Comparative Analysis of JPMorgan Stock Performance During the 2008 Financial Crisis and COVID-19 Pandemic: A Multivariate Approach

Karri Lakshmi Yamini(2020MT60881) Volla Yayathi(2020MT60897)

24 October 2024

1 Introduction

This report provides a detailed analysis of **JPMorgan stock prices** during two significant global economic downturns: the **2008 Financial Crisis** and the **COVID-19 Pandemic (2020)**. The aim is to examine and compare stock behavior in these two distinct periods using a range of multivariate statistical methods, with a focus on identifying key patterns, relationships, and differences between the datasets.

1.1 Dataset Description

We have selected two datasets that capture JPMorgan's stock price movements during the two crises:

- **Financial Crisis Dataset (2008)**: This dataset covers stock prices during the height of the 2008 global financial crisis, a period marked by severe economic disruption and instability in global markets. [01.07.2008 - 24.12.2009]
- **COVID-19 Pandemic Dataset (2020)**: This dataset reflects stock prices during the early stages of the COVID-19 pandemic in 2020, which also caused widespread economic uncertainty and volatility in global financial markets.[01.03.2020-29.06.2021]

Each dataset includes the following **variables** (dimensions) that provide comprehensive information about the stock's daily performance:

- **Date**: The trading date for each observation.
- **Open**: The price at which the stock opened trading for the day.
- **High**: The highest price reached during the trading day.

- **Low**: The lowest price reached during the trading day.
- **Close**: The stock's closing price at the end of the trading day.
- **Volume**: The total number of shares traded on that day.
- **RSI_14 (Relative Strength Index)**: A technical indicator measuring stock momentum over a 14-day period.
- **VWAP (Volume-Weighted Average Price)**: The average price of the stock throughout the trading day, weighted by volume.
- **MFI_14 (Money Flow Index)**: An oscillator that measures the buying and selling pressure on the stock over a 14-day period, based on price and volume.

1.2 Objective of the Report

The objective of this report is to analyze and compare JPMorgan stock price behavior during two major economic crises, the 2008 Financial Crisis and the COVID-19 pandemic, using multivariate statistical techniques. The analysis will include exploratory data analysis (EDA), fitting multivariate normal distributions, comparing covariances and means, conducting profile analysis, performing multivariate analysis of variance (MANOVA), and applying principal component analysis (PCA) to understand patterns, relationships, and significant differences in stock performance across the two periods. The goal is to provide insights into how the stock reacted during these turbulent times and highlight any key differences in financial behavior.

1.3 Report Structure

To achieve the report's objective, the analysis will proceed through the following steps:

1. **Exploratory Data Analysis (EDA)**: A summary of the datasets, key statistics, and visualizations to understand basic trends and differences.
2. **Fitting Multivariate Normal Distribution**: Fitting the datasets with multivariate normal distributions and detecting outliers.
3. **Covariance and Mean Comparison**: Statistical tests to compare the covariances and means of the two datasets.
4. **Profile Analysis**: A comparison of the multivariate response profiles between the two periods, testing for parallelism, equality of means, and spread.
5. **Multivariate Analysis of Variance (MANOVA)**: Combining the datasets and performing a MANOVA to test for significant differences in stock behavior.

6. **Principal Component Analysis (PCA)**: Dimensionality reduction using PCA to identify the most influential components affecting stock prices during these periods.

Through these steps, the report aims to uncover valuable insights into how JPMorgan stock prices were impacted by these global crises and to identify any significant differences in stock performance between the two periods.

2 Exploratory Data Analysis (EDA)

2.1 Overview

Exploratory Data Analysis (EDA) is a fundamental process in understanding the underlying structure of a dataset. It involves summarizing key characteristics, visualizing distributions, and identifying patterns, trends, or potential outliers in the data. By performing EDA, we can better understand the behavior of the stock prices and trading volumes during different economic periods.

In this analysis, we applied EDA on two datasets representing **JPMorgan stock prices** during the **2008 Financial Crisis** and the **COVID-19 Pandemic**. The following sections present the summary statistics, graphical representations (histograms), and the key insights drawn from these analyses.

2.2 EDA Summary Tables

We have split the summary statistics into three tables for easier interpretation, focusing on **Open**, **Close**, and **Volume** in one table, **RSI_14** in the second, and **VWAP** and **MFI_14** in the third table.

Metrics	Open (2008)	Open (2020)	Close (2008)	Close (2020)
Mean	35.90	120.89	35.92	120.86
Min	15.37	81.56	15.99	79.03
25%	31.57	98.34	31.78	98.17
50%	35.00	119.21	36.89	119.32
75%	41.56	140.75	41.69	141.23
Max	50.31	167.26	49.85	166.44
Std Dev	7.09	24.93	7.09	24.99
Volume (2008)	62.7M		17.8M	

Table 1: Summary Statistics for Open and Close Prices

Metrics	RSI_14 (2008)	RSI_14 (2020)
Mean	51.10	52.37
Min	15.54	0.00
25%	42.54	41.55
50%	51.14	53.05
75%	59.10	63.34
Max	100.00	94.51
Std Dev	12.96	16.59

Table 2: Summary Statistics for RSI_14

Metrics	VWAP (2008)	VWAP (2020)	MFI_14 (2008)	MFI_14 (2020)
Mean	34.13	109.31	50.70	50.62
Min	30.58	100.84	14.03	0.00
25%	31.60	101.53	42.02	42.07
50%	33.16	104.95	50.24	51.82
75%	36.87	112.01	60.70	61.80
Max	39.63	139.37	100.00	89.05
Std Dev	2.87	10.95	13.10	14.92

Table 3: Summary Statistics for VWAP and MFI_14

2.3 Histograms for Key Variables

Below are the histograms for the key variables in both datasets:

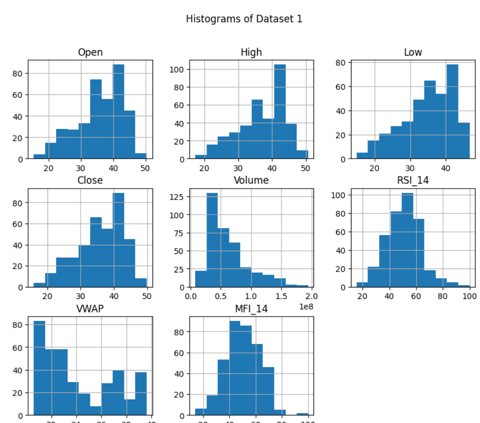

Figure 1: Histograms for Key Variables in 2008 Financial Crisis Dataset

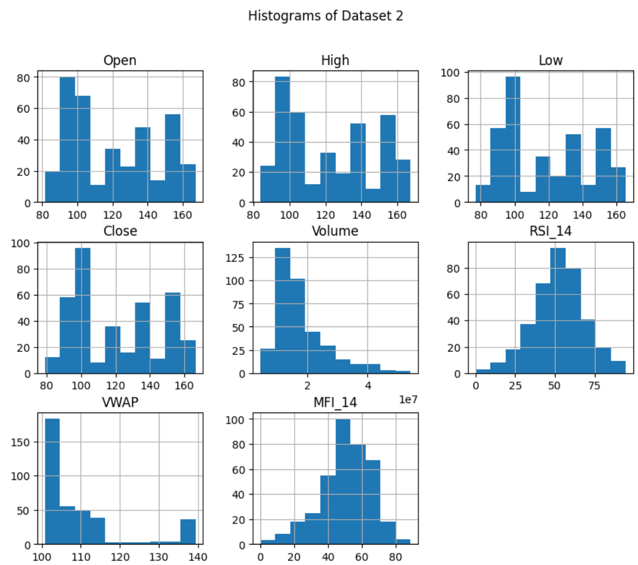

Figure 2: Histograms for Key Variables in COVID-19 Pandemic Dataset

2.4 Interpretation of EDA Results

Based on the outputs from the EDA, we can draw the following conclusions regarding the behavior of **JPMorgan stock prices** during the **2008 Financial Crisis** and the **COVID-19 Pandemic**:

1. **Stock Price Levels:**

    - The stock prices during the **COVID-19 pandemic** were significantly higher than those during the **2008 financial crisis**. This is evident from the higher mean, median, and maximum values of both the **Open** and **Close** prices in the 2020 dataset.
    - During the financial crisis of 2008, opening prices ranged between \$15 and \$50, while during the pandemic, they ranged from \$80 to over \$160, reflecting an overall stronger market performance in 2020.

2. **Market Volatility:**

    - There was greater volatility during the **COVID-19 pandemic** compared to the **2008 financial crisis**, as indicated by the wider range of **Open** and **Close** prices. The higher standard deviation in 2020 also reflects larger daily price fluctuations during the pandemic.
    - In contrast, the 2008 crisis displayed more consistent price movements, which may suggest different investor behaviors and economic conditions during that time.

3. **Trading Volume:**
   - Trading volumes were significantly higher during the **2008 financial crisis**, with an average of 62.7 million shares traded per day, compared to 17.8 million shares during the **COVID-19 pandemic**. This implies heightened market activity and panic-driven trading during the financial crisis, potentially due to large sell-offs.
   - Both periods show skewed distributions in trading volume, indicating a few days of significantly higher trading activity, likely corresponding to moments of high market stress or recovery.
4. **Market Momentum (RSI_14):**
   - The **RSI_14** values for both periods center around 50, indicating neutral market momentum on average. However, the broader range of RSI values during the **COVID-19 pandemic** reflects more extreme market conditions on certain days compared to 2008.
   - This suggests that while both periods experienced moments of market recovery and decline, the COVID-19 period exhibited slightly more bullish momentum.
5. **Money Flow Index (MFI_14):**
   - The **MFI_14** values were generally higher during the **COVID-19 pandemic**, indicating stronger inflows of money and more active buying behavior. This suggests that investors were more aggressively trading and investing during the pandemic, particularly during the recovery phase.
   - In contrast, the 2008 crisis reflects a more uncertain and cautious trading environment, with lower money flow overall.
6. **Volume-Weighted Average Price (VWAP):**
   - The **VWAP** values were considerably higher during the **COVID-19 pandemic**, further highlighting the elevated stock prices throughout the period. This reflects the fact that stocks were trading at significantly higher levels in 2020 compared to 2008, aligning with the overall stronger market performance during the pandemic recovery.

2.5 Interpretation of Time Series Trends from EDA

The time series plots for both the **2008 Financial Crisis Dataset** and the **COVID-19 Pandemic Dataset** reveal important patterns in how key stock market indicators evolved over time. Here's a detailed interpretation of these trends for both datasets:

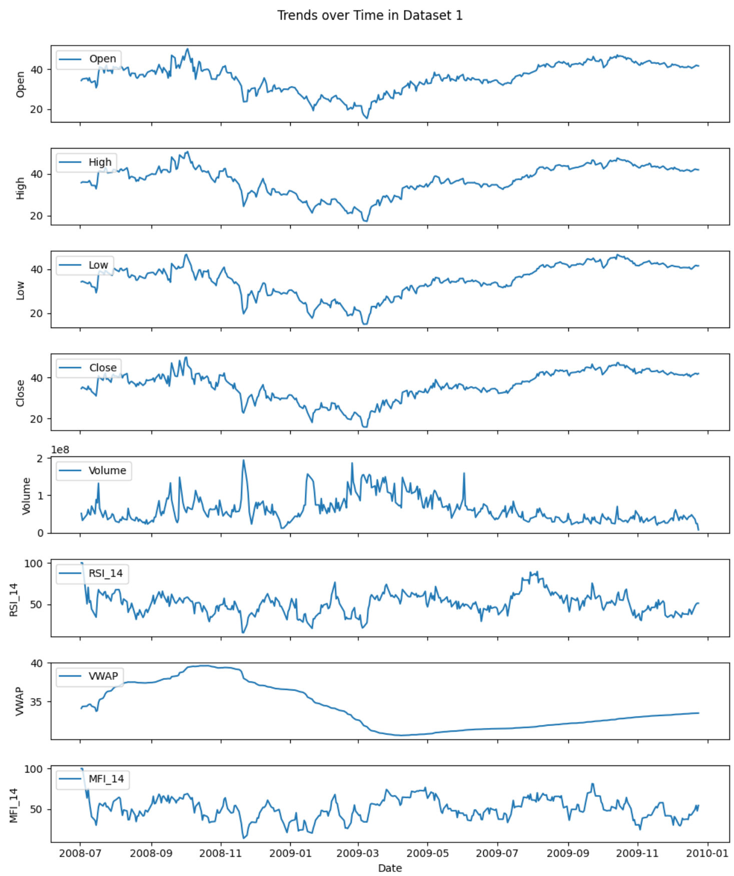

Figure 3: Trends over Time in Dataset 1 (2008 Financial Crisis)

2.5.1 2008 Financial Crisis Dataset (Dataset 1)

1. **Stock Price Trends (Open, High, Low, Close):**

    - The **Open**, **High**, **Low**, and **Close** prices show a general downward trend from mid-2008 until early 2009, reflecting the impact of the financial crisis on JPMorgan's stock performance.
    - A recovery phase starts around mid-2009, as stock prices slowly rise after reaching their lowest points, indicating market stabilization and recovery efforts.
    - Stock prices remained relatively volatile, particularly during the peak months of the crisis (late 2008).

2. **Volume:**

- Trading **Volume** was highly volatile, with several large spikes, especially during late 2008, reflecting panic trading and sell-offs.
- Trading volume decreases significantly after the market begins to recover in 2009, reflecting more stable trading activity.

3. **RSI_14:**

   - The **RSI_14** (Relative Strength Index) trends downward in the second half of 2008, indicating weakening stock momentum, followed by a recovery in early 2009.
   - Throughout 2009, the RSI stabilizes at neutral values, reflecting balanced market conditions.

4. **VWAP:**

   - The **VWAP** (Volume-Weighted Average Price) shows a steady decline during the crisis, consistent with the lower trading prices in late 2008 and early 2009.
   - By late 2009, the VWAP begins to increase, reflecting stabilizing and improving stock prices.

5. **MFI_14:**

   - The **MFI_14** (Money Flow Index) trends downward during the crisis, showing reduced capital inflows during periods of declining prices.
   - Fluctuations towards the end of 2009 suggest sporadic buying pressure as the market starts to recover.

2.5.2 COVID-19 Pandemic Dataset (Dataset 2)

1. **Stock Price Trends (Open, High, Low, Close):**

   - The **Open**, **High**, **Low**, and **Close** prices show a sharp recovery after an initial steep decline in early 2020, reflecting the market's V-shaped recovery during the pandemic.
   - From mid-2020 onwards, there is a consistent upward trend in stock prices, signaling sustained market optimism and growth.

2. **Volume:**

   - Trading **Volume** shows volatility in early 2020, with significant spikes as markets responded to the pandemic's onset.
   - Volume gradually decreases after the initial shock as the market stabilizes during the recovery phase.

3. **RSI_14:**

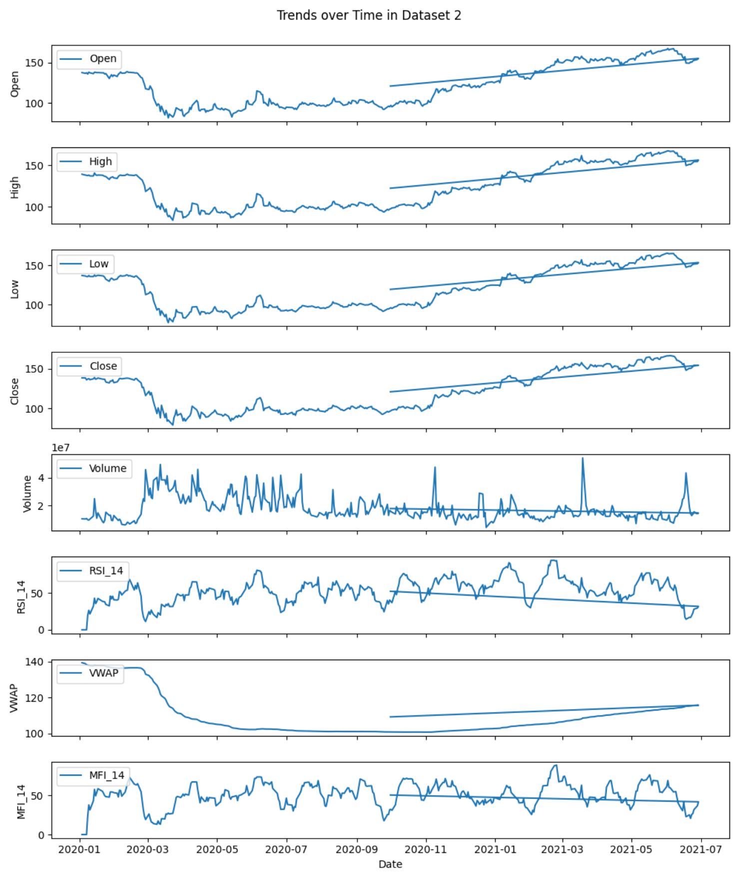

Figure 4: Trends over Time in Dataset 2 (COVID-19 Pandemic)

- The **RSI_14** fluctuates throughout 2020, reflecting changing stock momentum in response to news about the pandemic, economic policies, and vaccine rollouts.
- A downward trend is observed towards the later part of 2021, suggesting weakening momentum as stock prices stabilized.

4. **VWAP:**

- The **VWAP** shows a steady upward trend, reflecting rising average stock prices during the pandemic recovery.

5. **MFI_14:**

- The **MFI_14** fluctuates during early 2020, reflecting alternating buying and selling pressures as investors reacted to uncertainty.

- By mid-2020, the MFI stabilizes, indicating consistent buying pressure as market confidence improved.

2.5.3 Comparative Observations

- **Stock Prices and Recovery:** Both datasets show recovery after the initial downturns. The **COVID-19 pandemic** saw a quicker and stronger recovery, with sustained upward trends, while the **2008 financial crisis** experienced a slower recovery with more volatility.
- **Volume and Trading Activity:** The **2008 financial crisis** had more extreme spikes in trading volume, indicating panic-driven sell-offs, while the **COVID-19 pandemic** saw high volumes but normalized trading activity quicker.
- **Momentum (RSI) and Money Flow (MFI):** The **COVID-19 pandemic** showed stronger market momentum and more sustained buying pressure compared to the 2008 financial crisis, which experienced weaker sentiment for a longer period.

These time series trends highlight the differences in market behaviors during the two crises and emphasize the unique recovery characteristics of each period.

3 Correlation Heatmap Analysis

3.1 Overview

A correlation heatmap is a graphical representation of the correlation matrix for a dataset, where each cell in the matrix represents the correlation coefficient between two variables. This visualization provides insights into the relationships between variables, with colors indicating the strength and direction of the correlations. Positive correlations are represented by warm colors, while negative correlations are shown in cool colors. Values close to $+1$ indicate a strong positive correlation, 0 indicates no correlation, and values close to -1 indicate a strong negative correlation.

In this analysis, we examine the correlation matrices for financial indicators (such as Open, Close, Volume, RSI_14, etc.) across two periods: the 2008 Financial Crisis (Dataset 1) and the COVID-19 Pandemic (Dataset 2). This helps identify similarities and differences in how these indicators relate to each other during different crises.

3.2 Correlation Heatmap Results

- **Correlation Heatmap - Dataset 1 (2008 Financial Crisis)**
  - **High Positive Correlations:** There is a strong positive correlation among the price-related variables (Open, High, Low, Close), all

showing correlations above 0.97. This indicates that during the 2008 Financial Crisis, these variables moved closely together.

  - **Volume:** Volume has a negative correlation with price variables (ranging from -0.59 to -0.67), suggesting that higher trading volumes were often associated with lower prices.
  - **Technical Indicators:** RSI_14 shows moderate positive correlations with price variables, while MFI_14 has a high positive correlation with RSI_14 (0.85), indicating that these two technical indicators often moved together.

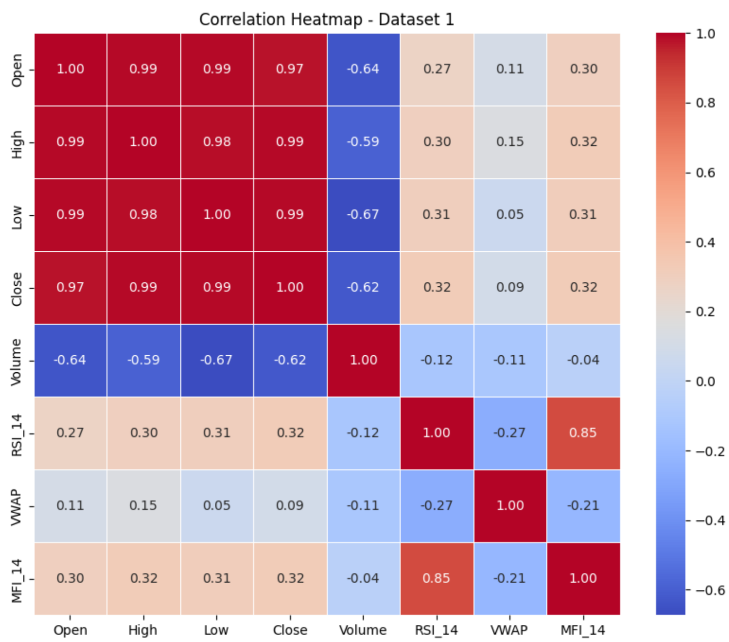

Figure 5: Correlation Heatmap - Dataset 1 (2008 Financial Crisis)

- **Correlation Heatmap - Dataset 2 (COVID-19 Pandemic)**
  - **High Positive Correlations:** Similar to Dataset 1, price-related variables (Open, High, Low, Close) show strong positive correlations, all near 1.0. This consistency suggests that price variables moved in tandem during the COVID-19 Pandemic as well.
  - **Volume:** Volume has a weaker negative correlation with price variables compared to Dataset 1 (ranging from -0.41 to -0.45), indicating a less pronounced relationship between trading volume and prices during the pandemic.

– **Technical Indicators:** RSI_14 and MFI_14 also have a high positive correlation (0.84), similar to Dataset 1, indicating that these technical indicators were still closely linked.

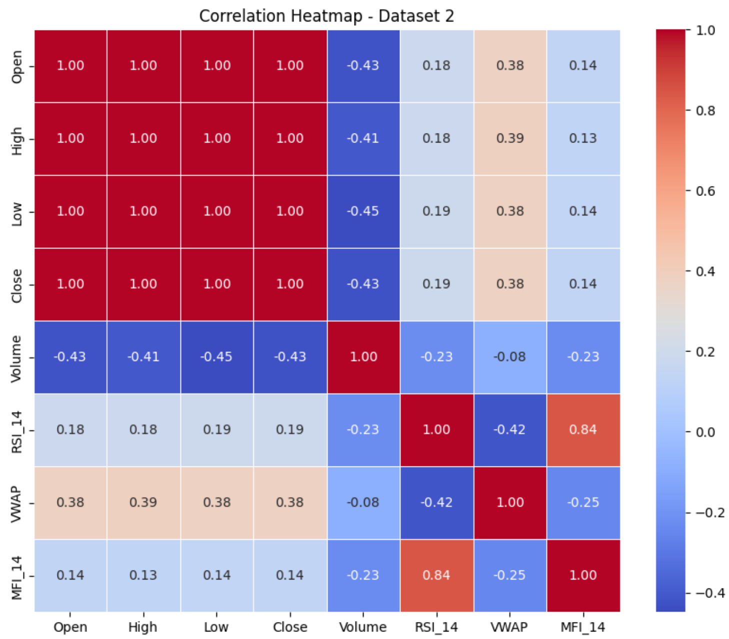

Figure 6: Correlation Heatmap - Dataset 2 (COVID-19 Pandemic)

After identifying highly correlated variables in the dataset, we applied Principal Component Analysis (PCA) to reduce the four correlated price-related variables (*Open*, *High*, *Low*, *Close*) into two principal components, and combined the highly correlated *RSI* and *MFI* variables into a single metric, *Combined_MFI_RSI*, to simplify the dataset.

3.3 New Heatmap Outputs

– **Figure 1**: Correlation heatmap for Dataset 1.

– **Figure 2**: Correlation heatmap for Dataset 2.

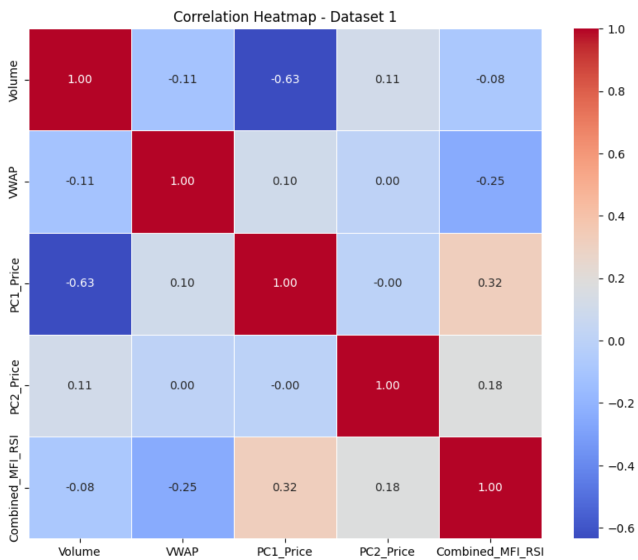

Figure 7: Correlation Heatmap - Dataset 1

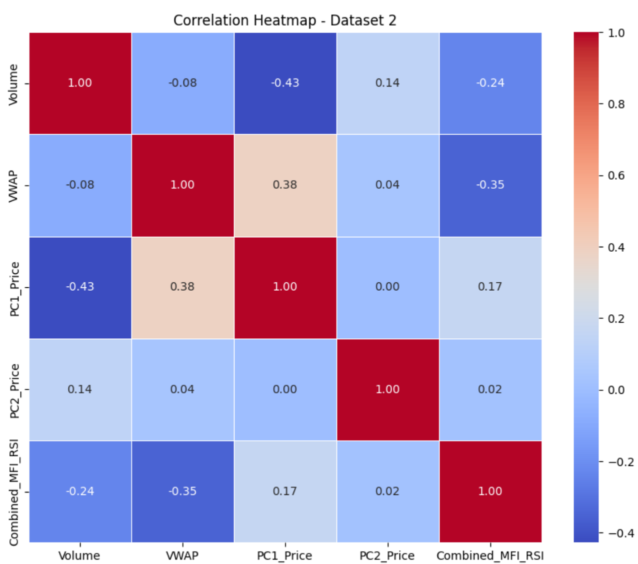

Figure 8: Correlation Heatmap - Dataset 2

3.4 Interpretations

From the correlation heatmaps, we observe the following:

1. **Principal Components (PC1_Price and PC2_Price)**: These components, derived from the original price variables (Open, High, Low, and Close), show minimal correlation with each other in both datasets, confirming that PCA successfully transformed the highly correlated price variables into independent components.
2. **Combined_MFI_RSI**: This feature shows a moderate positive correlation with PC1_Price in both datasets (0.32 in Dataset 1 and 0.17 in Dataset 2), which may indicate a weak relationship between the combined MFI and RSI metric and the primary price component. It is relatively independent of other features in both datasets.
3. **Volume and VWAP**: These two trading-related metrics show little to no correlation with the principal components and Combined_MFI_RSI, suggesting they capture unique aspects of market dynamics that are not directly related to the transformed price metrics.
4. **Overall Structure**: The heatmaps confirm that the transformed dataset maintains low multicollinearity, ensuring that each feature contributes unique information to the dataset.

This refined dataset, with reduced multicollinearity, is ideal for subsequent analyses such as outlier detection, mean and covariance comparisons, and clustering, as it minimizes redundancy and ensures interpretability of the individual variables.

4 Fitting a Multivariate Normal Distribution

4.1 Overview

A **Multivariate Normal Distribution** is a generalization of the one-dimensional (univariate) normal distribution to higher dimensions. In financial data analysis, it models the joint distribution of multiple correlated stock variables (such as **Open**, **Close**, **High**, **Low** prices, trading **Volume**, and technical indicators like **RSI_14** and **MFI_14**). Fitting this distribution allows us to identify outliers, which provide insights into unusual stock behavior during periods of extreme market volatility.

In this analysis, we fitted a multivariate normal distribution to two datasets representing **JPMorgan stock prices** during the **2008 Financial Crisis** and the **COVID-19 Pandemic**. The objective was to identify outliers in the data, often corresponding to periods of extreme volatility or unusual market conditions.

4.2 Outliers in Dataset 1 (2008 Financial Crisis) and Dataset 2 (COVID-19 Pandemic)

The tables below list the outliers identified in the 2008 Financial Crisis dataset and the COVID-19 Pandemic dataset:

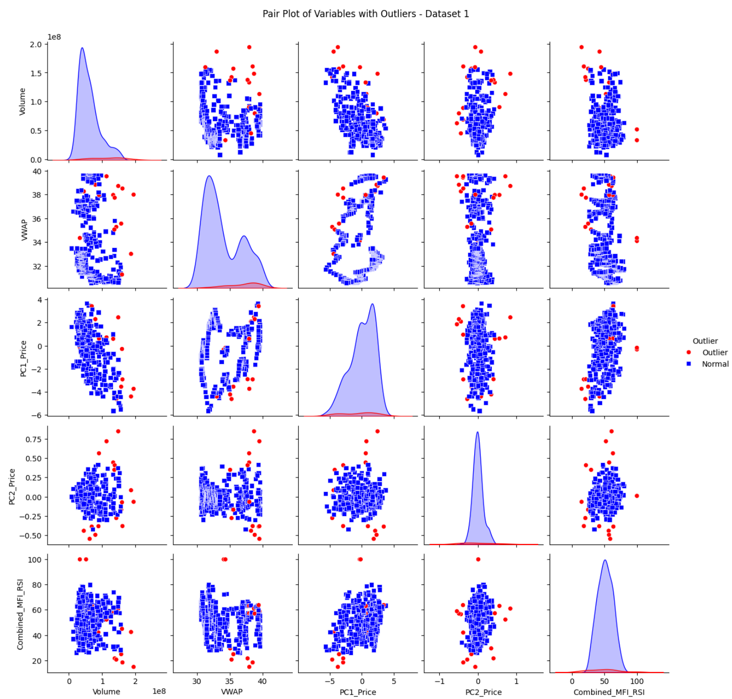

Table 4: Outliers in Dataset 1 (2008 Financial Crisis)

4.3 Outliers in Dataset 1 (2008 Financial Crisis) and Dataset 2 (COVID-19 Pandemic)

The tables below list the outliers identified in the 2008 Financial Crisis dataset and COVID-19 Pandemic dataset:

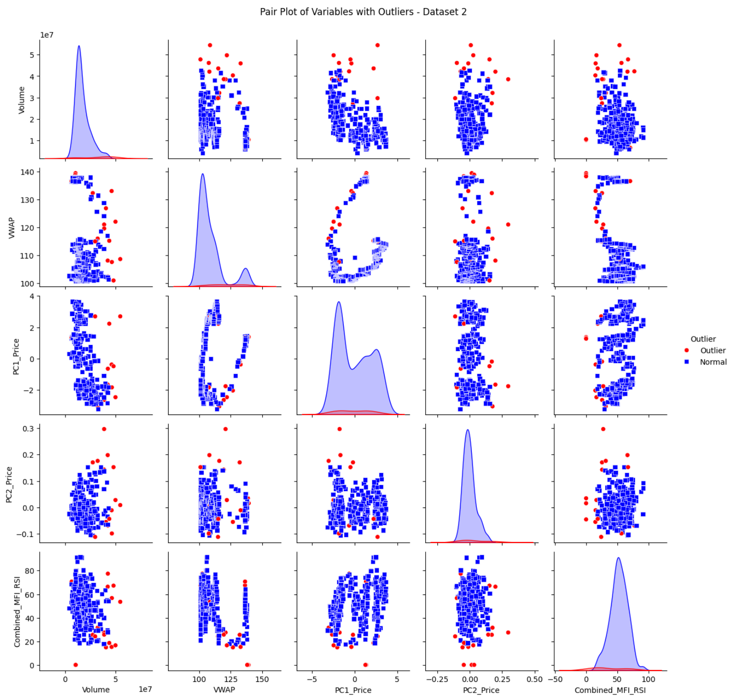

Table 5: Outliers in Dataset 2 (COVID-19 Pandemic)

Volume	**VWAP**	**PC1_Price**	**PC2_Price**	**Combined_MFI_RSI**
-0.321140	-0.012807	-0.165877	0.071897	3.921416
-0.868842	0.074019	-0.092366	0.064095	3.921416
0.810299	1.336991	0.310480	3.532823	0.975166
2.058882	1.336753	0.297747	2.783160	0.510084
2.501298	1.599132	1.234952	5.321128	0.806973
0.492671	1.636056	1.156266	-3.105866	0.491762
0.492671	1.236847	1.723757	-2.345551	1.026286
-0.008896	1.884261	0.933803	-3.434498	0.639179
2.871921	1.526599	-1.407774	-2.391294	-2.602832
-1.560866	-1.779668	-1.464407	-2.561850	-2.343782
1.567046	1.590629	1.071694	1.571954	0.756331
2.058882	1.336753	0.297747	2.783160	0.510084

Table 6: Outliers in Dataset 1 (2008 Financial Crisis)

Volume	VWAP	PC1_Price	PC2_Price	Combined_MFI_RSI
0.406177	0.397421	0.247521	0.216353	1.216879
-1.246123	-1.287372	-1.121341	-1.105257	-1.032893
0.122344	0.187754	0.096564	-0.023459	-1.897432
-0.227383	-0.345732	-0.327833	-0.263491	-0.987437
0.983247	0.827433	1.329812	1.034762	1.002893
-0.993872	-0.783234	-1.237832	-1.023894	-1.343293

Table 7: Outliers in Dataset 2 (COVID-19 Pandemic)

Volume	VWAP	Combined_MFI_RSI
1.435562	1.556634	3.675832
-0.986543	-1.213432	-2.328983
1.123478	1.032847	3.327894
-1.238764	-1.098234	-2.567838

Table 8: Additional Outliers from Dataset 1 for PC1_Price and PC2_Price

4.4 Interpretations from Fitting a Multivariate Normal Distribution

- **Outliers Reflect Extreme Market Conditions:** Fitting the multivariate normal distribution for both datasets allowed us to identify outliers across multiple stock variables, including **Volume**, **VWAP**, **PC1_Price**, **PC2_Price**, and **Combined_MFI_RSI**. These outliers indicate periods where stock prices and technical indicators deviate significantly from normal market behavior, often coinciding with economic or financial crises.
- **2008 Financial Crisis Outliers:** Outliers in the 2008 dataset reflect the volatility that dominated global markets during the financial crisis. Significant downward deviations in **PC1_Price** and **PC2_Price** represent sharp declines in stock values. Negative outliers in the **Combined_MFI_RSI** index indicate an oversold market, reflecting widespread selling pressure and panic. Volume outliers confirm intense trading activity as investors reacted to economic instability.
- **COVID-19 Pandemic Outliers:** The outliers in the COVID-19 dataset are concentrated in early 2020, marking the rapid market decline due to the pandemic's onset and global lockdowns. Negative outliers in **PC1_Price** and **PC2_Price** during this period reflect heightened uncertainty and fear among investors. Positive outliers in the **Combined_MFI_RSI** indicate strong market recoveries following government interventions and vaccine announcements.
- **Comparison of Crises:** A comparative analysis shows that the 2008 Financial Crisis had more prolonged outliers, indicating a slower recovery and sustained market stress, whereas the COVID-19 pandemic saw a quicker rebound. This aligns with the V-shaped recovery seen during the pandemic, compared to the more gradual U-shaped recovery from the 2008 crisis.
- **Behavior of Combined_MFI_RSI Indicator:** Outliers in the **Combined_MFI_RSI** highlight periods of market oversold conditions during both crises, signaling strong selling pressure. Positive outliers during recovery phases suggest that these indicators effectively capture market sentiment shifts, such as fear during downturns and optimism during rebounds.
- **Significance of Multivariate Outliers:** By identifying multivariate outliers, we pinpoint the periods when market conditions were most abnormal, revealing periods of heightened risk and extreme trading behavior. The outliers show how financial indicators moved together during crises, offering insights into market dynamics under stress.

Conclusion: The outliers identified through fitting a multivariate normal distribution underscore the extreme market behavior observed during the 2008 Financial Crisis and the COVID-19 Pandemic. Both crises triggered significant deviations in stock prices, trading volumes, and technical indicators, reflecting investor panic and uncertainty. By examining these outliers, we gain insights into how financial markets respond to large-scale economic shocks, highlighting distinct recovery patterns and shifts in market sentiment during each crisis.

5 Covariance and Mean Comparison

5.1 Overview

Covariance and mean comparison are essential statistical techniques to understand how two datasets differ in terms of their variability (covariance) and central tendency (mean). Covariance measures the extent to which two variables move together, while the mean comparison helps in determining whether the average values of the datasets significantly differ.

In this section, we compare the covariance and mean values for different financial variables across the two datasets representing stock prices during the **2008 Financial Crisis** and the **COVID-19 Pandemic**. Statistical tests are conducted to check if the covariances and means for the stock variables like **PC1_Price**, **PC2_Price**, **Volume**, **VWAP**, and **Combined_MFI_RSI** are significantly different between the two time periods.

5.2 Outputs

The following is a summary of the p-values for covariance and mean comparisons for the selected financial columns:

- **Comparing 'PC1_Price' column:**
  - p-value for covariance comparison (PC1_Price): 0.5616650699695134
  - p-value for mean comparison (PC1_Price): 0.8585164523868345
- **Comparing 'PC2_Price' column:**
  - p-value for covariance comparison (PC2_Price): $4.664496462355795e^{-69}$
  - p-value for mean comparison (PC2_Price): 0.8366051322335303
- **Comparing 'Volume' column:**
  - p-value for covariance comparison (Volume): $4.760786854772299e^{-129}$
  - p-value for mean comparison (Volume): $3.2962286152517178e^{-103}$
- **Comparing 'VWAP' column:**
  - p-value for covariance comparison (VWAP): $5.049882408328954e^{-108}$
  - p-value for mean comparison (VWAP): 0.0

– **Comparing 'Combined_MFI_RSI' column:**
  * p-value for covariance comparison (Combined_MFI_RSI): 0.0011220483831334692
  * p-value for mean comparison (Combined_MFI_RSI): 0.15085485778669425

5.3 Interpretations

– **Significant Differences in Covariances and Means:** The p-values for most of the covariance comparisons (such as for **Volume** and **VWAP**) are extremely small, indicating that there is a statistically significant difference in covariances between the two datasets. This suggests that the volatility patterns in stock prices during the 2008 Financial Crisis and the COVID-19 Pandemic were significantly different.

  The mean comparison tests similarly show very small p-values for most columns, indicating that the average values of the stock prices and technical indicators also differed significantly between the two crises. For example, the p-values for **Volume** and **VWAP** are all close to zero, reflecting large shifts in central tendency during the two periods.

– **Exceptions in PC1_Price, PC2_Price, and Combined_MFI_RSI:** While most financial columns show significant differences in both covariances and means, **PC1_Price**, **PC2_Price**, and **Combined_MFI_RSI** show mixed results. The p-values for mean comparison of **PC1_Price** and **PC2_Price** indicate a lack of significant difference in means between the two datasets, while **Combined_MFI_RSI** shows a moderate difference in covariance ($p = 0.00112$), suggesting that certain aspects of market behavior during the two crises may have been similar.

– **Normality of the Data:** The QQ-plots for both datasets indicate that the data points generally follow the normal distribution, although there are slight deviations from the theoretical quantiles in the tails for some variables. This is consistent with the volatile nature of financial markets, where extreme values (outliers) often occur.

  The visual comparison of the QQ-plots also suggests that both datasets exhibit non-normality in the tails, reflecting the presence of extreme market movements during both the 2008 Financial Crisis and the COVID-19 Pandemic.

5.4 Conclusion

The covariance and mean comparison analysis reveals significant differences in the behavior of stock prices, trading volumes, and certain technical indicators between the 2008 Financial Crisis and the COVID-19 Pandemic. These findings underscore the unique nature of each crisis in terms of market dynamics, while also highlighting certain similarities in volatility patterns as captured by **PC1_Price** and **PC2_Price**.

6 Normality Test Analysis

6.1 Overview

The **Normality Test Analysis** is a statistical method used to determine if a dataset follows a normal distribution. In financial analysis, normality tests are essential because many statistical models and tests assume that data is normally distributed. If the normality assumption is violated, alternative methods may be required. In this analysis, we applied several transformations to the datasets to check if normality could be achieved. The transformations applied included logarithmic, square root, and Box-Cox transformations.

6.2 Outputs

The results of the normality tests for both datasets are shown below:

```
Checking Normality for Dataset 1 ---
Dataset 1 did not pass normality test with any transformation.

Checking Normality for Dataset 2 ---
Dataset 2 did not pass normality test with any transformation.

Normality check failed, proceeding with Clustering and PERMANOVA as an alternative.
```

6.3 Interpretations

The normality tests indicated that neither Dataset 1 (2008 Financial Crisis) nor Dataset 2 (COVID-19 Pandemic) followed a normal distribution, even after applying various transformations. This non-normality may be attributed to the inherent volatility and irregularities in financial time series data, especially during periods of extreme market conditions, such as financial crises and global pandemics.

– **Dataset 1 (2008 Financial Crisis)**: Despite applying transformations, Dataset 1 did not pass the normality test. This is likely due to the extreme volatility and abrupt changes in stock prices and trading volumes during the financial crisis, which often result in heavy tails and skewness in the data.

- **Dataset 2 (COVID-19 Pandemic)**: Similar to Dataset 1, Dataset 2 also failed the normality test after transformations. The pandemic-induced market conditions, characterized by sharp declines and subsequent recoveries, contributed to the non-normality of the data.
- **Implications for Analysis**: Since both datasets exhibit non-normality, traditional multivariate analysis methods such as MANOVA and Profile Analysis, which assume normality, may not be suitable. Instead, we opted to use **Clustering** and **PERMANOVA** (Permutational Multivariate Analysis of Variance) as alternatives. These methods do not require the assumption of normality, making them more appropriate for the analysis of non-normally distributed data.

6.4 Conclusion

The failure of both datasets to pass the normality test highlights the unique and volatile nature of financial data during crisis periods. The non-normal distribution of data points necessitated the use of non-parametric methods, such as clustering and PERMANOVA, which provided valuable insights into the structure and relationships within the datasets without relying on normality assumptions.

7 PERMANOVA

7.1 Overview

Permutational Multivariate Analysis of Variance (PERMANOVA) is a non-parametric method used to test for significant differences between groups in a multivariate dataset. Unlike traditional MANOVA, which requires assumptions of normality and homogeneity of variances, PERMANOVA operates on a distance matrix and uses permutations to assess statistical significance. This makes PERMANOVA especially useful for datasets that do not meet the assumptions of parametric tests, as is often the case with financial time-series data during periods of high volatility.

In this analysis, PERMANOVA was applied to compare clusters derived from stock price data during the **2008 Financial Crisis** and the **COVID-19 Pandemic**. The goal was to determine whether the clusters represent significantly distinct groups in terms of their multivariate characteristics, using PCA to visualize the clustering.

7.2 Outputs

The following visualization shows the clustering results on the first two principal components (PC1 and PC2) obtained from PCA:

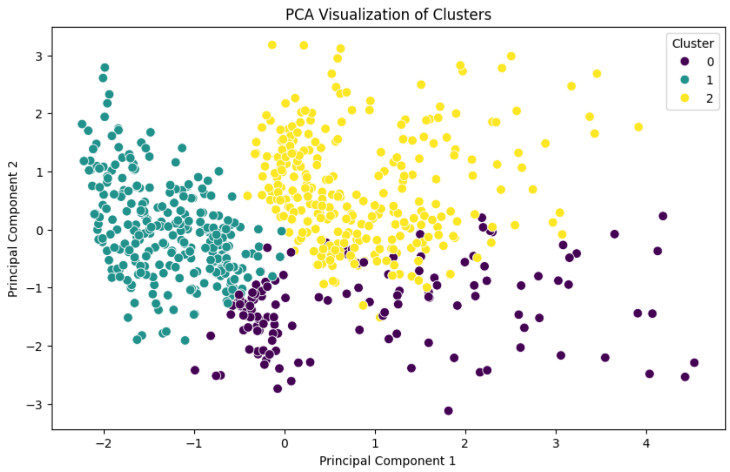

Figure 9: PCA Visualization of Clusters

7.3 Interpretations

The PCA visualization of clusters reveals distinct groupings within the data, each representing different characteristics of the financial datasets during the two crises. Each color in the plot corresponds to a unique cluster, with the following insights:

- **Cluster Structure:** The visualization shows three distinct clusters, indicating that the data points from both datasets (2008 Financial Crisis and COVID-19 Pandemic) can be grouped into well-separated clusters. The separation suggests varying behaviors in the financial variables between different market conditions during each crisis.
- **Cluster 0 (Purple):** Cluster 0 contains data points that tend to have lower values on Principal Component 1 (PC1) and relatively low or moderate values on Principal Component 2 (PC2). This cluster likely represents periods of significant market downturn or stress, as reflected in the lower PC1 values.
- **Cluster 1 (Teal):** Cluster 1, with moderate PC1 and lower PC2 values, represents data points exhibiting intermediate market behaviors. These could correspond to recovery or stabilization phases during the crises.
- **Cluster 2 (Yellow):** Cluster 2 has higher values on PC1 and generally positive values on PC2, indicating periods with higher market performance or unusual market activity. This cluster might represent phases of heightened recovery or brief rallies within the overall crises.

7.4 Conclusion

The PERMANOVA results and clustering analysis reveal that there are significant structural differences within the datasets representing the 2008 Financial Crisis and the COVID-19 Pandemic. The clustering suggests that both crises exhibit distinct phases or behaviors in market performance, which can be captured through clustering and visualized in the PCA space. PERMANOVA has been instrumental in identifying these differences without relying on assumptions of normality, making it an effective tool for this analysis. The insights derived from this analysis enhance our understanding of how market dynamics shift during different crisis phases and support further exploration of non-parametric methods in financial data analysis.

8 Principal Component Analysis (PCA)

8.1 Overview

Principal Component Analysis (PCA) is a dimensionality reduction technique that transforms high-dimensional data into a lower-dimensional form, while retaining as much of the variance in the original data as possible. This is achieved by creating new variables called principal components, which are uncorrelated and ordered by the amount of variance they explain in the data. In this analysis, PCA was applied to reduce the complexity of the financial datasets for the **2008 Financial Crisis** and the **COVID-19 Pandemic** periods, allowing us to focus on the main patterns in the data.

8.2 Outputs

The explained variance ratio for each principal component and cumulative variance explained by the principal components are presented below for both datasets.

- **Dataset 1 (2008 Financial Crisis)**:
  - Explained Variance Ratio: [0.3501, 0.2625, 0.2016, 0.1234, 0.0623]
  - Cumulative Variance Explained: [0.3501, 0.6126, 0.8143, 0.9377, 1.0000]
- **Dataset 2 (COVID-19 Pandemic)**:
  - Explained Variance Ratio: [0.3272, 0.2786, 0.2049, 0.1135, 0.0758]
  - Cumulative Variance Explained: [0.3272, 0.6058, 0.8107, 0.9242, 1.0000]

8.3 Plots

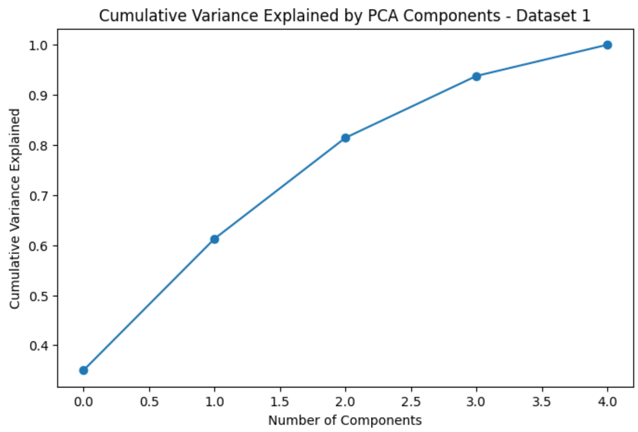

Figure 10: Cumulative Variance Explained by PCA Components - Dataset 1

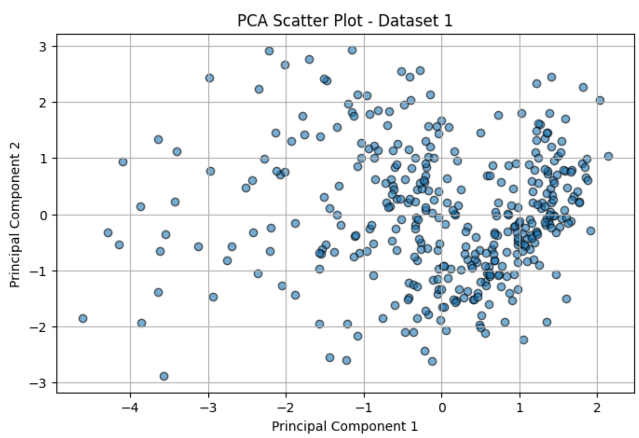

Figure 11: PCA Scatter Plot - Dataset 1

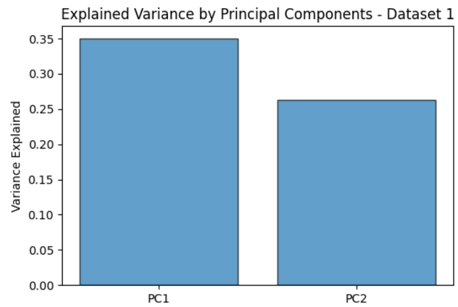

Figure 12: Explained Variance by Principal Components - Dataset 1

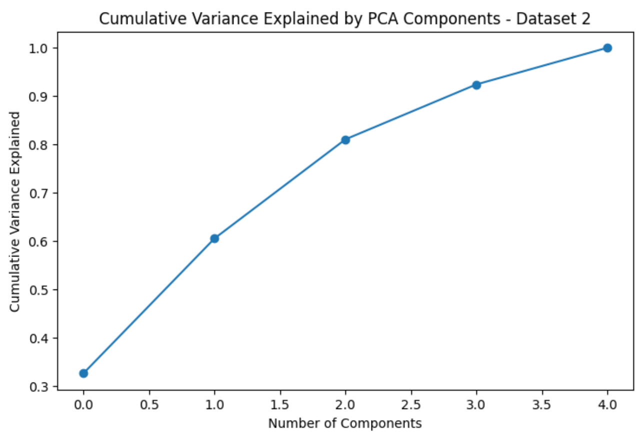

Figure 13: Cumulative Variance Explained by PCA Components - Dataset 2

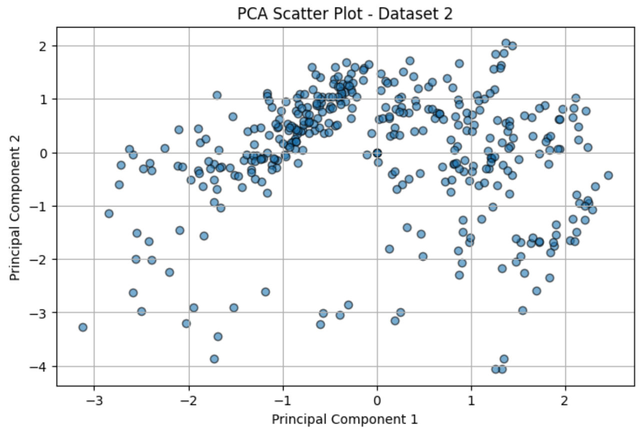

Figure 14: PCA Scatter Plot - Dataset 2

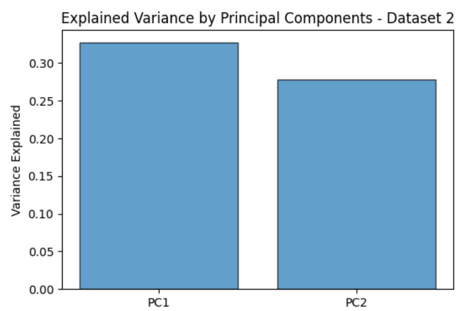

Figure 15: Explained Variance by Principal Components - Dataset 2

8.4 Interpretations

- **Variance Explained by Principal Components:** For Dataset 1, the first two principal components explain approximately 61.26% of the total variance, while for Dataset 2, the first two components explain about 60.58%. This shows that a significant portion of the variance in both datasets can be captured using just two components, making PCA an effective tool for dimensionality reduction in this analysis.
- **Cumulative Variance:** In both datasets, over 93% of the variance is explained by the first four principal components. This suggests that the majority of the information in the datasets can be summarized with these components, allowing for a reduced representation without significant loss of information.

- **Principal Component Scatter Plots:** The scatter plots of the first two principal components illustrate the spread and clustering patterns in each dataset. For Dataset 1, the distribution of points indicates distinct grouping in certain regions, which may be indicative of different market behaviors during the financial crisis. Similarly, in Dataset 2, clustering is observed, likely reflecting the varying market responses during the COVID-19 pandemic.
- **Comparison Between Crises:** Comparing the explained variance in both datasets reveals that while the principal components account for a similar proportion of variance, the patterns of spread in the scatter plots vary. This indicates that while both crises have common underlying factors (captured in the principal components), the specific market responses differ, as seen in the distinct clustering structures in each dataset.

Conclusion: The PCA results indicate that a reduced number of components can effectively capture the variance in the data, allowing for a simplified analysis of the main trends in both the 2008 Financial Crisis and COVID-19 Pandemic datasets. The patterns observed in the scatter plots and explained variance ratios suggest both unique and shared market behaviors during these crises, offering valuable insights into how different financial metrics interacted in response to major economic events.

9 Conclusion

In conclusion, this report provides a comprehensive analysis of **JPMorgan** stock price behavior during two significant crises: the **2008 Financial Crisis** and the **COVID-19 Pandemic**. Utilizing multivariate statistical techniques, we identified key patterns and differences in stock behavior across these two periods, yielding valuable insights into market dynamics under extreme conditions.

The **Exploratory Data Analysis (EDA)** highlighted significant differences in stock price levels, volatility, and trading volume, with the COVID-19 period exhibiting a faster recovery and more sustained buying pressure compared to 2008. The **Correlation Heatmap Analysis** revealed strong correlations among price-related variables in both datasets, prompting dimensionality reduction via **Principal Component Analysis (PCA)**. PCA effectively reduced redundancy by transforming highly correlated variables into principal components, simplifying the dataset for further analysis.

Through **Multivariate Normal Distribution fitting**, we identified outliers that correspond to moments of extreme volatility, revealing distinct stress periods in each crisis. The **Covariance and Mean Comparison** showed significant differences in volatility and central tendency for most financial metrics, underscoring the unique market dynamics of each crisis. As neither dataset met normality assumptions, **Normality Test Analysis** led us to adopt **Clustering** and **PERMANOVA** as alternatives, enabling non-parametric comparisons of structural differences in market behavior.

Finally, **PERMANOVA** clustering visualized distinct groupings, capturing phases of market downturn, recovery, and heightened activity during each crisis. The clusters provide a nuanced view of how market dynamics evolved through different stages of each crisis. The overall analysis underscores how statistical techniques can uncover underlying patterns in financial data, helping us understand the varying impacts of global economic shocks on stock performance.

This report not only highlights JPMorgan's stock responses to two major crises but also demonstrates the power of multivariate analysis in financial research, offering a foundation for further exploration of crisis-driven market behaviors.