Analysis Report

Introduction

This analysis report details the process and findings of clustering cryptocurrencies to determine if they are affected by 24-hour or 7-day price changes. The process involved data preparation, applying the K-means algorithm, performing PCA, and reapplying the K-means algorithm to the reduced dataset. The results were analyzed to determine the best clustering approach.

Data Preparation

The cryptocurrency market data was standardized using the StandardScaler function to ensure that all features contributed equally to the distance calculations in the clustering algorithm.

K-Value Algorithm

Finding the Best Value for k Using the Original Scaled Data

The elbow method was used to determine the optimal number of clusters (k). The best value for k was found to be 4.

Applying the K-means Algorithm

The K-means algorithm was applied with k=4, and the resulting clusters were plotted. Some overlap was observed, suggesting that the clusters might not be well-separated.

PCA & K-Value Algorithm

Performing PCA

PCA was performed to reduce the dataset to three principal components, capturing approximately 89.5% of the variance.

Finding the Best Value for k After PCA

The elbow method was applied again to the PCA data, confirming that the best value for k was still 4.

Applying the K-means Algorithm After PCA

The K-means algorithm was re-applied to the PCA data with k=4. The resulting clusters were well-separated, showing much less overlap compared to the original scaled data.

Results

- The optimal number of clusters (k) is 4 for both the original scaled data and the PCA data.
- PCA effectively reduced the dimensionality and improved cluster separation.
- The total explained variance for the PCA model is 89.5%.

Conclusion

The analysis indicates that PCA is a valuable step in the clustering process, as it helps reduce the dimensionality and improve the separation of clusters. The best approach involves using PCA to transform the data before applying the K-means algorithm with k=4.

Future Work

Future work could involve exploring other dimensionality reduction techniques or clustering algorithms to further improve the model's performance. Additionally, incorporating other features or external data sources might provide further insights into the factors affecting cryptocurrency price changes.