**Crypto Currencies**

**Overview**

The purpose of this project is to use unsupervised machine learning to analyze a database of cryptocurrencies and create a report including the traded cryptocurrencies classified by group according to their features.  
This classification report could be used by an investment bank to propose a new cryptocurrency investment portfolio to its clients.  
We use the following methods for the analysis:

* preprocessing the database
* reducing the data dimension using Principal Component Analysis
* clustering cryptocurrencies using K-Means
* visualizing classification results with 2D and 3D scatter plots

**Results**

Following the preprocessing and cleaning phase we have a total of 532 tradable cryptocurrencies.

**Clustering Cryptocurrencies using K-Means - Elbow Curve**

We don't know what the output of the analysis would be, so we are using unsupervised machine learning to identify clusters of the cryptocurrencies.  
We produced the elbow curve below using the K-Means method iterating on k values from 1 to 10.

[Chart, line chart

Description automatically generated](https://user-images.githubusercontent.com/104115586/189263192-b427706e-41ae-4d62-afe4-3f0cf945851d.png)

The best k value appears to be 4 so we would conclude on an output of 4 clusters to categorize the crypto currencies.

**Visualizing Cryptocurrencies Results**

**3D-Scatter plot with clusters**

[Chart, radar chart

Description automatically generated](https://user-images.githubusercontent.com/104115586/189263369-63506090-57d5-417b-8349-313c54df3326.png)

This 3-D scatter plot was obtained using the PCA algorithm to reduce the crypto currencies dimensions to three principal components.

**Tradable Cryptocurrencies Table**

[Table

Description automatically generated](https://user-images.githubusercontent.com/104115586/189263258-3e6aeaa8-6d37-41d2-8c94-407b9809d490.png)

Most of the cryptocurrencies are part of class #0 and #1.  
The snapshot above shows that BitTorrent is the only cryptocurrency in class #2.

**2D-Scatter plot with TotalCoinMined vs TotalCoinSupply**

[Chart

Description automatically generated](https://user-images.githubusercontent.com/68669675/101305764-008c7c00-3809-11eb-8bd7-85bc2c8e953e.png)

Plotting the scatter plot from two cryptocurrency features directly does not efficiently segregate the different classes. Then using the PCA algorithm is the right method for better visualizations.

**Summary**

We have identified the classification of 532 cryptocurrencies based on similarities of their features.  
Particularities of each group need to be analyzed to determine their performance and potential interest for the investment bank's clients.