Clustering of Cryptocurrency Price Movements by K-Means and Hierarchical Clustering Techniques

Thanmaya Sri Sigireddi (50559907)

Lakshmi Mounika Masimukku (50560106)

Venkata Sai Bhargav Dhara (50559703)

Uday Rohit Guntupalli (50534298)

Motivation:

- The cryptocurrency market has grown into a juggernaut within the modern global financial landscape, having in its nature high volatility and sudden shifts. Its volatility stems from a variety of factors, including market psychology, technological advances, and shifting regulatory environments. The stakeholders, which include investors, analysts, and financial institutions, need to be enlightened on these trends in order to arrive at prudent decisions. This, therefore, calls for classification of cryptocurrencies into categories based on their price fluctuation, thus informing risk management and investment strategies.

Significance:

- Classification of cryptocurrency is important for a better understanding of market dynamics and improvement in predictive modeling. The insight to be gained from clustering will help **Investors** to optimize portfolio diversification and risk management; **Financial Analysts** to identify trends and anomalies to guide research and decision-making; **Companies** to create financial products that focus on market segmentation.

Methods used:

Software Description:

- R language is used as it works best for statistical and visualization problems.
- Shiny package is used to develop interactive web application.
- The application is deployed on http://Shinyapps.io

Data source:

- Gathered historical price data for multiple cryptocurrencies from CSV files on Kaggle.
- Data source Link: https://www.kaggle.com/datasets/sudalairajkumar/cryptocurrencypricehistory

Data preprocessing:

- Date columns are converted to Date type and numerical columns are properly formatted
- Missing values are handled by removing incomplete records.

Techniques used:

- Principal Component Analysis
- K-Means Clustering
- Hierarchical Clustering

Principal Component Analysis (PCA):

- PCA allows for dimensionality reduction in a data set by summarizing related variables into a smaller number of unrelated components. It cleans up the data, therefore retaining most of the variance inherent in it. The analysis becomes more efficient since it entails less work. The next step is to extract important elements of the cryptocurrency data. PCA does this by finding the principal elements that explain the most variation. This step improves understanding of the data and assists further analysis based on classification.

Forming clusters using K-Means:

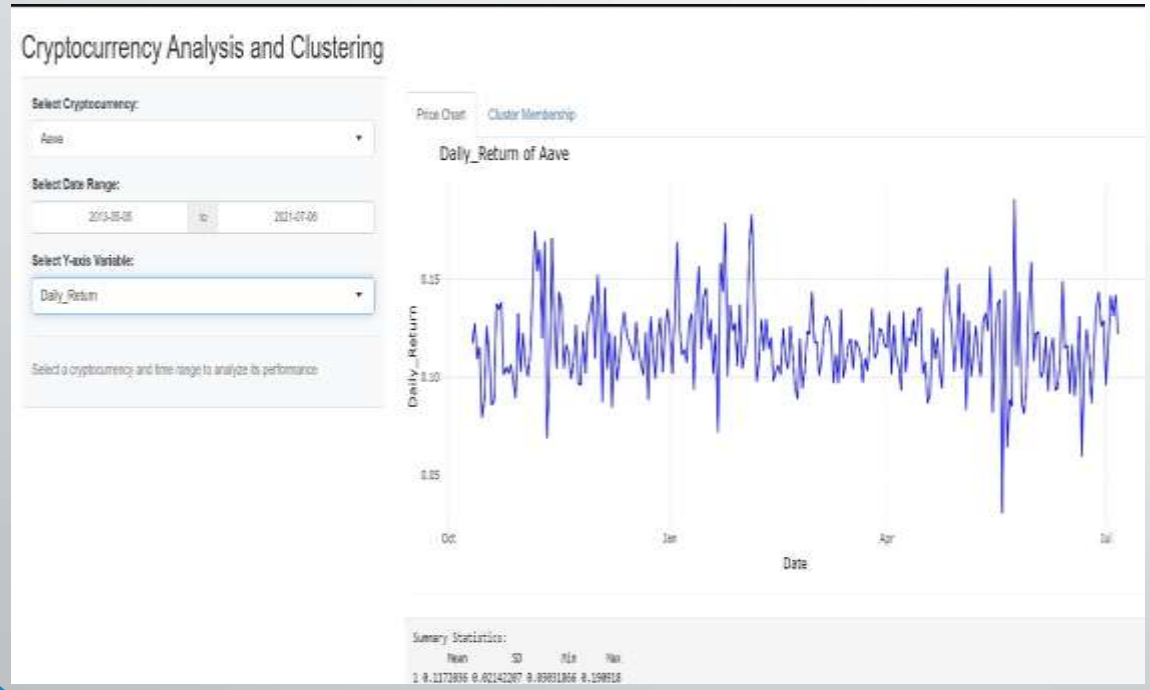
- The classification of cryptocurrencies by use of K-Means is based on the clustering process that will group them into predefined dynamic categories according to their features in operation. This way, it assigns each digital currency to the cluster whose centroid is closest to it, hence grouping similar entities into one class.
- Coin grouping through the K-Means algorithm to find out trends and correlations. This could help to possibly achieve a deep understanding of market dynamics and formulate business strategy decisions.

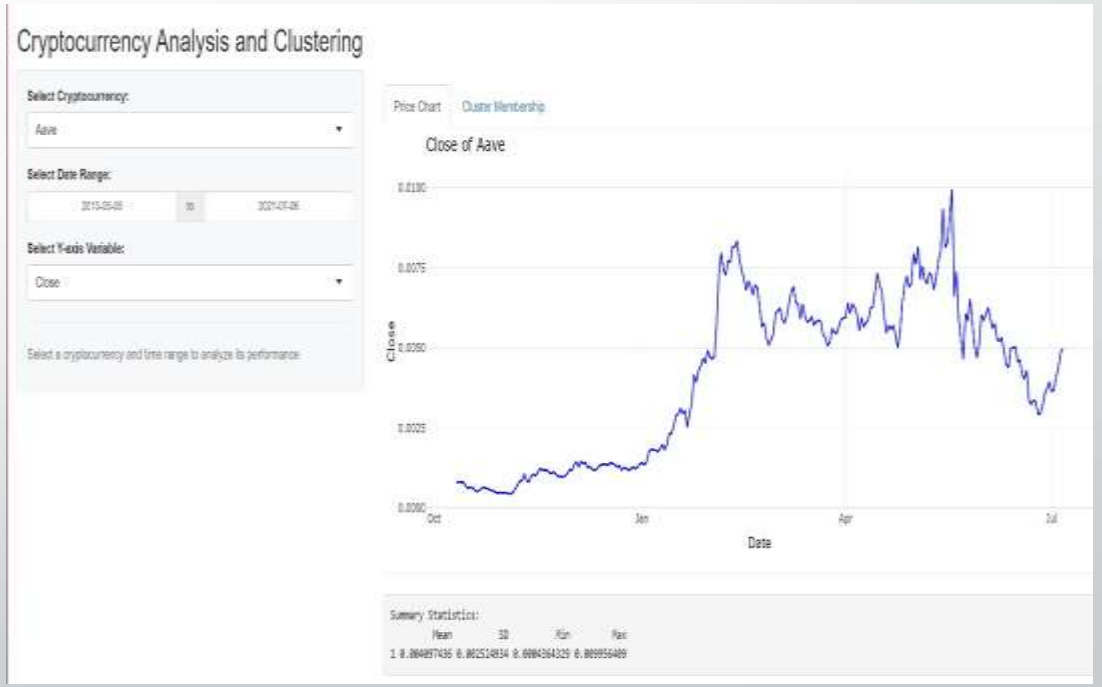
Clustering based on hierarchy:

- Dendrogram construction: Using Hierarchical Clustering, we construct a dendrogram for the purpose of visualizing the grouping of coins. This tool allows us to study the structure of groups and how close the data sets are to each other. It gives a detailed analysis of the information architecture.
- Cluster Structure Analysis: This describes the relationship among coins and shows how they are clustered into smaller groups, given a certain way of cluster formation. Research of this nature will help you to segment your market and view the performance of each segment relative to others.

Results:

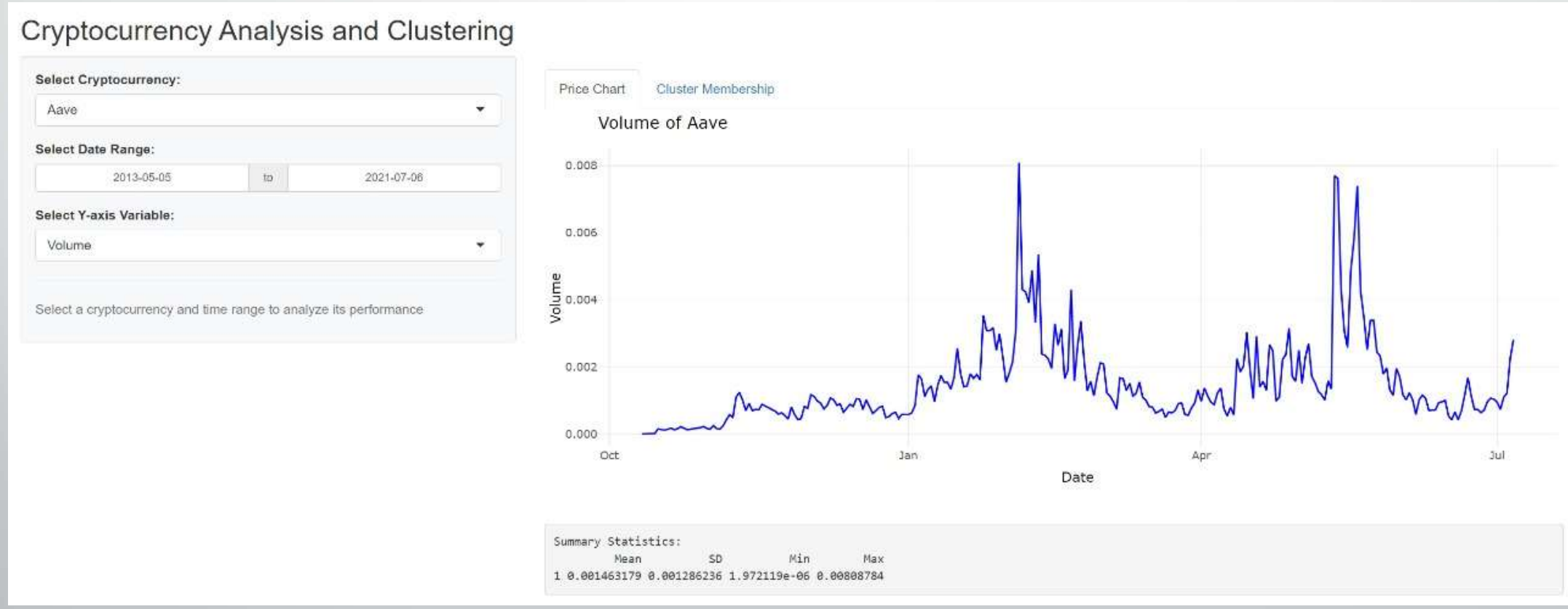
- Historical cryptocurrency data was cleaned and analyzed in order to find patterns. Clustering revealed various categories classified according to market behavior. The Shiny application provides dynamic visual representations of trends, fluctuations, and groupings.
- The below line graphs represents Aave's closing price, trading volume, and daily returns over a defined date range, highlights fluctuations in the cryptocurrency's price and market activity over time. Using the "select Cryptocurrency" dropdown type of cryptocurrency can be selected and get the market activity. Other details can also be changed to know various relations.





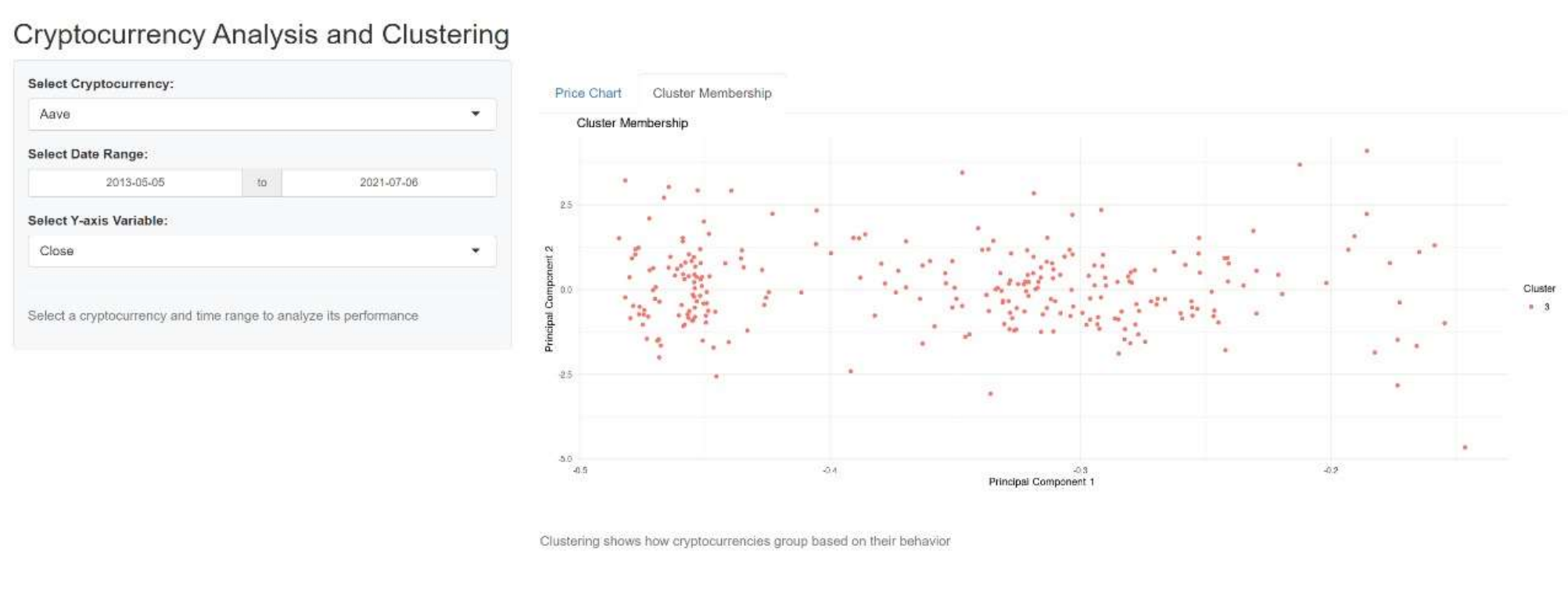
Results Contd.

Every graph shows basic statistical information including mean, standard deviation, minimum, and maximum, thus providing a quick summary of the distribution and variability of the data.



Results Contd.

Each point in the scatter plot of clustered data produced from principal components below indicates a data entry of the chosen cryptocurrency, therefore showing its relationship with others. This helps one to spot behavioral trends or similarities across many time periods or other cryptocurrencies.



Discussions:

Apart from a thorough classification of cryptocurrencies, the study emphasizes the main developments in the markets. The interactive tool enhances data exploring capability. Among the limitations are feature-dependent clustering results and historical data utilization.

Impact :

The project functions as an instructional tool for the bitcoin market and encourages investing decisions.Its open-source foundation and promise of regular upgrades ensure its ongoing relevance and usefulness.

Sustainability:

The project promotes investing decision-making and serves as an educational tool for the cryptocurrency market. Its open-source foundation and promise of regular upgrades ensure its ongoing relevance and usefulness.

Conclusion:

Both methods have successfully categorized the cryptocurrencies and provide benefits of their own.

Hierarchical Clustering gives hierarchical relationships whereas K-Means clusters huge quantities quickly.

Thank you!