2024

Internship Project

MBA BA
II Year I Trimester

IMPACT OF GOLD PRICE ON GDP

ABSTRACT

This project explores the relationship between gold price fluctuations and India's GDP growth from 1999 to 2024. It analyzes historical data to understand how changes in gold prices impact economic activities and overall economic performance. The study aims to identify trends, correlations, and key factors linking gold prices with GDP. By using statistical methods, the project provides insights into the economic significance of gold in India's economy. The findings could help policymakers and investors make informed decisions.

Contents

Project Cover Sheet

Project Title	Impact of Gold price on GDP	**Course**	MBA BA II Year – I Trimester
Start Date	28-10-2024	**SME/Supervisor**	Prof P.S.S.Murthy/ Dr.Seema
End Date		**Team Lead**	G.Anvitha

Team Members

SNO	Name
1.	V. Sreeya Reddy
2.	G. Anvitha

Signature:

SME: Supervisor:

Date:

EXECUTIVE SUMMARY

This study investigates the relationship between gold price fluctuations and India's GDP, focusing on the period from 1999 to 2024. Leveraging a rich dataset, the research examines key economic indicators, including gold prices, imports and exports of gold, GDP growth rates, inflation, and currency exchange rates. Gold's unique position as a cultural and financial asset in India underscores its critical influence on household savings, investment portfolios, and trade balance. By analysing over two decades of historical data, the project identifies patterns and correlations to explore whether fluctuations in gold prices have an associative or causative impact on GDP.

The analysis reveals how shifts in gold prices influence consumer behaviour, gold trade dynamics, and broader macroeconomic indicators, such as inflation and currency exchange rates. Insights gained from the research aim to assist policymakers, economists, and financial analysts in understanding the macroeconomic implications of commodity price volatility. These findings contribute to informed fiscal and monetary policies, investment strategies, and trade decisions, equipping stakeholders with tools to navigate economic complexities effectively. This study provides a nuanced perspective on the interplay between commodity markets and national development, enhancing strategic responses to India's evolving economic landscape.

OBJECTIVE

The objective of this project is to analyze the relationship between gold price fluctuations and changes in India's GDP from 1999 to 2024. By examining historical data, it aims to identify correlations between gold price trends and economic growth. This analysis seeks to understand how changes in gold prices may impact or reflect shifts in India's economic performance, offering insights into any causative or associative effects of gold price movements on GDP. The findings could reveal valuable connections between commodity price trends and national economic health.

PROBLEM STATEMENT

This project examines whether fluctuations in gold prices have a measurable impact on India's GDP, focusing on the period from 1999 to 2024. With gold as a significant asset in India, its price changes may influence economic factors like consumer spending, investment, and trade balance. The study aims to analyze historical data to determine if variations in gold prices correlate with GDP shifts, assessing both potential causative and associative effects. Understanding this relationship could provide insights into the broader economic impact of commodity price fluctuations on national growth.

PROJECT MILESTONES

MILESTONE	KEY ACTIVITIES	TARGET DATE
1) Domain Study	• Develop Knowledge about the domain • Basic Terminology used in this domain	1 Week
2) Data Collection	• Understanding Client Requirements • Identify Relevant Data Sources • Import the file	3 Weeks
3) Data Preprocessing	• Data Cleaning • Data Transforming • Data Integration • EDA • Trend Analysis	1Week
4) Data Modelling	• Model Building • Statistical Models	1 Week
5) Data Visualization	• Graphs • Dashboards	1 Week
6) Presentation	• Report Submission • Project Presentation	1 Week

REQUIREMENT SPECIFICATIONS:

1. User Requirement
2. Functional Specifications
3. System Specifications

➢ User Requirements:

1. 10 Years GDP data
2. Imports of Gold
3. How Imports of Gold affect GDP
4. Percentage of Imports of Gold
5. Trends in GDP for last 10 years
6. Sector wise Contribution on GDP
7. How is Gold Valued
8. Relationship between Gold and GDP
9. Gold Exports have major impact on GDP
10. Gold Value for 10 years
11. Impurities in Gold
12. GDP Definition
13. GDP Growth Rate
14. Indian production of Gold 10 years
15. Indian consumption of Gold
    I. Production
    II. Imports
16. Impact of Gold price on GDP
17. Correlation
    i. Gold price vs GDP
    ii. Gold price vs Imports
    iii. Gold price vs Exports
18. Visualization of Correlation Analysis
19. Interpretation of Correlation Analysis
20. Gold Imports value as Percentage on GDP

➢ Functional Specifications:

- **Data Collection and Import**
  Input Data:
  Gold Prices: The system will import historical gold price data (e.g., Annual prices) from reliable sources

GDP Data: The system will import quarterly or yearly GDP data for India from government sources

- **Data Preprocessing**

  Cleaning Data: The system will handle missing values by filling them with the median or mean of the respective columns.
  - The system will remove duplicate entries and handle any inconsistencies in the dataset.

  Feature Engineering:
  - **Gold Price Percentage Change**: Calculate the percentage change in gold prices over time.
  - **GDP Growth Rate**: Compute the GDP growth rate over time.
  - **Outlier Detection and Removal:**
    The system will use Z-scores to identify and remove outliers (e.g., values with Z-scores greater than 3).
- **Statistical Analysis**

  Correlation Analysis:
  - The system will calculate the Pearson correlation between gold price fluctuations and GDP.
  - The system will identify potential causal relationships using statistical methods
- **Data Normalization**

  Min-Max Scaling:
  The system will apply Min-Max scaling to standardize numerical columns (e.g., gold prices, GDP) to a range of [0,1].
- **Visualization and Reporting**

  Dashboard:
  - Trends visualizations for gold prices and GDP growth.
  - Scatter plots to visualize correlations.
  - A correlation matrix displaying relationships between multiple economic indicators.

    Reports:
  - The strength of correlation between gold prices and GDP.
  - The impact of gold price volatility on GDP growth.
  - Potential periods of significant influence.
- **User Interface (UI)**
  - Upload and manage their datasets.
  - View and interact with visualizations.
  - Generate and export reports.
- **Data Export**

  Users will be able to export the analysis results and reports in CSV, Excel, and PDF formats.

➢ System Specifications

1. Hardware Requirements:

- ✓ **Processor:** Intel Core i5 or higher / AMD equivalent
- ✓ **RAM:** Minimum 8 GB (16 GB recommended for large datasets)
- ✓ **Storage:**
  - 100 GB free space for dataset storage and backups
  - SSD recommended for faster data processing
- ✓ **Graphics Card:** Not mandatory, but a basic GPU (e.g., NVIDIA GTX 1650 or equivalent) can enhance visualizations in some tools.
- ✓ **Display:** Monitor with a resolution of 1920x1080 or higher

2. Software Requirements:

- ✓ **Operating System:**
  - Windows 10/11 (64-bit)
  - macOS 11+
  - Linux (Ubuntu 20.04 or newer)
- ✓ **Programming Tools:**
  - Python (3.8 or later) with libraries such as pandas, NumPy, matplotlib, seaborn, and stats models for statistical analysis.
  - R (optional, for advanced statistical modelling).
- ✓ **Data Analysis Tools:**
  - Jupyter Notebook for interactive coding and analysis.
  - Microsoft Excel or Google Sheets for preliminary data review.
- ✓ **Database Tools (if needed):**
  - SQLite or MySQL for managing large datasets.
- ✓ **Visualization Tools:**
  - Tableau or Power BI for creating dashboards.
  - Python visualization libraries (e.g., matplotlib, seaborn, plotly).

3. Data Requirements:

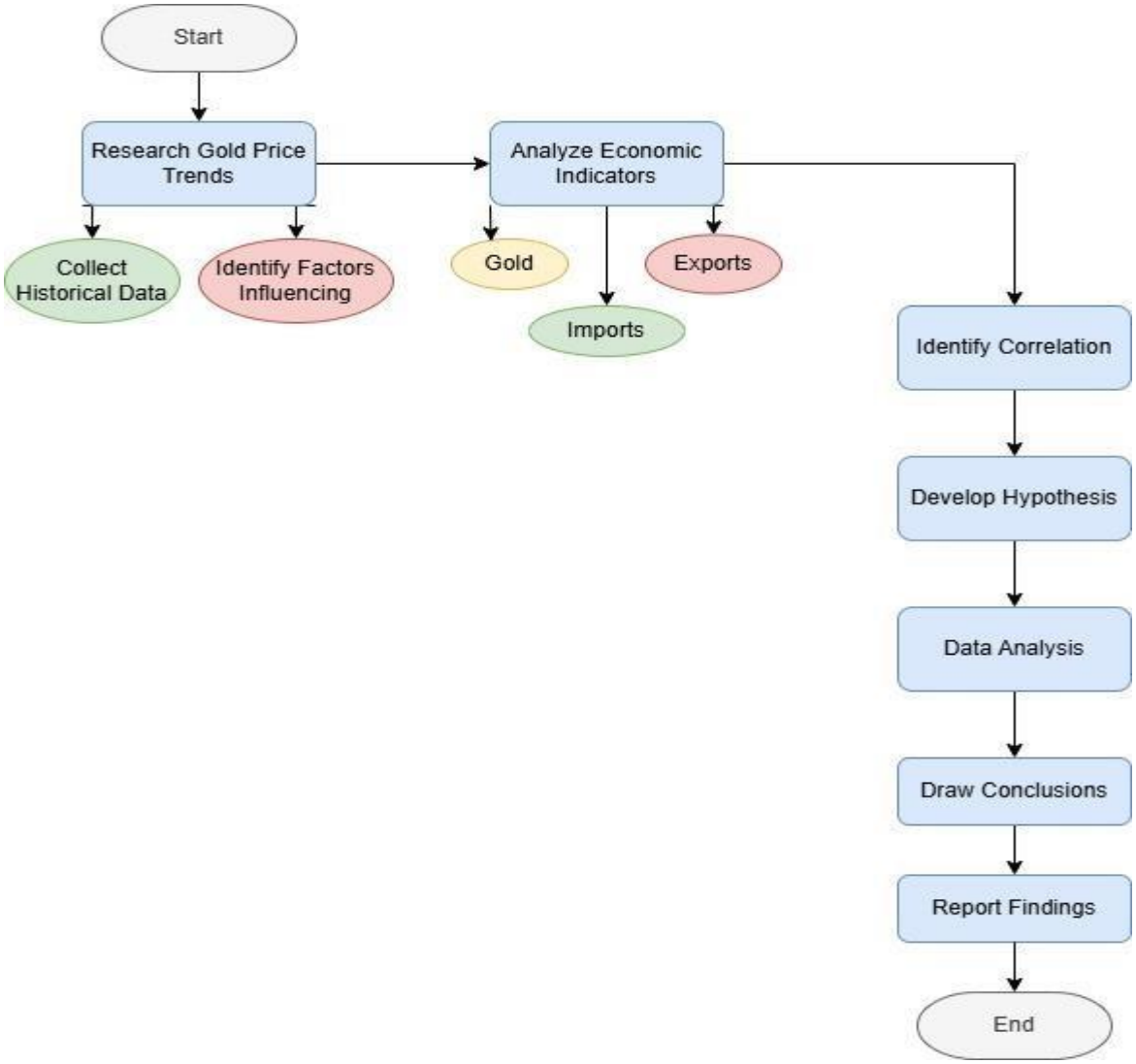
- ✓ **Sources:**
  - Historical gold price data from 1999 to 2024.
  - GDP and macroeconomic indicators (e.g., trade balance, inflation rate) from trusted sources like World Bank, IMF, or Reserve Bank of India.
- ✓ **Format:** CSV, Excel, or SQL database.

4. Network Requirements:

- ✓ **Internet Connection:** Stable internet for accessing real-time data, research papers, and cloud-based tools like Google Collab or AWS.
- ✓ **Cloud Storage:** For collaboration and secure data backup (e.g., Google Drive, OneDrive).

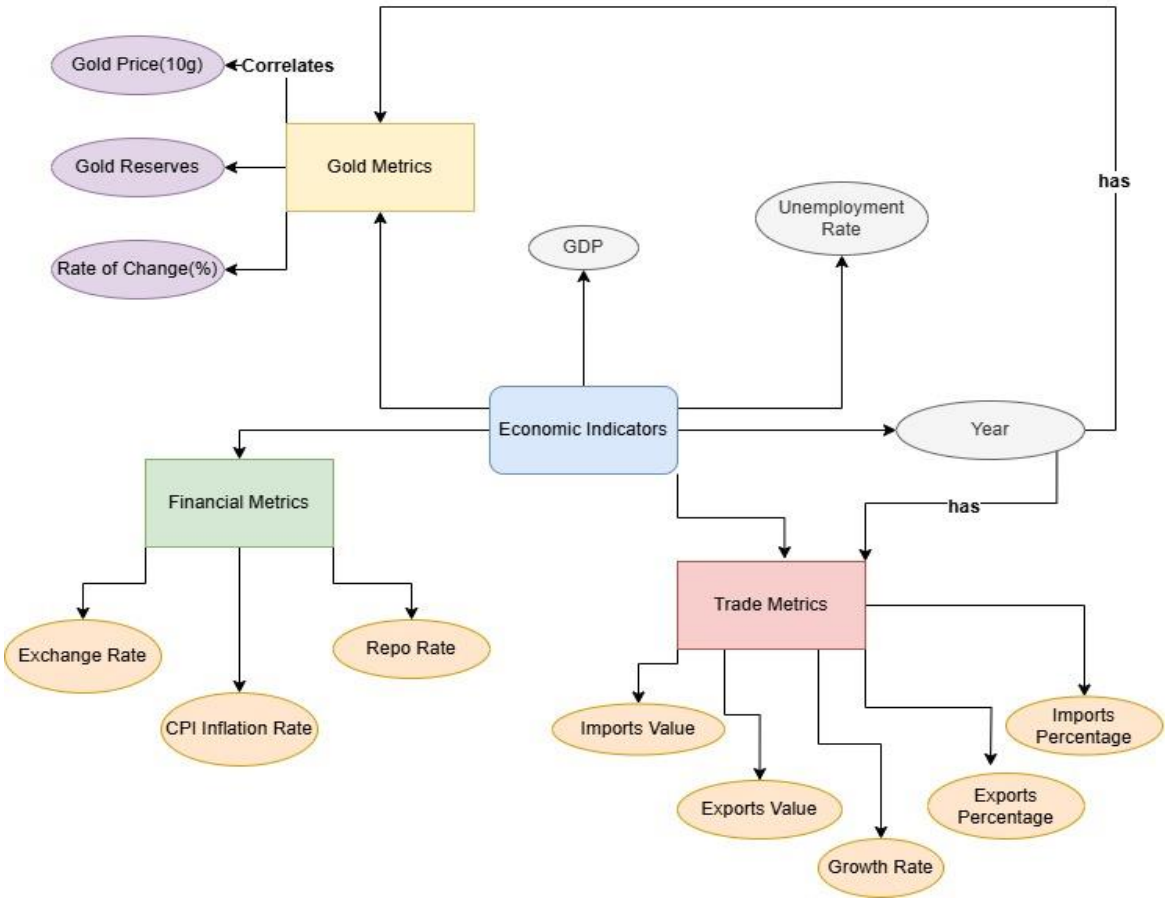
BUSINESS PROCESS CHARTS

1. Activity Chart



This activity diagram depicts the process of determining the influence of gold prices on GDP. It starts with investigating gold price patterns, which includes gathering historical data and determining impacting elements. Economic data, such as gold imports and exports, are then examined. Following that, potential correlations are detected, resulting in hypothesis building. Data analysis is carried out to draw conclusions, which are then documented in the results report to complete the process.

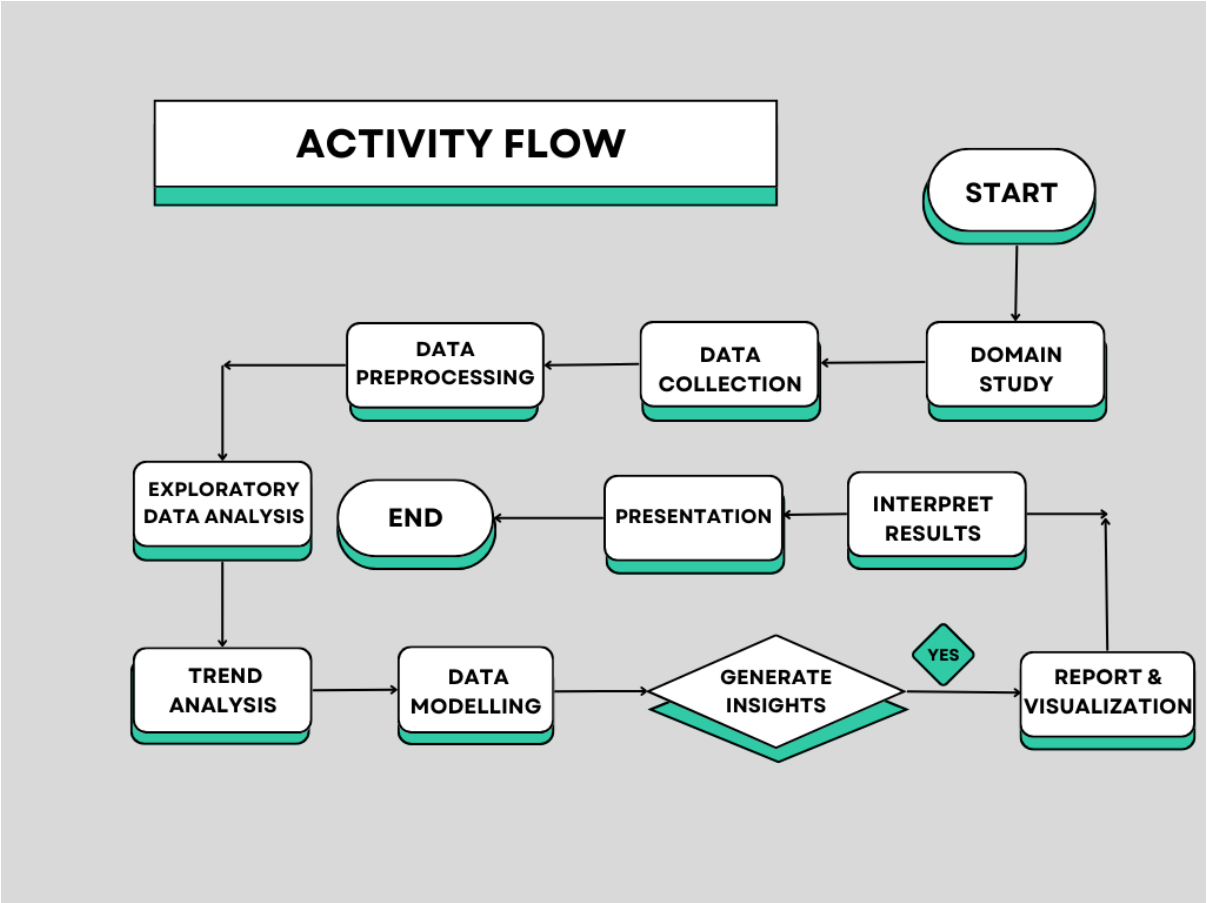
2.ER Diagram



The ER diagram illustrates the relationships among various economic indicators, categorized into Gold Metrics, Financial Metrics, and Trade Metrics, all connected to GDP and the Unemployment Rate over different years. Gold Metrics include attributes such as Gold Price (10g), Gold Reserves, and Rate of Change (%), which correlate with broader economic trends. Financial Metrics encompass indicators like Exchange Rate, Repo Rate, and CPI Inflation Rate, reflecting monetary and inflationary dynamics. Trade Metrics cover details of Imports and Exports, including their values, percentages, and overall Growth Rate, emphasizing trade's role in the economy.

GDP and Unemployment Rate serve as overarching indicators that influence and are influenced by these metrics. The "Year" entity links these variables over time, enabling temporal analysis. The diagram highlights dependencies and correlations among key factors driving the economy. It offers a structured framework to study economic interactions comprehensively.

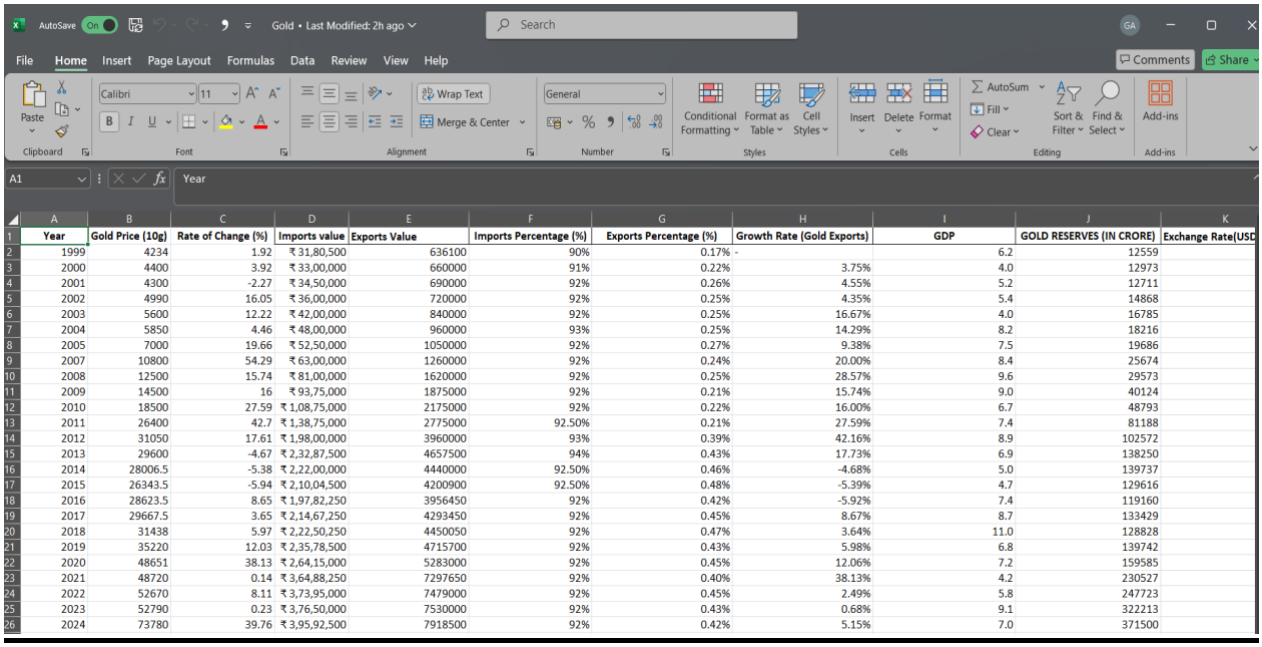
3.State Chart



The State Chart outlines a workflow for data-driven projects, starting with the Start phase. The first step, Domain Study, involves defining objectives and understanding the problem scope. Next, the Data Collection phase gathers relevant data, followed by Data Preprocessing, where data is cleaned and organized for analysis.

Exploratory Data Analysis (EDA) uses statistical methods and visualizations to identify patterns and insights, which guide Trend Analysis to study data trends. In Data Modelling, mathematical or machine learning models are applied to derive predictions or test hypotheses. Results lead to actionable insights in the Generate Insights phase, which are then communicated through the Report & Visualization phase.

DATA SET



Year	Gold Price (10g)	Rate of Change (%)	Imports value	Exports Value	Imports Percentage (%)	Exports Percentage (%)	Growth Rate (Gold Exports)	GDP	GOLD RESERVES (IN CRORE)	Exchange Rate(USD
1999	4234	1.92	₹31,80,500	636100	90%	0.17%	-	6.2	12559	
2000	4400	3.92	₹33,00,000	660000	91%	0.22%	3.75%	4.0	12973	
2001	4300	-2.27	₹34,50,000	690000	92%	0.26%	4.55%	5.2	12711	
2002	4990	16.05	₹36,00,000	720000	92%	0.25%	4.35%	5.4	14868	
2003	5600	12.22	₹42,00,000	840000	92%	0.25%	16.67%	4.0	16785	
2004	5850	4.46	₹48,00,000	960000	93%	0.25%	14.29%	8.2	18216	
2005	7000	19.66	₹52,50,000	1050000	92%	0.27%	9.38%	7.5	19686	
2007	10800	54.29	₹63,00,000	1260000	92%	0.24%	20.00%	8.4	25674	
2008	12500	15.74	₹81,00,000	1620000	92%	0.25%	28.57%	9.6	29573	
2009	14500	16	₹93,75,000	1875000	92%	0.21%	15.74%	9.0	40124	
2010	18500	27.59	₹1,08,75,000	2175000	92%	0.22%	16.00%	6.7	48793	
2011	26400	42.7	₹1,38,75,000	2775000	92.50%	0.21%	27.59%	7.4	81188	
2012	31050	17.61	₹1,98,00,000	3960000	93%	0.39%	42.16%	8.9	102572	
2013	29600	-4.67	₹2,32,87,500	4657500	94%	0.43%	17.73%	6.9	138250	
2014	28006.5	-5.38	₹2,22,00,000	4440000	92.50%	0.46%	-4.68%	5.0	139737	
2015	26343.5	-5.94	₹2,10,04,500	4200900	92.50%	0.48%	-5.39%	4.7	129616	
2016	28623.5	8.65	₹1,97,82,250	3956450	92%	0.42%	-5.92%	7.4	119160	
2017	29667.5	3.65	₹2,14,67,250	4293450	92%	0.45%	8.67%	8.7	133429	
2018	31438	5.97	₹2,22,50,250	4450050	92%	0.47%	3.64%	11.0	128828	
2019	35220	12.03	₹2,35,78,500	4715700	92%	0.43%	5.98%	6.8	139742	
2020	48651	38.13	₹2,64,15,000	5283000	92%	0.45%	12.06%	7.2	159585	
2021	48720	0.14	₹3,64,88,250	7297650	92%	0.40%	38.13%	4.2	230527	
2022	52670	8.11	₹3,73,95,000	7479000	92%	0.45%	2.49%	5.8	247723	
2023	52790	0.23	₹3,76,50,000	7530000	92%	0.43%	0.68%	9.1	322213	
2024	73780	39.76	₹3,95,92,500	7918500	92%	0.42%	5.15%	7.0	371500	

The dataset spanning the years 1999 to 2024 regarding the Project of Impact of Gold price on GDP

Economic Indicators:

- **Year:** The year for which the data is recorded.
- **GDP:** India's Gross Domestic Product (GDP) in billions of rupees.
- **Exchange Rate (USD):** The exchange rate of the Indian Rupee against the US Dollar.

Gold Market Indicators:

- **Gold Price (10g):** The average price of 10 grams of gold in Indian Rupees.
- **Rate of Change (%):** The percentage change in the gold price compared to the previous year.
- **Imports Value:** The total value of gold imports into India in rupees.
- **Exports Value:** The total value of gold exports from India in rupees.
- **Imports Percentage (%):** The percentage of gold imports in relation to total gold imports and exports.
- **Exports Percentage (%):** The percentage of gold exports in relation to total gold imports and exports.
- **Growth Rate (Gold Exports):** The annual growth rate of gold exports.

Other Indicators:

- **Gold Reserves (in Crore):** The amount of gold reserves held by India in crores of rupees.

Data Preprocessing and Analysis Description

1. **Data Loading and Initial Inspection:**
   The dataset was loaded, and key columns, including Gold Price (10g) and GDP, were selected for analysis. Initial rows were reviewed to understand the structure of the data.
2. **Handling Missing Values:**
   - Missing values were addressed using forward-fill (fill), which carries forward the last known value.
   - Any residual missing values were dropped to ensure the dataset's completeness.
3. **Removing Duplicates:**
   Duplicate rows in the Gold Price (10g) and GDP columns were identified and removed to avoid redundancy.
4. **Data Transformation:**
   - Percentage changes for Gold Price (10g) and GDP were calculated to analyse trends and growth rates effectively.
   - These percentage changes were added as new features (gold_price_pct_change and GDP_growth) for further analysis.
5. **Outlier Detection and Removal:**
   - Z-scores were calculated for numerical columns to detect outliers.
   - Data points with Z-scores greater than 3 (indicating extreme values) were removed. This ensured that outliers did not distort analysis results.
6. **Scaling and Normalization:**
   - Numerical columns like Gold Price (10g) and GDP were scaled using StandardScaler for consistency.
   - Scaling ensured that features with varying units and ranges were brought to a comparable scale.
7. **Final Cleaning and Export:**
   - After cleaning and processing, the dataset was saved to a CSV file named cleaned_gold_gdp_data.csv for further analysis.
   - The cleaned dataset was inspected to verify the shape and quality.

Descriptive Analysis

1. Data Loading and Overview:

- The dataset was loaded using pandas and converted into a DataFrame for analysis.
- The head() method displayed the first few rows to inspect the structure and content of the data.
- The info() function provided an overview of data types, non-null counts, and memory usage.

2. Missing Value Analysis:

- Missing values in each column were identified using isnull().sum(). This ensured that any gaps in the dataset were accounted for during preprocessing.

3. Statistical Summary:

- Descriptive statistics for numerical columns, including mean, median, standard deviation, and percentiles, were computed using describe().
- For categorical columns, the describe() method was applied to review the count, unique values, and most frequent entries.

4. Categorical Data Analysis:

- Unique values in each categorical column were displayed to understand the diversity of entries and identify any anomalies.

5. Duplicate Analysis:

- The dataset was checked for duplicate rows, and the count was displayed to assess data quality.

6. Visualization of Trends:

- **Gold Price Trend:**
  A line chart showed the trend of gold prices (Gold Price (10g)) over the years, highlighting changes visually with a purple line.
- **GDP Trend:**
  A similar line chart depicted the GDP trend over the years with exact GDP values annotated on the graph for clarity.
- **Import Percentage on GDP:**
  The percentage of imports relative to GDP was calculated and plotted over the years, providing insights into how imports contributed to economic performance.
- **Exports Trend:**
  The export revenue trend was visualized using a line chart, offering a perspective on India's export growth.
- **Inflation Rate Trend:**
  The inflation rate (CPI Inflation Rate (%)) was plotted, with values annotated for better interpretation of the annual rate of change.

7. Additional Insights:

- The calculated Import Percentage on GDP highlighted the proportion of imports relative to economic output, providing key insights into trade balance dynamics.
- Visualizations effectively demonstrated the relationships and fluctuations across variables such as gold price, GDP, exports, imports, and inflation over the years.

```
Summary Statistics (Categorical Columns):
       Growth Rate (Gold Exports)
count                          25
unique                         25
top                             -
freq                            1
```

Exploratory Data Analysis

1. Dataset Overview:

- **Loading Data:**
  The dataset was imported using `pandas` and converted to a DataFrame. The `head()` function displayed the first few rows for a quick overview, and `info()` provided information about column data types, non-null counts, and memory usage.
- **Missing Values:**
  Missing values were checked using `isnull().sum()` to identify gaps in the dataset. This step is essential for handling incomplete data during preprocessing.

2. Descriptive Statistics:

- The `describe()` function computed summary statistics for both numerical and categorical columns, such as count, mean, median, and unique values, providing insights into the dataset's distribution and variability.

3. Visualization of Categorical Data:

- **Count Plots:**
  For categorical columns, count plots were created using `seaborn`, highlighting the frequency distribution of each category. These visualizations help detect imbalances or anomalies in the categorical data.

4. Distribution Analysis:

- **Histograms with KDE:**
  For each numerical column, histograms overlaid with Kernel Density Estimation (KDE) curves were plotted. These visualizations helped examine the data's distribution (e.g., normality, skewness) and detect unusual patterns.

5. Outlier Detection:

- **Box Plots:**
  Boxplots for each numerical column were created to identify outliers. These plots visually depict the data spread, interquartile range (IQR), and extreme values.

6. Correlation Analysis:

- **Correlation Matrix:**
  A correlation matrix was computed to quantify relationships between numerical variables.
- **Heatmap Visualization:**
  The correlation matrix was visualized as a heatmap with annotations. This helped identify strong positive or negative correlations (e.g., values close to +1 or -1), indicating potentially significant relationships between variables.

CLASSIFCATION

1. Data Collection

- **Description**: Gather the data that is relevant to the problem. This data should include both input features (predictors) and labels (the classes you are trying to predict). For instance, in a classification task to predict whether a person will buy a product or not, the features might include age, income, etc., and the label might be "Yes" or "No."

2. Data Preprocessing

- **Description**: Clean the data to ensure it is in a format suitable for the model. Common preprocessing tasks include:
  - **Handling missing values**: Impute missing values or remove rows with missing data.
  - **Data normalization/standardization**: Scale numeric data to a standard range (e.g., between 0 and 1, or with a mean of 0 and standard deviation of 1).
  - **Encoding categorical variables**: Convert categorical data into a numeric format, such as using one-hot encoding or label encoding.
  - **Feature engineering**: Create new features based on the existing data that might be useful for the model.

3. Data Splitting

- **Description**: Split the dataset into two (or more) parts: a training set and a testing set. Typically, 70-80% of the data is used for training, and the remaining 20-30% is used for testing. This helps in evaluating the model's performance on unseen data.

4. Model Selection

- **Description**: Choose a suitable classification algorithm based on the nature of the problem and data. Some commonly used classification algorithms are:
  - **Logistic Regression**: Good for binary classification problems.
  - **XG-Boost**: create a strong predictive model, minimizing errors through iterative improvements.
  - **Random Forests**: An ensemble method that combines multiple decision trees.

5. Model Training

- **Description**: Use the training data to train the selected model. During this step, the algorithm learns to map the input features to the correct output labels. The learning process involves optimizing the model’s parameters (e.g., weights in a neuralnetwork) to minimize errors.

6. Model Evaluation

- **Description**: Evaluate the model's performance using the test set. Common evaluation metrics for classification tasks include:
  - **Accuracy**: The proportion of correct predictions.
  - **Precision**: The proportion of true positives out of all predicted positives.
  - **Recall**: The proportion of true positives out of all actual positives.
  - **F1-Score**: The harmonic mean of precision and recall.

7. Hyperparameter Tuning

- **Description**: Tune the model's hyperparameters (e.g., the number of trees in a random forest, the learning rate in a neural network) to improve its performance. This is typically done using methods like grid search or random search, combined with cross-validation to ensure the model generalizes well.

8. Model Validation

- **Description**: Validate the model's performance using cross-validation, where the training set is divided into multiple folds and the model is trained and evaluated multiple times. This helps assess the model's stability and robustness.

9. Model Deployment

- **Description**: Once the model is trained and validated, it can be deployed to make predictions on new, unseen data. This can involve setting up the model in a production environment, where it can receive input data and output predictions in real time or batch mode.

10. Model Monitoring and Maintenance

- **Description**: After deployment, it is important to continuously monitor the model's performance. Over time, the model might degrade in accuracy due to changes in the data distribution (concept drift). In such cases, the model may need to be retrained or updated.

```
Model: XGBoost
Accuracy: 0.90
Classification Report:
              precision    recall  f1-score   support

           0       0.75      0.75      0.75         4
           1       0.94      0.94      0.94        16

    accuracy                           0.90        20
   macro avg       0.84      0.84      0.84        20
weighted avg       0.90      0.90      0.90        20

----------------------------------------
Model: Logistic Regression
Accuracy: 0.95
Classification Report:
              precision    recall  f1-score   support

           0       0.80      1.00      0.89         4
           1       1.00      0.94      0.97        16

    accuracy                           0.95        20
   macro avg       0.90      0.97      0.93        20
weighted avg       0.96      0.95      0.95        20

----------------------------------------
Model: Random Forest
Accuracy: 0.95
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.75      0.86         4
           1       0.94      1.00      0.97        16

    accuracy                           0.95        20
   macro avg       0.97      0.88      0.91        20
weighted avg       0.95      0.95      0.95        20

----------------------------------------
```

REGRESSION

1. Define the Research Problem and Hypothesis

- **Goal**: Clearly define the objective of your regression analysis, such as predicting a value or understanding relationships between variables.
- **Hypothesis**: Formulate a hypothesis regarding the relationship between variables. For example, "The price of gold affects GDP."

2. Collect and Prepare Data

- **Data Collection**: Gather the necessary data for both the dependent and independent variables.
- **Data Cleaning**: Handle missing values, duplicates, and outliers.
- **Data Transformation**: If necessary, transform variables (e.g., logarithmic transformations for skewed data).

3. Visualize the Data

- **Exploratory Data Analysis (EDA)**: Before fitting a regression model, visualize the data using scatter plots or correlation matrices to check the potential relationship between variables.
- **Check for Multicollinearity**: Ensure that the independent variables are not highly correlated with each other.

4. Split the Data (Training and Testing)

- **Training Set**: Use a subset of the data to train the regression model.
- **Testing Set**: Use a separate subset to evaluate the model's performance.

5. Select the Regression Model

- **Linear Regression**: For modelling a linear relationship.
- **Multiple Regression**: If there are multiple predictors.
- **Logistic Regression**: For binary outcomes (classification problems).
- **Polynomial Regression**: For nonlinear relationships.
- **Regularized Regression**: Use techniques like Lasso or Ridge for handling overfitting.

6. Fit the Model

- Use statistical software (e.g., Python's `sklearn`, R, SPSS) to train the model on the training data. This involves calculating the coefficients (weights) of the regression equation.

7. Evaluate the Model

- **Goodness of Fit**: Check how well the model fits the data. Common metrics include:
  - **R-squared**: Proportion of variance explained by the model.
  - **Adjusted R-squared**: Takes into account the number of predictors.
  - **RMSE (Root Mean Squared Error)**: Measures the average prediction error.
  - **p-values**: Test the significance of each predictor.
- **Residual Analysis**: Examine the residuals (the differences between actual and predicted values) for patterns. Residuals should be randomly distributed.

8. Model Interpretation

- **Interpret Coefficients**: Understand the impact of each independent variable on the dependent variable. For example, in linear regression, a coefficient represents how much the dependent variable changes for each unit change in the predictor.
- **Check for Assumptions**: Verify the assumptions of regression, such as linearity, normality, homoscedasticity (constant variance of errors), and independence of errors.

9. Model Improvement (Optional)

- **Variable Selection**: If the model is overfitting or underfitting, consider removing insignificant variables or adding relevant ones.
- **Regularization**: Apply Lasso or Ridge regression if you face multicollinearity or overfitting.

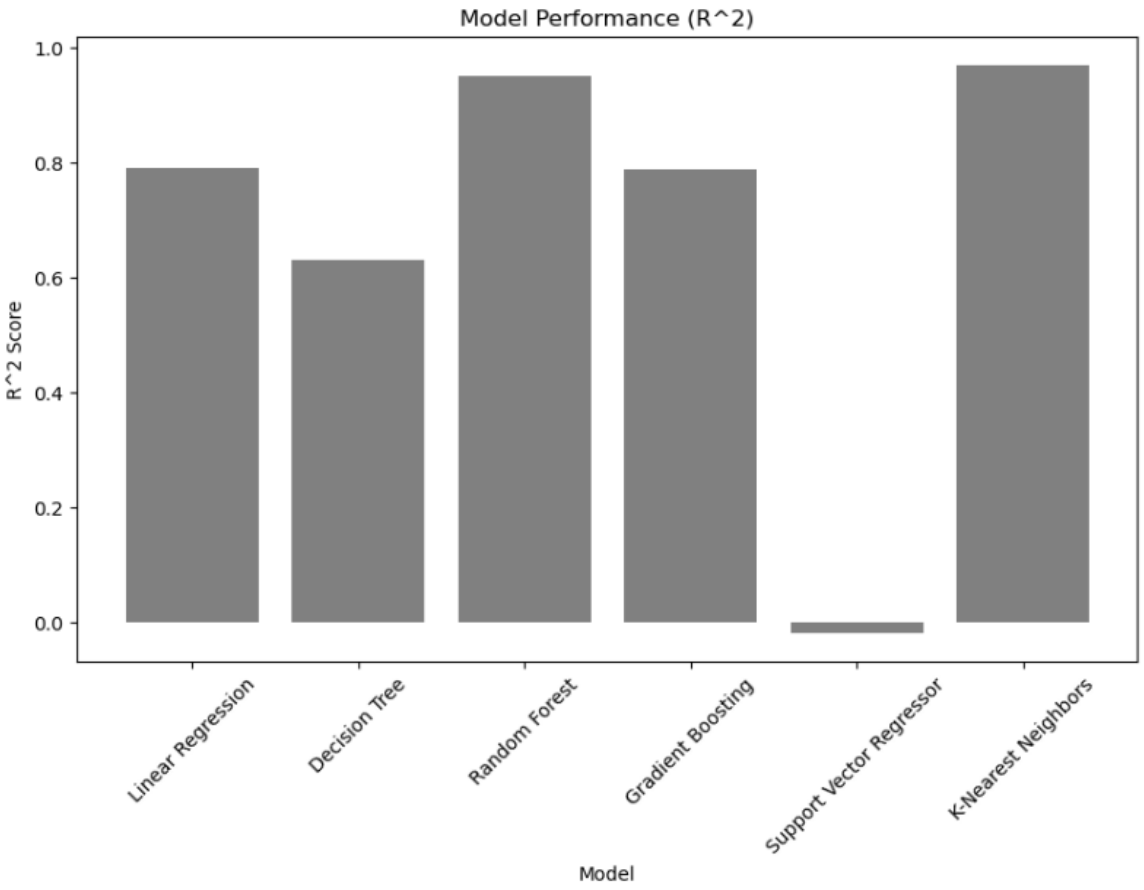
10. Make Predictions

- Once the model is finalized, use it to make predictions on new or unseen data.

11. Communicate Results

- **Model Summary**: Present the regression equation, coefficients, p-values, R-squared, and other relevant metrics.
- **Visualization**: Use visual aids like regression plots, scatter plots, and residual plots to communicate findings clearly.
- **Implications**: Discuss the practical implications of your results based on your hypothesis.

	Model	MSE	RMSE	R^2
0	Linear Regression	5.775852e+07	7599.902788	0.789940
1	Decision Tree	1.015083e+08	10075.133468	0.630828
2	Random Forest	1.345701e+07	3668.379604	0.951059
3	Gradient Boosting	5.805586e+07	7619.439934	0.788859
4	Support Vector Regressor	2.799339e+08	16731.224509	-0.018082
5	K-Nearest Neighbors	8.276805e+06	2876.943750	0.969898



DECISION TREE

1. Collect and Prepare Data

- Gather the data you want to analyse.
- Clean and preprocess the data by handling missing values, normalizing or encoding categorical data, and ensuring the data is in the right format.

2. Choose the Target Variable (Output)

- Identify the variable you want to predict (dependent variable). For example, in a marketing campaign, it could be whether a customer will buy a product or not.

3. Select Features (Inputs)

- Choose the features (independent variables) that you believe influence the target variable. These are the attributes or characteristics you'll use to make predictions.

4. Split the Data

- Divide the data into training and test sets. The training set will be used to build the tree, while the test set will evaluate how well the tree predicts outcomes on new data.

5. Choose a Splitting Criterion (Metric)

- The decision tree algorithm splits the data at each step based on certain criteria. Common criteria include:
    - **Gini Index**: Measures the impurity of a dataset. Lower values mean the data is more pure.
    - **Entropy**: Measures the uncertainty in the dataset. Lower entropy indicates better splits.
    - **Information Gain**: Based on how much information is gained by splitting the data at each step.

6. Build the Tree (Recursive Process)

- The algorithm recursively splits the data at each node based on the feature that provides the best split according to the chosen criterion (like Gini or entropy).
- This process continues until:
    - All data points in a node belong to the same class (pure node).
    - There are no more features to split on.
    - A pre-defined tree depth or minimum sample size is reached.

7. Pruning the Tree (Optional)

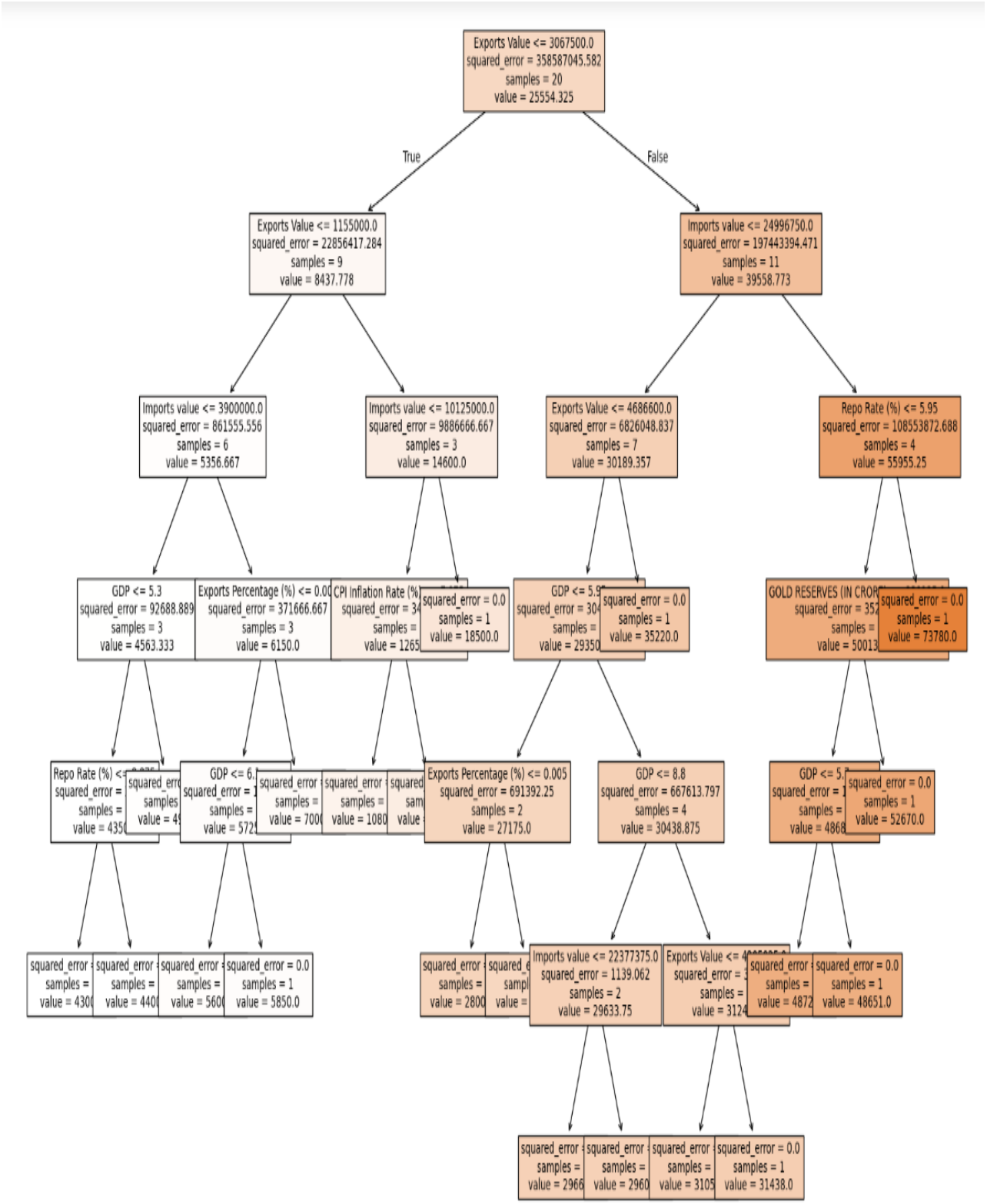
- After the tree is built, it may be too complex, leading to overfitting. Pruning involves cutting back branches that provide little predictive power to improve generalization on new data.

8. Evaluate the Model

- Use the test set to check how well the decision tree performs. Evaluate accuracy, precision, recall, or other metrics based on the nature of the problem.

9. Deploy and Use the Model

- Once the tree performs well, you can use it to make predictions on new data.


Exports Value <= 3067500.0
squared_error = 358587045.582
samples = 20
value = 25554.325
True
False
Exports Value <= 1155000.0
squared_error = 22856417.284
samples = 9
value = 8437.778
Imports value <= 24996750.0
squared_error = 197443394.471
samples = 11
value = 39558.773
Imports value <= 3900000.0
squared_error = 861555.556
samples = 6
value = 5356.667
Imports value <= 10125000.0
squared_error = 9886666.667
samples = 3
value = 14600.0
Exports Value <= 4686600.0
squared_error = 6826048.837
samples = 7
value = 30189.357
Repo Rate (%) <= 5.95
squared_error = 108553872.688
samples = 4
value = 55955.25
GDP <= 5.3
squared_error = 92688.889
samples = 3
value = 4563.333
Exports Percentage (%) <= 0.0
squared_error = 371666.667
samples = 3
value = 6150.0
CPI Inflation Rate (%)
squared_error = 0.0
samples = 1
value = 18500.0
GDP <= 5.9
squared_error = 0.0
samples = 1
value = 35220.0
GOLD RESERVES (IN CRORE
squared_error = 0.0
samples = 1
value = 73780.0
Repo Rate (%) <=
GDP <= 6.
Exports Percentage (%) <= 0.005
squared_error = 691392.25
samples = 2
value = 27175.0
GDP <= 8.8
squared_error = 667613.797
samples = 4
value = 30438.875
GDP <= 5.
squared_error = 0.0
samples = 1
value = 52670.0
squared_error = 0.0
samples = 1
value = 5850.0
Imports value <= 22377375.0
squared_error = 1139.062
samples = 2
value = 29633.75
squared_error = 0.0
samples = 1
value = 48651.0
squared_error = 0.0
samples = 1
value = 31438.0

CORRELATION

1. Understand the Data

- Ensure that the two columns you're comparing are numerical (e.g., integers or floats). Correlation is typically calculated between continuous variables.
- Example: You might want to find the correlation between the "Age" and "Income" columns in a dataset.

2. Prepare the Data

- Clean the data by checking for any missing values, outliers, or incorrect entries in the columns you're comparing.
- If there are missing values, decide how to handle them (e.g., by removing rows with missing values or filling them with the mean/median).

3. Select the Two Columns

- Identify the two columns you want to calculate the correlation between. These columns should be numerical and related in some way (e.g., "Height" vs. "Weight").

4. Choose a Correlation Method

- The most common method to measure correlation is **Pearson Correlation**, which measures the linear relationship between two variables.
- If the relationship is not linear, you might consider using **Spearman's Rank Correlation** (for ordinal or non-linear data) or **Kendall's Tau** (for smaller datasets with ranked variables).

5. Calculate the Correlation

- **Pearson Correlation Formula**: r=∑(xi−x¯)(yi−y¯)∑(xi−x¯)2∑(yi−y¯)2r = \frac{\sum{(x_i - \bar{x})(y_i - \bar{y})}}{\sqrt{\sum{(x_i - \bar{x})^2}\sum{(y_i - \bar{y})^2}}}r=∑(xi−x¯)2∑(yi−y¯)2∑(xi−x¯)(yi−y¯) Where:
  - rrr is the correlation coefficient (between -1 and 1).
  - xi,yix_i, y_ixi,yi are individual data points.
  - x¯,y¯\bar{x}, \bar{y}x¯,y¯ are the means of the x and y variables.
- **Interpretation**:
  - r=1r = 1r=1: Perfect positive correlation.
  - r=−1r = -1r=−1: Perfect negative correlation.
  - r=0r = 0r=0: No correlation.

6. Use Statistical or Data Analysis Tools

- **Using Python**: You can use `pandas` or `numpy` to calculate correlation in a dataset.

7. Interpret the Result

- A positive value indicates a positive relationship (as one variable increases, the other increases).
- A negative value indicates a negative relationship (as one variable increases, the other decreases).
- A value near 0 means there is little to no linear relationship between the variables.

8. Visualize the Correlation (Optional)

- You can plot a scatter plot to visually inspect the relationship between the two columns.

Correlation matrix:

- **Gold Price vs GDP**

```
Correlation between Gold Price and GDP:
                  Gold Price (10g)       GDP
Gold Price (10g)          1.000000  0.119595
GDP                       0.119595  1.000000
```

- **Gold Price vs Imports**

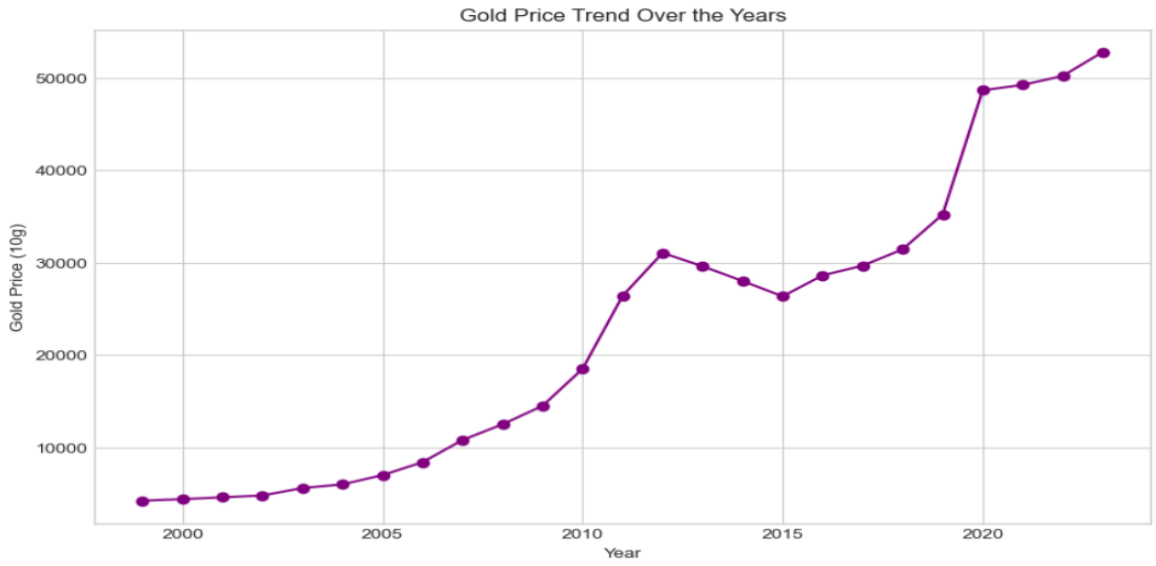
```
Correlation between Gold Price and Imports:
                  Gold Price (10g)  Imports value
Gold Price (10g)          1.000000       0.971102
Imports value             0.971102       1.000000
```

- **Gold Price vs Exports**

```
Correlation between Gold Price and Exports Value:
                     Rate of Change (%)  Exports Percentage (%)
Rate of Change (%)             1.000000              -0.294239
Exports Percentage (%)        -0.294239               1.000000
```

Visualizations

Graph 1:



Description:

The graph is a line graph with data points marked, showing the trend of gold prices over time. The graph shows a long-term upward trend in gold prices from 1999 to 2023, with periods of fluctuation. The gold price starts low in 1999, rises steadily until around 2012, dips slightly between 2013 and 2015, and then rises sharply again after 2018.

Dimensions:

- X-axis: The years ranging from 1999 to 2023
- Y-axis: The gold price (per 10g)

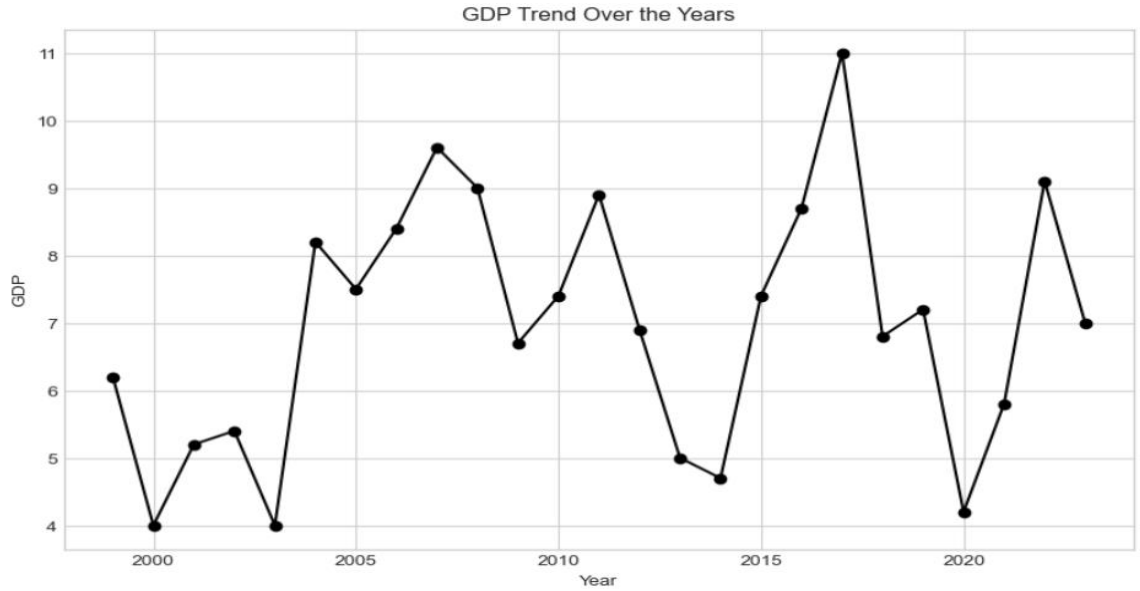
Measures:

- x-axis: Measures years.
- y-axis: Measures gold price per 10g

Graph type:

The above graph is a line graph showing the trend of gold prices over time.

Graph 2:



Description:

The graph is titled "GDP Trend Over the Years." It illustrates the trend of GDP from the year 1999 to 2023. The GDP values fluctuate significantly over the years, showing periods of both growth and decline. Notable peaks are observed around the years 2005, 2010, and 2015, while significant drops are seen around 2002, 2009, and 2017. The highest GDP value is slightly above 11, occurring around 2015, and the lowest is around 4, occurring around 2002.

Dimensions:

- X-Axis: The years ranging from 1999 to 2023.
- Y-Axis: The GDP values ranging from 4 to 11.

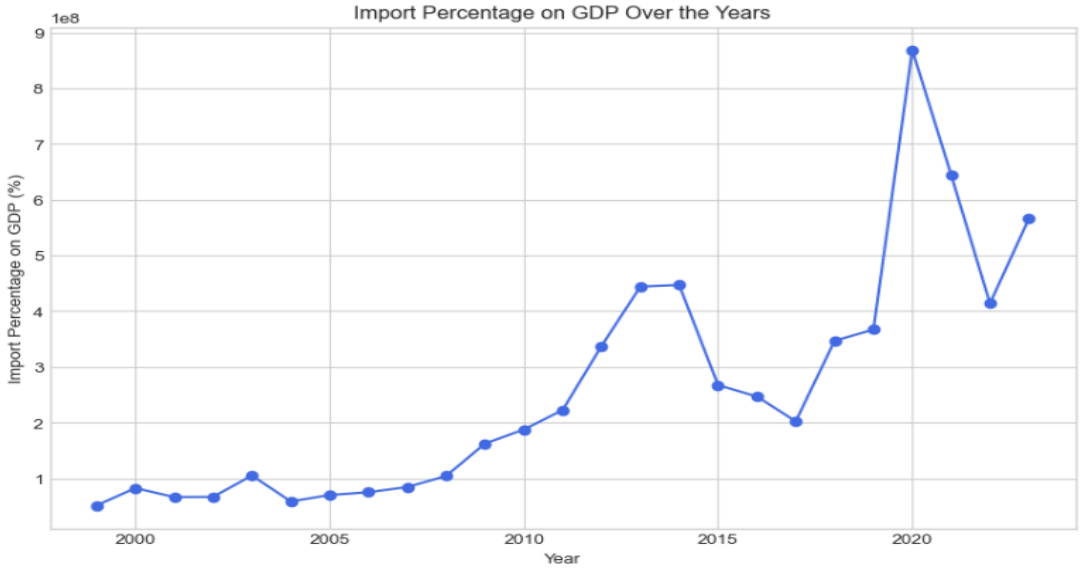
Measures:

- X-Axis: Measures time in years.
- Y-Axis: Measures GDP values.

Graph type:

The above graph is a Line graph showing GDP trend over years

Graph 3:



Description:

The graph shows the percentage of imports as a share of GDP over the years from 2000 to approximately 2023. Initially, the import percentage remained relatively stable until 2005, after which it started increasing steadily, peaking around 2015. A sharp decline followed, but there was a significant surge around 2020, reaching the highest point in the timeline. After 2020, the import percentage dropped and then showed signs of recovery.

Dimensions:

- X-axis: The years ranging from 1999 to 2023.
- Y-axis: The imports percentage on GDP for each year.

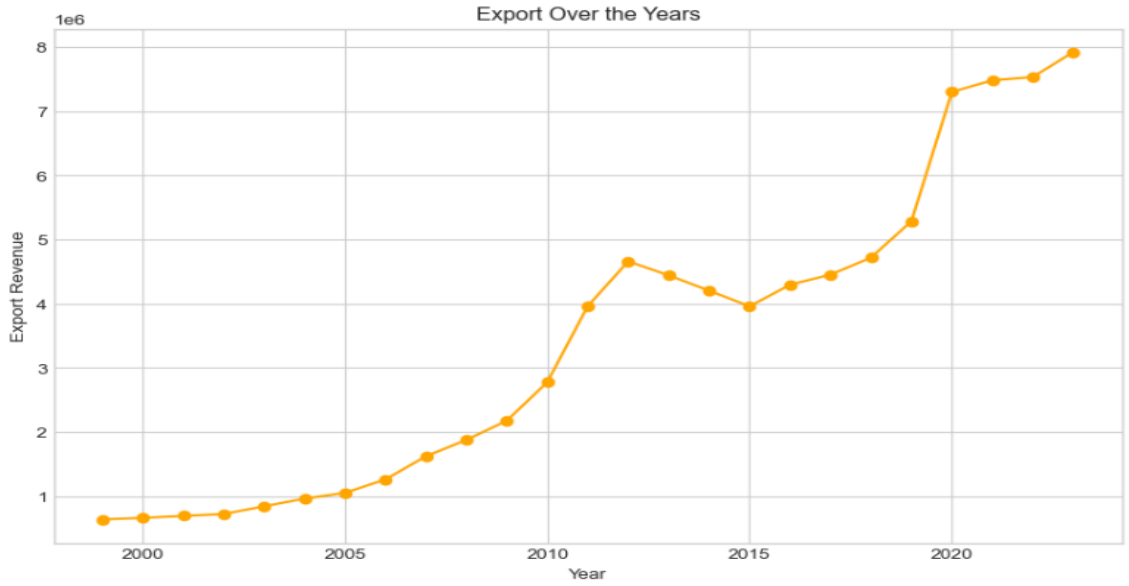
Measures:

- X-Axis: Measures time in Years
- Y-Axis: Measures Percentage (%)

Graph type:

The above graph is a line graph showing Import percentage on GDP over the years.

Graph 4:



Description:

The graph shows export revenue trends over the years from 2000 to 2023. Export revenue demonstrates a steady increase from 2000 to 2010, followed by a sharper rise from 2010 to 2015. After a slight dip around 2015, it resumes growth, accelerating significantly around 2020. The overall trend reflects consistent growth in export revenues, peaking in recent years.

Dimensions:

- X-Axis: The Years ranging from 1999 to 2023.
- Y-Axis: The Export Revenue for each year.

Measures:

- X-Axis: Measures time in years
- Y-Axis: Measures representing the export revenue values

Graph type:

The above graph is a line graph showing exports revenue over the years

Graph 5:

Description:

The graph depicts the inflation rate over the years from 2000 to 2022. It shows fluctuations, with a peak of 12% around 2009 and a significant decline afterward, reaching a low of 3.3% in 2017. After 2017, the inflation rate began to rise again, showing moderate fluctuations in recent years, ranging between 5% and 6.7%. The trend illustrates periods of both rapid increases and decreases in inflation rates over the given timeline.

Dimensions:

- X-Axis: The Years ranging from 1999 to 2023.
- Y-Axis: The Inflation rate over the years

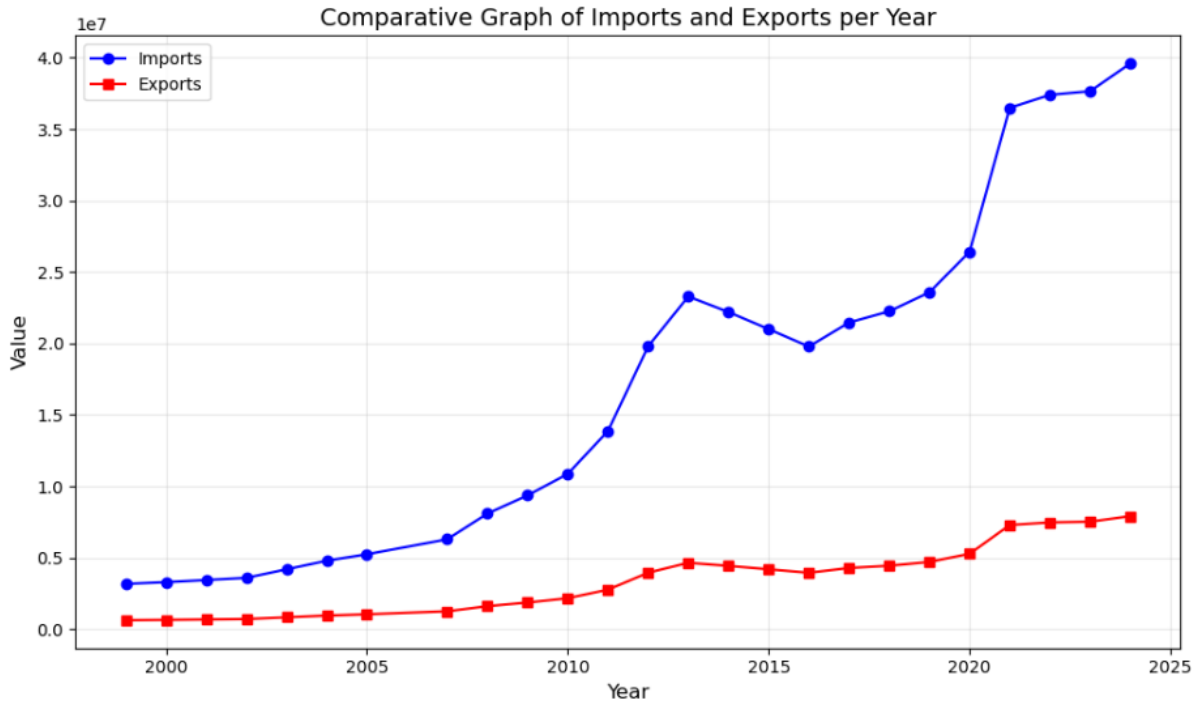
Measures:

- X-Axis: Measures time in years
- Y-Axis: Measures percentage of inflation rate for each year.

Graph type:

The above graph is a line graph showing Inflation rate over the years

Graph 6:



Description:

This graph shows a comparative trend between imports and exports over the years. Imports demonstrate a sharp and consistent upward trajectory, peaking around 2023 at close to 40 million units. Exports have grown steadily but at a much slower pace, remaining below 10 million units even in 2023.

Dimensions:

- X-Axis: The Years ranging from 1999 to 2023.
- Y-Axis: The value of Exports and Imports

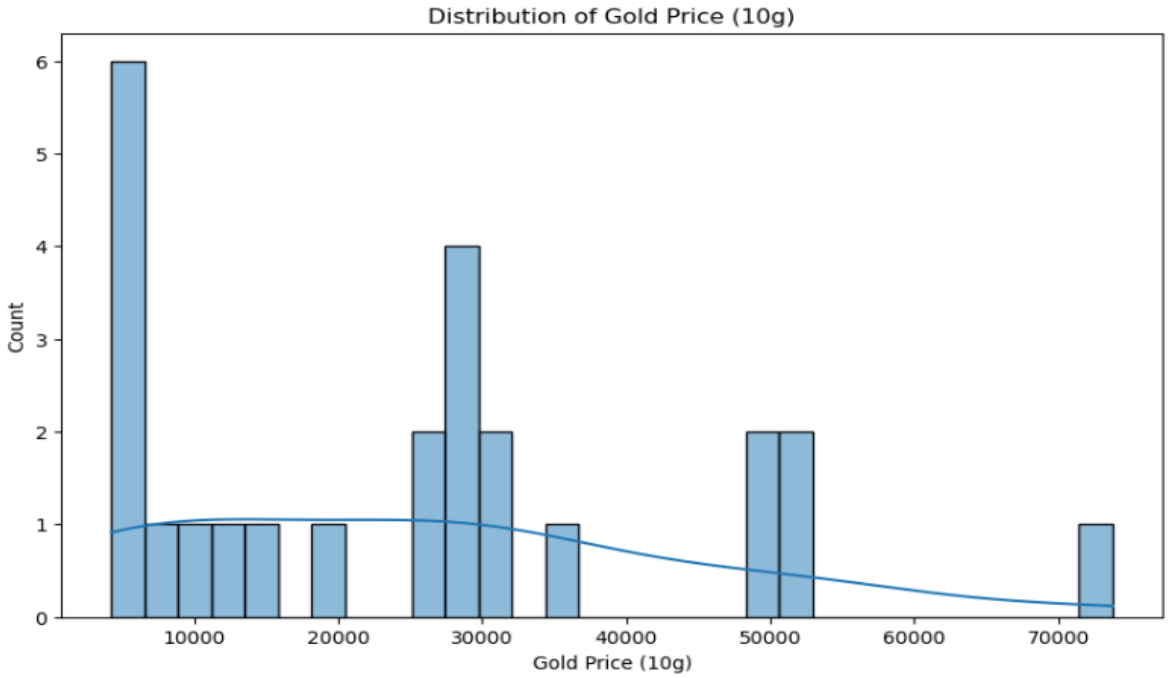
Measures:

- X-Axis: Measures time in years
- Y-Axis: Measures Economic Value (e.g., monetary units)

Graph type:

The above graph is a line graph showing Comparative Graph of Imports and Exports per Year

Graph 7:



Description:

This graph illustrates the distribution of gold prices for 10 grams over a dataset. Histogram bars indicate the frequency of gold prices within specific price ranges. The density curve shows a smoothed trend of the distribution, highlighting the probability density of gold prices. Most of the observations are concentrated around the lower price ranges (below 10,000), with a noticeable peak at the 30,000 range and smaller peaks around 50,000 and 70,000.

Dimensions:

- X-Axis: The Gold Price (10g)
- Y-Axis: The Count, representing the frequency of occurrences for a given price range.

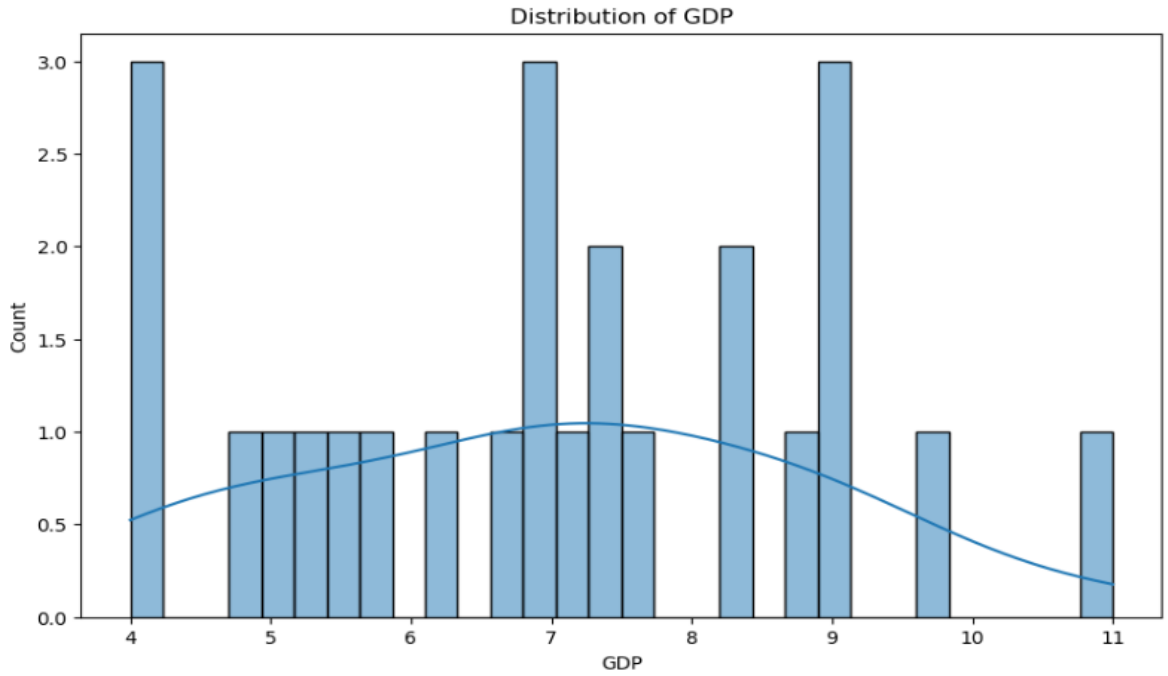
Measures:

- X-Axis: Measures Price (likely in currency units, e.g., INR, USD)
- Y-Axis: Measures Frequency (number of occurrences)

Graph type:

The above graph is a Histogram showing Distribution of Gold Price (10g)

Graph 8:



Description:

The graph depicts the distribution of GDP using a histogram combined with a Kernel Density Estimation (KDE) curve. The histogram shows the frequency of GDP values grouped into bins along the x-axis. The KDE curve overlays a smooth, continuous line that represents the estimated probability density of the GDP distribution, helping to visualize its underlying pattern. Peaks in the KDE curve indicate where GDP values are more concentrated, while gaps or valleys represent less frequent ranges of GDP.

Dimensions:

- X-axis: The GDP (likely in logarithmic scale or ranges of GDP values)
- Y-axis: The Count (the frequency of occurrences within each bin)

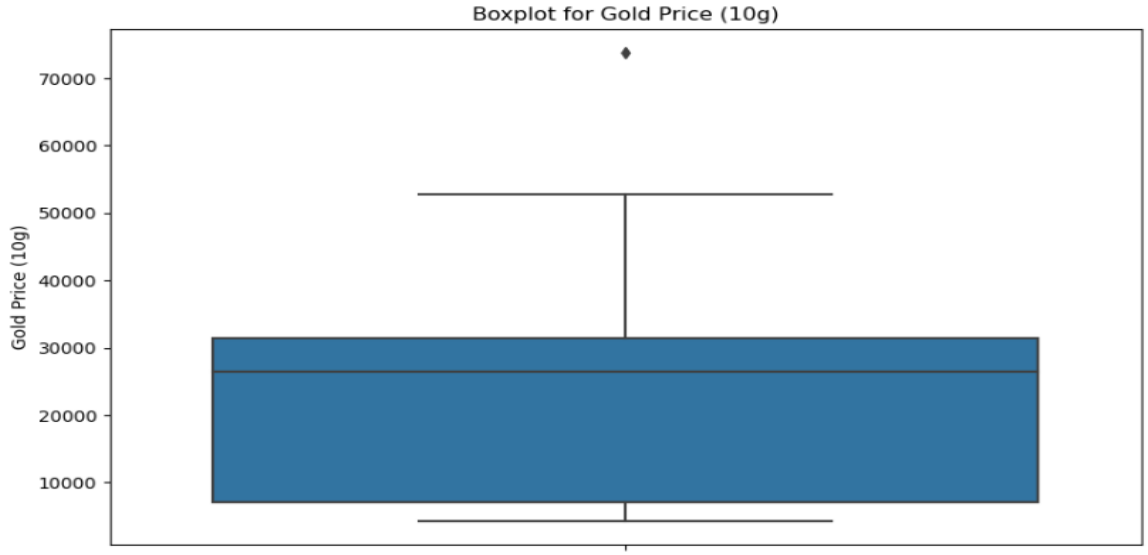
Measures:

- X-axis: Measures GDP
- Y-axis: Measures Frequency (number of data points falling within each GDP bin)

Graph type:

The above graph is a Histogram showing Distribution of GDP

Graph 9:



Description:

The boxplot depicts the distribution of gold prices for 10 grams, summarizing key statistics such as the median, interquartile range (IQR), and outliers. The box spans from the 25th percentile (Q1) to the 75th percentile (Q3), with the median represented by a horizontal line inside the box. The whiskers extend to data points within 1.5 times the IQR, while any points beyond this range are outliers (e.g., a value above 70,000). This visualization highlights the spread, central tendency, and variability of the data. It also identifies potential anomalies, such as the single high outlier.

Dimensions:

- X-axis: The plot does not explicitly use the x-axis for categorization, as this is a univariate boxplot (focused on one variable).
- Y-axis: The Gold Price for 10 grams (measured in currency units, e.g., rupees or dollars).

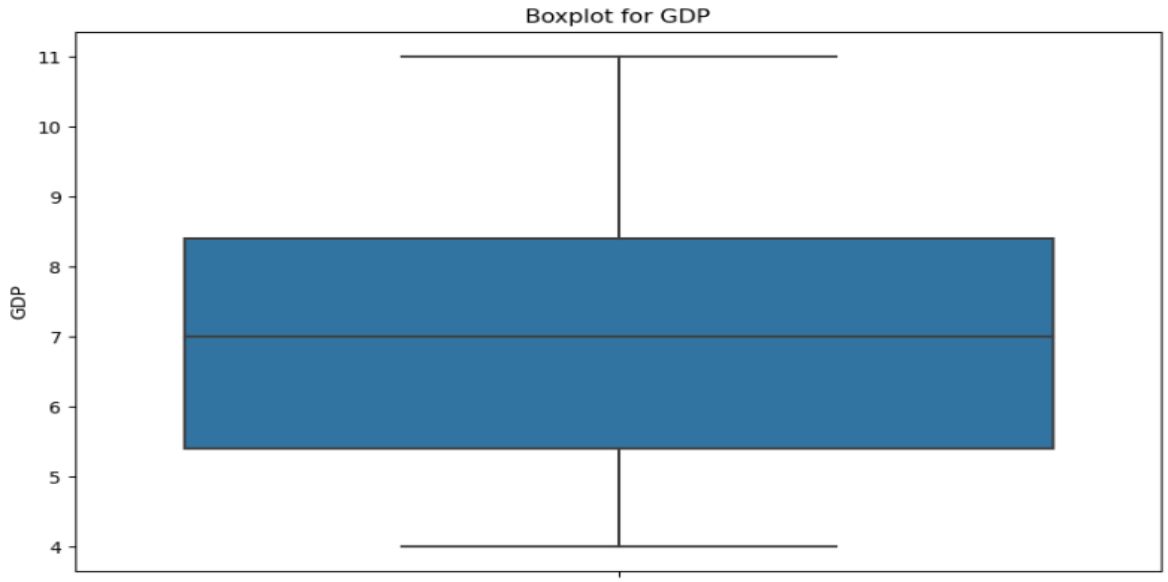
Measures:

- X-axis: Measures this boxplot does not utilize the x-axis for a specific measure since it represents a single variable (gold price).
- Y-axis: Measures Gold price for 10 grams (measured in currency units such as rupees, dollars, etc.)

Graph type:

The above graph is a Boxplot showing Gold Price (10g)

Graph 10:



Description:

The graph is a boxplot visualizing the distribution of GDP values. It summarizes the dataset by highlighting the median, interquartile range (IQR), and overall spread. The box represents the range from the 25th percentile (Q1) to the 75th percentile (Q3), with the median depicted as a horizontal line within the box. The whiskers extend to the minimum and maximum data points within 1.5 times the IQR, showing the range of non-outlier values.

Dimensions:

- X-axis: This boxplot does not actively use the x-axis for any variable; it serves as a label for the data being analysed (GDP).
- Y-axis: GDP values, likely in logarithmic form or specific numerical range.

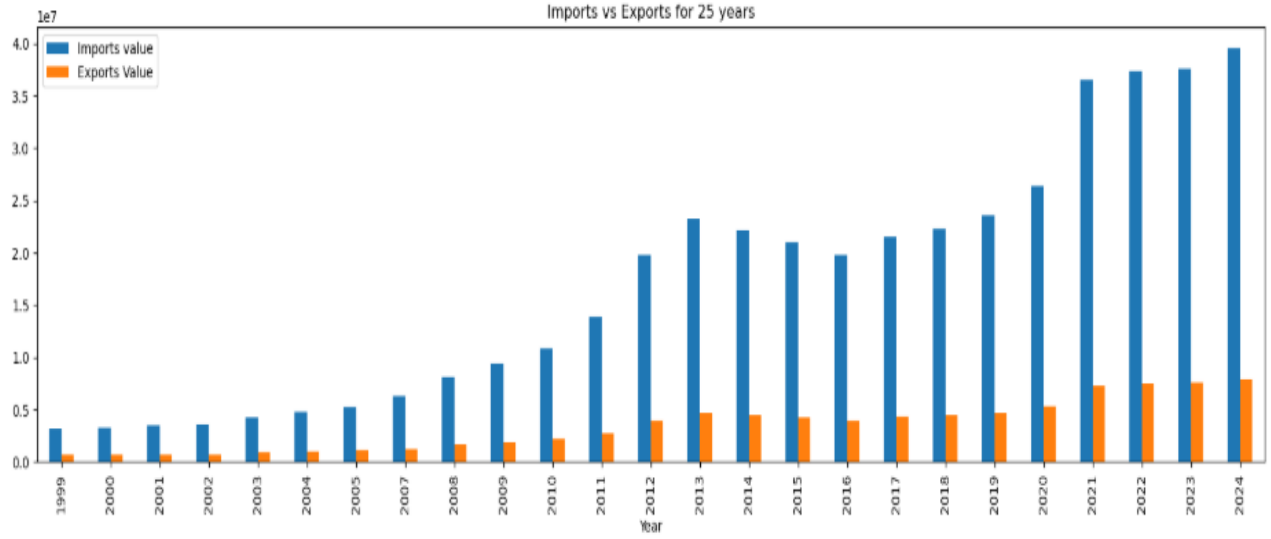
Measures:

- X-axis: Measures label for the variable (GDP)
- Y-axis: Measures GDP values (likely in logarithmic form or specific numerical range)

Graph type:

The above graph is a Boxplot showing GDP

Graph 11:



Description:

The graph depicts the comparison of "Imports" and "Exports" values over a period of 25 years, from 1999 to 2024. The blue bars represent the "Imports Value," and the orange bars represent the "Exports Value" for each year. The values are plotted along the vertical axis, and the years are shown along the horizontal axis.

Dimensions:

- X-axis: The Years ranging from 1999 to 2023
- Y-axis: The Imports and Exports for 25 years

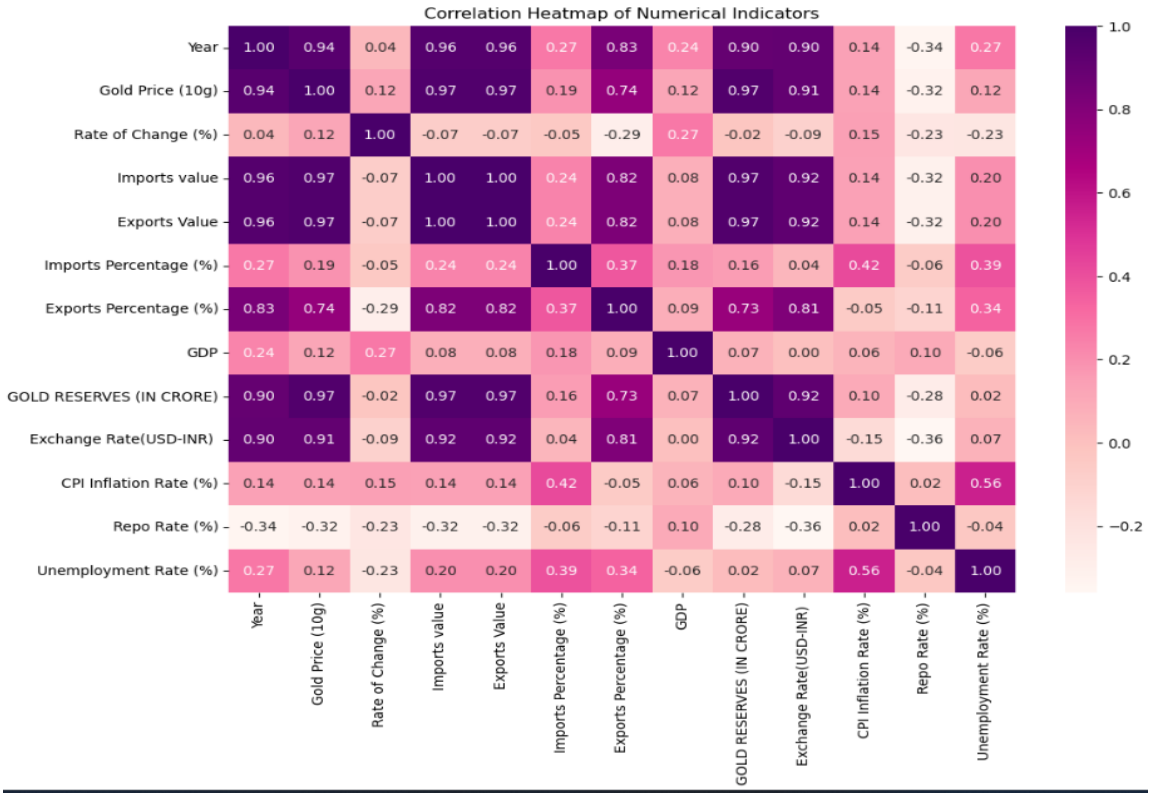
Measures:

- X-axis: Measures time in years
- Y-axis: Measures representing the monetary value of imports and exports, scaled in units

Graph type:

The above graph is a Bar Chart showing Imports Vs Exports for 25 years

Graph12:



Description:

The graph is a correlation heatmap displaying the relationships between various economic and financial indicators. It uses a color gradient to represent correlation strength, with darker hues indicating stronger positive or negative correlations and lighter shades showing weaker correlations. The indicators include metrics like Gold Price, GDP, Exchange Rate, and CPI Inflation Rate. High correlations are evident between variables such as Gold Price and Imports Value, while weaker or negative correlations exist with factors like Repo Rate. This visualization helps identify interdependencies among the indicators for analysis and decision-making.

Dimensions:

X-Axis and Y-Axis: Both axes represent the same set of numerical indicators:

- Year
- Gold Price (10g)
- Rate of Change (%)
- Imports Value
- Exports Value

- Imports Percentage (%)
- Exports Percentage (%)
- GDP
- Gold Reserves (in Crores)
- Exchange Rate (USD-INR)
- CPI Inflation Rate (%)
- Repo Rate (%)
- Unemployment Rate (%)

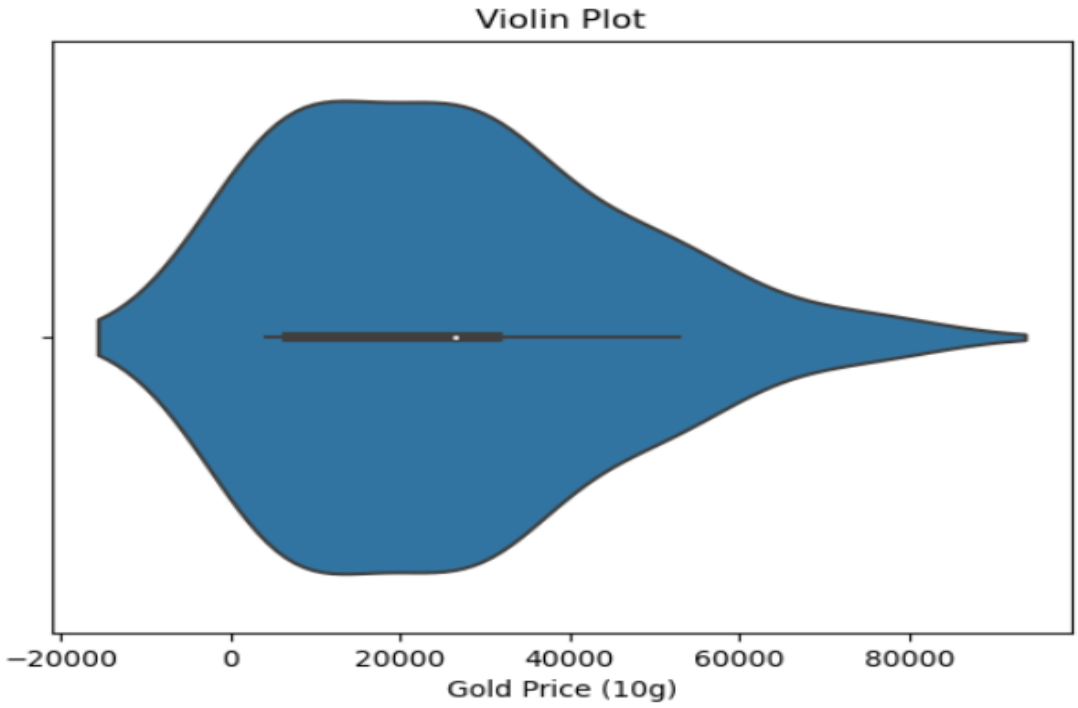
Measures:

1. Correlation Strength:
   - High Correlation: Values closer to 1 or -1 (e.g., Gold Price and Imports Value at 0.97).
   - Low Correlation: Values closer to 0 (e.g., Gold Reserves and Rate of Change at 0.08).
2. Positive Correlations (high values in pink/magenta regions):
   - Gold Price (10g) shows a strong positive correlation with Imports Value (0.97) and Gold Reserves (0.97).
   - Exchange Rate (USD-INR) is positively correlated with Imports Value (0.92).
   - CPI Inflation Rate and Unemployment Rate have a moderate positive correlation (0.56).
3. Negative Correlations (dark purple regions with values < 0):
   - Repo Rate has a notable negative correlation with Gold Price (-0.32) and Imports Percentage (-0.32).
   - Unemployment Rate has a slight negative correlation with Rate of Change (-0.23).
4. Weak Correlations:
   - Indicators like GDP and Rate of Change have very low correlations (close to 0) with many other variables, suggesting limited linear relationships.
5. Trend Over Time:
   - Year correlates positively with indicators like Gold Price (0.94) and Imports Value (0.96), suggesting upward trends over time.

Graph type:

The above graph is a Correlation Heatmap showing Numerical Indicators

Graph 13:



Description:

The graph is a violin plot showing the distribution of gold prices (10g) along the x-axis. It combines a box plot (median and interquartile range) with a density plot, illustrating the spread and concentration of data. The majority of the data is concentrated around lower gold price ranges, as indicated by the wider section in the middle. Sparse data points exist at higher prices, with unusual values extending into the negative range.

Dimensions:

- X-axis: Gold Price (10g)
- Y-axis: Represents a categorical variable or indices

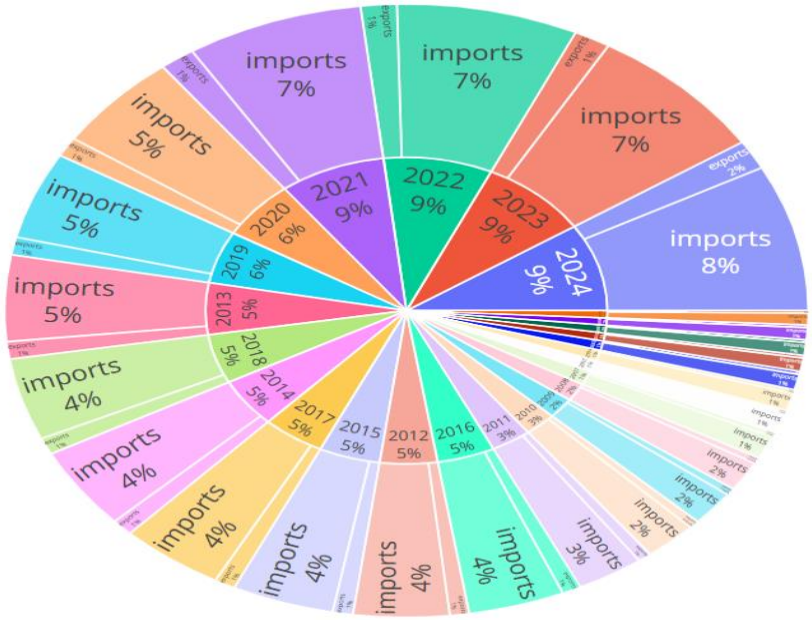
Measures:

- X-axis: Approximately from -20,000 to 90,000
- Y-axis: Represents a categorical variable or index

Graph type:

The above graph is a Violin Plot showing the distribution of data and its probability density, highlighting the spread and clustering of values.

Graph 14:



Description:

The graph is a sunburst chart displaying hierarchical data on yearly import and export percentages from 2011 to 2024. The inner layer represents the years, while the outer layer depicts categories like "imports" and "exports." Each segment's size reflects the percentage contribution, with higher percentages (e.g., 9% for 2022-2024) taking up larger portions. This visualization effectively highlights trade trends over the years.

Dimensions:

- **Year:** Categories such as 2011, 2012, ... up to 2024.
- **Category Type:** Labels like "imports" and "exports."

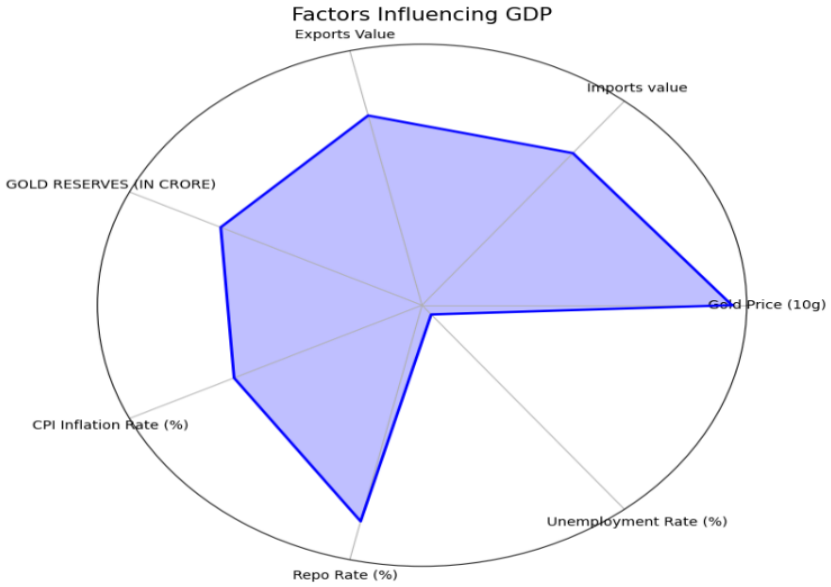
Measures:

Percentage values associated with each year and category type (e.g., 9%, 7%, 4%, etc.).

Graph type:

The above graph is a Sunburst Chart showing Imports and Exports of Gold

Graph 15:



Description:

The radar chart visualizes factors influencing GDP, with each axis representing an economic indicator such as exports, imports, gold price, unemployment rate, repo rate, CPI inflation, and gold reserves. The shaded region shows the relative values of these indicators, forming a distinct polygon. This type of chart helps compare multidimensional data in a single view. It highlights the contribution and variation of each factor in relation to others.

Dimensions:

- Exports Value
- Imports Value
- Gold Price (10g)
- Unemployment Rate (%)
- Repo Rate (%)
- CPI Inflation Rate (%)
- Gold Reserves (in crore)

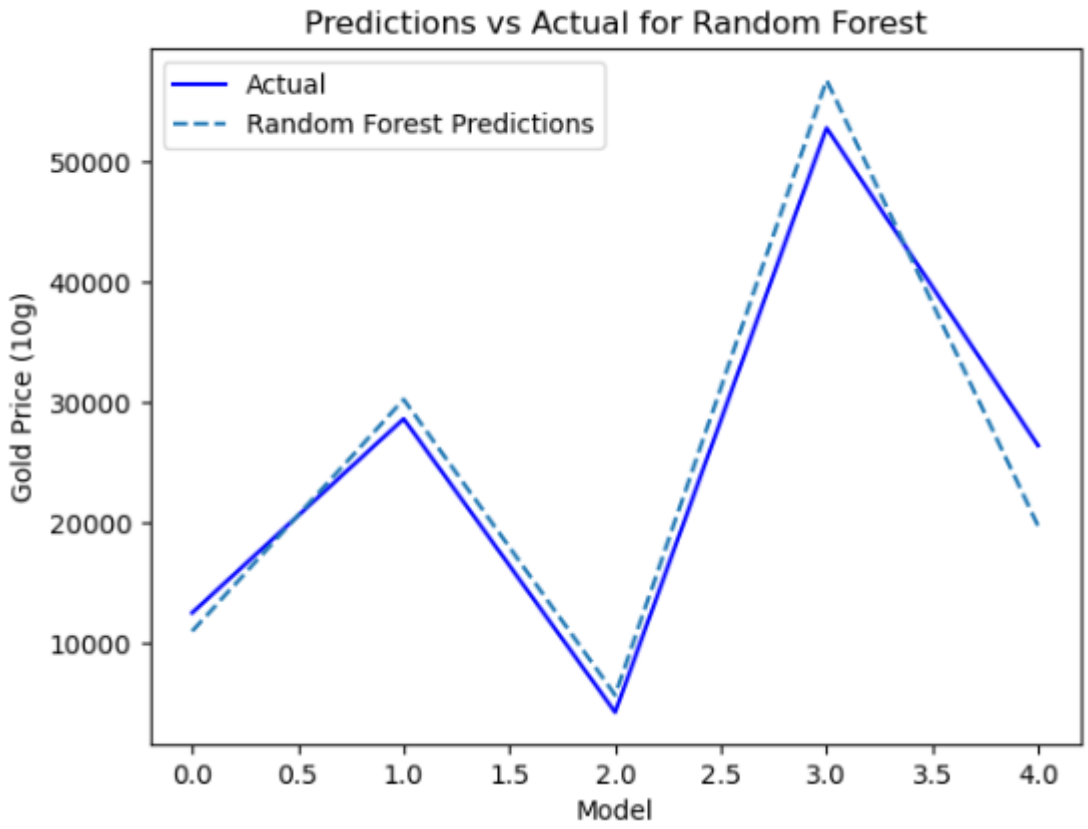
Measures:

Quantitative values corresponding to each of the above factors. These measures are plotted on the respective axes to form the enclosed shape.

Graph type:

The above graph is a radar chart showing multiple variables on a radial grid, with each variable represented by one axis.

Graph 16:



Description:

The graph shows the comparison between actual gold prices and the predictions made by a Random Forest model. The solid blue line represents the actual gold prices, while the dashed blue line represents the predicted values.

Dimensions:

- X-axis: The "Model" index (0.0 to 4.0)
- Y-axis: The "Gold Price (10g)" in units ranging from 0 to 50,000+.

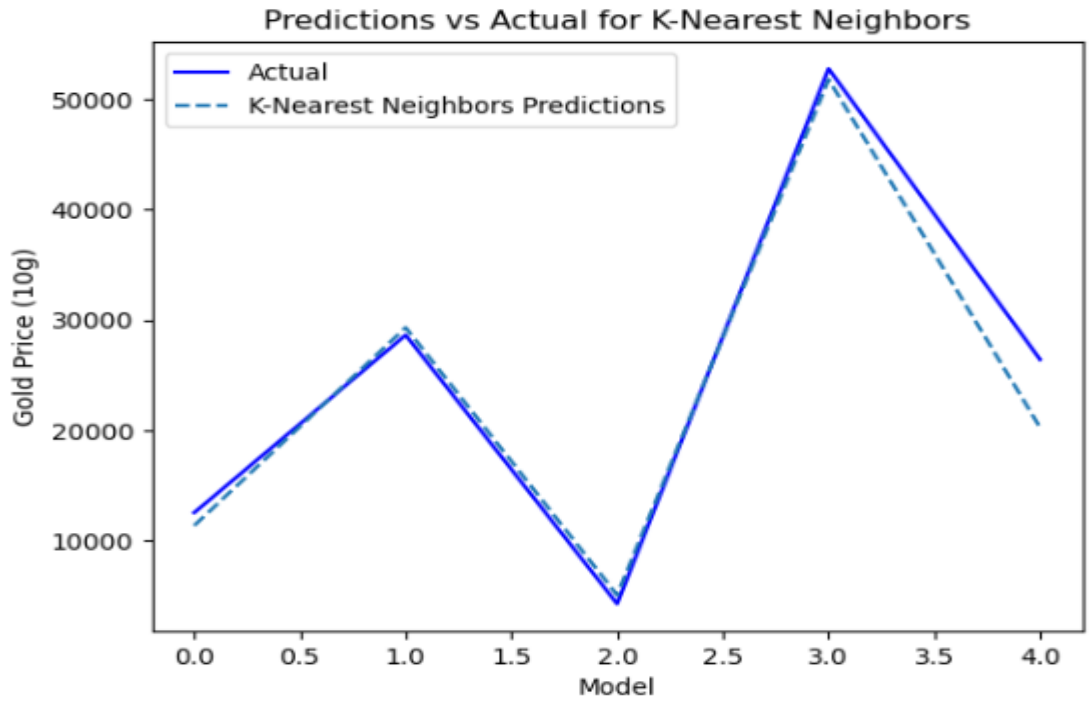
Measures:

- X-axis: Represents the index of observations or instances, likely numerical identifiers ranging from 0 to 4
- Y-axis: Represents the gold price for 10 grams, measured in monetary units

Graph type:

The above graph is a line graph showing Predictions vs Actual for Random Forest

Graph 17:



Description:

The graph compares the actual gold prices with predictions made by a K-Nearest Neighbors (KNN) model. The x-axis represents different models or data points, while the y-axis shows gold prices (10g) in numeric values, ranging up to 50,000. The solid blue line indicates actual values, while the dashed blue line represents the predicted values. The lines closely follow each other, indicating a good alignment between actual and predicted values.

Dimensions:

- X-axis: The "Model" (likely a sequence or index of different models or data points).
- Y-axis: The Gold Price (10g)

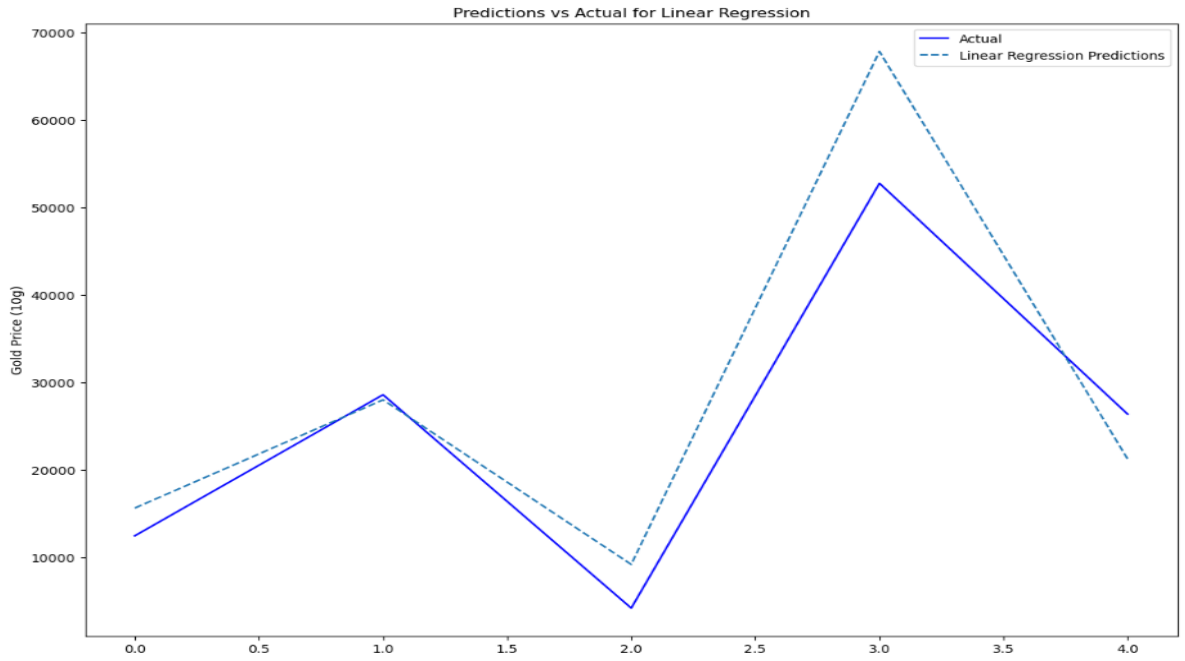
Measures:

- X-axis: Measures no explicit units provided but appears to be numeric (e.g., 0 to 4).
- Y-axis: Measures Prices are in units of a numeric value, scaling from approximately 0 to 50,000.

Graph type:

The above graph is a line graph showing Predictions vs Actual for K-Nearest Neighbor

Graph 18:



Description:

The graph shows a comparison of actual gold prices and predicted values from a linear regression model. The X-axis represents observation indices, while the Y-axis displays gold prices in units of 10 grams. The solid blue line indicates the actual values, and the dashed blue line represents the predictions. Deviations between the lines highlight the model's prediction accuracy and areas where it diverges from actual values.

Dimensions:

- X-axis: Represents an index or some unitless measure (e.g., time steps, observation indices)
- Y-axis: The gold price (in units of 10 grams)

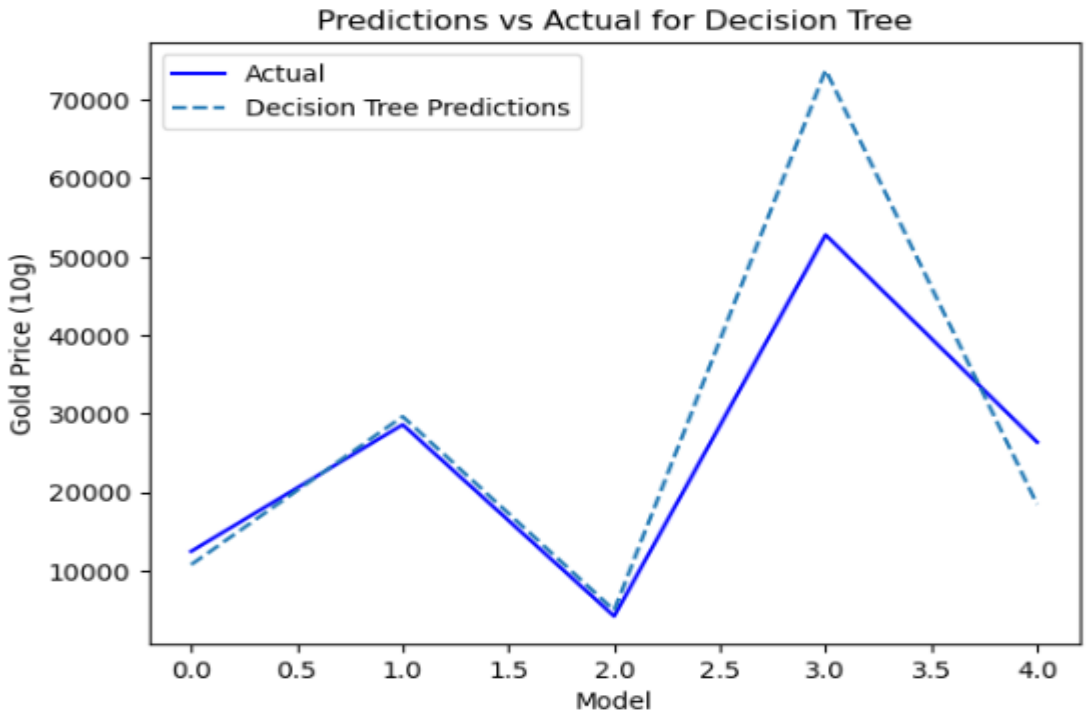
Measures:

- X-axis: The observation indices (e.g., time steps or data points) with values ranging from approximately 0 to 4.
- Y-axis: The gold price in units of 10 grams, with values ranging from 0 to 70,000.

Graph type:

The above graph is a line graph showing Predictions vs Actual for Linear Regression

Graph 19:



Description:

The graph compares the actual gold prices (in units of 10g) to the predicted prices using a decision tree model. The x-axis represents the model indices (arbitrarily ranging from 0 to 4), and the y-axis shows the gold price values. Two lines are present: a solid blue line for actual values and a dashed blue line for the decision tree predictions.

Dimensions:

- X-axis: Model index (arbitrary numerical representation, ranging from 0 to 4).
- Y-axis: Gold price (in units of 10g, ranging from approximately 0 to 75,000).

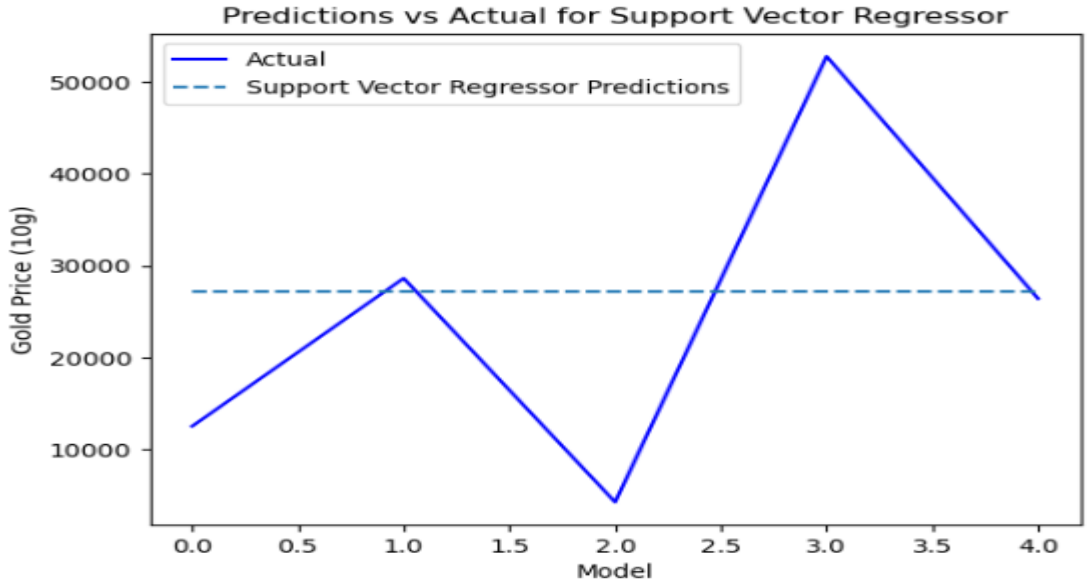
Measures:

- X-axis: Model (categorical or ordinal representation, possibly indicating different data points, instances, or time steps).
- Y-axis: Gold Price (measured in units of 10 grams, with values ranging approximately from 0 to 75,000).

Graph type:

The above graph is a line graph showing Predictions vs Actual for Decision Tree

Graph 20:



Description:

The graph compares actual gold prices (solid blue line) with predictions from a Support Vector Regressor (dashed blue line). The x-axis represents model indices, while the y-axis shows gold prices (10g) ranging up to 60,000. The actual prices fluctuate significantly, while predictions remain constant. The graph highlights a clear gap between predictions and actual values.

Dimensions:

- X-axis: The index of models or observations
- Y-axis: The gold price in monetary units for 10 grams

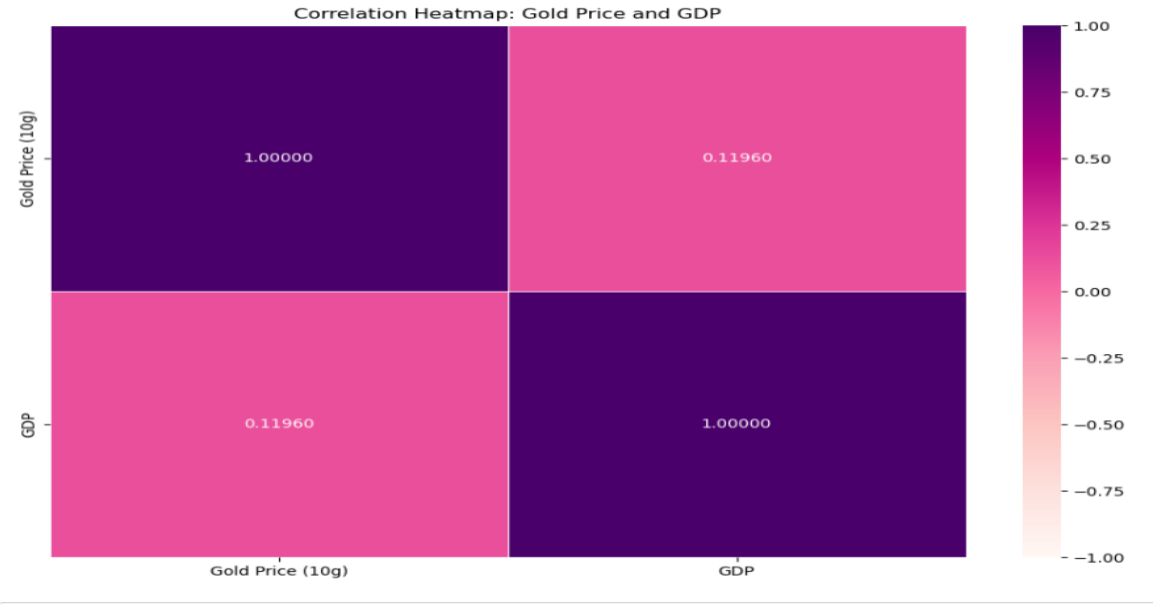
Measures:

- X-axis: Measures numerical values representing data points or observations
- Y-axis: Measures numerical values in monetary units, ranging from approximately 0 to 60,000

Graph type:

The above graph is a line graph showing Predictions vs Actual for Support Vector Regressor

Graph 21:



Description:

The graph is a correlation heatmap showing the relationship between Gold Price (10g) and GDP. It uses colors to represent the correlation coefficient, where dark purple indicates low or no correlation and bright pink indicates higher correlation. The diagonal elements show a perfect correlation (1.00000) of each variable with itself, while the off-diagonal value (0.11960) suggests a weak positive correlation between Gold Price and GDP. This helps visualize the linear relationship between these two variables.

Dimensions:

- X-axis: Gold Price (10g) and GDP
- Y-axis: Gold Price (10g) and GDP

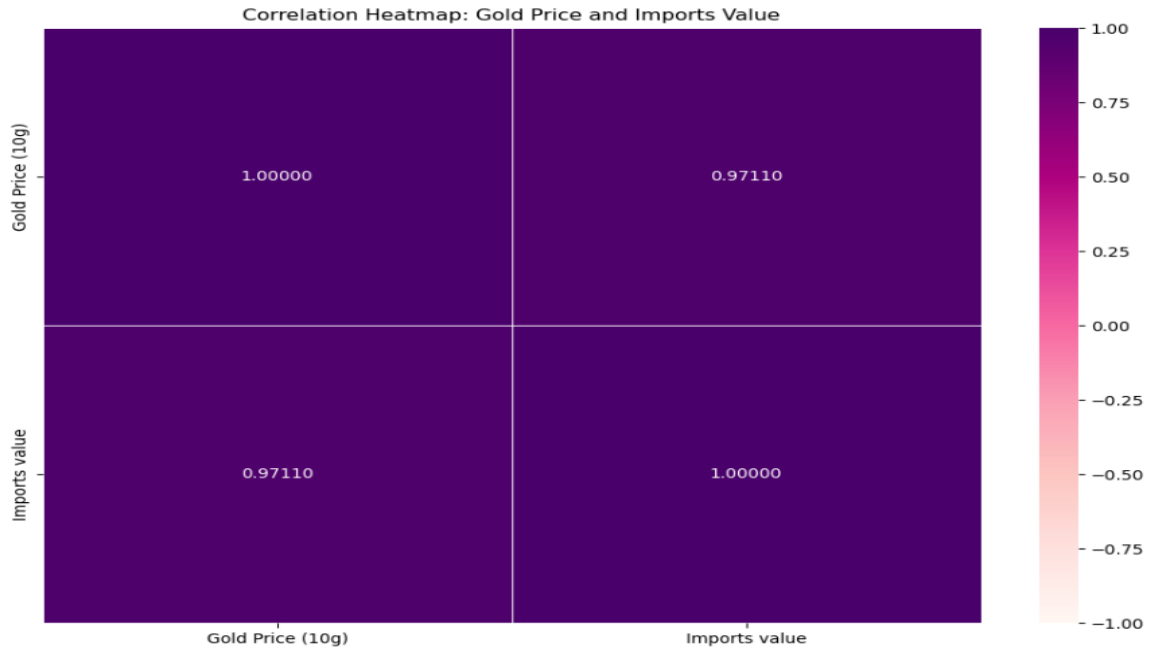
Measures:

- The diagonal elements measure the self-correlation of variables (always 1.00000).
- The off-diagonal element (0.11960) measures the weak positive correlation between Gold Price (10g) and GDP.

Graph type:

The above graph is a Correlation Heatmap showing Gold Price and GDP

Graph 22:



Description:

The heatmap shows the correlation between gold price (per 10g) and imports value. The diagonal values are 1.0, indicating perfect self-correlation. The correlation between gold price and imports value is 0.9711, representing a strong positive relationship. The color bar ranges from -1.0 to 1.0, with lighter colors signifying higher correlations.

Dimensions:

- X-axis: Variables - Gold Price (10g), Imports Value
- Y-axis: Variables - Gold Price (10g), Imports Value

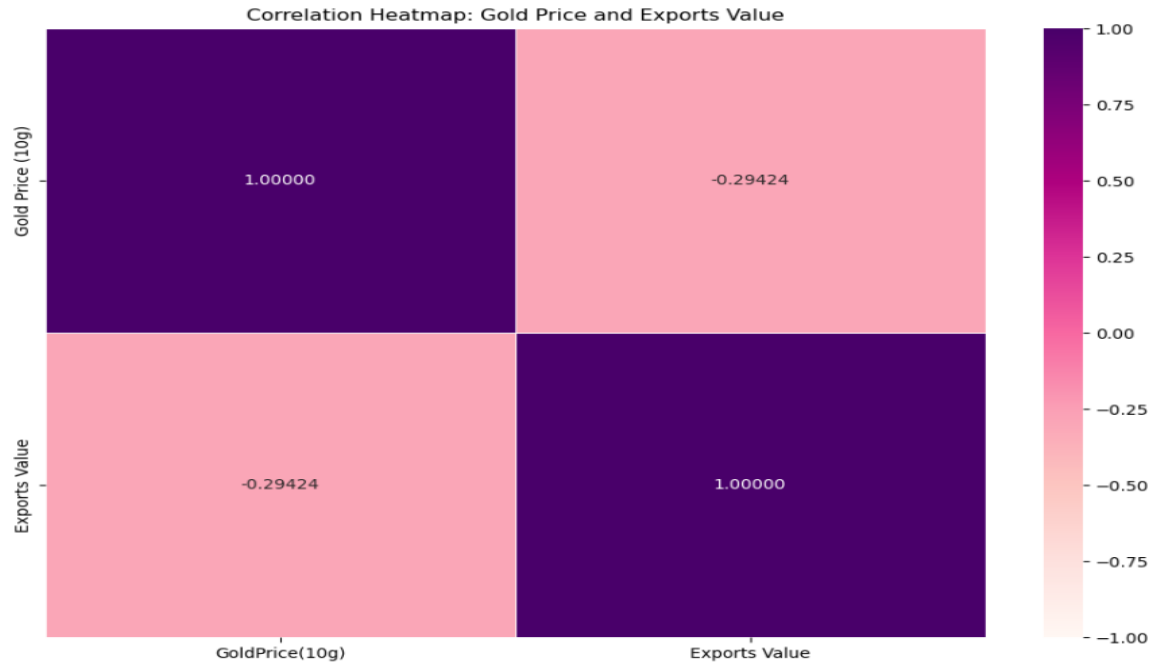
Measures:

- A value of **1.000** (diagonal) indicates perfect correlation of a variable with itself.
- A value of **0.971** shows a very strong positive correlation between Gold Price (10g) and Imports Value.

Graph type:

The above graph is a Correlation Heatmap showing Gold Price and Imports Value

Graph 23:



Description:

The graph is a correlation heatmap showing the relationship between "Gold Price (10g)" and "Exports Value." It uses numerical values and a colour gradient to indicate the strength and direction of correlations. The diagonal values are 1.00, showing perfect self-correlation, while the off-diagonal value (-0.29424) represents a weak negative correlation between the two variables. Darker shades indicate stronger correlations, while lighter shades indicate weaker ones.

Dimensions:

- X-axis: Gold Price (10g) and Exports Value
- Y-axis: Gold Price (10g) and Exports Value

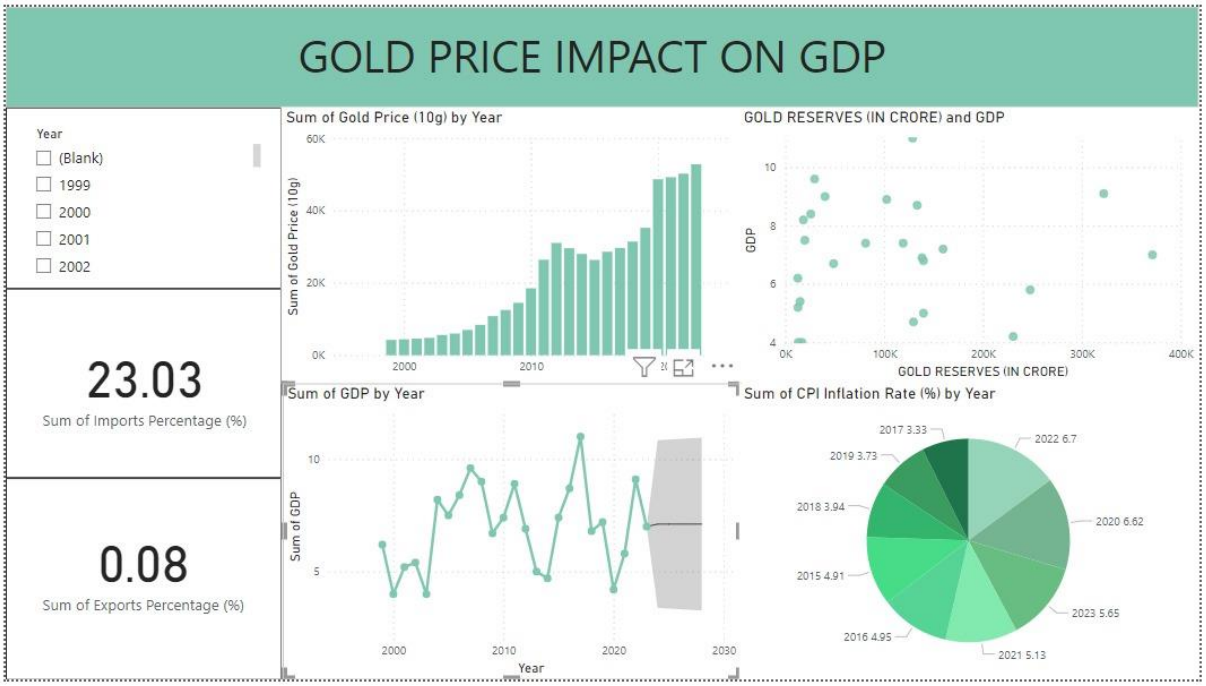
Measures:

- The correlation between "Gold Price (10g)" and itself is **1.00** (perfect correlation).
- The correlation between "Exports Value" and itself is also **1.00**.
- The correlation between "Gold Price (10g)" and "Exports Value" is **-0.29424**, indicating a weak negative relationship.

Graph type:

The above graph is a Correlation Heatmap showing Gold Price and Exports value

DASH BOARD



The dashboard titled **"Gold Price Impact on GDP"** visualizes the relationship between gold prices, reserves, and economic indicators such as GDP, imports, exports, and CPI inflation rates over time. Here's a breakdown of its key components:

1. **Year Filter:**
   Allows users to filter data by specific years (1999, 2000, 2001, 2002, or other years with blank entries).

2. **Import and Export Percentages:**

   - **Sum of Imports Percentage:** Displays the percentage contribution of imports to the economy, showing a value of 23.03%.

   - **Sum of Exports Percentage:** Displays the percentage contribution of exports, with a value of 0.08%.

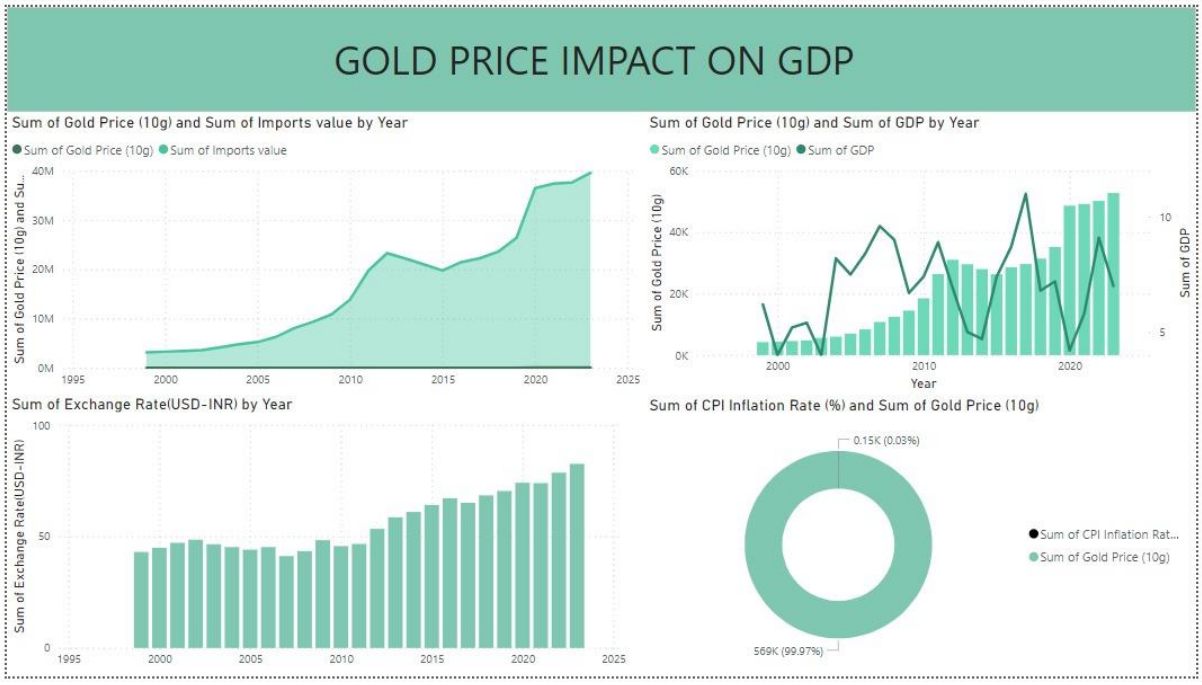
3. **Gold Price by Year:**
   A bar chart illustrating the trend in the price of 10 grams of gold over the years. The chart shows a steady increase, especially after the 2000s.

4. **Gold Reserves and GDP:**
   A scatterplot visualizing the relationship between gold reserves (in crores) and GDP. It helps identify potential correlations between the two metrics.

5. **GDP by Year:**
   A line chart showing fluctuations in GDP over time. The chart highlights trends, peaks, and troughs in economic performance.

6. **CPI Inflation Rate by Year:**
   A pie chart displaying the annual CPI inflation rates (%) for specific years, highlighting the changes in inflation rates over time.

DASHBOARD



This dashboard provides an analysis of the impact of gold prices on GDP, showing trends and correlations across multiple dimensions. The visualizations include the following:

1. **Sum of Gold Price (10g) and Sum of Imports Value by Year** (Top Left):
   - This line chart displays the trend of gold prices and the value of imports over the years. It shows how both metrics have evolved, with a noticeable increase in recent years.

2. **Sum of Gold Price (10g) and Sum of GDP by Year** (Top Right):
   - This combined line and bar chart illustrates the relationship between gold prices and GDP. The bar represents the GDP, while the line tracks the gold price trend, allowing for a visual correlation between these variables over time.

3. **Sum of Exchange Rate (USD-INR) by Year** (Bottom Left):

- This bar chart shows the yearly trend of the exchange rate between the US Dollar and Indian Rupee. The exchange rate's movement provides context for gold price fluctuations.

4. **Sum of CPI Inflation Rate (%) and Sum of Gold Price (10g)** (Bottom Right):
    - Represented as a donut chart, this visualization compares the cumulative contribution of the Consumer Price Index (CPI) inflation rate and gold prices. It highlights the overwhelming contribution of gold price growth relative to inflation rates.