Abstract

This research aims to dive into the complex dynamics of global financial markets, with an emphasis on detecting market anomalies and assessing the impact of inflation on global indices and precious metals. The study uses a thorough predictive modelling method to identify optimal hedging options for investment portfolios by expanding the diversification opportunities to include commodities such as gold and silver. In a time where economic trends have a significant impact on investing decisions, understanding these trends is critical for successful portfolio management and risk mitigation. The study employs a comprehensive analytical approach to achieve these objectives, utilising advanced data analytics techniques. The methods include K-means and hierarchical clustering to find anomalies, whereas Support Vector Machines and Naïve Bayes Classification are used to assess the impact of inflation. The study is expected to provide valuable insights into the behaviour of global indices and precious metals in response to inflation. Anomalies detected through clustering algorithms provide a deeper understanding of market movements, while the classification models reveal the nuanced relationship between inflation and asset prices. These findings provide useful insights for diversification in various asset classes, emphasising the necessity of a diversified portfolio to hedge against economic uncertainty.

Table of Contents

Introduction

The constant rise in global inflation rates emphasises the need for strategic investment in the modern financial world. Inflation causes a consistent increase in the overall price level of goods and services, and gradually reduces the buying power of money, posing a significant challenge to both individual and institutional investors. "People invest in stocks for two opposite reasons - in hope and confidence in the future of an enterprise or in fear that the value of their capital will be lost through inflation" (Bernard Baruch)**.** The primary objective of this research is to understand the complexity of the equity and commodity markets, with an additional emphasis on identifying market anomalies and understanding the impact of inflation on important global indices and precious metals. Additionally, developing effective risk management techniques that can study economic volatilities.

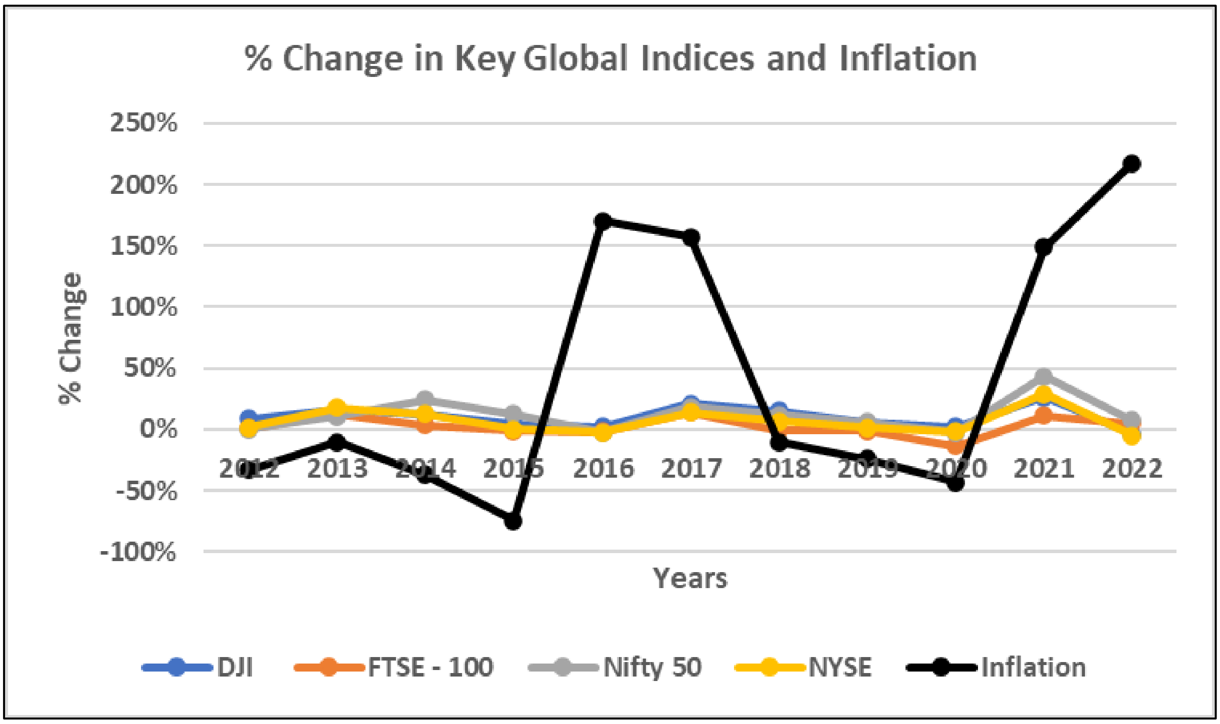

Figure: 1.1 Percentage Change in Key Indices and Inflation 2012 – 2022 (Created by Author)

Figure 1.1 The line graph displays the percentage change in key global indices from 2012 to 2022, demonstrating a consistent pattern in their movement over the years. At the same time, the line graph of inflation shows an inverse relationship link with the stock market, i.e. rise in inflation often causes a downfall in the financial market.

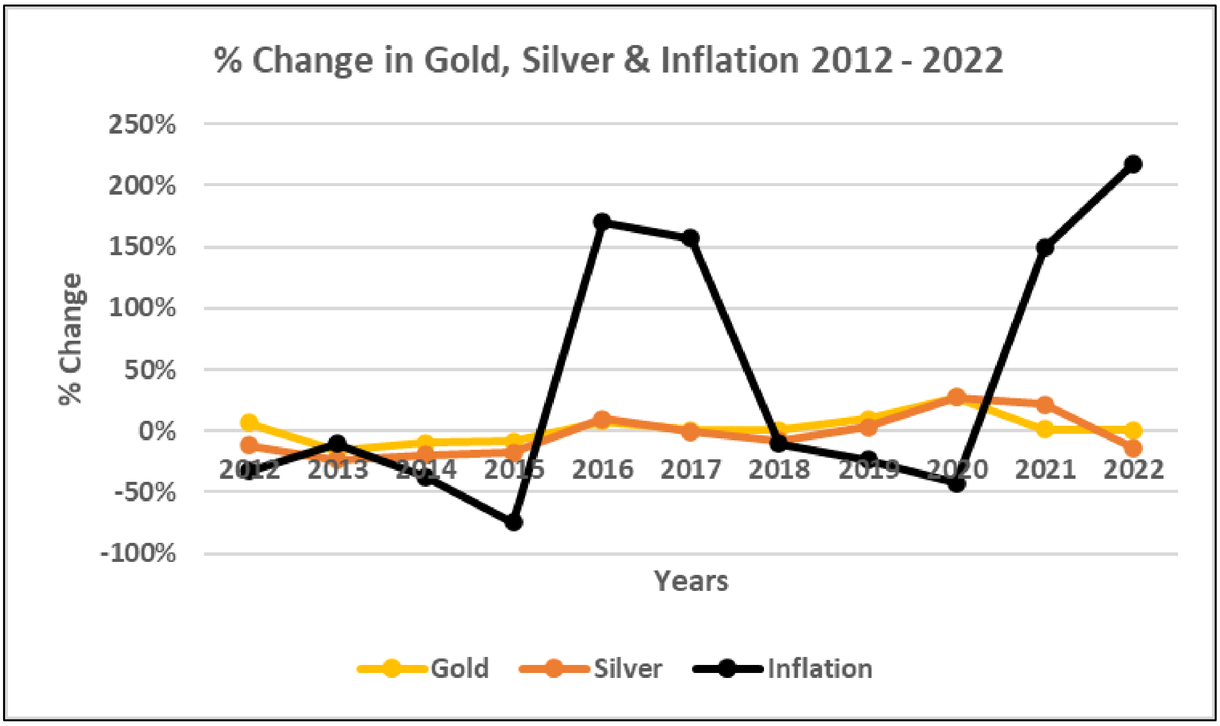

Figure: 1.2 Percentage Change in Gold, Silver and Inflation 2012 – 2022 (Created by Author)

Figure 1.2 The line graph displays the percentage change in Gold and Silver from 2012 to 2022, demonstrating a consistent pattern in their movement over the years. At the same time, the line graph of inflation shows an inverse relationship link with the stock market, i.e. rise in inflation often causes a downfall in the financial market.

Research questions and objectives

Research question: 01

Exploring the effectiveness of advanced clustering algorithms for detecting market anomalies in equity and commodity markets.

Objective: 01

This objective involves implementing and critically evaluating advanced clustering approaches, including K-means and hierarchical clustering. The goal is to discover and analyse anomalies in stock markets and precious metals, to provide risk management strategies customised to a variety of market scenarios.

Research question: 02

Creating a reliable, adaptive classification model to evaluate and forecast the effects of inflation on global indices and precious metals.

Objective: 02

The purpose is to create a classification framework using Support Vector Machines, and Naïve Bayes Classification. This framework is designed to perform a thorough assessment of historical price movements of equity and commodity market data to determine the impact of changes in inflation, especially while utilising various asset classes as inflation hedges.

Research question: 03

Optimised portfolio diversification techniques using predictive modelling, with a specific focus on hedging through Gold and Silver.

Objective: 03

This objective necessitates conducting a thorough analysis of past performance of gold, silver, and key financial indices, as well as risk-free rates. The purpose is to use advanced predictive modelling approaches like Non-Linear and Ridge Regression to estimate future market trends and identify the optimised portfolio diversification opportunities for various economic scenarios.

Feature selection, Data gathering and analysis:

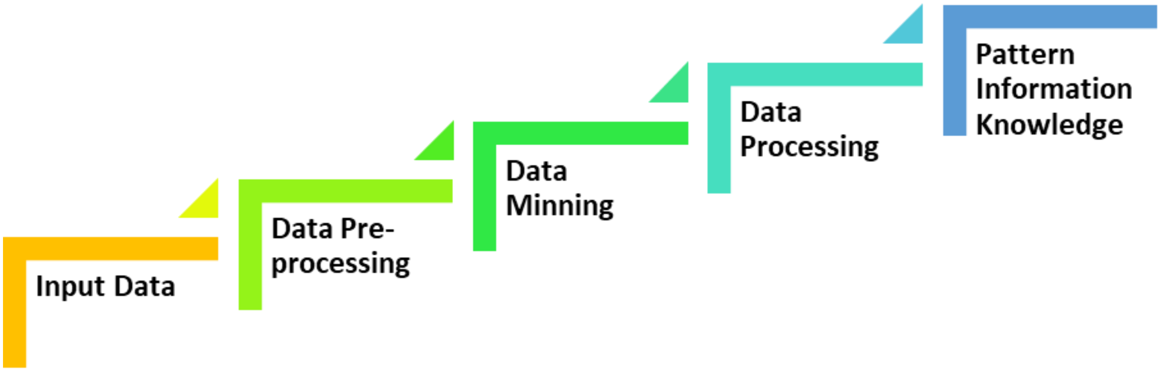

Figure 1.3: process knowledge discovery in databases (Created by Author)

Figure 1.1 depicts the proposed process of the process knowledge discovery in databases.

1. Anomalies detection in price movement:

Data Type: String data (investment instrument names). Numeric data (integers and floats) - Daily closing prices and trading volumes of indices and commodities during the previous decade (2012-2022).

Data Structure: Cross-sectional and time-series data for clustering by price and volume (Chang and Cai, 2023). Columns include investment instrument name, date, opening and closing prices, day's high and low, and trading volume.

Data Gathering: Online financial databases include websites for various indices, Yahoo Finance, and Bloomberg. Open-source database: Kaggle (Narne, 2023).

Data Analysis Approach: Use of K-means and hierarchical clustering because of their effectiveness in discovering trends in financial data (Strimpel, 2023).

2. The Effect of Inflation on Indices and Precious Metals:

Data Type: String data (investment instrument names). Numeric data includes inflation rates, price fluctuations in indices and precious metals, and trade volume (Chiang, 2022).

Data Structure: Time-series data is used to analyse the long-term impact of inflation. The data structure includes columns for name, date, inflation rates, price movements, quantities of indices and precious metals, and exponential moving averages.

Data Gathering: Economic databases include the World Bank and IMF. (Hossainds,2022), (www.imf.org, 2023). Financial market databases include Bloomberg and Yahoo Finance.

Data Analysis Approach: SVM and Naïve Bayes classification were chosen for their accuracy in financial trend analysis (Tang and Chou, 2015).

3. Portfolio Diversification Strategies:

Data Type: String data (investment instrument names). Numeric data includes historical prices, returns, volatility, and risk-free rates.
Data Structure: Cross-sectional and time-series data with emphasis on previous performance and prediction indicators. Comprehensive financial dataset includes columns for asset names, dates, prices, returns, volatility indicators, and risk-free rates.

Data Gathering: Online financial database – World Bank, International Monetary Fund, Website of the respective indices, Yahoo Finance, Bloomberg (World Bank, 2023), (Macrotrends, 2023). Open-source database: Kaggle.

Data Analysis Approach: Non-linear regression and Ridge regression are effective at financial forecasting and addressing several correlations. (Niu, Wang and Zhang, 2023).

Table: 1.1 Data Features Overview (Created by Author)

No.	Feature	Description	Data Type	Data Analysis Relevance
1	**Year/Month**	Time reference for data points	Time series data	Time-series analysis, trend identification
2	**Investment instrument name**	Name of the financial instrument	String data	Categorising data; essential for comparative analysis
3	**DJI**	Avg. closing price of the Dow Jones Industrial Average	Numeric (Float) data	Key in analyzing market trends, identifying patterns, and performance evaluation of this specific index.
4	**FTSE 100**	Avg. closing price of the Financial Times Stock Exchange 100 Index	Numeric (Float) data	To study UK market trends, comparison with other indices, and economic impact analysis.
5	**Nifty 50**	Avg. closing price of the Nifty 50 index (India)	Numeric (Float) data	To examine Indian market dynamics, global comparison
6	**NYSE**	Avg. closing price of the New York Stock Exchange Composite Index	Numeric (Float) data	To understand US market behaviour, comparative analysis with other indices.
7	**Gold**	Avg. closing price of gold per ounce	Numeric (Float) data	Analysing gold's role as a safe-haven asset, inflation hedge, and its correlation with market indices.
8	**Silver**	Avg. closing price of silver per ounce	Numeric (Float) data	To study silver's market performance and correlation with gold and other commodities.
9	**Avg. inflation rate**	Avg. inflation rate over a specific period	Numeric (Percentage) data	Impact of economic conditions on investment instruments, for studying the correlation between inflation and asset prices.
10	**Avg. risk-free rate**	Average rate of return of risk-free securities over a specific period	Numeric (Percentage) data	Evaluating investment performance against a risk-free benchmark and optimizing portfolio diversification strategies.

Table 1.1 presents a comprehensive list of features from the dataset used in the study, including their structure, data type, and relevance in significance to the data analysis objectives.

1. **Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models** (Niu, Wang and Zhang, 2023)

This article's study of geopolitical risks and stock market volatility aligns closely in terms of detecting market anomalies and evaluating external variables impacting the volatility of the stock market. It emphasises the significant impact of military buildups and wars on market volatility. The significance of machine learning-based volatility projections is shown by several machine learning and forecast combination approaches. The findings about the influence of geopolitical events on market volatility have immediate implications for our research into external economic issues, such as inflation. Key findings include the efficacy of machine learning algorithms over traditional approaches for forecasting volatility. The study contributes significantly to risk management and volatility forecasting, giving useful insights into the complexities of geopolitical influence on financial markets.

2. **Stock return anomalies identification during the COVID-19 with the application of a grouped multiple comparison procedure** (Chang and Cai, 2023)

This study analyses stock return anomalies during the COVID-19 pandemic, focusing on the Chinese stock market, the process of using multiple hypothesis testing is strongly related to this research of advanced statistical methods to find anomalies. The results that various industries react differently to natural disasters, such as a pandemic, give important insight into how external factors like geopolitical issues or economic crises may affect different sectors of the market. This knowledge is critical to this study, which seeks to investigate the complicated dynamics of market behaviour under various external influences, such as inflation and geopolitical conflicts.

Data Pre-processing

Data pre-processing is an essential stage in the analysis of financial data. It includes a wide range of activities targeted at improving raw data's usefulness for generating insights and developing prediction models.

Descriptive Analysis

Table: 1.2 Descriptive Statistics on Avg. Monthly closing for the year 2018 (Created by Author)

Descriptive Statistics on Avg. Monthly closing for the year 2018							
Descriptive Statistics	**DJI**	**FTSE - 100**	**Nifty 50**	**NYSE**	**Gold**	**Silver**	**Inflation**
Count	12	12	12	12	12	12	12
Mean	25045.75	7357.36	10750.33	12647.54	1271.95	15.72	2.30
Std.	693.40	304.48	365.46	410.96	53.59	1.05	0.18
Min.	23805.55	6790.90	10232.62	11624.26	1202.58	14.28	2.00
25%	24579.76	7125.30	10518.07	12536.56	1220.85	14.77	2.20
50%	24979.89	7337.07	10703.71	12647.27	1270.40	16.10	2.30
75%	25614.50	7652.54	10838.14	12818.76	1327.39	16.54	2.33
Max	26232.67	7695.64	11498.44	13286.10	1338.38	17.17	2.70

In 2018, the DJI was noticeably stable, with a mean of **25,045.75** and a comparatively low std. of **693.40**, indicating sustaining performance with small volatility. The FTSE-100 likewise exhibited a small variance **(std. 304.48)**, with a mean of **7,357.37**, indicating a generally stable market in contrast to the more volatile Nifty 50, which had a std. of **365.46**.

Table: 1.3 Descriptive Statistics on Avg. Monthly closing for the year 2019 (Created by Author)

Descriptive Statistics on Avg. Monthly closing for the year 2019							
Descriptive Statistics	**DJI**	**FTSE - 100**	**Nifty 50**	**NYSE**	**Gold**	**Silver**	**Inflation**
Count	12	12	12	12	12	12	12
Mean	26378.37	7277.35	11433.87	12864.17	1396.57	16.22	1.75
Std.	1071.47	173.66	431.74	469.03	96.67	1.20	0.20
Min.	24157.80	6861.59	10804.89	11879.58	1285.18	14.65	1.40
25%	25739.25	7204.30	11087.34	12629.41	1298.71	15.21	1.65
50%	26280.84	7297.62	11503.06	12877.34	1390.22	15.80	1.80
75%	26947.46	7370.40	11709.59	13054.26	1489.66	17.21	1.90
Max	28167.01	7549.26	12096.88	13714.16	1515.91	18.19	2.00

In 2019, the NYSE and Gold's mean increased to **12,864.17** and **1,396.57**, respectively, indicating a bullish market trend. Silver (std. 1.20) began to exhibit an upward trend, with a mean price of **16.22**, from **15.72** in 2018, supporting portfolio diversification in precious metals.

Table: 1.4 Descriptive Statistics on Avg. Monthly closing for the year 2020 (Created by Author)

Descriptive Statistics on Avg. Monthly closing for the year 2020							
Descriptive Statistics	**DJI**	**FTSE - 100**	**Nifty 50**	**NYSE**	**Gold**	**Silver**	**Inflation**
Count	12	12	12	12	12	12	12
Mean	26906.89	6276.17	11118.03	12634.56	1777.74	20.68	1.00
Std.	2409.43	589.38	1423.81	1217.60	141.64	4.44	0.43
Min.	22637.42	5729.35	9063.58	10726.11	1562.13	14.94	0.50
25%	25614.46	5912.64	9943.92	11881.89	1680.72	17.53	0.70
50%	27777.39	6120.57	11327.64	12837.64	1797.97	19.46	0.85
75%	28609.80	6314.01	12007.38	13681.83	1876.91	24.56	1.20
Max	30148.58	7558.00	13550.44	14370.38	1980.27	27.10	1.80

In 2020, Gold had a significant mean increase to **1,777.74**, indicating potential as a hedge against economic uncertainties. In contrast, the Nifty 50 **(std. 1,423.81)** and other indices had a higher std., indicating more volatility throughout the COVID-19 pandemic.

Table: 1.5 Descriptive Statistics on Avg. Monthly closing for the year 2021 (Created by Author)

Descriptive Statistics on Avg. Monthly closing for the year 2021							
Descriptive Statistics	**DJI**	**FTSE - 100**	**Nifty 50**	**NYSE**	**Gold**	**Silver**	**Inflation**
Count	12	12	12	12	12	12	12
Mean	34009.89	6999.97	16006.06	16243.80	1800.32	25.20	2.49
Std.	1657.12	226.28	1325.84	706.48	40.38	1.70	1.38
Min.	30821.35	6583.58	14284.60	14835.83	1719.99	22.52	0.70
25%	33445.79	6870.71	14926.41	15914.82	1777.98	23.84	1.45
50%	34489.17	7056.36	15758.39	16532.87	1799.48	25.72	2.25
75%	35102.63	7144.66	17257.86	16727.93	1823.36	26.20	3.20
Max	35848.57	7280.23	18020.22	17079.32	1866.37	27.61	4.80

2021, saw a significant recovery, DJI with a mean of **34,009.89** and a lower std., suggesting a more reliable market. In particular, the mean inflation increased to **2.49**, highlighting the inflationary impact on markets after the pandemic.

Figure: 1.6 Descriptive Statistics on Avg. Monthly closing for the year 2022 (Created by Author)

Descriptive Statistics on Avg. Monthly closing for the year 2022							
Descriptive Statistics	**DJI**	**FTSE - 100**	**Nifty 50**	**NYSE**	**Gold**	**Silver**	**Inflation**
Count	12	12	12	12	12	12	12
Mean	32911.74	7360.27	17249.42	15463.89	1805.82	21.82	7.90
Std.	1601.78	172.88	778.98	895.70	88.63	2.27	1.55
Min.	30570.68	6971.96	15933.16	14091.56	1674.36	18.88	4.90
25%	31513.17	7263.09	16723.26	14889.27	1734.08	19.54	7.40
50%	33213.76	7407.72	17404.12	15344.39	1812.31	21.68	8.40
75%	34101.05	7480.67	17626.95	16347.05	1850.19	23.54	8.90
Max	35456.14	7554.64	18385.13	16863.79	1952.32	25.46	9.60

In 2022, the Nifty 50's high mean of **17,249.42** and a significant rise in the mean of inflation to **7.90** emphasise the importance of diversification, to mitigate risks.

Table: 1.7 Descriptive Statistics on Avg. Closing for the years 2018 – 2022 (Created by Author)

Descriptive Statistics on Avg. closing for the years 2018 - 2022							
Descriptive Statistics	**DJI**	**FTSE - 100**	**Nifty 50**	**NYSE**	**Gold**	**Silver**	**Inflation**
Count	5	5	5	5	5	5	5
Mean	29055.36	7054.97	13320.42	13971.84	1610.68	19.91	3.09
Std.	4111.49	459.80	3062.93	1748.33	256.25	3.96	2.75
Min.	25053.95	6275.52	10750.67	12624.35	1271.26	15.71	1.00
25%	26379.54	7002.53	11153.00	12653.07	1397.71	16.23	1.75
50%	26890.67	7276.49	11435.85	12864.52	1779.31	20.70	2.30
75%	32897.35	7357.44	16019.11	15455.16	1799.03	21.76	2.49
Max	34055.29	7362.89	17243.47	16262.10	1806.07	25.16	7.90

Overall, from 2018 to 2022, the DJI's mean was **29,055.36**, with std. of **4,111.49**, indicating significant volatility. The FTSE-100 and Nifty 50 displayed high volatility with a wider range. Gold and silver displayed less volatility, silver had a broader range compared to its mean than gold. Inflation rose sharply on average, with the period's peak greatly skewing the mean, indicating volatility. Descriptive statistics provide a foundation for advanced clustering algorithms to detect uncertainties and develop responsive classification models for forecasting inflation's impact on markets (Hayes, 2023).

Visualizations

1. Line graph for Historical data and Forecasting

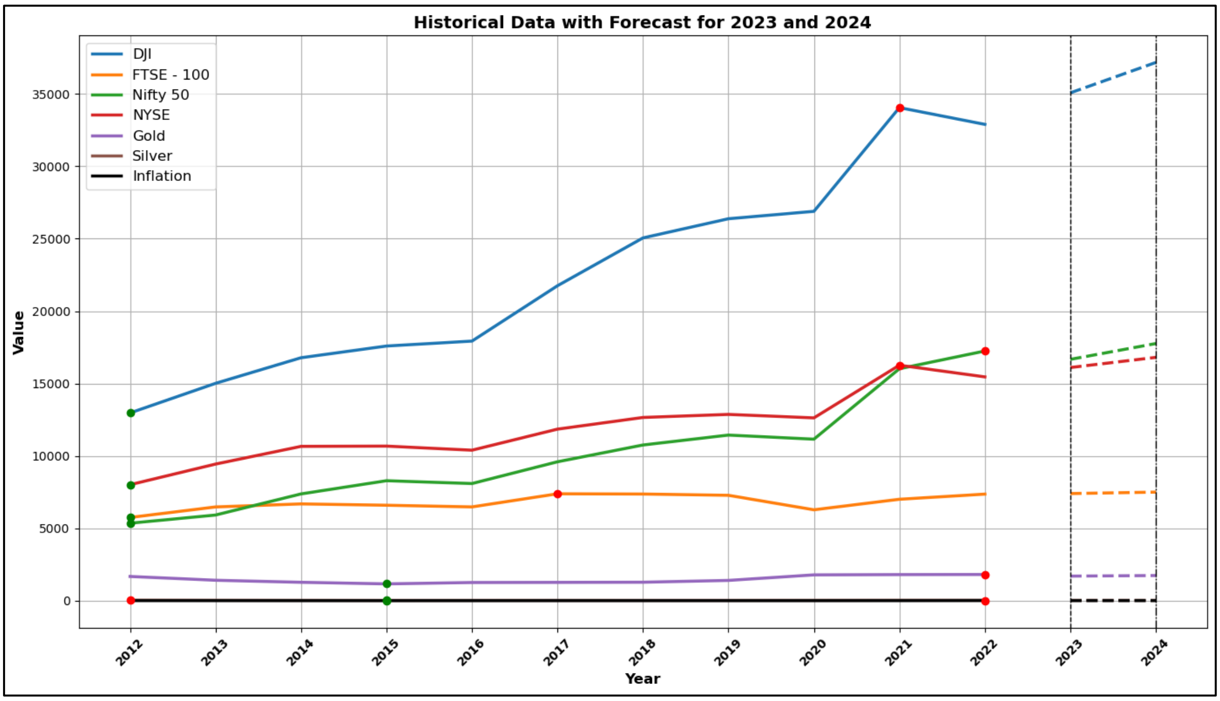

Figure: 1.4 Line graph on historical data and Forecasting (Created by Author)

This line chart displays historical data and forecast, DJI's steady rise from **13,000** in 2012 to **35,000** in 2022, with forecasted uptrend into 2023 and 2024. However, this may not be accurate enough as it is based on absolute numbers, with some variables measured in the thousands and others in the tens. To rectify this disparity, percentage changes throughout the years are studied.

2. Line graph on percentage change in Indices and Inflation

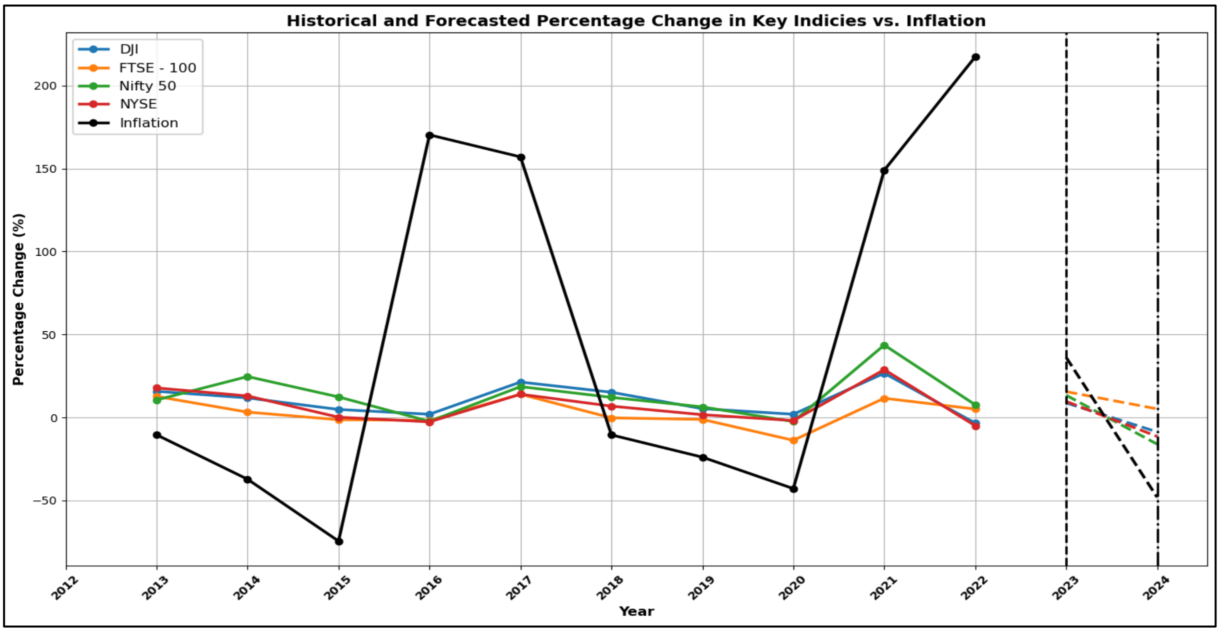

Figure: 1.5 Line graph on percentage change of indices and Inflation (Created by Author)

The line graph on percentage change shows inflation's rise to about **8%** in 2022, compared to an average of roughly **2.3%** previously. This inflation spike is an important variable for developing adaptive classification models, to forecast the impact on global indices as it's visible that an increase in inflation has an inverse impact on indices.

3. Line graph on percentage change in Metals and Inflation

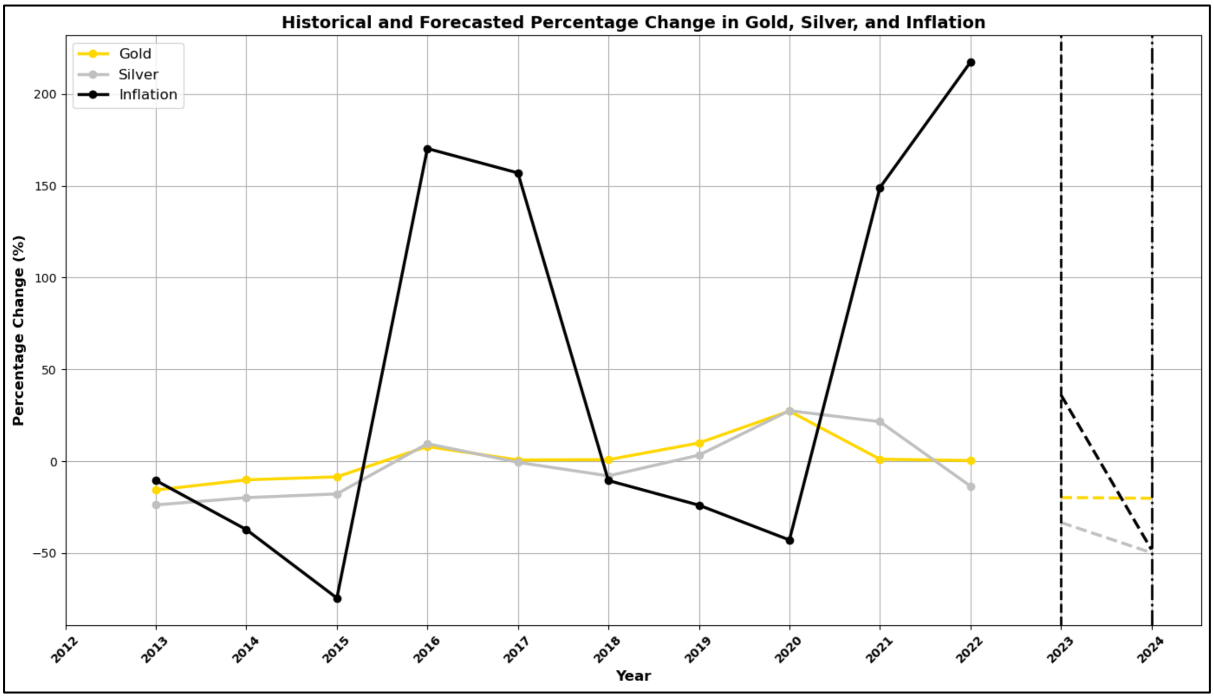

Figure: 1.6 Line graph on percentage change of Metals and Inflation (Created by Author)

Metals, unlike indices, do not have a significant correlation with inflation. In 2016, despite a substantial spike in inflation, gold and silver prices kept on rising. Also, in 2022, fluctuating inflation rates proved to have minimal impact on metals.

4. Cumulative Returns: A Case Study

A theoretical scenario is developed to demonstrate diversification as a hedge against inflation. An initial £1000 investment in all instruments in 2012 and tracking the returns of that investment based on their respective returns. Along with the key instruments, one diversification strategy is added to highlight the benefits of hedging and diversification (LIOUDIS, 2022).

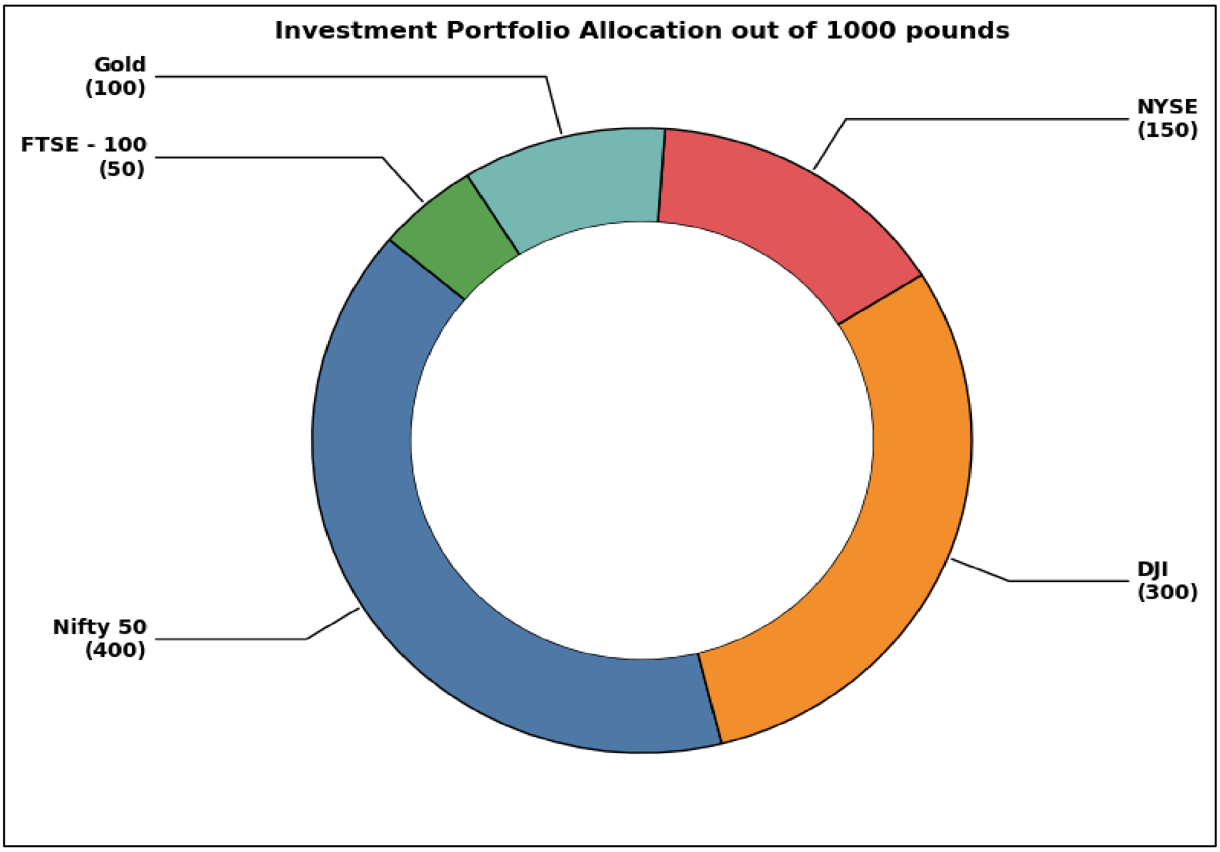

Figure: 1.7 Diversification using Strategy 1 into various asset classes (Created by Author)
This Donut chart shows portfolio diversification of £1000, **£400** in Nifty 50, and **£300** in DJI suggesting a preference for equities. The allocations of **£100** to Gold and **£50** to the FTSE-100 represent balancing growth and security.

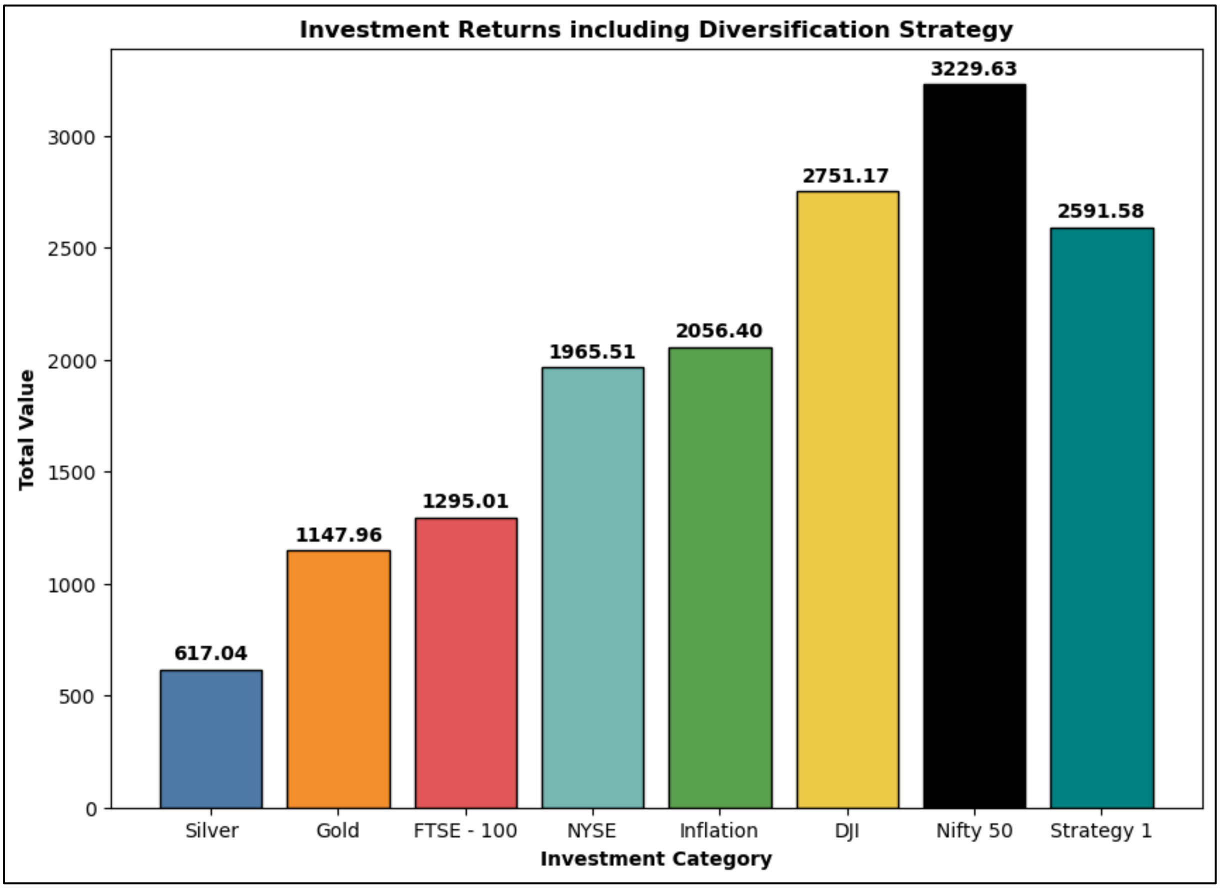

Figure: 1.8 Bar graph Return on Investment (Created by Author)

The bar chart illustrates the Return on Investment (ROI), Nifty 50 leading at **£3229.63** and outperforming inflation (£2056.40 in 2022 = £1000 in 2012). This underlines the importance of optimised portfolio diversification, as Strategy 1 is valued at **£2591.58**, making it one of the best performers and beating inflation by a fair margin. On the other side, silver has negative returns, while other instruments such as gold, FTSE-100, and NYSE have positive returns but fail to beat inflation.

Proximity analysis

1. Correlation Heatmap of Indices, Metals, and Inflation

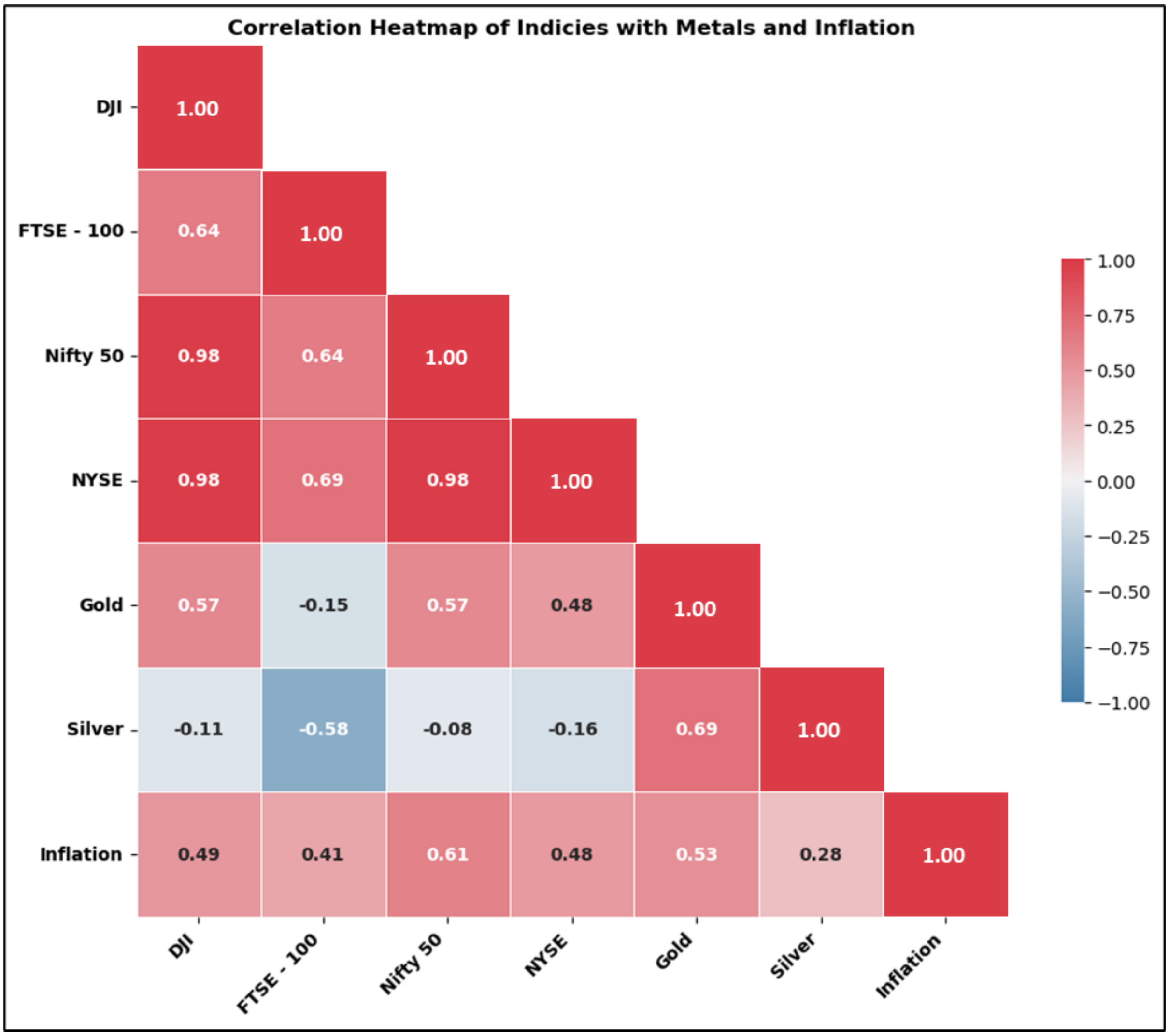

Figure: 1.9 Correlation Heatmap of Indices, Metals and Inflation (Created by Author)

The heatmap shows both positive and negative correlations (Szabo, 2020). The high correlation of **0.98** between DJI, NYSE and Nifty 50 reflects similar market patterns. In contrast, FTSE-100 and Silver have a negative correlation of **-0.58**, indicating that they move in opposing directions, offering a strategic insight into diversification for hedging against market uncertainties.

2. Euclidian Distance Heatmap of Indices, Metals, and Inflation

The heatmap represents the Euclidean distance between financial products, indicating their similarity (Gorthy, 2021). Lower values reflect comparable movements, as shown with the FTSE-100 and Nifty 50 **(1.53)**, implying connected behaviour. Higher values indicate dissimilarity; for example, DJI and Inflation **(9.80)** are likely to respond to different dynamics. Moderate levels exhibit partial similarity, showing both shared influences and separate causes.

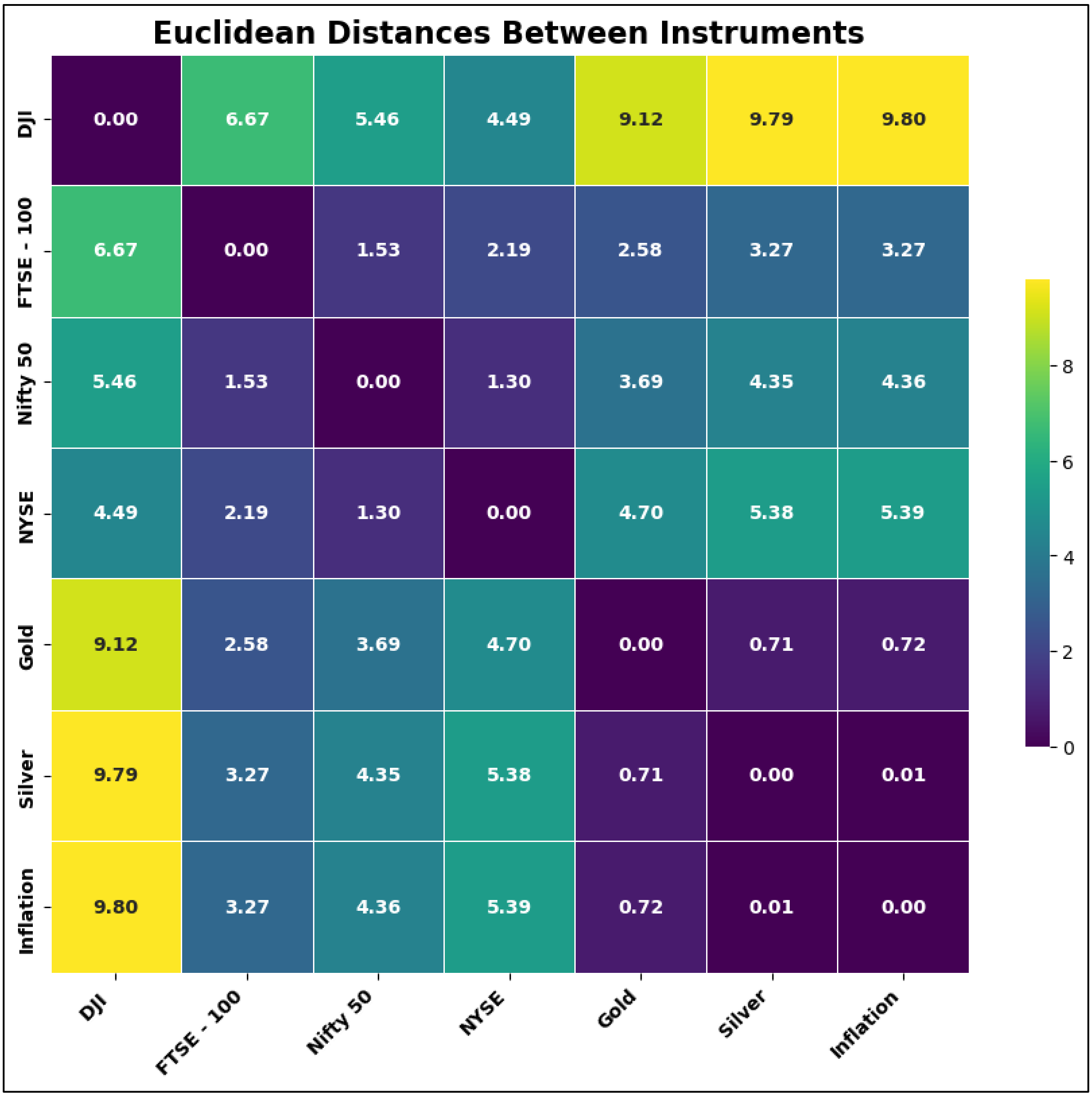

Figure: 1.10 Euclidian distance of Indices, Metals, and Inflation (Created by Author)

These proximity strategies were chosen to capture the complexity of financial data. Correlation analysis identifies directional movement across instruments, which is critical for understanding market synchronisation. Euclidean distances provide a clear measure of similarity that is useful for discovering diversification possibilities.

Data Pre-processing steps

Data Cleaning

The data set contained columns like Date, Open, High, Low, Close, Volume, and currency of the selected instruments. Outliers can suggest market abnormalities; thus, they should be carefully analysed rather than eliminated. Due to limitations in data collection of Inflation and Risk-free rate, as only monthly data is available, the data for our instruments was scaled up to a monthly and yearly basis instead of daily data. Also, for ease of analysis of price movements, only Avg. closing price is considered along with the volume.

Data Integration

Integration involves gathering data from various sources, including IMF, World Bank, Yahoo Finance, Bloomberg, and Kaggle. Also, the data from the official websites of the respective index was taken to ensure the credibility of the data. To maintain coherence and facilitate accurate forecasts and portfolio diversification modelling, the data is synchronised by date, standardised the currency to USD, and reconciled various data formats into a uniform structure. In the case of inflation, the dataset contained data since 1989, whereas in the case of Indices and metals the dataset contained data since 2000, so to focus only on years of our study i.e. 2012 to 2022 data for previous years was removed.

Data Reduction

Considering the large volume of data, daily data for 10 years, it's critical to reduce it to a manageable size without sacrificing any important characteristics. Principal Component Analysis (PCA) was used to minimise dimensionality while keeping volatility in data (Parte, 2020).

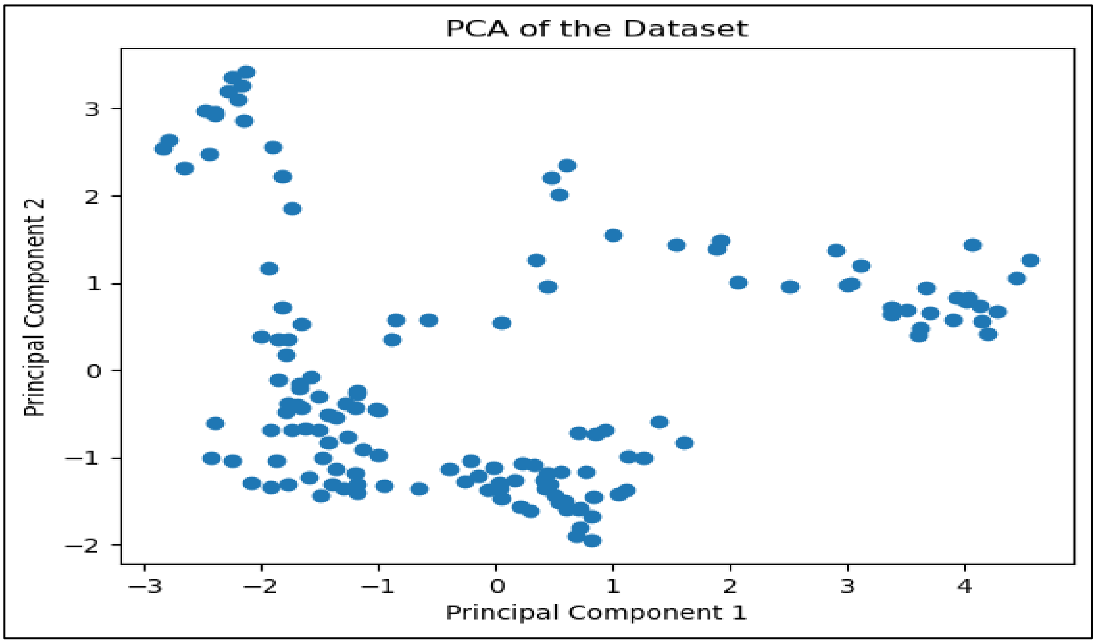

Figure: 1.11 PCA output (Created by Author)

The PCA results demonstrate that the top two main components account for more than 85% of the variance, showing a strong representation of the data's structure in two dimensions.

Index	Date	Open	High	Low	Close	Adj Close	Volume
NYSE	03/01/2012	7477.03	7659.97	7459.45	7624.32	7624.32	3943710000
FTSE - 100	03/01/2012	5572.3	5699.9	5572.3	5699.9	5699.9	778529200
Nifty 50	03/01/2012	4675.8	4773.1	4675.8	4765.3	4765.3	0
DJI	03/01/2012	12221.19	12479.65	12221.19	12397.38	12397.38	152560000

Figure: 1.12 Snapshot of data before cleaning & reduction (Created by Author)

Year / Month	DJI	FTSE - 100	Nifty 50	NYSE	Gold	Silver	Inflation
2012	**12,965.31**	**5,744.81**	**5,348.78**	**8,008.24**	**1,670.55**	**31.18**	**2.58**
Jan-12	12,550.99	5,694.46	4,942.83	7,737.67	1,661.08	30.97	3.20
Feb-12	12,889.05	5,893.36	5,409.09	8,071.44	1,745.62	34.25	3.10
Mar-12	13,079.47	5,875.40	5,295.58	8,166.76	1,676.32	32.96	3.10

Figure: 1.13 Snapshot of data after cleaning & reduction (Created by Author)

Data Transformation

Data should be normalised or standardised to remove the disparity, which is essential for clustering and predictive modelling. Some manual transformation was done to standardise the formatting errors and to arrange the data in a suitable format for analysis. Along with that standardisation, Normalisation, and Box-cox were done to identify outliers (Soni, 2021).

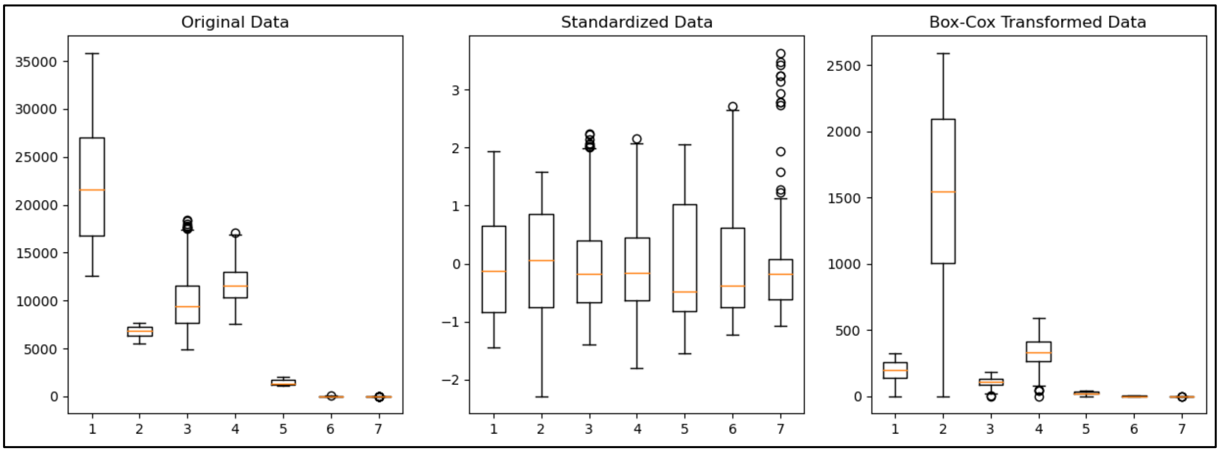

Figure: 1.14 Normalisation, standardisation, and Box-cox transformation (Created by Author)

Original data varied levels of skewness, with inflation having a high positive skewness **(2.19)**, suggesting a lengthy tail to the right. Silver also exhibits significant positive skewness **(0.99)**. Box-Cox the skewness values for all variables are significantly closer to zero, showing that the Box-Cox transformation successfully decreased skewness and made the distributions more symmetric. Inflation had the biggest skewness at first, but now has a skewness close to zero **(0.003)**, indicating a roughly symmetrical distribution.

Table: 1.8 Skewness data (Created by Author)

Index	Original Data	Normalized Data	Standardized Data	Box-Cox Transformed Data
DJI	0.378	0.378	0.378	-0.257
FTSE - 100	-0.423	-0.423	-0.423	-0.403
Nifty 50	0.719	0.719	0.719	-0.148
NYSE	0.400	0.400	0.400	-0.157
Gold	0.435	0.435	0.435	-0.079
Silver	0.989	0.989	0.989	-0.017
Inflation	2.186	2.186	2.186	-0.003

Data Discretisation

The data was transformed into categorical bins, reducing the complexity of financial data for predictive modelling. It's critical for optimising portfolio diversification methods since it allows for the segmentation of data points into separate investment categories or risk levels. The data was divided into 3 groups which are Equity, Metals, and Inflation.

Data Processing

Identifying optimum numbers of clusters

Elbow method

The Elbow Method graph displays the number of clusters vs inertia. Inertia is a measure of how internally coherent clusters are, which reduces as the number of clusters grows. The "elbow" on the graph is often when the rate of drop changes dramatically, showing the appropriate number of clusters (Franklin, 2019). Here, the inertia reduces rapidly until roughly three clusters, at which point the rate of drop slows dramatically. This shows that the ideal number of clusters for this dataset is about three.

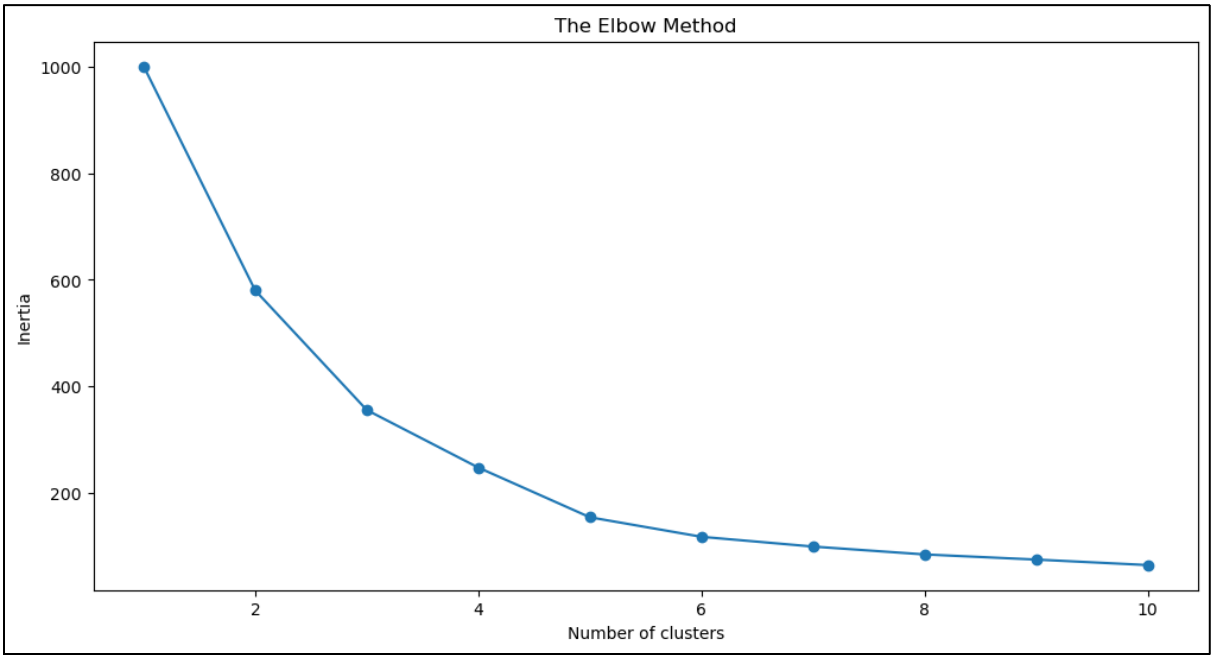

Figure: 1.15 Elbow method to define the number of clusters (Created by Author)

Clustering for detecting Market Anomalies

1. K – Means clustering

	Cluster Label	mean	median	std	count
0	0.0	10411.252166	7138.510	889.990886	31.0
1	1.0	4960.738413	5450.510	285.676949	18.0
2	2.0	7078.896231	6828.255	1115.421668	94.0

Figure: 1.16 Summary Statistics of K-Means Clustering (Created by Author)

K-Means clustering revealed three clusters on the relationship between indices and Metals. Cluster 0 has an average of **10,411**, a median of **7,138**, and a std. of **890**, indicating variable financial indicators. Cluster 1, the smallest group (mean: **4,960**, median: **5,458**, std.: **285**), indicates more steadiness amid market volatility. Finally, Cluster 2 (mean: **7,078**, median: **6,828**, standard deviation: **1,115**) is the largest and most diversified, possibly representing indicators with different market implications. This clustering indicates diverse market dynamics, with certain indicators constantly moving together while others exhibiting separate behaviour (Arvai, 2019).

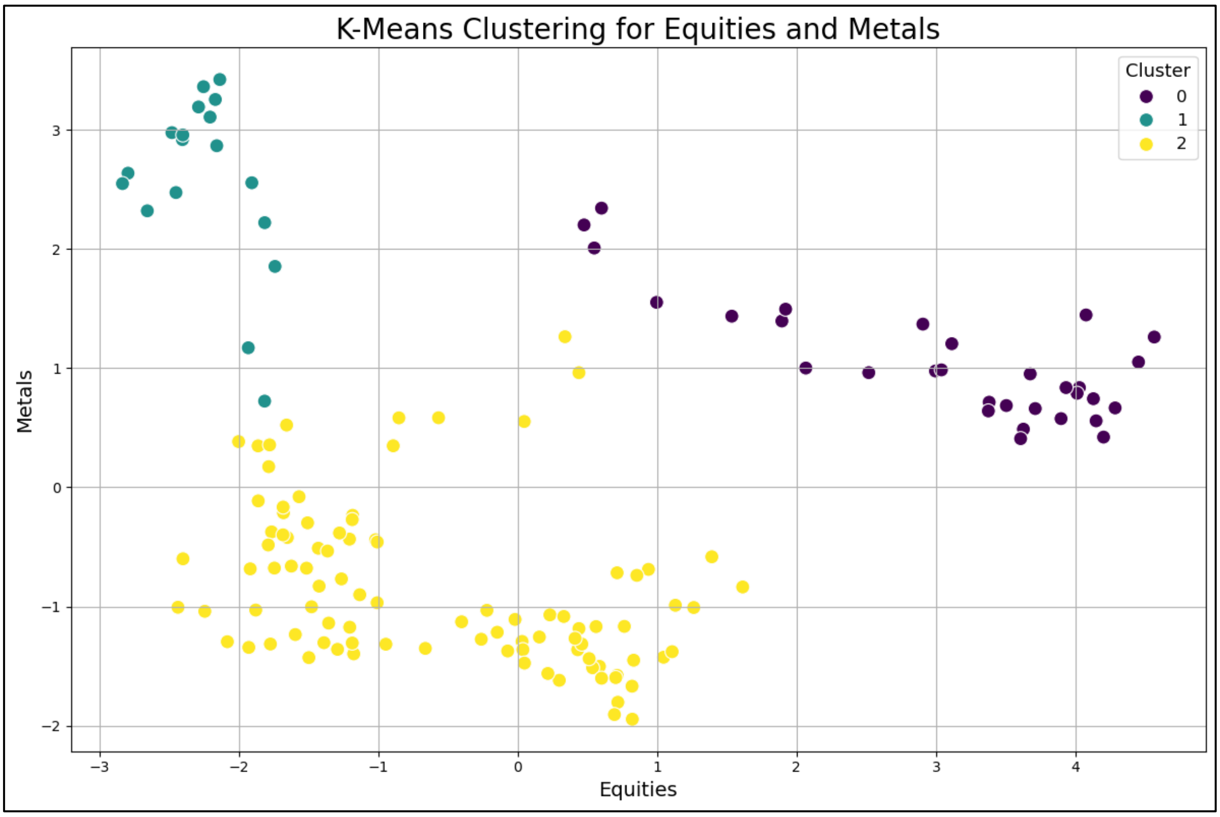

Figure: 1.17 Box plot of K-Means clusters of Equities & Metals (Created by Author)

2. Hierarchical clustering

	Cluster_Label	mean	median	std	count
0	0	20921.971250	19838.655	4711.572476	8
1	1	32992.392857	33213.760	1504.854319	14
2	2	12965.310000	12965.310	NaN	1

Figure: 1.18 Summary Statistics of Hierarchical Clustering (Created by Author)

Hierarchical clustering revealed two major clusters, emphasising the market's multilayered structure. The first cluster (mean: **20,921**, median: **19,838**, std.: **4,711**) contains indicators that respond similarly to market movements, demonstrating a shared sensitivity. The second (mean: **32,992**, median: **33,213**, std.: **1,504**) includes indicators that may be resistant to market volatility. A single data point in the third cluster may be an outlier, exhibiting distinct behaviour. The dendrogram depicts the progressive merging of indicators based on similarities, providing a more detailed understanding of market linkages and the possibility of indicators being driven by a shared set of economic causes (Sharma, 2019).

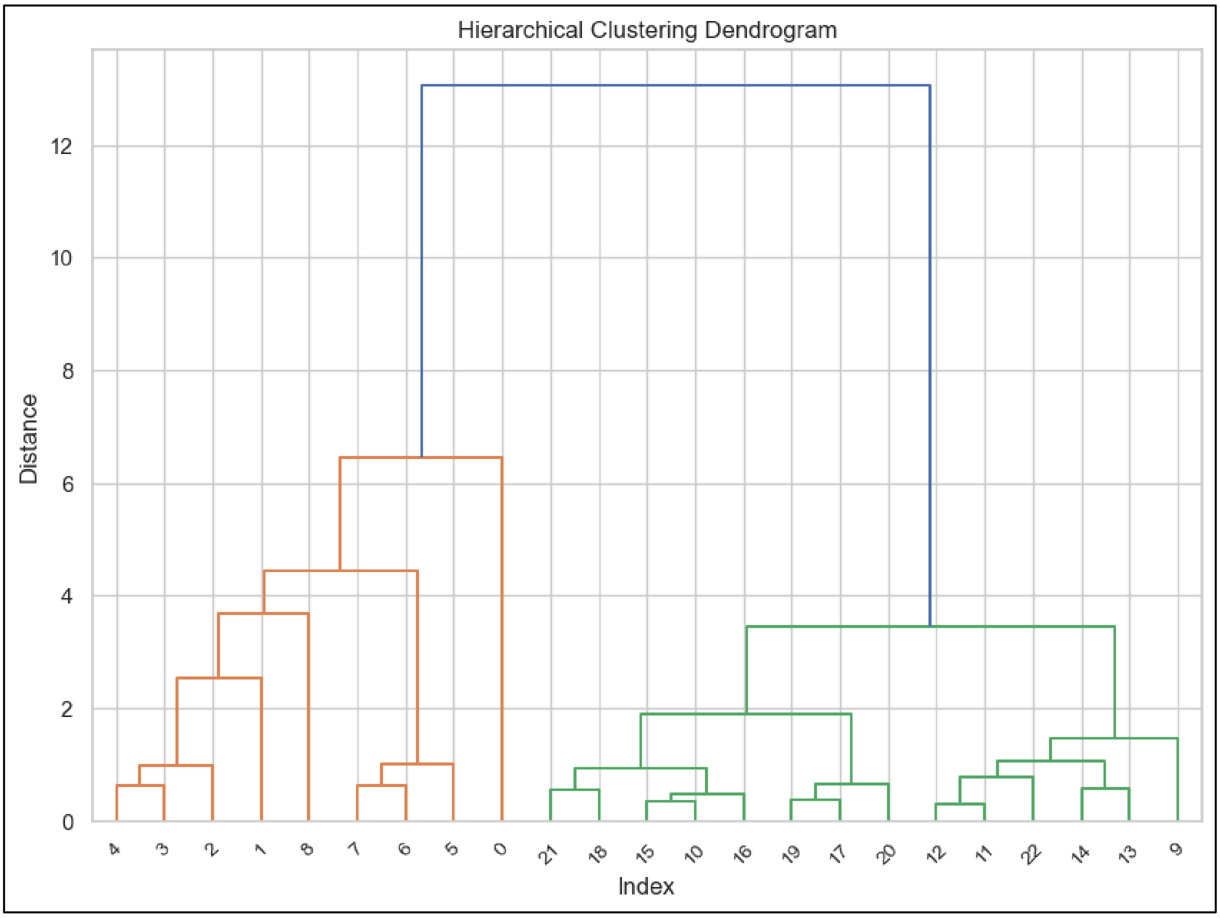

Figure: 1.19 Dendrogram of Hierarchical clusters of Equities & Metals (Created by Author)

Clustering Validity

```
Silhouette Score for K-Means: 0.5577471388725275
Silhouette Score for Hierarchical Clustering: 0.5559880853390966
```

Figure: 1.20 Silhouette score for Clustering validity (Created by Author)

The silhouette score ranges from -1 to 1, with higher values indicating better-defined clusters (Koli, 2023). The silhouette scores show that K-Means and Hierarchical clustering algorithms separated the dataset's clusters similarly, with scores of **0.5577** and **0.5559**, respectively. These scores indicate that, while clusters are not very different, and relatively well-separated, with members being more similar to one another than to those in other clusters. The small difference in scores suggests that both techniques are nearly equally successful for this dataset.

Classifications to Forecast Impacts of Inflation

1. Support Vector Machine

	precision	recall	f1-score	support
0	1.00	0.50	0.67	2
1	0.75	1.00	0.86	3
accuracy			0.80	5
macro avg	0.88	0.75	0.76	5
weighted avg	0.85	0.80	0.78	5

Accuracy: 0.8

Figure: 1.21 SVM Classification Report (Created by Author)

Support Vector Machine is a supervised machine learning algorithm that is commonly used for classification. It operates by determining the hyperplane that best separates various classes in the feature space by the greatest margin (Saini, 2021). Here, SVM has an overall accuracy of **80%**. Class 0 (low inflation) has perfect precision **(1.00)**, suggesting no false positives, but recall is just **0.50**, implying that half of the low inflation cases were missed. Class 1 (high inflation) had a precision of **0.75** and a perfect recall of **1.00**, suggesting despite some false positives, all high inflation cases were accurately identified. The F1-score, is **0.67** for class 0 and **0.86** for class 1, demonstrating superior performance in the high inflation class. The confusion matrix demonstrates that one case of low inflation was incorrectly identified as high, whereas all instances of high inflation were accurately classified.

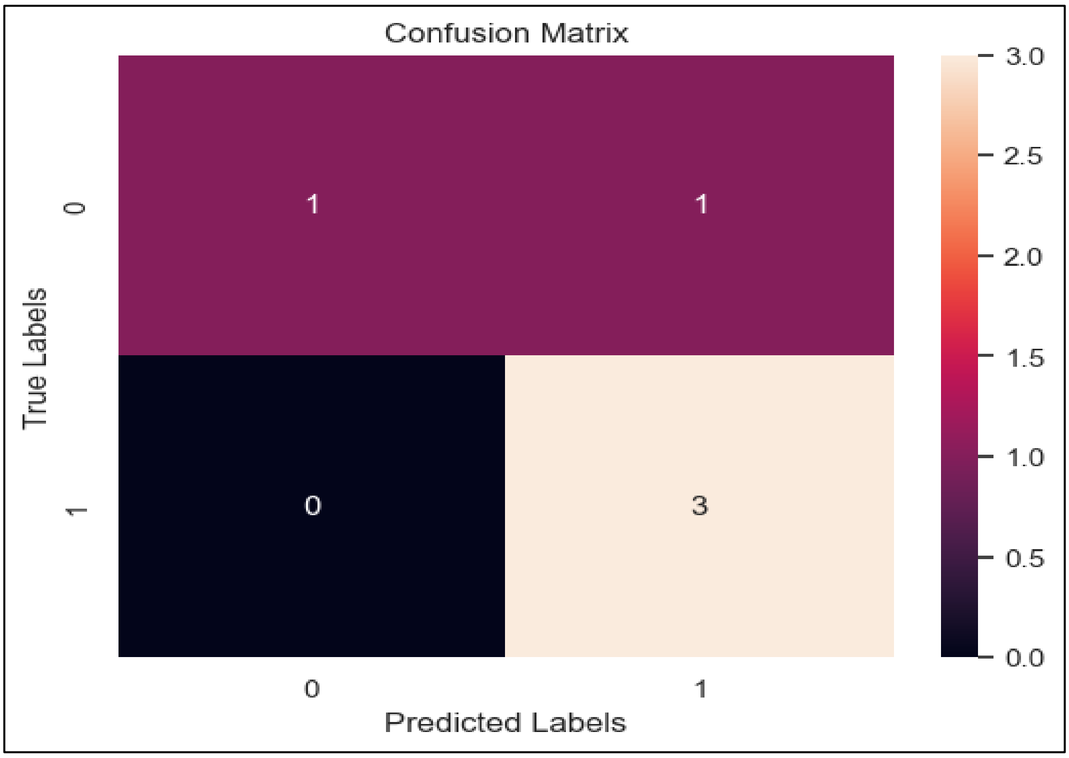

Figure: 1.22 SVM Confusion Matrix (Created by Author)

2. Naïve Bayes Classification

	precision	recall	f1-score	support
0	1.00	0.97	0.99	36
1	0.80	1.00	0.89	4
accuracy			0.97	40
macro avg	0.90	0.99	0.94	40
weighted avg	0.98	0.97	0.98	40

Accuracy: 0.97

Figure: 1.23 Naïve Bayes Classification Report (Created by Author)

The Naïve Bayes classification has shown high accuracy in separating high and low inflation groups, with an overall accuracy of **97%**. The perfect accuracy (**1.00**) in the low inflation class, indicating no false positives, and a high recall (**0.97**), suggesting that it accurately detected most of the low inflation occurrences. The precision for the high inflation class is lower (**0.80**), indicating some false positives; but, with a perfect recall (**1.00**), it accurately detected all high inflation events. The F1-score is outstanding in both classes (**0.99** for low and **0.89** for high). Naïve Bayes classification is a probabilistic machine learning technique that uses Bayes' Theorem and the assumption of feature independence to predict a data point's category. Known for its simplicity, efficiency, and efficacy (Ray, 2017).

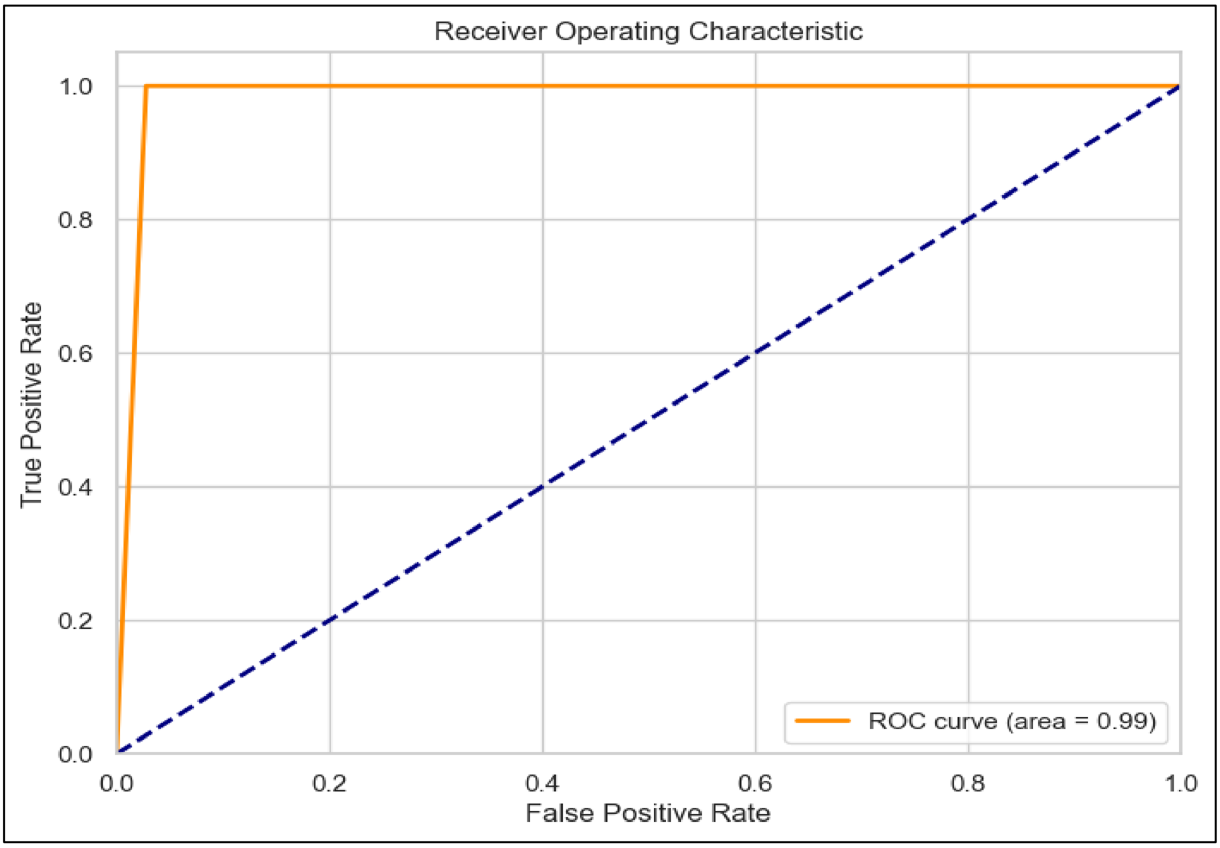

Figure: 1.24 ROC curve for Naïve Bayes Classification (Created by Author)

Classification Validation

The Naïve Bayes classifier outperforms the SVM, with an overall accuracy of **97%** vs the SVM's **80%**. Naïve Bayes achieves flawless class identification for 'High' with a recall of **1.00** and good precision for 'Low' at **1.00**. SVM has a perfect recall for 'High' (**1.00**) but a lower precision for 'Low' (**0.75**). The Naive Bayes ROC curve shows good class discrimination, with an AUC of **0.99**, demonstrating its strong predictive potential.

Statistical Modelling and Analysis

1. **Non-Linear Regression**

```
RMSE: 30.614018088540277
R-squared: 0.9860788687457553
```

Figure: 1.25 Model Evaluation for Non-Linear Regression (Created by Author)

The Root Mean Squared Error (RMSE) of **30.61** represents the residuals' standard deviation. It indicates that, on average, the model's forecasts are around **30.61** units away from real gold prices. R-squared of **0.986** indicates that the model accounts for about **98.6%** of the variation in observed gold prices, indicating a very good match.

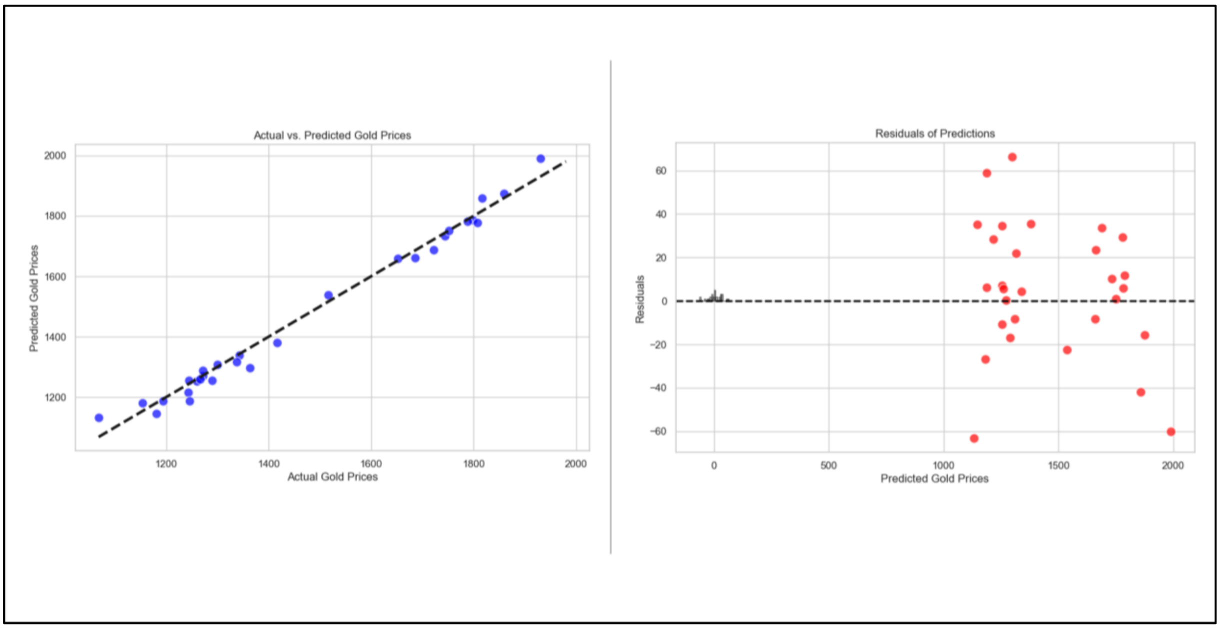

Figure: 1.26 Actual Vs. Predicted Gold Price and Residual Vs. Predicted Plot for Non-Linear Regression (Created by Author)

The actual vs. Predicted Plot shows the actual gold prices on the x-axis and the predicted gold prices on the y-axis. The closer the points are to the diagonal line, the better the model's predictions. Most points seem to cluster near the line, suggesting a good fit. Residuals vs. Predicted plots are plotted against the predicted values. The horizontal line at zero represents perfect predictions. Points are scattered above and below this line without any discernible pattern, indicating no systematic error in the model's predictions (Kenton, 2022).

Overall, the model looks to be quite effective in capturing gold price patterns and volatility. The high R-squared value indicates a model that fits the historical data well. However, the RMSE also reveals that there is an average forecast error, which, depending on the size of the gold price changes, may be considerable.

2. **Non-Linear Regression**

```
Ridge Regression RMSE: 50.52301120380605
Ridge Regression R-squared: 0.962084890635635
```

Figure: 1.27 Model Evaluation for Ridge Regression (Created by Author)

The Root Mean Squared Error (RMSE) is **50.52**, which represents the average forecast error. This figure should be interpreted in context; for high-value commodities such as gold, a lower RMSE is preferred. R-squared of **0.962** shows that the model explains **96.2%** of the variation in real gold prices, which is fairly high and indicates a good model fit.

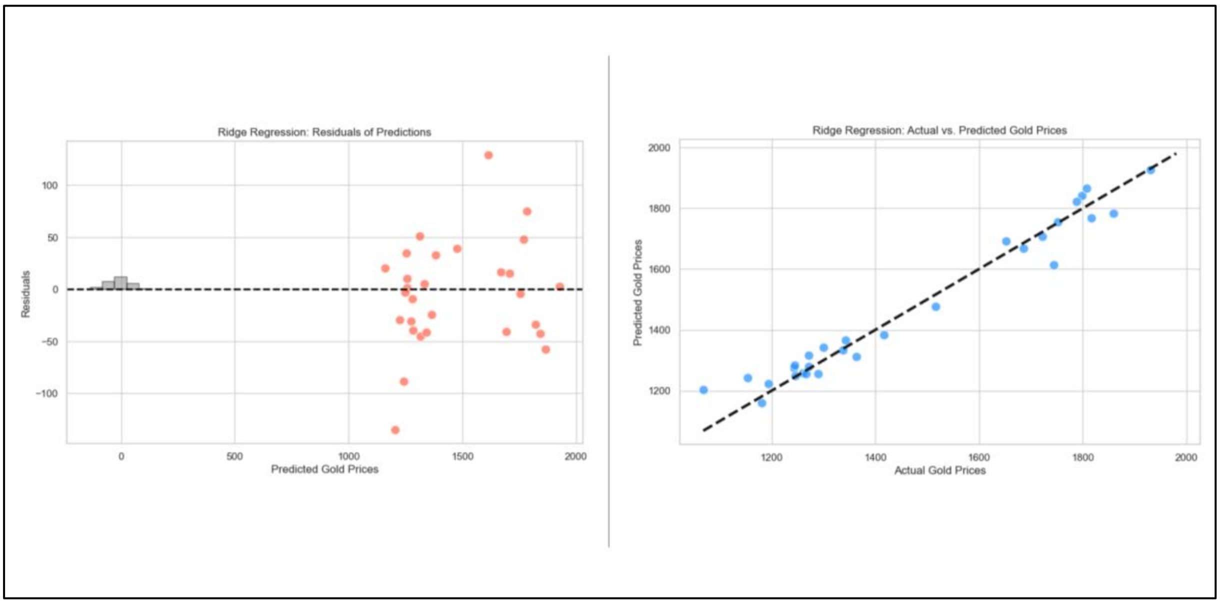

Figure: 1.28 Actual Vs. Predicted Gold Price and Residual Vs. Predicted Plot for Ridge Regression (Created by Author)

Actual vs. projected Plot shows a strong linear association between actual and projected gold prices, as indicated by the proximity of points to the dashed diagonal line (ideal prediction line). The residuals are evenly distributed around the horizontal line at zero, with no discernible patterns, which is good. However, the distribution of residuals indicates significant variation in the model's prediction accuracy, particularly at higher values. Overall, the model performs well, although the RMSE suggests potential for improvement. Ridge regression uses a regularisation term to penalise big coefficients and prevent overfitting. This strategy is very beneficial when dealing with multicollinearity or when a model has more predictor variables than data. (Jain, 2016)

Predictive Model Validation

Both the Nonlinear and Ridge regression models show great predictive power for gold prices, with R-squared values of **0.9867** and **0.9621**, suggesting that they explain a significant percentage of the variance in real prices. The nonlinear model has a lower RMSE of **30.61** than the ridge regression, which has an RMSE of **50.52**, implying that it predicts more accurately on average.

Conclusion

Recap

This study investigated advanced clustering algorithms for identifying market anomalies and developing adaptive classification models to assess the impact of inflation on global indices and precious metals. The primary objectives were to implement and critically evaluate clustering approaches like K-means and hierarchical clustering for anomaly detection in equity and commodity markets. Additionally, a classification framework using Support Vector Machines, and Naïve Bayes Classification was developed to assess the impact of inflation. The study also attempted to use predictive modelling tools to optimise portfolio diversification strategies by analysing previous performance data for gold, silver, and important financial indices.

Discussion and Implications

The findings of advanced clustering algorithms revealed a significant number of anomalies in the equities and commodity markets. The application of K-means and hierarchical clustering provides detailed insights into market movements, revealing the nuanced links between various financial instruments. The classification models developed to measure the impact of inflation used Support Vector Machines and Naïve Bayes, providing an in-depth understanding of how different asset classes react to rising inflation. These models were successful in projecting market changes, showing the value of multiple asset classes as inflation hedges.

The study additionally emphasised the need for optimised portfolio diversification. The analysis of historical performance data using Non-linear and Ridge Regression models demonstrated that gold, which is commonly regarded as a safe-haven asset, might play an important role in hedging strategies against market volatility. This understanding is especially relevant in today's economic context, where inflation and market uncertainty are serious concerns for investors.

Limitations

1. **Dependency on Historical data**

One of the primary limitations of the study is dependence on historical data, which may not completely represent future market behaviours or unexpected economic events. Furthermore, the model's fundamental assumptions, such as market efficiency and the absence of irrational investment behaviour, may restrict the scope of the results.

2. **Data Restrictions**

Data collection limited the scope of the research as the data on the inflation rate is available monthly. As the base currency for all the specified variables was different, the conversion into uniform currency may miss minute details must be converted. The daily data may be inconsistent due to different operating patterns and timings of the global indices. Also, while converting the data to a monthly and yearly basis some of the market movement might have been overlooked.

3. Proficiency

The aim was to build a highly responsive predictive model to explore hedging opportunities and forecast the impact of inflation on all the selected variables, but due to constraints of resources and time, only the forecasting for gold could be done. The model is built for study purposes, and it does have some limitations in terms of the ability to handle and process large data, which can hamper the accuracy of the outcome.

4. Economic uncertainties

As Equity and commodities markets are influenced by various economic factors, and the selected variables belonged to different geographies, capturing all the economic trends is impossible. Only large-scale trends or pandemics like Covid-19 could be considered as other regional trends are difficult to record and may not affect all instruments equally.

Recommendations

1. Expand the scope with Real-time Data

To reduce dependency on historical data, future research should include real-time data analysis to better capture market dynamics. Furthermore, expanding the scope of the research to incorporate behavioural finance theories might lead to a more complete knowledge of market behaviour. Exploring a broad range of financial products and worldwide marketplaces may improve understanding of global financial trends.

2. Improving accuracy with deep learning models

Incorporating machine learning techniques such as deep learning algorithms may improve the accuracy of the models. These strategies might be particularly beneficial for detecting nuanced trends in large datasets that standard models might overlook.

References

Arvai, K. (2019). *K-Means Clustering in Python: A Practical Guide – Real Python*. [online] realpython.com. [Accessed 13 Jan. 2024]. Available at: https://realpython.com/k-means-clustering-python/#:~:text=The%20k%2Dmeans%20clustering%20method

BEATTIE, A. (2022). *How to Calculate Return on Investment (ROI)*. [online] Investopedia. [Accessed 12 Jan. 2024]. Available at: https://www.investopedia.com/articles/basics/10/guide-to-calculating-roi.asp#:~:text=Key%20Takeaways

Chang, C.-L. and Cai, Q. (2023). Stock return anomalies identification during the Covid-19 with the application of a grouped multiple comparison procedure. *Economic Analysis and Policy*, [online] 79, pp.168–183. [Accessed 9 Jan. 2024]. Available at: https://www.sciencedirect.com/science/article/pii/S0313592623001297#sec4

Chiang, T.C. (2022). Real stock market returns and inflation: Evidence from uncertainty hypotheses. *Finance Research Letters*, [online] 53(103606), p.103606. [Accessed 6 Jan. 2024]. Available at: https://www.sciencedirect.com/science/article/pii/S1544612322007826

Franklin, S.J. (2019). *Elbow method of K-means clustering algorithm*. [online] Analytics Vidhya. [Accessed 11 Jan. 2024]. Available at: https://medium.com/analytics-vidhya/elbow-method-of-k-means-clustering-algorithm-a0c916adc540

Gorthy, S. (2021). *Euclidean Distance*. [online] MLearning.ai. [Accessed 12 Jan. 2024]. Available at: https://medium.com/mlearning-ai/euclidean-distance-8fae145ef5f3

Hayes, A. (2023). *Descriptive Statistics*. [online] Investopedia. [Accessed 13 Jan. 2024]. Available at: https://www.investopedia.com/terms/d/descriptive_statistics.asp#:~:text=The%20three%20main%20types%20of

Hossainds, B. (2022). *Global Inflation Dataset - (1970~2022)*. [online] www.kaggle.com. [Accessed 24 Oct. 2023]. Available at: https://www.kaggle.com/datasets/belayethossainds/global-inflation-dataset-212-country-19702022

Investing.com India. (2023). *Live Index Chart*. [online] [Accessed 31 Oct. 2023]. Available at: https://in.investing.com/charts/indices-charts

Jain, A. (2016). *A Complete Tutorial on Ridge and Lasso Regression in Python*. [online] Analytics Vidhya. [Accessed 11 Jan. 2024]. Available at: https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial/

Kenton, W. (2022). *What Is Nonlinear Regression?* [online] Investopedia. [Accessed 11 Jan. 2024]. Available at: https://www.investopedia.com/terms/n/nonlinear-regression.asp

Koli, S. (2023). *How to Evaluate the Performance of Clustering Algorithms Using Silhouette Coefficient*. [online] Medium. [Accessed 11 Jan. 2024]. Available at: https://medium.com/@MrBam44/how-to-evaluate-the-performance-of-clustering-algorithms-3ba29cad8c03

LIOUDIS, N. (2022). *The Importance of Diversification*. [online] Investopedia. [Accessed 11 Jan. 2024]. Available at: https://www.investopedia.com/investing/importance-diversification/#:~:text=Diversification%20is%20a%20common%20investing

Macrotrends (2023). *10 Year Treasury Rate - 54 Year Historical Chart*. [online] Macrotrends.net. [Accessed 28 Oct. 2023]. Available at: https://www.macrotrends.net/2016/10-year-treasury-bond-rate-yield-chart

Narne, P. (2023). *Global Stock Market (2008-2023)*. [online] www.kaggle.com. [Accessed 25 Oct. 2023]. Available at: https://www.kaggle.com/datasets/pavankrishnanarne/global-stock-market-2008-present?select=2023_Global_Markets_Data.csv

Niu, Z., Wang, C. and Zhang, H. (2023). Forecasting stock market volatility with various geopolitical risks categories: New evidence from machine learning models. *International Review of Financial Analysis*, [online] 89(89), p.102738. [Accessed 7 Jan. 2024]. Available at: https://www.sciencedirect.com/science/article/abs/pii/S1057521923002545

Parte, K. (2020). *Dimensionality Reduction: Principal Component Analysis*. [online] Analytics Vidhya. [Accessed 13 Jan. 2024]. Available at: https://medium.com/analytics-vidhya/dimensionality-reduction-principal-component-analysis-d1402b58feb1

Plastun, A., Sibande, X., Gupta, R. and Wohar, M.E. (2020). Price gap anomaly in the US stock market: The whole story. *The North American Journal of Economics and Finance*, [online] 52, p.101177. [Accessed 3 Jan. 2024]. Available at: https://www.sciencedirect.com/science/article/pii/S1062940820300747

Ray, S. (2017). *Naive Bayes Classifier Explained: Applications and Practice Problems of Naive Bayes Classifier*. [online] Analytics Vidhya. [Accessed 12 Jan. 2024]. Available at: https://www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/#:~:text=The%20Na%C3%AFve%20Bayes%20classifier%20is

Saini, A. (2021). *Support Vector Machine(SVM): A Complete guide for beginners*. [online] Analytics Vidhya. [Accessed 12 Jan. 2024]. Available at: https://www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/

Sharma, P. (2019). *Hierarchical Clustering | Hierarchical Clustering Python*. [online] Analytics Vidhya. [Accessed 11 Jan. 2024]. Available at: https://www.analyticsvidhya.com/blog/2019/05/beginners-guide-hierarchical-clustering/

Soni, H. (2021). *A Guide To Data Standardization & Transformation*. [online] Medium.com. [Accessed 13 Jan. 2024]. Available at: https://soniheet6498.medium.com/a-guide-to-data-standardization-transformation-1bf98d9b13b4

Statista. (2023). *Domestic equity market capitalization globally 2023*. [online] [Accessed 27 Oct. 2023]. Available at: https://www.statista.com/statistics/274490/global-value-of-share-holdings-since-2000/#:~:text=The%20value%20of%20global%20domestic

Strimpel, J. (2023). *Machine learning technique for financial data analysis*. [online] PyQuantnews.com. [Accessed 12 Jan. 2024]. Available at: https://pyquantnews.com/machine-learning-technique-financial-data-analysis/#:~:text=KMeans%20clustering%20can%20be%20used

Szabo, B. (2020). *How to Create a Seaborn Correlation Heatmap in Python?* [online] Medium. [Accessed 11 Jan. 2024]. Available at: https://medium.com/@szabo.bibor/how-to-create-a-seaborn-correlation-heatmap-in-python-834c0686b88e

Tang, Y. and Chou, J. (2015). The performance of PSO-SVM in inflation forecasting. [online] 2015 12th International Conference on Service Systems and Service Management (ICSSSM). doi: https://doi.org/10.1109/ICSSSM.2015.7170251

World Bank. (2023). *A Global Database of Inflation*. [online] [Accessed 28 Oct. 2023]. Available at: https://www.worldbank.org/en/research/brief/inflation-database

www.imf.org. (2023). *https://www.imf.org/external/datamapper/PCPIEPCH@WEO/WEOWORLD*. [online] [Accessed 28 Oct. 2023]. Available at: https://www.imf.org/external/datamapper/PCPIEPCH@WEO/WEOWORLD

www.kaggle.com. (2022a). *Daily Gold Price Historical Data*. [online] [Accessed 26 Oct. 2023]. Available at: https://www.kaggle.com/datasets/psycon/daily-gold-price-historical-data

www.kaggle.com. (2022b). *Daily Silver Price Historical Data*. [online] [Accessed 26 Oct. 2023]. Available at: https://www.kaggle.com/datasets/psycon/daily-silver-price-historical-data/