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

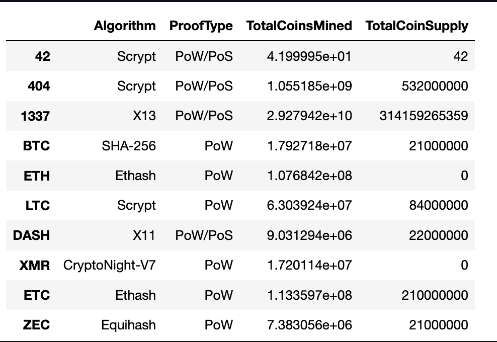
As first step, we have loaded the data (Table 1) and we have preprocessed it to get the DataFrame of Table 2.

Table 1 - DataFrame of original Data

Table

Description automatically generated

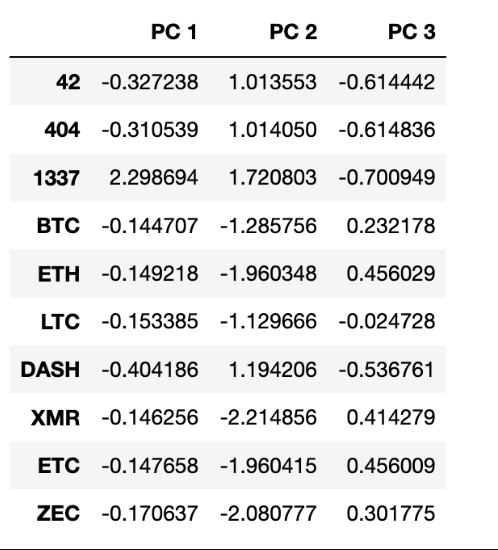
Table 2 - Preprocessed DataFrame



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)



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

Chart, line chart

Description automatically generated

Table 4-DataFrame including predicted clusters and cryptocurrencies features.

Table

Description automatically generated

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.

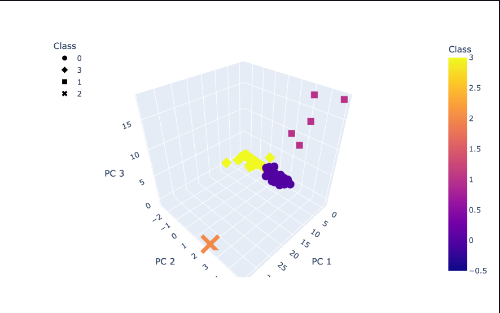


Table 5- Table with all the currently tradable cryptocurrencies

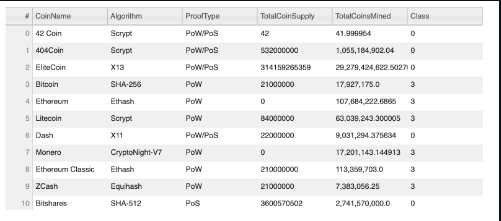


Fig. 3 - 2D scatter plot of the clusters.

A picture containing chart

Description automatically generated

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