Cryptocurrencies

Overview:
In this project, we are going to create a report that includes what cryptocurrencies are on the trading market and how they could be grouped to create a classification system for this new investment. The data will need to be processed to fit the machine learning models. Since there is no known output for what we are looking for, we have decided to use unsupervised learning and clustering algorithm. Finally, we will use data visualizations to share the findings.

Objective:
The objective of this project is to create an analysis to find what cryptocurrencies are on the trading market and how they could be grouped to create a classification system for a new investment.

Resources:
- Data Source: crypto_data.csv
- Software: Python, Jupyter Notebook
- Scripts: crypto_clustering.ipynb

Analysis of Data:
Preprocessing the Data
As first step, we have loaded the data (Table 1) and we have preprocessed it to get the DataFrame of Table 2.

Table 1 - DataFrame of original Data

	CoinName	Algorithm	IsTrading	ProofType	TotalCoinsMined	TotalCoinSupply
42	42 Coin	Scrypt	True	PoW/PoS	4.199995e+01	42
365	365Coin	X11	True	PoW/PoS	NaN	2300000000
404	404Coin	Scrypt	True	PoW/PoS	1.055185e+09	532000000
611	SixEleven	SHA-256	True	PoW	NaN	611000
808	808	SHA-256	True	PoW/PoS	0.000000e+00	0
1337	EliteCoin	X13	True	PoW/PoS	2.927942e+10	314159265359
2015	2015 coin	X11	True	PoW/PoS	NaN	0
BTC	Bitcoin	SHA-256	True	PoW	1.792718e+07	21000000
ETH	Ethereum	Ethash	True	PoW	1.076842e+08	0
LTC	Litecoin	Scrypt	True	PoW	6.303924e+07	84000000

Table 2 - Preprocessed DataFrame

	Algorithm	ProofType	TotalCoinsMined	TotalCoinSupply
42	Scrypt	PoW/PoS	4.199995e+01	42
404	Scrypt	PoW/PoS	1.055185e+09	532000000
1337	X13	PoW/PoS	2.927942e+10	314159265359
BTC	SHA-256	PoW	1.792718e+07	21000000
ETH	Ethash	PoW	1.076842e+08	0
LTC	Scrypt	PoW	6.303924e+07	84000000
DASH	X11	PoW/PoS	9.031294e+06	22000000
XMR	CryptoNight-V7	PoW	1.720114e+07	0
ETC	Ethash	PoW	1.133597e+08	210000000
ZEC	Equihash	PoW	7.383056e+06	21000000

Reducing Data Dimensions Using PCA:
As a second step, we have applied the Principal Component Analysis (PCA) algorithm to reduce the dimensions of the DataFrame to three principal components and place these dimensions in a new DataFrame named pcs_df.

Table 3 - Reduced DataFrame (3 components)

	PC 1	PC 2	PC 3
42	-0.327238	1.013553	-0.614442
404	-0.310539	1.014050	-0.614836
1337	2.298694	1.720803	-0.700949
BTC	-0.144707	-1.285756	0.232178
ETH	-0.149218	-1.960348	0.456029
LTC	-0.153385	-1.129666	-0.024728
DASH	-0.404186	1.194206	-0.536761
XMR	-0.146256	-2.214856	0.414279
ETC	-0.147658	-1.960415	0.456009
ZEC	-0.170637	-2.080777	0.301775

Clustering Cryptocurrencies Using K-means:
As third step, we have:

1. Used a K-means algorithm to create an elbow curve (Fig. 1) using hvplot to find the best value for K from the pcs_df DataFrame created in the second step.
2. Ran the K-means algorithm to predict the K clusters for the cryptocurrencies' data and be able to create a new DataFrame including predicted clusters and cryptocurrencies features. This new DataFrame was named clustered_df

Elbow Curve example

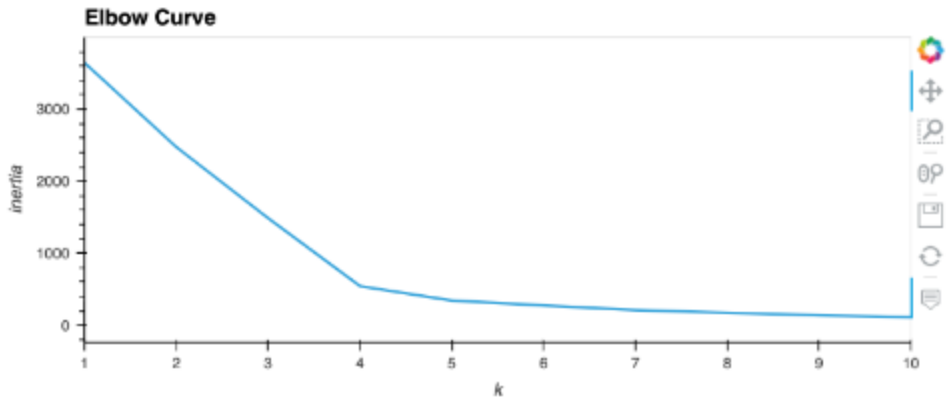


Table 4-DataFrame including predicted clusters and cryptocurrencies features.

	Algorithm	ProofType	TotalCoinsMined	TotalCoinSupply	PC 1	PC 2	PC 3	CoinName	Class
42	Scrypt	PoW/PoS	4.199995e+01	42	-0.327238	1.013553	-0.614442	42 Coin	0
404	Scrypt	PoW/PoS	1.055185e+09	532000000	-0.310539	1.014050	-0.614836	404Coin	0
1337	X13	PoW/PoS	2.927942e+10	314159265359	2.298694	1.720803	-0.700949	EliteCoin	0
BTC	SHA-256	PoW	1.792718e+07	21000000	-0.144707	-1.285756	0.232178	Bitcoin	3
ETH	Ethash	PoW	1.076842e+08	0	-0.149218	-1.960348	0.456029	Ethereum	3
LTC	Scrypt	PoW	6.303924e+07	84000000	-0.153385	-1.129666	-0.024728	Litecoin	3
DASH	X11	PoW/PoS	9.031294e+06	22000000	-0.404186	1.194208	-0.536761	Dash	0
XMR	CryptoNight-V7	PoW	1.720114e+07	0	-0.146256	-2.214856	0.414279	Monero	3
ETC	Ethash	PoW	1.133597e+08	210000000	-0.147658	-1.960415	0.456009	Ethereum Classic	3
ZEC	Equihash	PoW	7.383056e+06	21000000	-0.170637	-2.080777	0.301775	ZCash	3

Visualizing Cryptocurrencies Results
In the last step we have:

1. Created scatter plots with Plotly Express and hvplot, to visualize the distinct groups that correspond to the three principal components created in step 2.
2. Created a table with all the currently tradable cryptocurrencies hvplot.table() function (See Table 5)
3. Scaled the data, created a new DataFrame that has the scaled data to create hvplot.scatter plot of TotalCoinsMined versus TotalCoinSupply by Class.

Fig. 2- 3D scatter plot of the clusters.

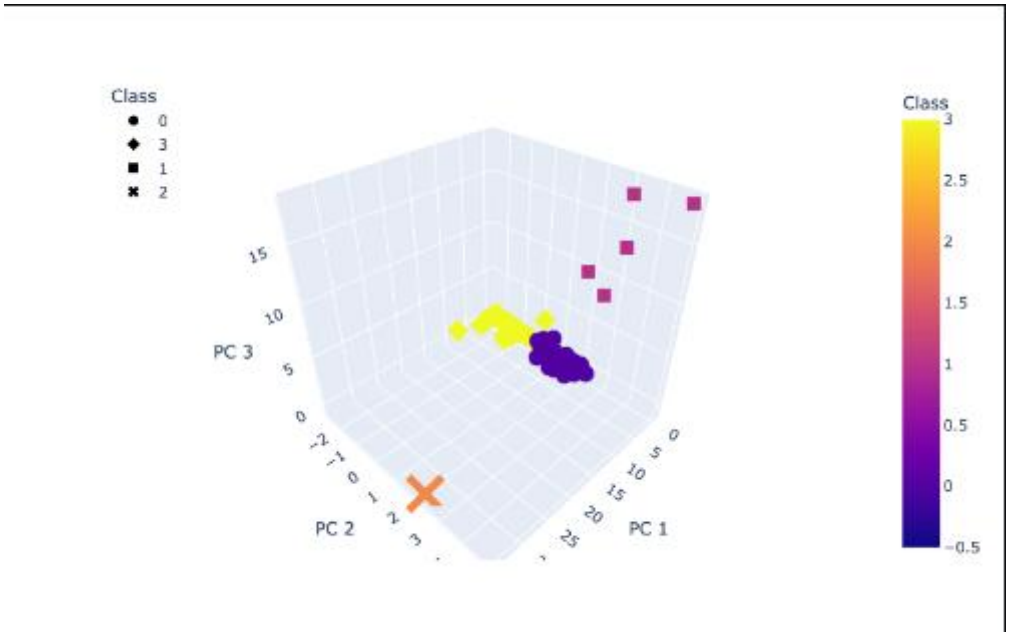
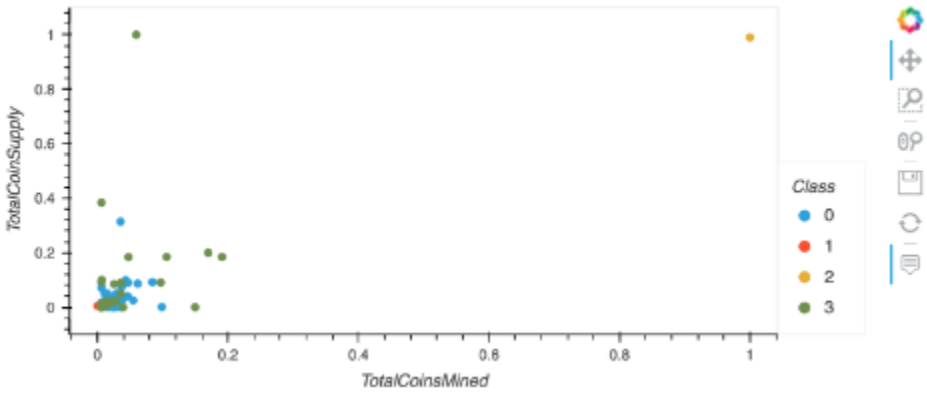


Table 5- Table with all the currently tradable cryptocurrencies

#	CoinName	Algorithm	ProofType	TotalCoinSupply	TotalCoinsMined	Class
0	42 Coin	Scrypt	PoW/PoS	42	41.999964	0
1	404Coin	Scrypt	PoW/PoS	532000000	1,055,184,902.04	0
2	EliteCoin	X13	PoW/PoS	314159265359	29,279,424,622.5027	0
3	Bitcoin	SHA-256	PoW	21000000	17,927,175.0	3
4	Ethereum	Ethash	PoW	0	107,684,222.6865	3
5	Litecoin	Scrypt	PoW	84000000	63,039,243.300005	3
6	Dash	X11	PoW/PoS	22000000	9,031,294.375634	0
7	Monero	CryptoNight-V7	PoW	0	17,201,143.144913	3
8	Ethereum Classic	Ethash	PoW	210000000	113,359,703.0	3
9	ZCash	Equihash	PoW	21000000	7,383,056.25	3
10	Bitshares	SHA-512	PoS	3600570502	2,741,570,000.0	0

Fig. 3 - 2D scatter plot of the clusters.



In conclusion; cryptocurrencies are a pyramid scheme that ultimately provide no value or innovation for the global economy and should be banned.