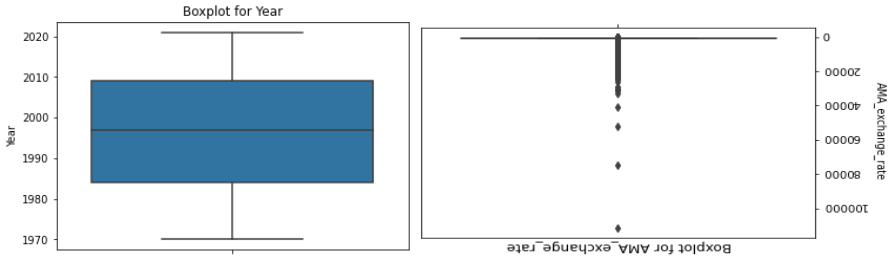
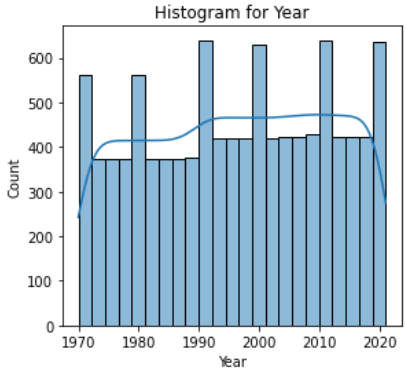
Data Collection and Preprocessing Phase

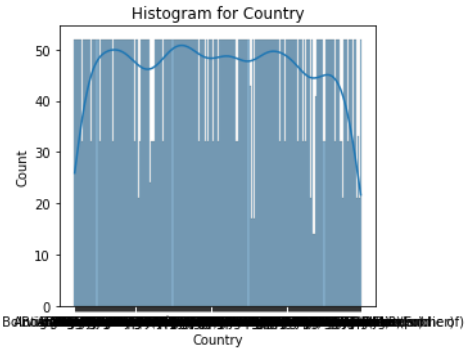
Data Exploration and Preprocessing Report

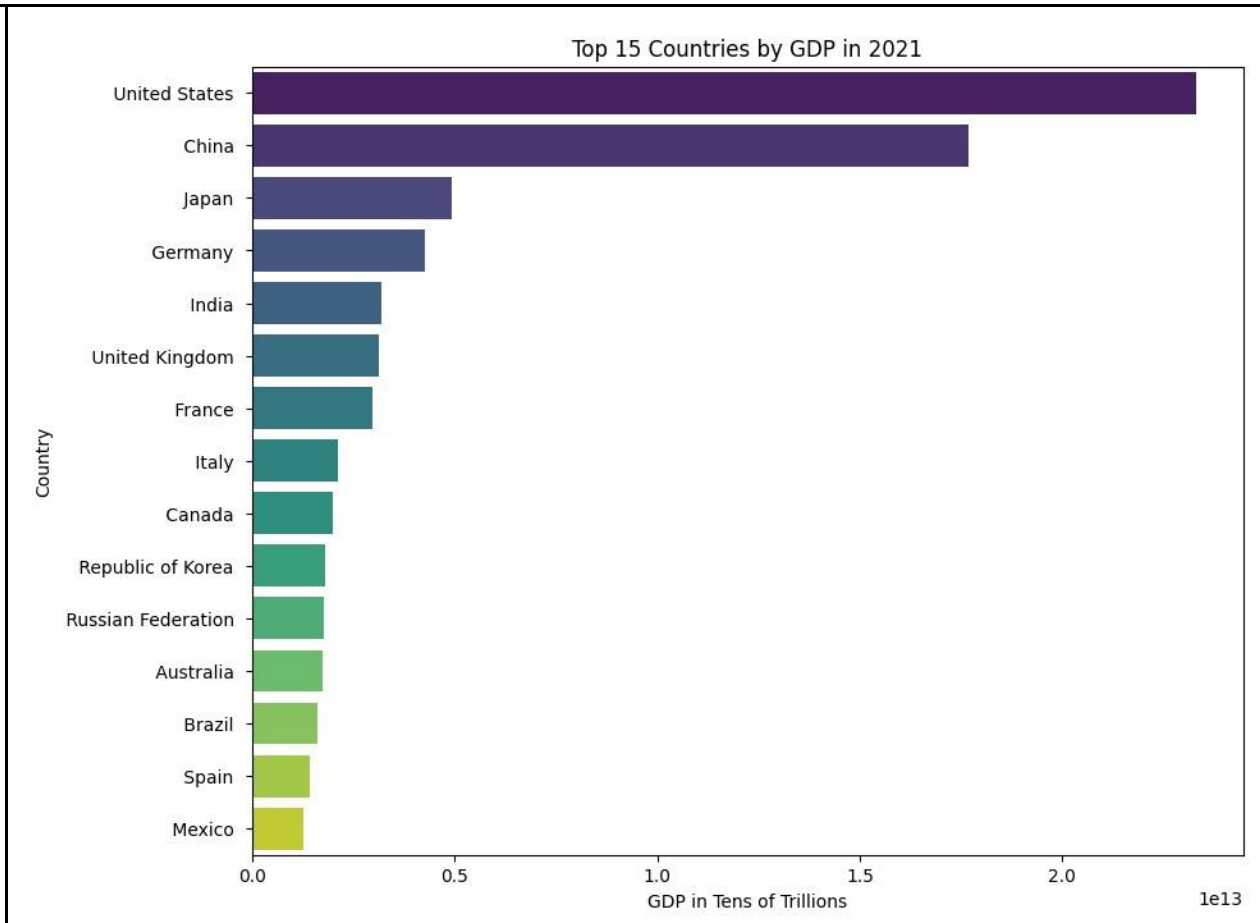
Date	15 March 2024
Team ID	740115
Project Title	**Predicting IMF-Based Exchange Rates: Leveraging Economic Indicators for Accurate Regression Modeling**
Maximum Marks	6 Marks

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.


Boxplot for Year
Year
2020
2010
2000
1990
1980
1970
0
20000
40000
60000
80000
100000
AMA_exchange_rate
Boxplot for AMA_exchange_rate


Histogram for Year
Count
600
500
400
300
200
100
0
1970
1980
1990
2000
2010
2020
Year


Histogram for Country
Count
50
40
30
20
10
0
Country

Bivariate Analysis	

Section	Description

Data Overview

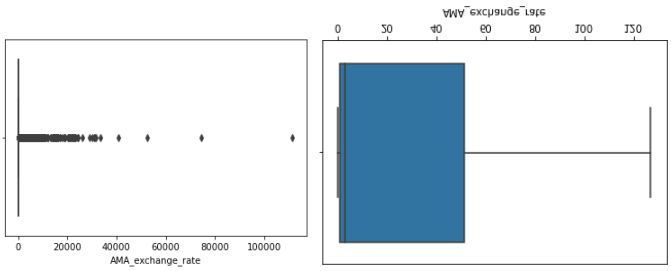
Dimension:

614 rows × 13 columns

Descriptive statistics:

	Year	AMA_exchange_rate	IMF_exchange_rate	Population	Per capita GNI	(ISIC A-B)	Changes_i
count	10512.000000	1.051200e+04	1.051200e+04	1.051200e+04	10512.000000	1.051200e+04	
mean	1996.262747	3.573959e+02	3.419846e+02	2.851523e+07	8965.564593	7.793212e+09	
std	14.900361	2.291128e+03	1.941857e+03	1.141296e+08	17070.205895	4.011060e+10	
min	1970.000000	4.300000e-14	4.300000e-14	4.359000e+03	34.000000	2.813900e+04	-
25%	1984.000000	1.000000e+00	1.000000e+00	6.330615e+05	730.000000	1.336557e+08	
50%	1997.000000	2.812895e+00	2.761315e+00	5.051556e+06	2316.500000	9.569466e+08	
75%	2009.000000	5.134316e+01	4.806684e+01	1.678862e+07	8965.750000	4.213059e+09	
max	2021.000000	1.116366e+05	4.200000e+04	1.425893e+09	234317.000000	1.350000e+12	

Univariate Analysis

Outliers and Anomalies	
Data Preprocessing Code Screenshots	

Loading Data

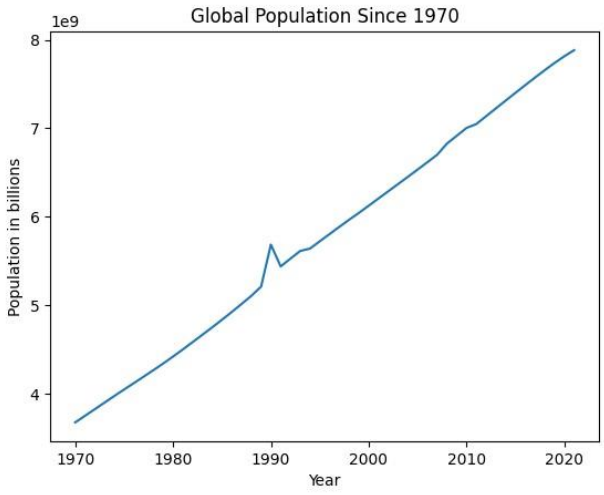
Loading the dataset

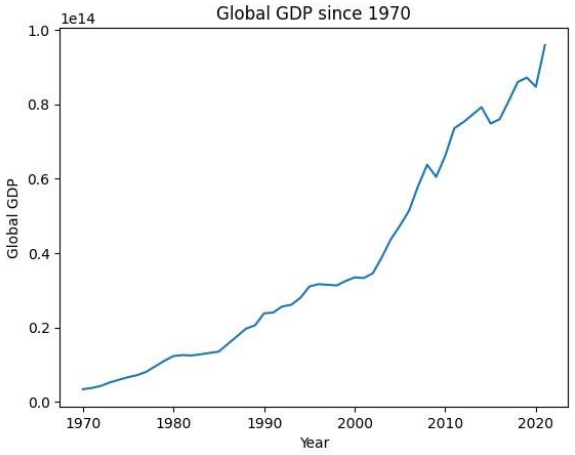
```
data=pd.read_csv("Global Economy Indicators.csv")
```

```
data
```

	CountryID	Country	Year	AMA exchange rate	IMF based exchange rate	Population	Currency	Per capita GNI	Agriculture, hunting, forestry, fishing (ISIC A-B)	Changes in inventories
0	4	Afghanistan	1970	0.044998	0.044998	10752971	Afghani	164	8.699174e+08	NaN
1	4	Afghanistan	1971	0.044998	0.044998	11015857	Afghani	168	9.108281e+08	NaN
2	4	Afghanistan	1972	0.044998	0.044998	11286753	Afghani	149	8.279453e+08	NaN
3	4	Afghanistan	1973	0.044998	0.044998	11575305	Afghani	150	8.554869e+08	NaN

Multivariate Analysis





Handling Missing Data	**Handling missing values** [14]: ``` data.isnull().sum() ``` [14]: ``` Country 0 Year 0 AMA_exchange_rate 0 IMF_exchange_rate 0 Population 0 Currency 0 Per capita GNI 0 (ISIC A-B) 121 Changes_in_inventories 1841 (ISIC F) 0 Exports 21 Final_expenditure 0 Govt_expenditure 52 GCF 52 GFCF 52 HCE 52 Imports 42 (ISIC D) 43 ``` [ ]: ``` data['(ISIC A-B)']=data['(ISIC A-B)'].fillna(data['(ISIC A-B)'].mean()) data['Changes_in_inventories']=data['Changes_in_inventories'].fillna(data['Changes_in_inventories'].mea data['Exports']=data['Exports'].fillna(data['Exports'].mean()) data['Govt_expenditure']=data['Govt_expenditure'].fillna(data['Govt_expenditure'].mean()) data['GCF']=data['GCF'].fillna(data['GCF'].mean()) data['GFCF']=data['GFCF'].fillna(data['GFCF'].mean()) data['HCE']=data['HCE'].fillna(data['HCE'].mean()) data['Imports']=data['Imports'].fillna(data['Imports'].mean()) data['(ISIC D)']=data['(ISIC D)'].fillna(data['(ISIC D)'].mean()) data['(ISIC I)']=data['(ISIC I)'].fillna(data['(ISIC I)'].mean()) data['(ISIC G-H)']=data['(ISIC G-H)'].fillna(data['(ISIC G-H)'].mean()) ```
Data Transformation	
Feature Engineering	Attached the codes in final submission

Save Processed Data	