Name: Etukuri Naveen

Student Number: 23017408

Git repository URL: [Releases · etukuri6/IMO-Rank · GitHub]

1. Introduction:

In an Exploratory data analysis report, we will analyze the dataset containing the International Mathematical Olympiad (IMO) ranks by country. The dataset provides information about the ranks achieved by different countries in the IMO competition over the years.

In [1]:
```
import pandas as pd

# Load the dataset
data = pd.read_csv("IMO Rank.csv")

# Display the first few rows of the dataset
print(data.head())

# Summary statistics of the numerical variables
print(data.describe())

# Check for missing values
print(data.isnull().sum())
```

```
   Year  ALBANIA  ARGENTINA  ARMENIA  AUSTRALIA  AUSTRIA  AZERBAIJAN  BELGIUM  \
0  2023     80.0         59       32         23       49        54.0       63
1  2022     77.0         42       26         29       56        34.0       63
2  2021     94.0         46       37         18       64        51.0       43
3  2020     77.0         50       37          8       65        65.0       50
4  2019     80.0         53       49         18       60        51.0       68

   BANGLADESH  BULGARIA  ...  TUNISIA  TURKEY  TAIWAN  UKRAINE  \
0        46.0        25  ...     75.0       8      10       31
1        57.0        16  ...     67.0      26      14       17
2        43.0        18  ...     56.0      35       9        6
3        38.0        38  ...     71.0      26      23       11
4        67.0        19  ...     74.0      29      21       11

   UNITED KINGDOM  URUGUAY  USA  UZBEKISTAN  VENEZUELA  VIETNAM
0              13     83.0    2        49.0      104.0        7
1              13     89.0    3        50.0      104.0        4
2               9     88.0    4        61.0       83.0       14
3               9     80.0    3        69.0       83.0       17
4              20     88.0    1        63.0      105.0        7
```

2. Data Exploration:

Data exploration refers to the process of examining and analyzing a dataset to understand its structure, characteristics, and patterns. The main goals of data exploration are:

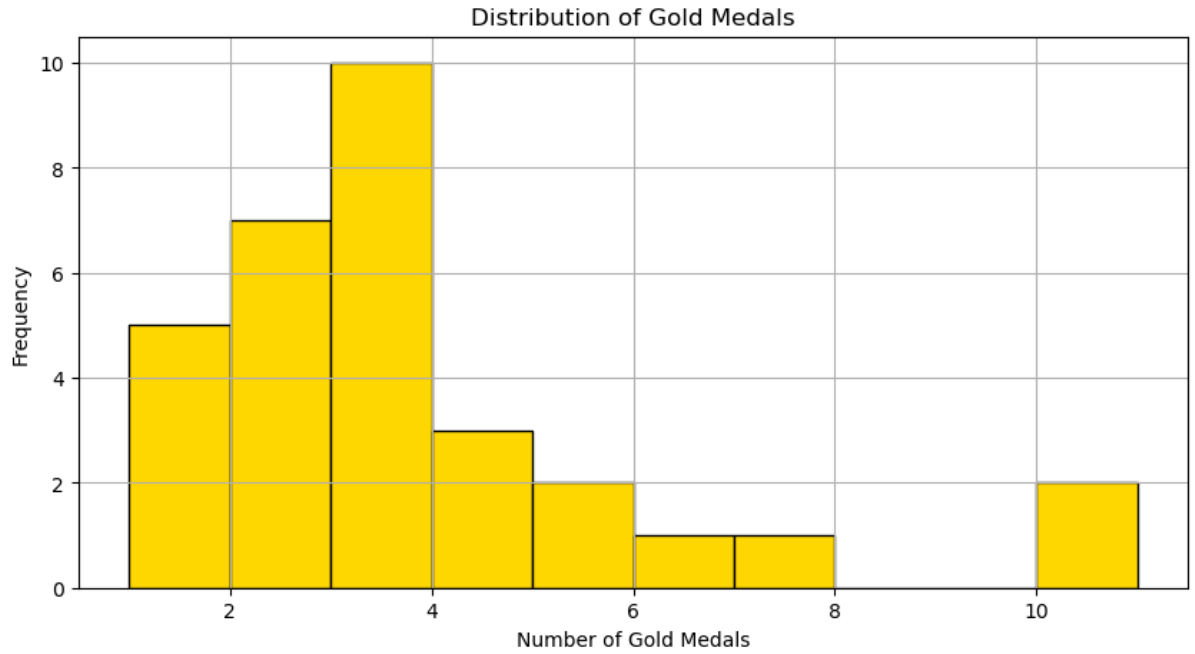
Understanding the Data: Exploring the dataset helps in gaining insights into the variables (columns) and observations (rows) it contains. It includes understanding the data types, variable names, and the overall structure of the dataset.

Identifying Patterns and Trends: Data exploration involves visualizing the data to identify any patterns, trends, or relationships between variables. It can help in uncovering interesting insights or correlations that may not be immediately apparent.

Detecting Anomalies or Outliers: By exploring the data, we can identify any unusual or unexpected observations that deviate from the norm. These anomalies, also known as outliers, may need further investigation as they can have a significant impact on the analysis or modeling process.

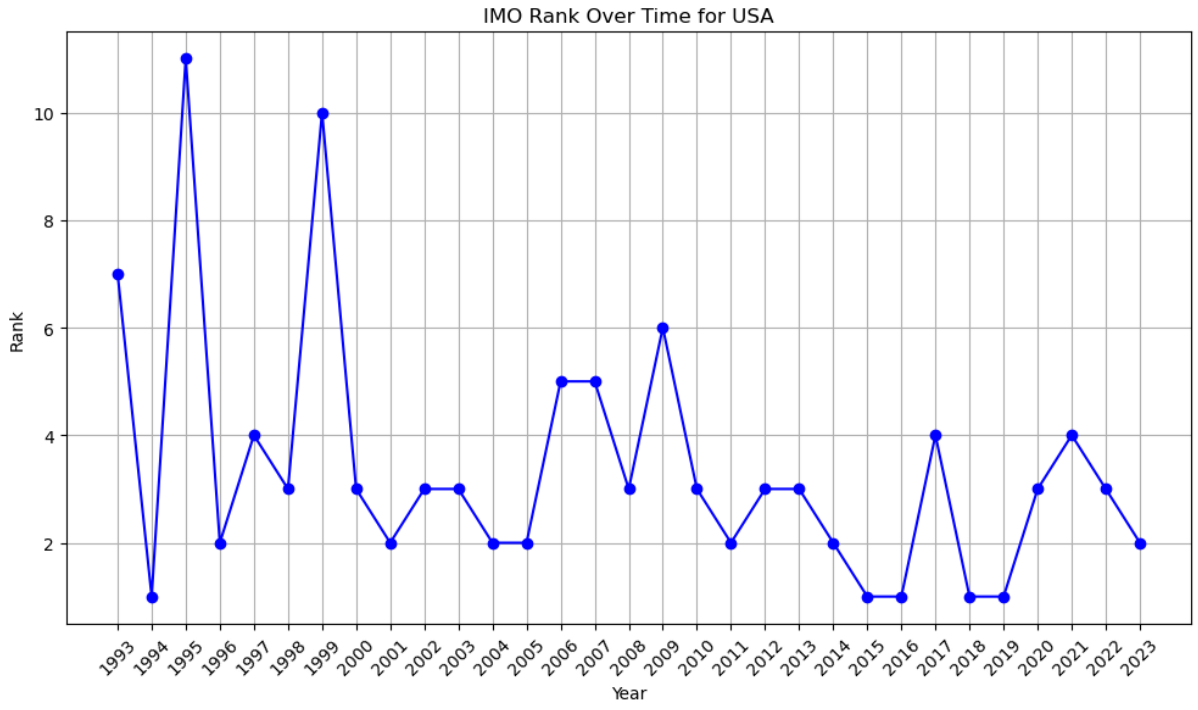
Preparing Data for Analysis: Data exploration also involves data preprocessing tasks such as removing duplicate datasets, and pointless columns from the dataset and scaling numerical features. By exploring any dataset, we can determine the appropriate preprocessing steps needed to clean and prepare the data for analysis or modeling.

2.1 Histogram



The above Histogram provides a clear visual representation of the distribution of the number of gold medals won by the country 'USA'. It is a visual representation of the data, which can help me in understanding the distribution, outliers, and skewness of the data. The bins in the histograms represent the range of values, and the height of the bin shows the frequency of occurrence of the values within that range.

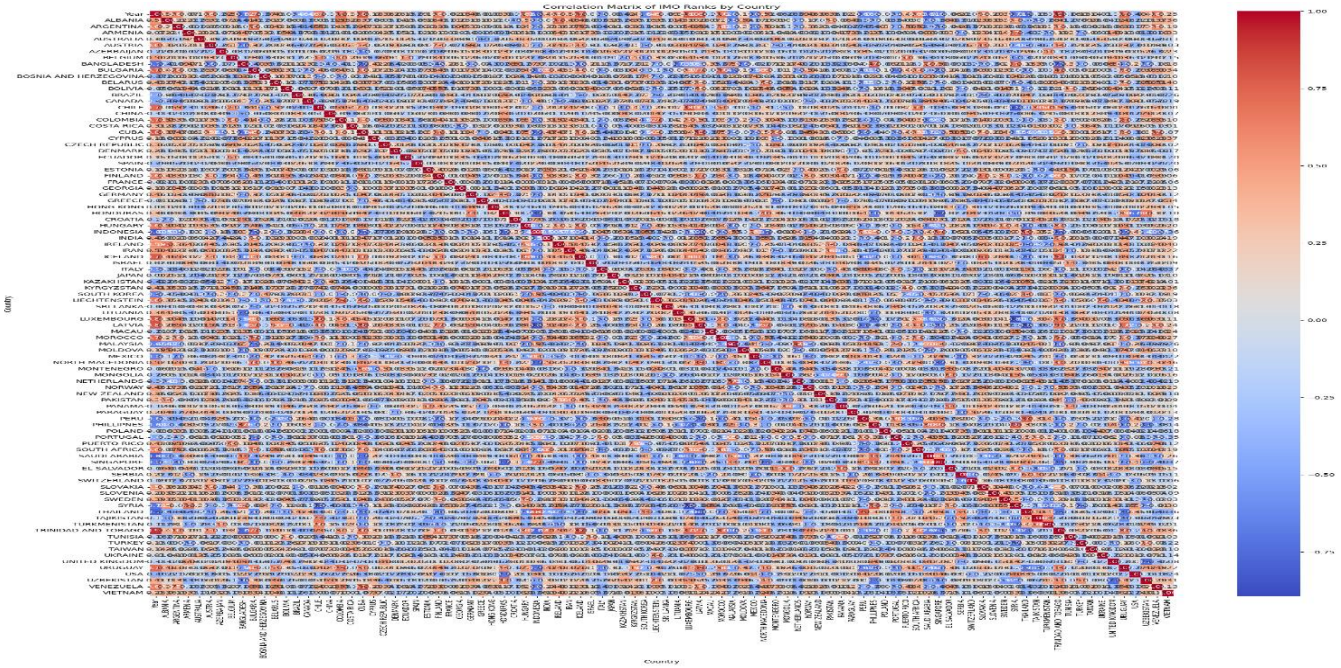
2.2 Line plot



The line graph shows the IMO rank over time for the USA. It provides a visual representation of the data, which can help me understand the performance of the country in the competition and identify the trend in the rank over the years. In the Graph, the X-axis represents the year, and the Y-axis represents the IMO rank. IMO rank in the year 1995, it had a higher rank indicated a poor performance, whereas a lower rank indicated a better performance.

2.3 Correlation Matrix:

It is a graphical representation of the relationships between several Country characteristics.



Lastly, we'll calculate the correlation matrix to explore the relationships between the ranks of different countries over the years. We'll visualize the correlation matrix using a heatmap.

3. Conclusion:

Through this Exploratory analysis, we gained insights into the distribution of ranks achieved by countries in the IMO competition over the years. We also explored how the ranks of a specific country changed over time and investigated the relationships between the ranks of different countries using correlation analysis.

Further analysis could involve examining trends in ranks for multiple countries simultaneously and identifying factors that may influence a country's performance in the IMO competition.

4. **References:**

Tukey, J.W(1977), Exploratory Data Analysis (Vol2).