 1   Open        2983 non-null   float64
 2   High        2983 non-null   float64
 3   Low         2983 non-null   float64
 4   Close       2983 non-null   float64
 5   Volume      2983 non-null   float64
 6   Market Cap  2983 non-null   float64
dtypes: float64(6), object(1)
memory usage: 163.3+ KB
```

	Open	High	Low	Close	Volume	Market Cap
Date						
2021-05-24	34700.363568	39835.139830	34551.080550	38705.978637	6.735958e+10	7.243321e+11
2021-05-25	38795.780250	39776.349798	36581.428916	38402.223851	5.621192e+10	7.188294e+11
2021-05-26	38392.623656	40782.078183	37905.834854	39294.197382	5.134674e+10	7.355564e+11
2021-05-27	39316.889678	40379.617672	37247.903676	38436.968535	4.321097e+10	7.195381e+11
2021-05-28	38507.083075	38856.967885	34779.039427	35697.606390	5.520019e+10	6.682839e+11

	Open	High	Low	Close	Volume	Market Cap
count	2953.000000	2953.000000	2953.000000	2953.000000	2.953000e+03	2.953000e+03
mean	6315.811956	6496.437203	6118.113348	6327.536586	1.054393e+10	1.136229e+11
std	10853.865920	11194.114680	10456.360555	10866.056805	1.872924e+10	2.026335e+11
min	68.504997	74.561096	65.526001	68.431000	0.000000e+00	7.784112e+08
25%	425.631989	433.743011	420.510010	425.190002	2.971700e+07	6.262845e+09
50%	1756.520020	1831.420044	1708.540039	1787.130005	7.680150e+08	2.917097e+10
75%	8290.759766	8470.987890	8110.770020	8293.867741	1.522041e+10	1.468189e+11
max	63523.754869	64863.098908	62208.964366	63503.457930	3.509679e+11	1.186364e+12

Methods & Algorithms:

- EDA; include any visuals you think are important to your project
- Data preparation
- Model building and evaluation

The visualization of features importance allows us to understand more the effect of some features that the model considers more important in its classification. Thus, more processes can be done to help the model reach a high performance level. We can also continue to fine tune the hyper-parameters of the model to gain some % in the accuracy measure.

I will also use the pipeline in this process as it helps to enforce desired order of application steps, creating a convenient work-flow. But, there is something more to the pipeline, as we will use\ StandardScaler for cross validation, we can understand data a bit better. .

I will also use Onehotencoder for categorical variables which used to turn categorical features into binary features that are “one-hot” encoded, meaning that if a feature is represented by that column, it receives a 1. Otherwise, it receives a 0.

Algorithms:

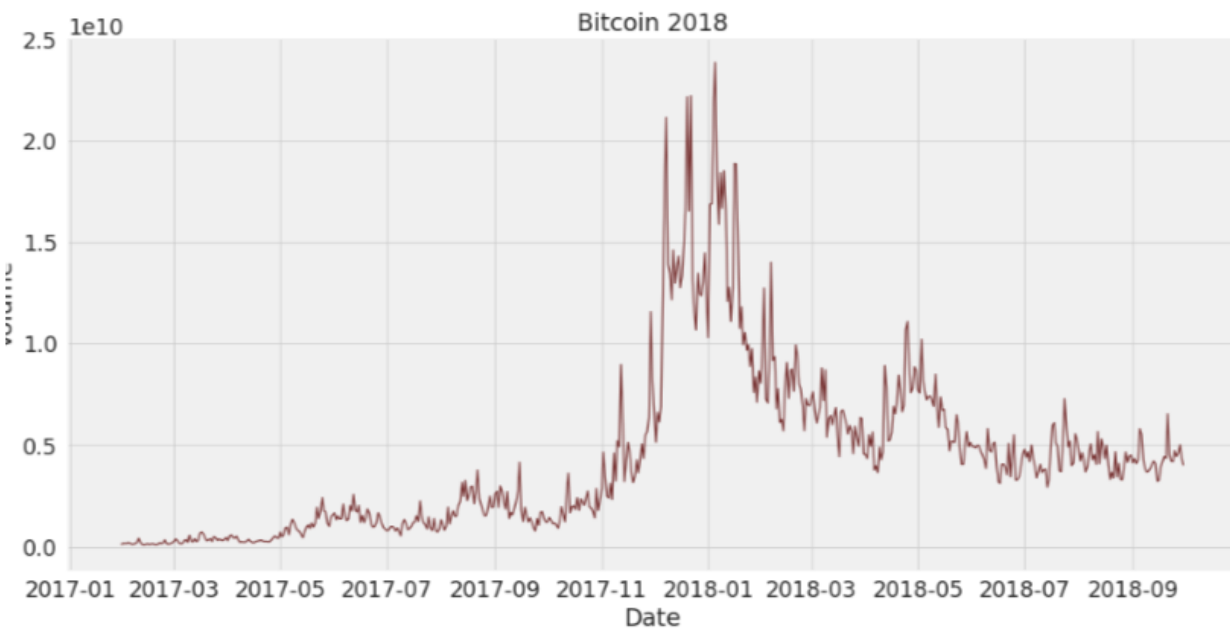
- Building LSTM model using Keras
- Normalizing data using MinMaxScaler from Scikit-Learn.
- Re-framing of data for supervised learning using Pandas
- Training LSTM model on training data set
- Testing/Predicting close prices

Given that we are dealing with time series data, LSTM is well suited. Long short-term memory (LSTM) units (or blocks) are a building unit for layers of a recurrent neural network (RNN). An RNN composed of LSTM units is often called an LSTM network.

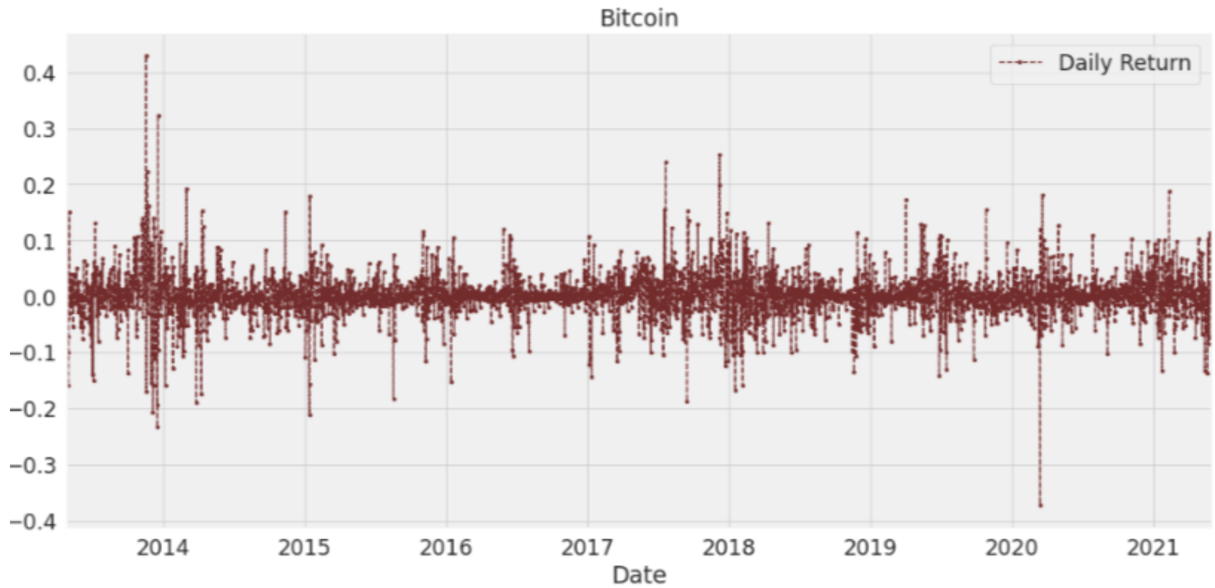
Analysis:

Here is some analysis work I have done so far:

One main thing is observed after looking at the volume of various currencies is 2018 had a huge spike in all currencies and a fall afterwards.



Here we see volume of the daily returns by bitcoin year over year


Bitcoin
Daily Return
0.4
0.3
0.2
0.1
0.0
−0.1
−0.2
−0.3
−0.4
2014
2015
2016
2017
2018
2019
2020
2021
Date


1e6
5.0
4.5
4.0
3.5
3.0
2.5
2.0
Price
Jan 02
Jan 22
Feb 11
Mar 03
Mar 23
Apr 11

How approach addresses the problem:

- Future Appreciation: Market valuation of use case vs current adoption can determine future growth potential Gold and Real Estate appreciation prediction will provide comparison to
- Risk factor: Institutional adoption data will provide stability data around cryptocurrency to under. Regional favored laws for cryptocurrency.
- Market capture: User and market volume provide indicators for future stability
- Investment Portfolio Mix: This will provide the minimum investment strategy in cryptocurrency

Important Variables:

Date : Date of observation

Open : Opening price on the given day

High : Highest price on the given day

Low : Lowest price on the given day

Close : Closing price on the given day

Volume : Volume of transactions on the given day

Market Cap : Market capitalization in USDPrice history is available on a daily basis from April 28, 2013.

Conclusion:

- Cryptocurrencies traded in public markets suffer from price volatility. Bitcoin has experienced rapid surges and crashes in its value, climbing to as high as $17,738 in 2018 before dropping to $7,575 in the following months.
- Bitcoin and Ethereum have many similarities but different long-term visions and limitations.
- From the graphs we can not clearly see that Ethereum, and Bitcoin are quite a lot safer than Dogecoin and Bitconnect.

Assumptions:

- There are various other factors impacts changes in cryptocurrencies not considering those for our analysis

- bitcoin blockchain supports a cryptocurrency which is not backed by any organization or banks
- Various platform could have impact on it

Limitations:

- This analysis is only considering a few coins available in the market which might unfold more mysteries.

Ethical Considerations:

During our Analysis, we haven’t used any PII data. All Data Sets are extracted from Public Websites. Used Datasets have no restrictions for Academic usage All used Datasets are Shared by respective government bodies for public benefits.

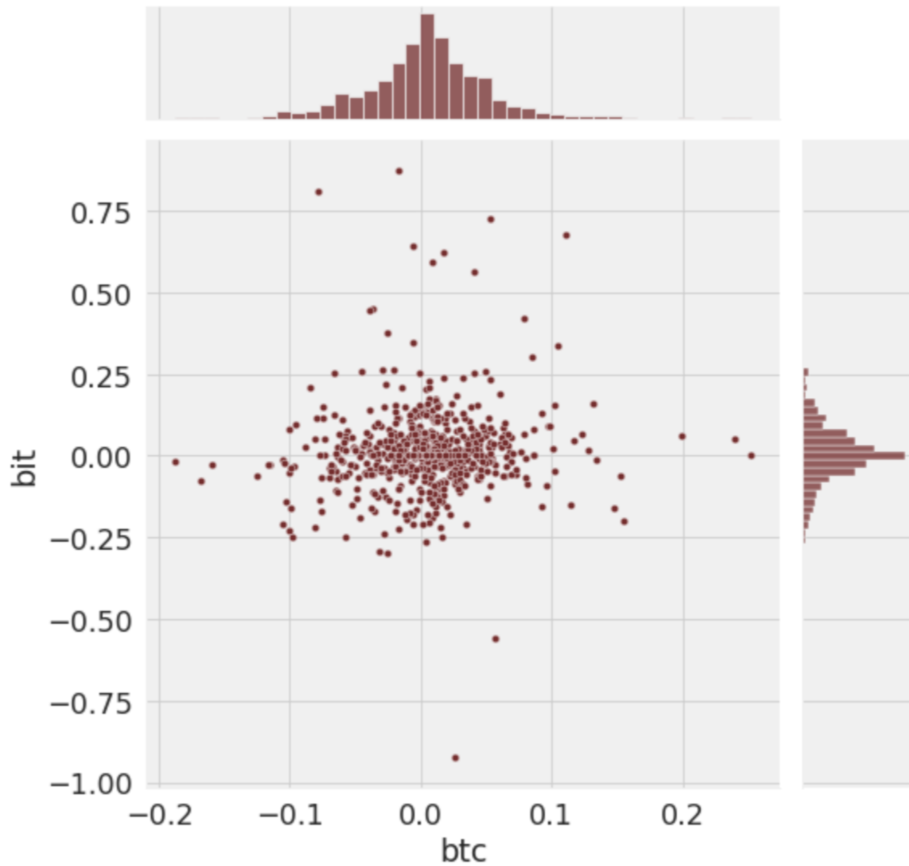
Challenges/Issues:

I think of some challenges around the Testing and choice of the test statistic.

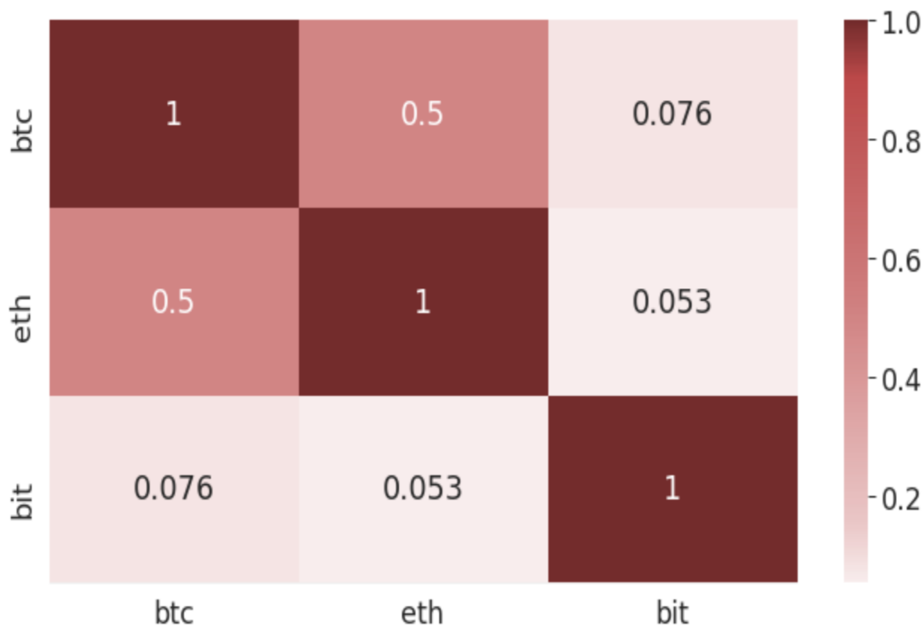
I still feel like a novice and learning as I read and practice with different datasets.

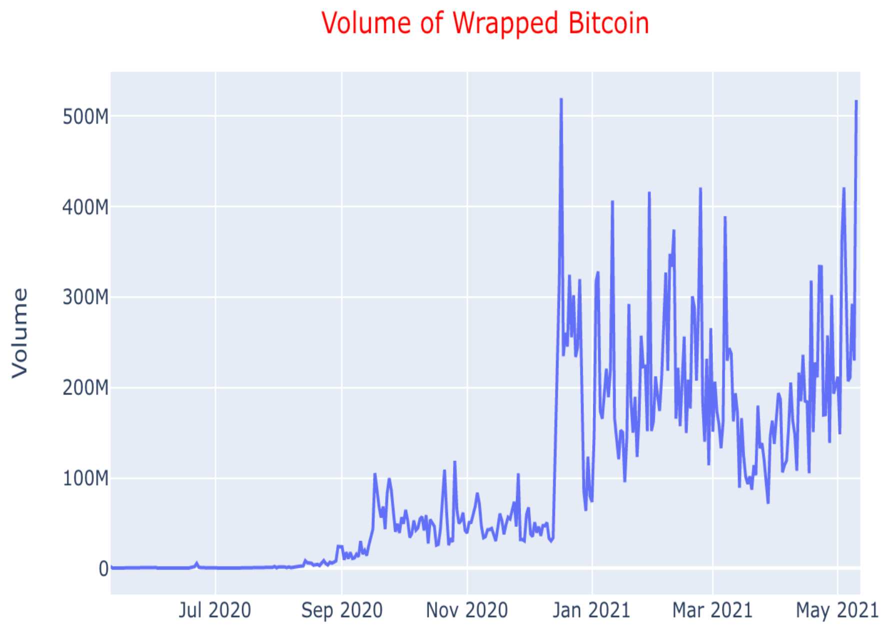
Appendix:

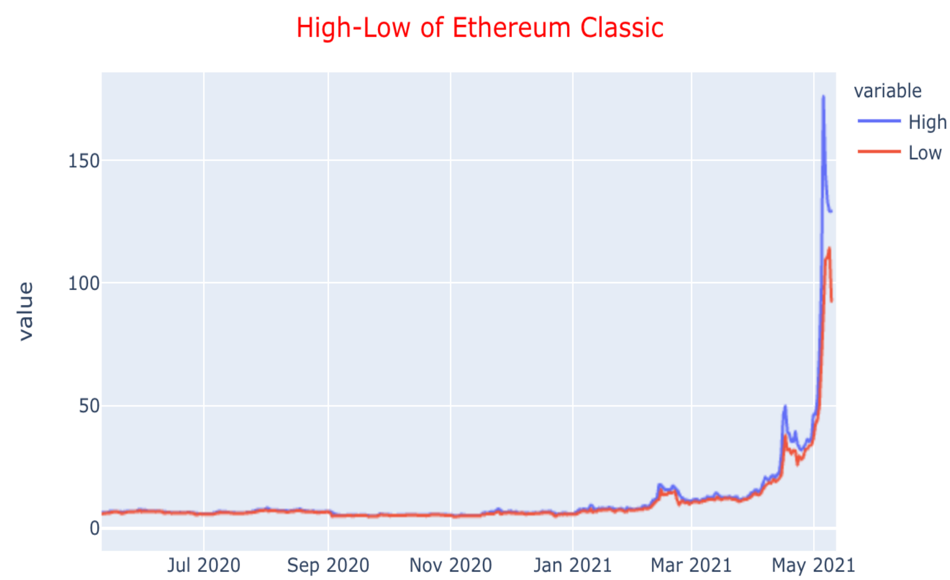
Compare the daily percentage return of two cryptocurrency:



Creating custom color map for heatmap







References:

https://www.kaggle.com/datasets/deepshah16/meme-cryptocurrency-historical-data

Drummond, Chris, and Robert C. Holte. "C4. 5, class imbalance, and cost sensitivity: why under-sampling beats over-sampling." Workshop on learning from imbalanced datasets. Vol. 11. Washington DC: Citeseer, 2003.

Eitrich, Tatjana, and Bruno Lang. "Efficient optimization of support vector machine learning parameters for unbalanced datasets." Journal of computational and applied mathematics 196.2 (2006): 425-436.

Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830, 2011.

https://pirimidtech.com/predicting-cryptocurrency-prices-using-ai-ml/

https://www.sciencedirect.com/topics/computer-science/crypto-currency