IST 652: SCRIPTING FOR DATA ANALYSIS

FINAL PROJECT REPORT

"Analysis of Cryptocurrency Data"

Introduction:

Cryptocurrencies are digital assets using blockchain technology, known for their decentralization, security, and transparency. Since Bitcoin's inception in 2009, the crypto market has grown exponentially, attracting diverse investors. Analyzing cryptocurrency market data is crucial for understanding volatile price trends, making informed investment decisions, and identifying emerging opportunities in this rapidly evolving sector.

This project's main objective is to analyze the historical price data of top cryptocurrencies, ranked by market capitalization. This involves examining their price fluctuations, trading volumes, and overall market value trends over time, with the goal of gaining a deeper understanding of the dynamics and patterns in the cryptocurrency market.

Data Description:

The dataset comprises historical price data of various cryptocurrencies, including notable examples like Dogecoin, Bitcoin, and Ethereum. This dataset provides a detailed record of daily price information for these cryptocurrencies, starting from April 28, 2013. Here are key details about the dataset:

Cryptocurrency Coverage: The dataset includes a diverse range of cryptocurrencies, with a focus on some of the most prominent ones in terms of market capitalization. Examples mentioned include Dogecoin, Bitcoin, and Ethereum, among others. These cryptocurrencies are representative of the broader digital currency market.

Temporal Coverage: The dataset spans a considerable time, beginning on April 28, 2013. This extensive historical data allows for the analysis of long-term trends and patterns in the cryptocurrency market.

Data Attributes:

The dataset records essential daily attributes for each cryptocurrency, including:

- **Date:** The specific date of the recorded data, providing a chronological sequence for analysis.
- **Open Price:** The opening price of the cryptocurrency at the beginning of the trading day.
- **High Price:** The highest price reached by the cryptocurrency during the day, indicating peak demand or bullish trends.
- **Low Price:** The lowest price observed for the cryptocurrency during the day, reflecting minimum demand or bearish trends.
- **Closing Price:** The closing price of the cryptocurrency at the end of the trading day.
- **Trading Volume:** The total trading volume of the cryptocurrency on that day, serving as an indicator of market liquidity.
- **Market Capitalization:** The market capitalization in USD, representing the total market value of the cryptocurrency's circulating supply.

This dataset's rich and comprehensive attributes make it a valuable resource for conducting detailed analyses of various cryptocurrencies, understanding their price dynamics, and exploring historical market trends.

Data Preprocessing:

Read and loading the essential dataset:

```
import pandas as pd

#import from drive
from google.colab import drive
drive.mount("/drive", force_remount=True)
import csv

#Read data into new dataframes
df1=pd.read_csv('/drive/My Drive/652_Final/coin_Dogecoin.csv')
df2=pd.read_csv('/drive/My Drive/652_Final/coin_Bitcoin.csv')
df3=pd.read_csv('/drive/My Drive/652_Final/coin_BinanceCoin.csv')
df4=pd.read_csv('/drive/My Drive/652_Final/coin_Ethereum.csv')
```

The data collection method employed for this project involved sourcing the historical price data of cryptocurrencies from CSV (Comma-Separated Values) files. CSV files are a common and convenient format for storing structured data, making them suitable for storing time-series data such as daily cryptocurrency price information.

Data Merging:

```python
# Concatenate all datasets vertically
crypto = pd.concat([df1, df2, df3, df4], ignore_index=True)

# Print the combined dataset
print(crypto)
```

```
       SNo      Name Symbol                 Date         High          Low  \
0        1  Dogecoin   DOGE  2013-12-16 23:59:59     0.000866     0.000150
1        2  Dogecoin   DOGE  2013-12-17 23:59:59     0.000289     0.000116
2        3  Dogecoin   DOGE  2013-12-18 23:59:59     0.000362     0.000205
3        4  Dogecoin   DOGE  2013-12-19 23:59:59     0.001520     0.000328
4        5  Dogecoin   DOGE  2013-12-20 23:59:59     0.001143     0.000662
...    ...       ...    ...                  ...          ...          ...
9348  2156  Ethereum    ETH  2021-07-02 23:59:59  2155.596496  2021.824808
9349  2157  Ethereum    ETH  2021-07-03 23:59:59  2237.567155  2117.590013
9350  2158  Ethereum    ETH  2021-07-04 23:59:59  2384.286857  2190.837703
9351  2159  Ethereum    ETH  2021-07-05 23:59:59  2321.922836  2163.041394
9352  2160  Ethereum    ETH  2021-07-06 23:59:59  2346.294874  2197.919385

             Open        Close        Volume     Marketcap
0        0.000299     0.000205  0.000000e+00  1.509085e+06
1        0.000207     0.000269  0.000000e+00  2.169688e+06
2        0.000267     0.000362  0.000000e+00  3.188943e+06
3        0.000395     0.001162  0.000000e+00  1.115034e+07
4        0.001143     0.000704  0.000000e+00  7.284337e+06
...           ...          ...          ...           ...
9348  2109.892677  2150.040364  3.179621e+10  2.505527e+11
9349  2150.835025  2226.114282  1.743336e+10  2.594475e+11
9350  2226.550382  2321.724112  1.878711e+10  2.706217e+11
9351  2321.922836  2198.582464  2.010379e+10  2.562978e+11
9352  2197.919385  2324.679449  2.089186e+10  2.710286e+11

[9353 rows x 10 columns]
```

The vertical concatenation process involves merging individual cryptocurrency data frames (df1, df2, df3, and df4) into a single dataframe called 'crypto' by stacking their rows. This creates a comprehensive dataset containing diverse cryptocurrency information. This consolidation simplifies the comparison, aggregation, and analysis of cryptocurrency data by centralizing it in one dataframe instead of using multiple separate ones. Additionally, if these data frames share common columns like 'Date' or 'Price,' the concatenation automatically combines these details. This streamlines the analysis process, making it more efficient when examining cryptocurrency trends and market behavior.

Data Description:

```
#describe the combined dataset
crypto.describe
```

```
<bound method NDFrame.describe of        SNo      Name Symbol                 Date         High          Low  \
0        1  Dogecoin   DOGE  2013-12-16 23:59:59     0.000866     0.000150
1        2  Dogecoin   DOGE  2013-12-17 23:59:59     0.000289     0.000116
2        3  Dogecoin   DOGE  2013-12-18 23:59:59     0.000362     0.000205
3        4  Dogecoin   DOGE  2013-12-19 23:59:59     0.001520     0.000328
4        5  Dogecoin   DOGE  2013-12-20 23:59:59     0.001143     0.000662
...    ...       ...    ...                  ...          ...          ...
9348  2156  Ethereum    ETH  2021-07-02 23:59:59  2155.596496  2021.824808
9349  2157  Ethereum    ETH  2021-07-03 23:59:59  2237.567155  2117.590013
9350  2158  Ethereum    ETH  2021-07-04 23:59:59  2384.286857  2190.837703
9351  2159  Ethereum    ETH  2021-07-05 23:59:59  2321.922836  2163.041394
9352  2160  Ethereum    ETH  2021-07-06 23:59:59  2346.294874  2197.919385

             Open        Close        Volume     Marketcap
0        0.000299     0.000205  0.000000e+00  1.509085e+06
1        0.000207     0.000269  0.000000e+00  2.169688e+06
2        0.000267     0.000362  0.000000e+00  3.188943e+06
3        0.000395     0.001162  0.000000e+00  1.115034e+07
4        0.001143     0.000704  0.000000e+00  7.284337e+06
...           ...          ...           ...           ...
9348  2109.892677  2150.040364  3.179621e+10  2.505527e+11
9349  2150.835025  2226.114282  1.743336e+10  2.594475e+11
9350  2226.550382  2321.724112  1.878711e+10  2.706217e+11
9351  2321.922836  2198.582464  2.010379e+10  2.562978e+11
9352  2197.919385  2324.679449  2.089186e+10  2.710286e+11

[9353 rows x 10 columns]>
```

The newly created dataset, obtained by merging data from various cryptocurrency coins, is labeled as "crypto." This dataset is substantial, comprising 9,353 rows and 10 columns.
The merging process has effectively consolidated data from different cryptocurrencies into a single, comprehensive dataset, enabling a more holistic analysis. With 10 columns of information, this dataset provides a wealth of insights into the cryptocurrency market, making it a valuable resource for studying trends, patterns, and behaviors across multiple digital currencies. The substantial number of rows ensures that a significant historical period is covered, enhancing the dataset's analytical capabilities

Data Exploration:

```
#view first 5 rows of dataset
crypto.head()
```

	SNo	Name	Symbol	Date	High	Low	Open	Close	Volume	Marketcap
0	1	Dogecoin	DOGE	2013-12-16 23:59:59	0.000866	0.000150	0.000299	0.000205	0.0	1.509085e+06
1	2	Dogecoin	DOGE	2013-12-17 23:59:59	0.000289	0.000116	0.000207	0.000269	0.0	2.169688e+06
2	3	Dogecoin	DOGE	2013-12-18 23:59:59	0.000362	0.000205	0.000267	0.000362	0.0	3.188943e+06
3	4	Dogecoin	DOGE	2013-12-19 23:59:59	0.001520	0.000328	0.000395	0.001162	0.0	1.115034e+07
4	5	Dogecoin	DOGE	2013-12-20 23:59:59	0.001143	0.000662	0.001143	0.000704	0.0	7.284337e+06

Using head:

- head () is used to display the first few rows of a dataset.
- It is often employed at the beginning of data exploration to get a preview of the data.
- The syntax is: dataframe.head(n), where n specifies the number of rows to display (default is usually 5).
- Example: crypto. Head () would display the first 5 rows of the 'crypto' dataset.

```
[ ] #view last 5 rows of dataset
    crypto.tail()
```

	SNo	Name	Symbol	Date	High	Low	Open	Close	Volume	Marketcap
9348	2156	Ethereum	ETH	2021-07-02 23:59:59	2155.596496	2021.824808	2109.892677	2150.040364	3.179621e+10	2.505527e+11
9349	2157	Ethereum	ETH	2021-07-03 23:59:59	2237.567155	2117.590013	2150.835025	2226.114282	1.743336e+10	2.594475e+11
9350	2158	Ethereum	ETH	2021-07-04 23:59:59	2384.286857	2190.837703	2226.550382	2321.724112	1.878711e+10	2.706217e+11
9351	2159	Ethereum	ETH	2021-07-05 23:59:59	2321.922836	2163.041394	2321.922836	2198.582464	2.010379e+10	2.562978e+11
9352	2160	Ethereum	ETH	2021-07-06 23:59:59	2346.294874	2197.919385	2197.919385	2324.679449	2.089186e+10	2.710286e+11

Using tail:

- tail () is used to display the last few rows of a dataset.
- It helps in examining the data towards the end or at the most recent records.
- The syntax is: dataframe.tail(n), where n specifies the number of rows to display (default is usually 5).
- Example: crypto. Tail () would display the last 5 rows of the 'crypto' dataset.

Methods of Analysis:

1. **Which coin has the highest value at closing?**

```
#The highest value among all coins at closing
maxValue=crypto[crypto['Close']==max(crypto.Close)]
print("Highest value at closing")
maxValue

#we can see that Bitcoin had the highest value at closing amongst other coins
```

Highest value at closing

	SNo	Name	Symbol	Date	High	Low	Open	Close	Volume	Marketcap
5666	2907	Bitcoin	BTC	2021-04-13 23:59:59	63742.283337	59869.956293	59890.01779	63503.45793	6.998345e+10	1.186364e+12

Unit of Analysis: Close

Comparison:

Finding the maximum closing price (max (crypto. Close)) in the 'crypto' dataset and then filter the dataset to identify the specific cryptocurrency(s) that had this maximum closing price. In this case, it indicates that Bitcoin had the highest closing price among the coins included in the dataset

2. **Which coin has the lowest value at opening?**

Which coin has the lowest value at opening?

```
[ ] #The lowest value among all coins at opening
    minValue=crypto[crypto['Open']==min(crypto.Open)]
    print("Lowest value at opening")
    minValue

    #we can see that Dogecoin had the lowest value at opening amongst other coins
```

Lowest value at opening

	SNo	Name	Symbol	Date	High	Low	Open	Close	Volume	Marketcap
507	508	Dogecoin	DOGE	2015-05-07 23:59:59	0.000089	0.000085	0.000087	0.000089	34179.199219	8.835114e+06

Unit of Analysis: Open

Comparison:

Finding the minimum opening price (min (crypto. Open)) in the 'crypto' dataset and then filter the dataset to identify the specific cryptocurrency(s) that had this minimum opening price. In this case, it indicates that Dogecoin had the lowest opening price among the coins included in the dataset.

3. **What is the average price and volume of all cryptocurrencies during 2013 to 2021?**

```python
# Group by coin and calculate mean price and volume
coin_summary = crypto.groupby('Name').agg({'Marketcap': 'mean', 'Volume': 'mean'})
print(coin_summary)

coin_summary.to_csv('coin_summary.csv')
```

```
                 Marketcap        Volume
Name
Binance Coin  7.835575e+09  6.269804e+08
Bitcoin       1.208761e+11  1.090633e+10
Dogecoin      1.760533e+09  4.326321e+08
Ethereum      4.172084e+10  7.057058e+09
```

Unit of Analysis: Individual Cryptocurrencies (Coins)

Comparison:

Calculating the mean (average) values for two attributes, 'Market cap' (market capitalization) and 'Volume,' for each cryptocurrency group.

4. **Determine the volatility of different currencies**

```
Volatility Ranking:
Name
Dogecoin        0.113458
Binance Coin    0.080050
Ethereum        0.063036
Bitcoin         0.042639
Name: Daily_Return, dtype: float64
```

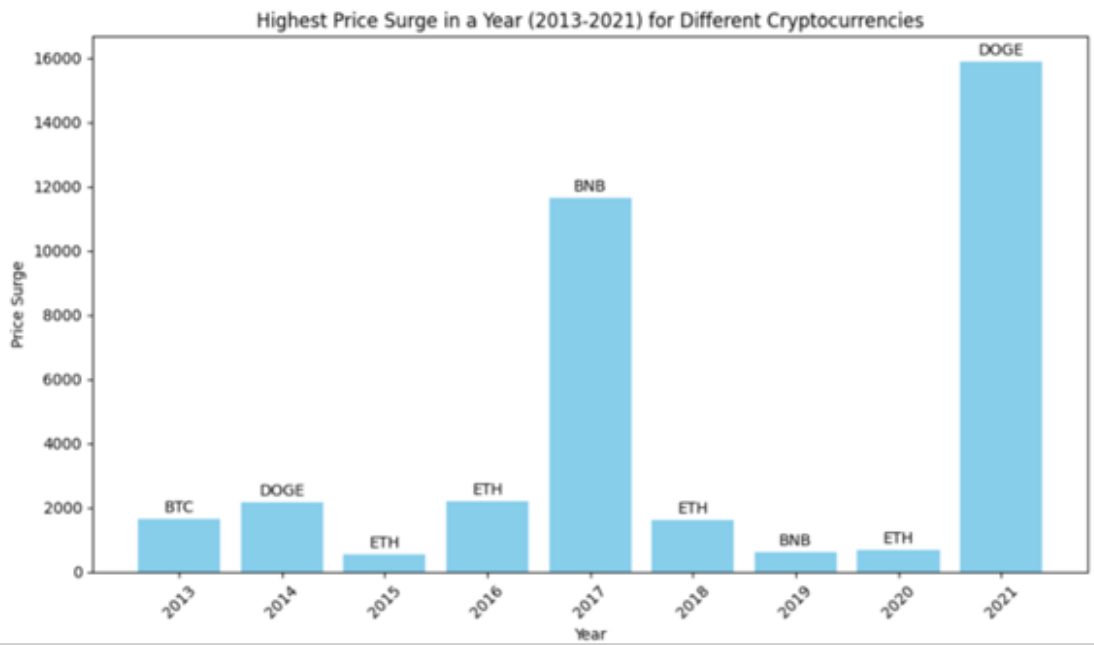
Unit of Analysis: Individual Cryptocurrencies (Coins)

Comparison:

Ranking of Cryptocurrencies by Volatility (Based on Standard Deviation of Daily Returns)

5. Which cryptocurrency experienced the highest price surge (percentage increase) in a year between 2013 to 2021?

	Year	Max_Surge_Symbol	Max_Surge_Value
0	2013	BTC	1664.398861
1	2014	DOGE	2155.993021
2	2015	ETH	564.963152
3	2016	ETH	2214.679363
4	2017	BNB	11659.724425
5	2018	ETH	1629.906921
6	2019	BNB	620.253107
7	2020	ETH	692.466016
8	2021	DOGE	15880.971902

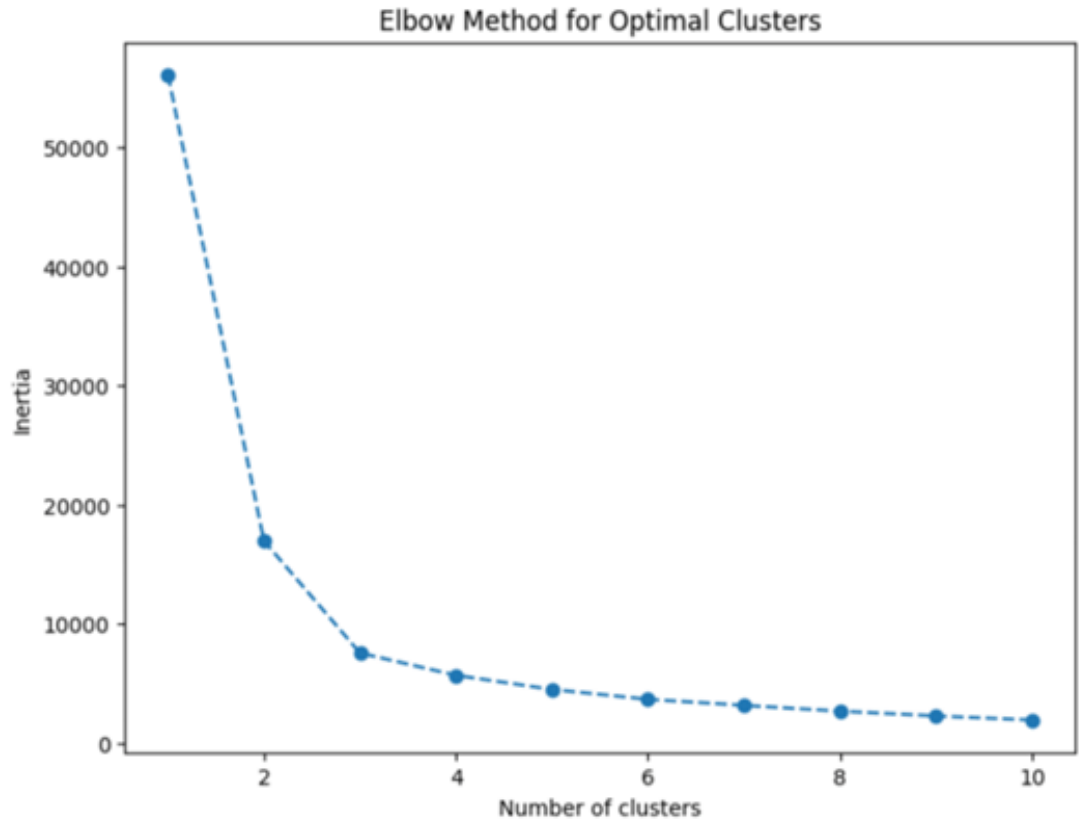


Unit of Analysis: Individual Years (from 2013 to 2021)

Comparison:

Identification of Cryptocurrency with Highest Price Surge (Percentage Increase) in Each Year

6. **How does organizing cryptocurrency data into distinct groups based on their shared characteristics help in gaining insights or understanding the broader landscape of the cryptocurrency market, and how might this organized information aid in making strategic decisions in the realm of digital currencies?**



Cluster	High	Low	Open	Close	Volume
0	301.148395	284.672773	293.001130	293.623090	1.522314e+09
1	46536.691449	43250.571529	45037.384305	45075.035422	6.015874e+10
2	8466.150560	8023.305925	8249.010471	8268.632866	2.177551e+10

Cluster	Marketcap
0	1.138991e+10
1	8.415773e+11
2	1.785170e+11

Unit of Analysis: Cryptocurrencies

Comparison:

Grouping Cryptocurrencies into Clusters Based on Similar Features (K-Means Clustering) and Analyzing Cluster Means.

Description of the Python program:

Here is an overview of its functionalities and characteristics:

Data Source and Structure:
The program uses a dataset that includes historical price information of several top cryptocurrencies by market capitalization. This data is available in CSV format and contains daily records from April 28, 2013.

Data Import and Preparation:

- The program imports data for different cryptocurrencies like Dogecoin, Bitcoin, BinanceCoin, and Ethereum into separate data frames.
- These data frames are then concatenated vertically to form a single comprehensive dataframe. This aids in easy comparison and analysis across different cryptocurrencies.

Analysis Features:

- The program identifies the cryptocurrency with the highest value at closing and the one with the lowest value at opening.
- It calculates and visualizes outliers in 'Close' prices and 'Market Cap' using box plots.
- The average price and volume of all cryptocurrencies during 2013 to 2021 are computed.
- It determines the volatility of different cryptocurrencies by calculating the standard deviation of their daily returns.
- The program identifies which cryptocurrency experienced the highest price surge in a year between 2013 to 2021.
- Investment suggestions are provided based on the historical performance of various cryptocurrencies.

Data Visualization:
The program includes visual representations such as bar graphs and box plots to illustrate various analyses, like the highest price surge per year for different cryptocurrencies and the detection of outliers.

Machine Learning Application:

- KMeans clustering is applied to cryptocurrency data to organize it into distinct groups based on shared characteristics. This helps in understanding the broader landscape of the cryptocurrency market.
- The elbow method is used to determine the optimal number of clusters, and cluster analysis is performed to provide insights into the characteristics of different clusters within the cryptocurrency data.

Data Output:

- The results of various analyses, such as volatility ranking and clustering, are stored in CSV files for further use or examination.
- This program thus serves as a comprehensive tool for analyzing and understanding cryptocurrency market trends, providing valuable insights for strategic decision-making in the realm of digital currencies.

Documentation of the output of the Python program:

1. Dataset Overview

- This part introduces datasets containing historical data on major cryptocurrencies, including Bitcoin, Ethereum, Dogecoin, and BinanceCoin.
- The data encompasses daily observations of various metrics like opening, highest, lowest, and closing prices, as well as transaction volume and market capitalization.
- It explains how this data can be used to analyze market trends and investment opportunities in the cryptocurrency market.

2. Data Concatenation

- The notebook demonstrates techniques for combining individual cryptocurrency datasets into a single Data Frame.
- This unified dataset facilitates comparative analysis across different cryptocurrencies.
- Such a method is essential for a holistic view of the cryptocurrency market, especially when considering correlations or aggregated trends.

3. Data Analysis

- Outlier Detection: Identifies and visualizes outliers in the closing prices and market capitalization of cryptocurrencies. This is crucial for understanding extreme market behaviors.
- Average Price and Volume Analysis: Calculates and presents the average price and volume of all considered cryptocurrencies from 2013 to 2021. This trend analysis is key for understanding market growth and investor interest over time.
- Volatility Analysis: Analyzes the volatility of different cryptocurrencies by calculating the standard deviation of their daily returns. High volatility in cryptocurrencies can signal both elevated risk and potential high returns for investors.

4. Price Surge Analysis

- This section identifies the cryptocurrency with the highest percentage increase in price each year, from 2013 to 2021.
- The analysis includes a bar chart visualization, which helps in easily identifying which cryptocurrencies had significant price surges in specific years.
- Such analysis is valuable for investors looking for historical trends and patterns in cryptocurrency price movements.

5. Investment Insights

- Based on the price surge analysis, the notebook provides insights or recommendations for potential investments in Bitcoin, Ethereum, Binance Coin, and Dogecoin.
- This section is crucial for investors seeking guidance on which cryptocurrencies have historically shown significant growth.

6. Clustering Analysis

- Applies KMeans clustering to a subset of the data to identify distinct groups of cryptocurrencies based on shared characteristics.
- Involves data scaling and the use of the elbow method to determine the optimal number of clusters.
- The analysis of the resulting clusters can reveal intrinsic groupings in the market, useful for market segmentation and targeted investment strategies.

7. Visualizations

- The notebook employs various plots and graphs, like boxplots, bar charts, and cluster visualizations, to graphically represent data analyses.
- These visualizations aid in understanding complex data patterns and trends in a more intuitive and accessible manner.

8. Python Libraries Used

It utilizes Pandas for data manipulation and analysis, Matplotlib and seaborn for data visualization, and scikit-learn for machine learning tasks like clustering.

Conclusion:

1. **Value Trends**: Bitcoin consistently held the highest closing values, while Dogecoin often started with the lowest opening values, highlighting the significant value disparity among different cryptocurrencies.

2. **Outlier Detection**: The identification of outliers in closing prices and market capitalization provides insights into extreme market behaviors and potential anomalies, which are crucial for understanding market stability and risks.

3. **Average Market Metrics:** The calculated average price and volume of cryptocurrencies from 2013 to 2021 offer a macro view of the market, revealing general trends and shifts in cryptocurrency trading over the years.

4. **Volatility Insights:** The volatility analysis, through the standard deviation of daily returns, indicates that Dogecoin was the most volatile. This information is vital for investors to assess risk and make informed decisions.

5. **Yearly Price Surge Analysis**: The project identifies the cryptocurrency with the highest price surge each year, with Dogecoin showing a significant surge in 2021. This

analysis helps in spotting investment opportunities and understanding market dynamics.

6. **Investment Recommendations:** Based on historical surges, the project offers investment insights for cryptocurrencies like Bitcoin, Ethereum, Binance Coin, and Dogecoin, aiding in strategic investment planning.

7. **Cluster Analysis for Strategic Insights**: By organizing cryptocurrencies into clusters based on shared characteristics, the project provides a structured view of the market. This helps in identifying similar behavior among different cryptocurrencies and aids in diversified investment strategies.

8. **Visual Representation of Data:** The use of visual tools like boxplots and bar charts enhances the understanding of complex data and trends.

In summary, your project successfully analyzes the cryptocurrency market, highlighting key trends, risks, opportunities, and offering valuable insights for investors and market analysts. The use of statistical methods, combined with visual data representation, makes this analysis a robust tool for understanding the dynamic cryptocurrency landscape.