ITW1 PYTHON PROJECT GROUP 19

TOKYO OLYMPICS 2021 ANALYSIS
GRAPHS, HISTOGRAMS & CHARTS

Problem Statement:

Data Analysis of the Tokyo Olympics 2021 Based on the Data set Received from the Official Website. We did this project to predict which country will take more medals from analyzing the data of Tokyo Olympics 2021 .Hence, we are going to plot graphs, histograms and charts to do so using python language and it's modules which are mentioned below.

Libraries we used

- **NumPy :** NumPy is a basic level external library in Python used for complex mathematical operations. Here, we used it to analyze the data and calculate the mathematical functions we included in our project .

- **Pandas :** Pandas is a useful library in data analysis. It can be used to perform data manipulation and analysis .

- **Matplotlib :** Matplotlib supports various types of graphical representations like Bar Graphs, Histograms, Line Graph, Scatter Plot, Stem Plots, etc. Matplotlib is a 2-D plotting library. Hence, we used this library to plot all our graphs .

- **Seaborn :** Seaborn is an open-source Python library built on top of matplotlib. It is used for data visualization and exploratory data analysis. Hence, we used the seaborn library to create charts and histograms .

- **Plotly :** Plotly allows users to import, copy and paste, or stream data to be analyzed and visualized .

Data Set we used

Here is the link of the dataset we used as a reference to plot our graphs, histograms and charts : https://rb.gy/uostls

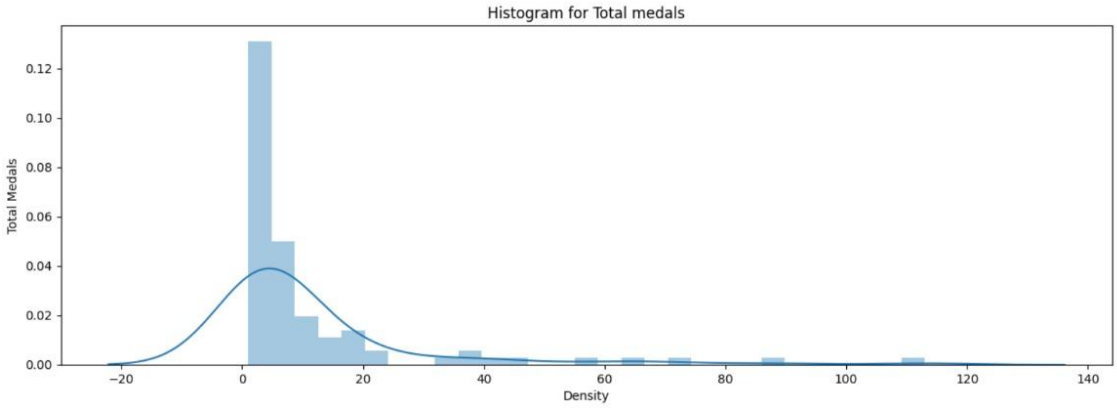
Each row corresponds to an individual athlete competing in an individual event, including the athlete's name, sex, age, height, weight, country, and medal, and the event's name, sport, games, year, and city.

Main Functions we used

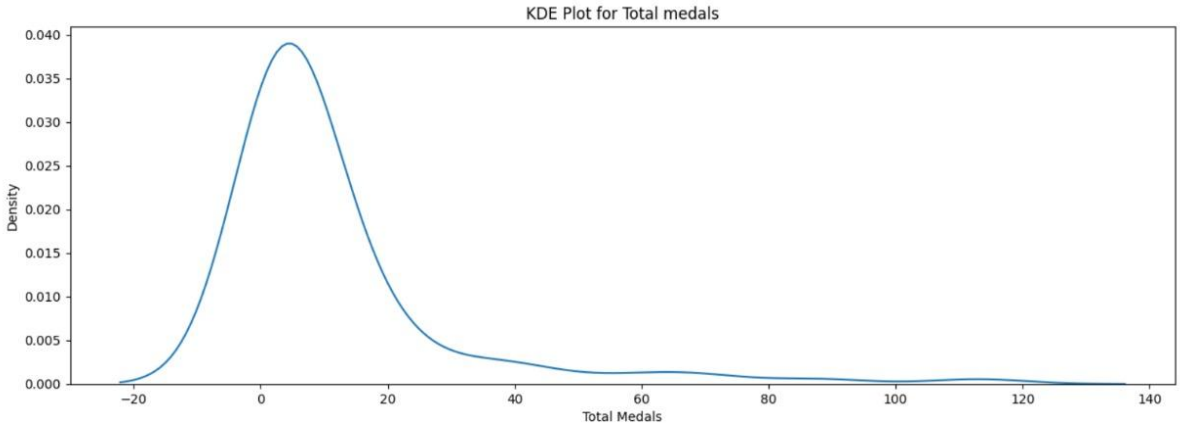
- .read_csv() - reads delimited text files .
- .head() - gets first n rows for the object based on position .
- .describe() -views some basic statistical details like percentile, mean, std of a data frame
- .value_counts() - returns object containing counts of unique values .
- .loc() -accesses a group of rows and columns by label(s) or a boolean array.
- .array() - processes data dynamically and stores a collection of data .
- .random.randint() - returns an integer number selected element from the specified range.
- .arrange() - returns an array with equally spaced elements .
- .distplot() -visualize the parametric distribution of a dataset.
- .title() - a title to the current plot .
- .xlabel() - sets the axis-labels of a Seaborn plot .
- .ylabel() - sets the axis-labels of a Seaborn plot .
- .subplots() -returns a figure object and a tuple containing axes objects equal to nrows*ncols .
- .scatter() - displays colored circles at the locations specified by the vectors X and Y .
- .swarmplot() - draws a categorical scatterplot with non-overlapping points .
- .figure() - creates a figure object .

Results and plots generated from our application

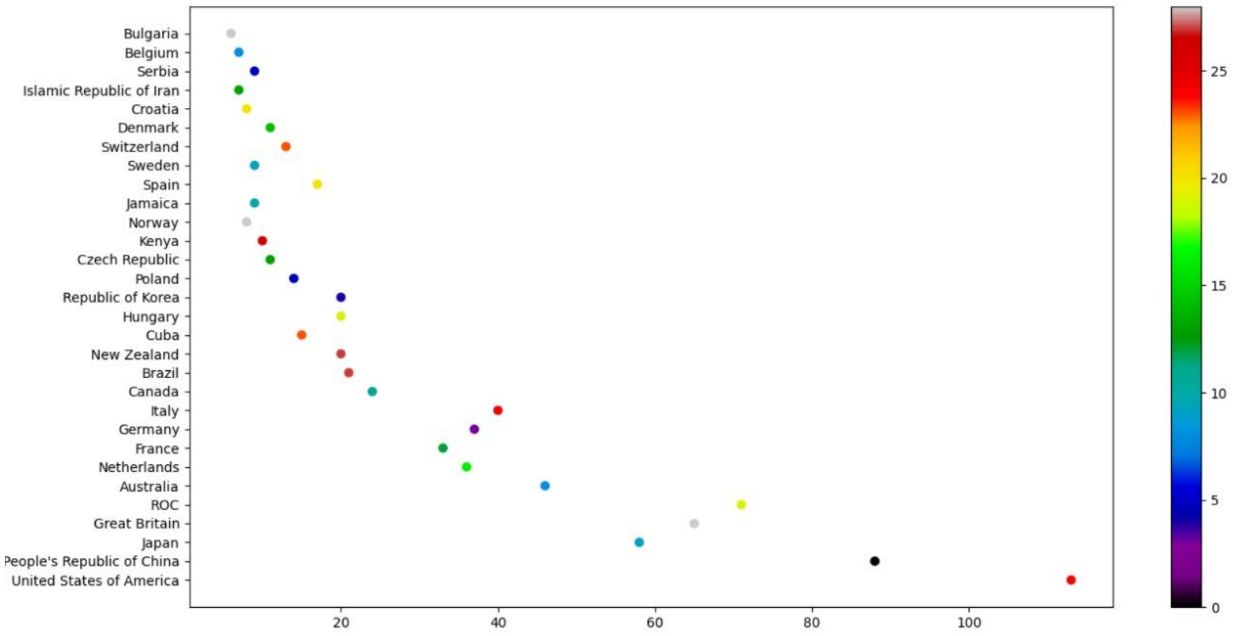
- Histogram of total medals and probability density :



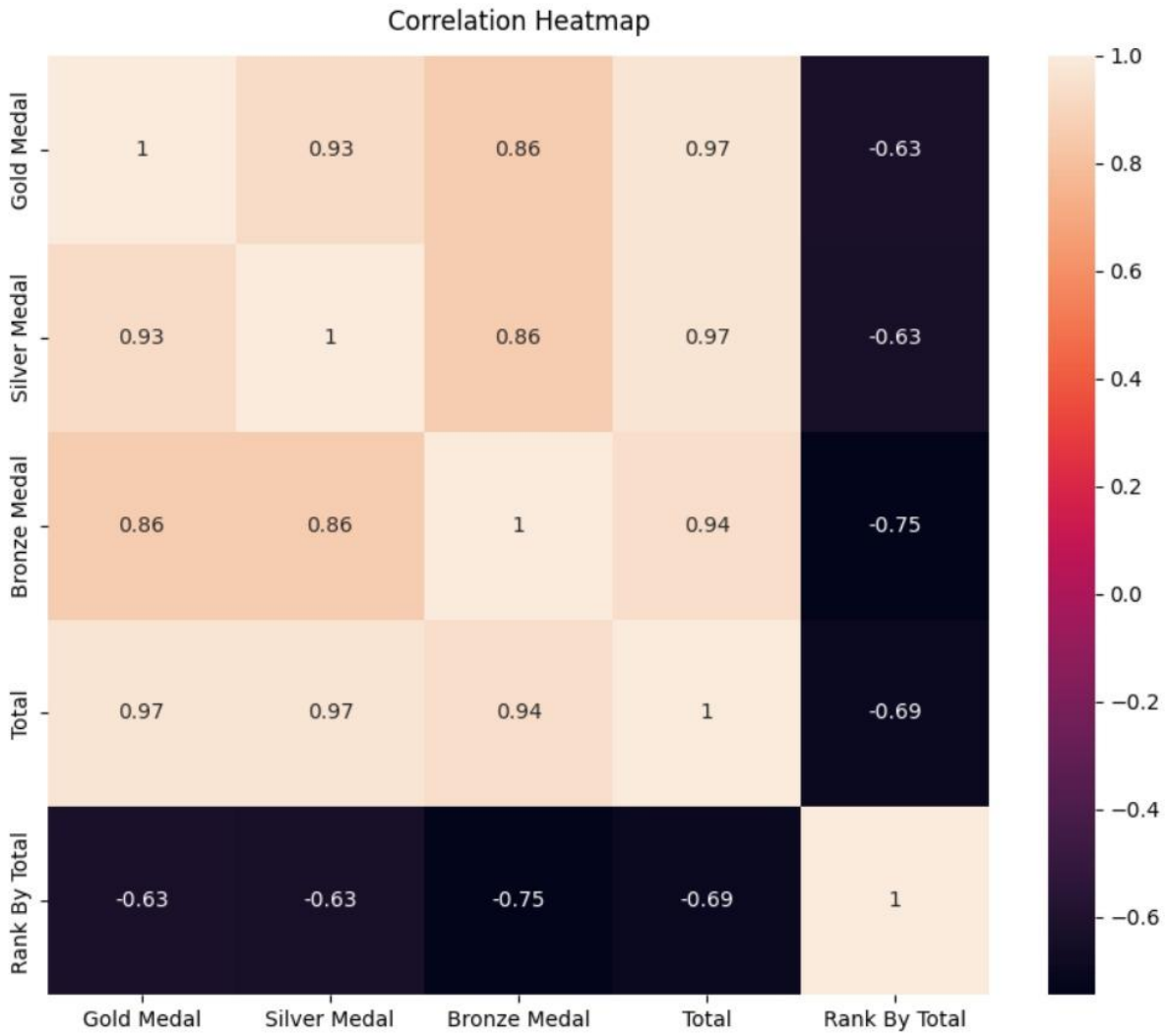
- KDE Plot for the total medals :



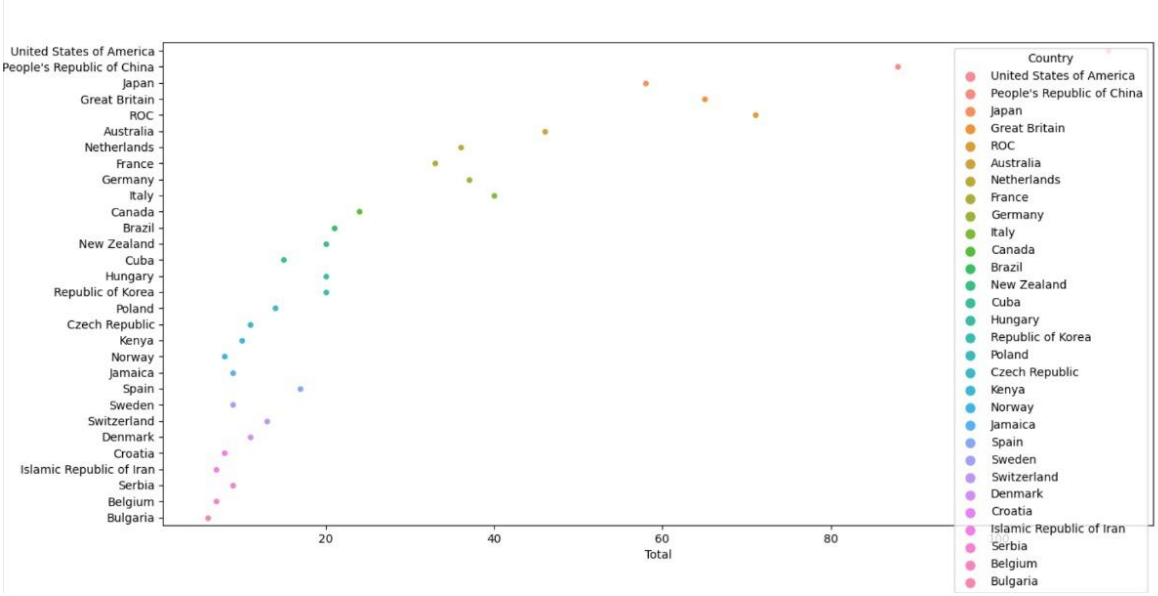
- Scatter Plot of the Data Set :



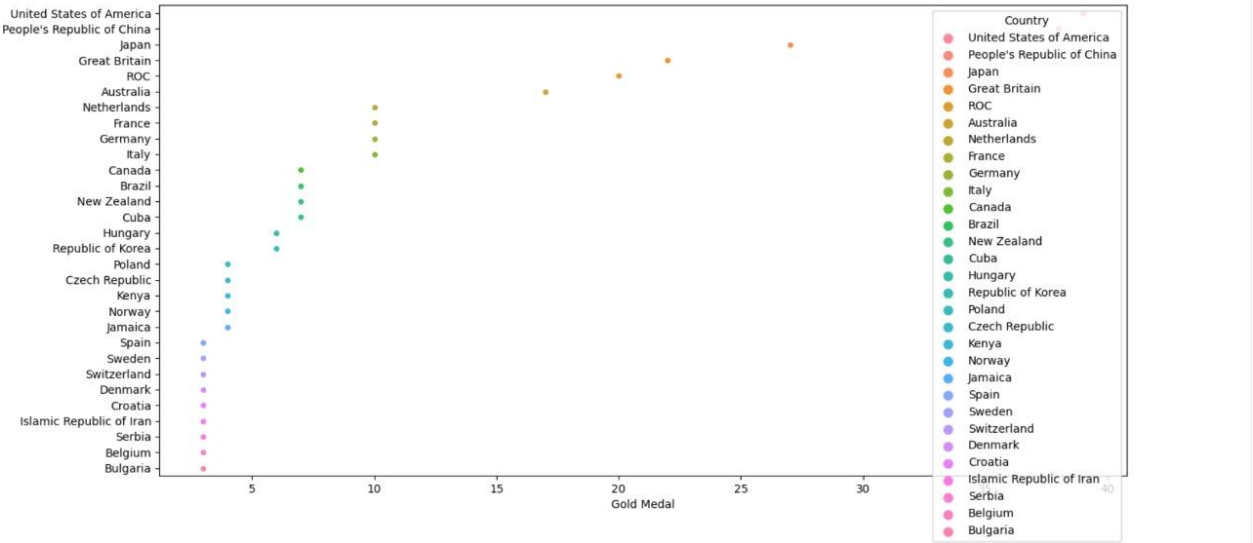
- Correlation heatmap of the Data Set :



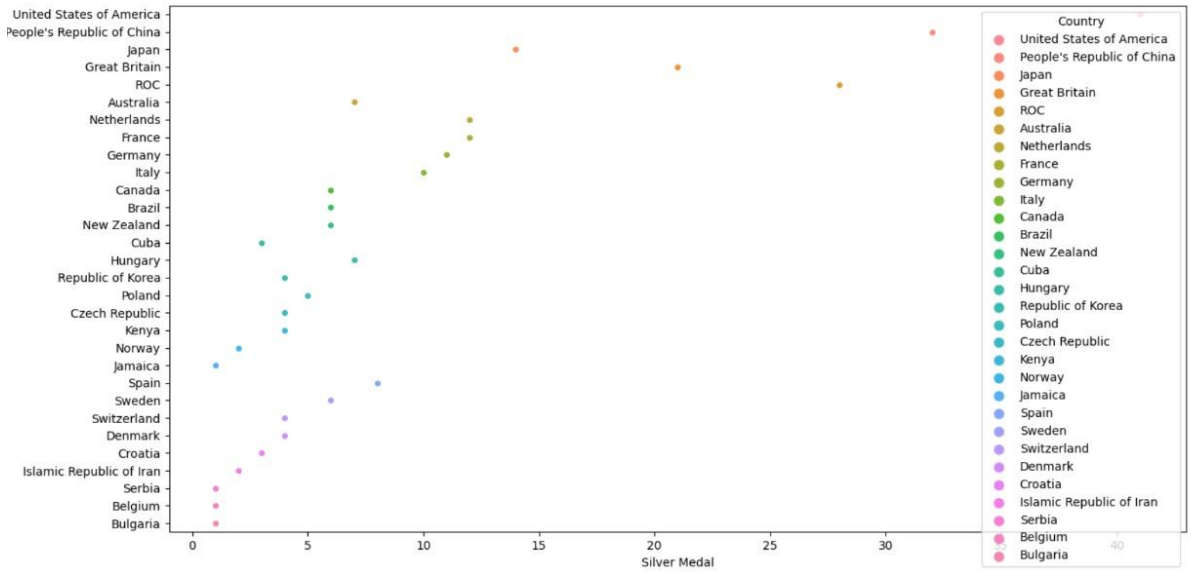
- Swarm plot of total medals :



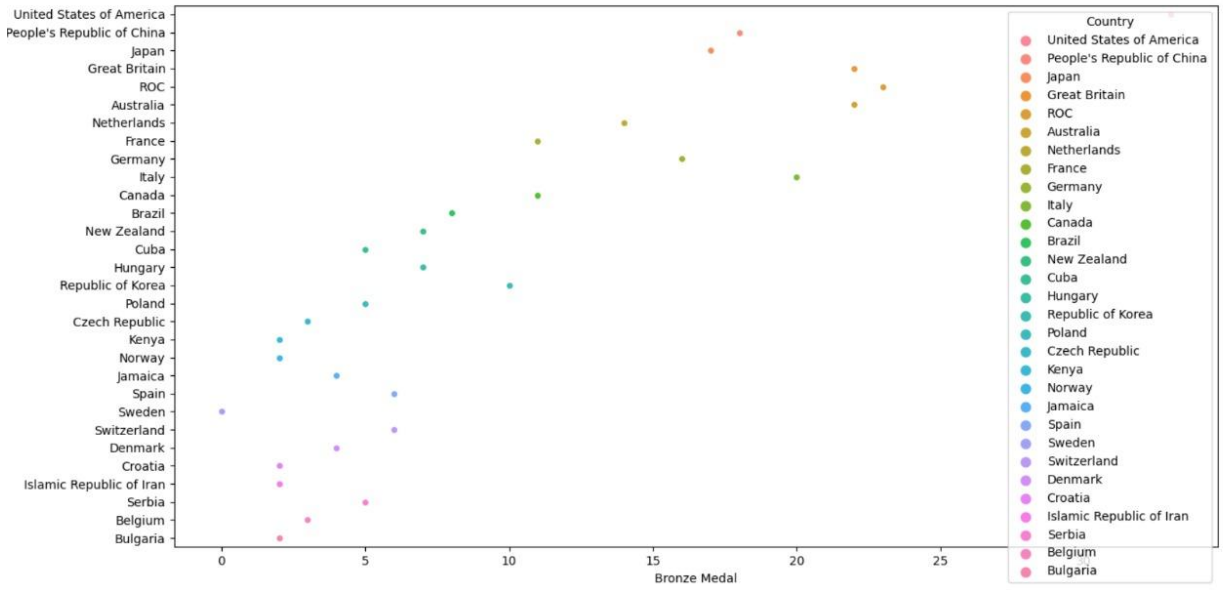
- Swarm plot of Gold Medals :



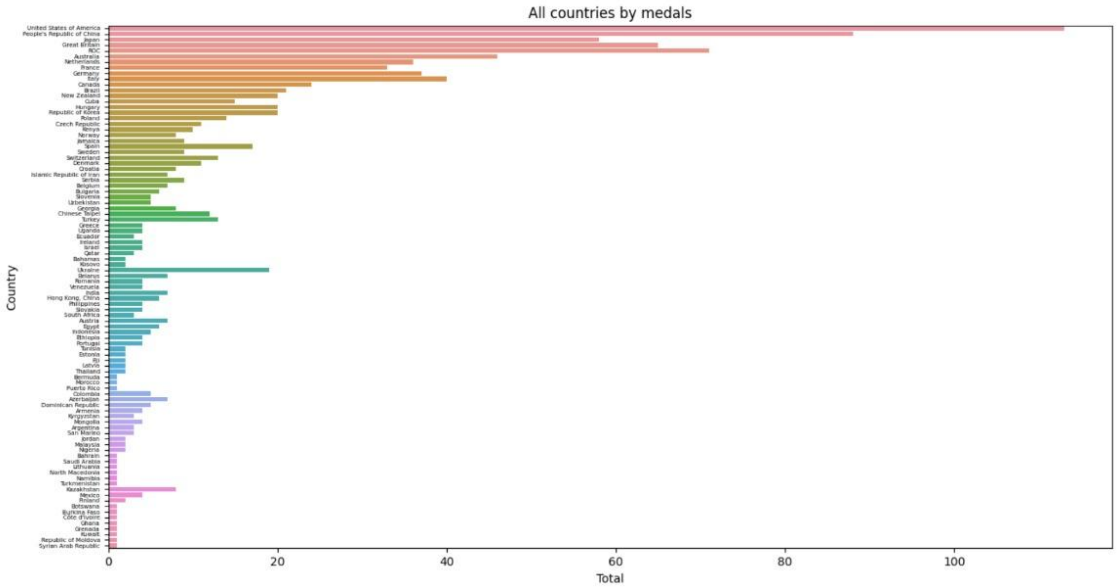
- Swarm Plot of Silver Medals :



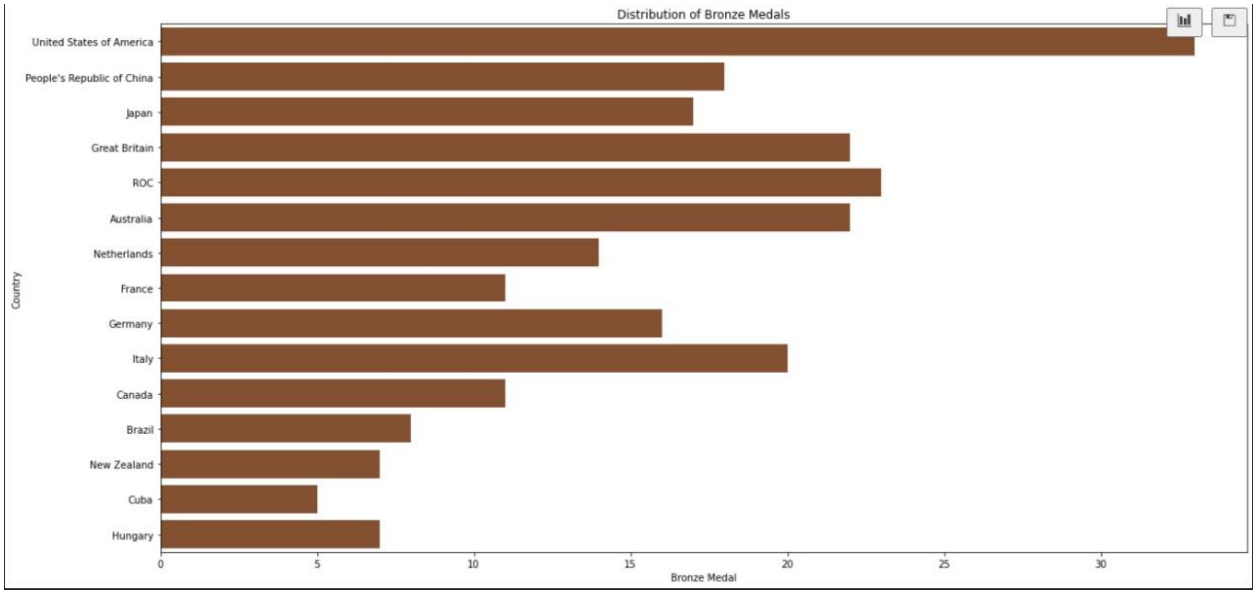
- Swarm Plot of Bronze Medals :



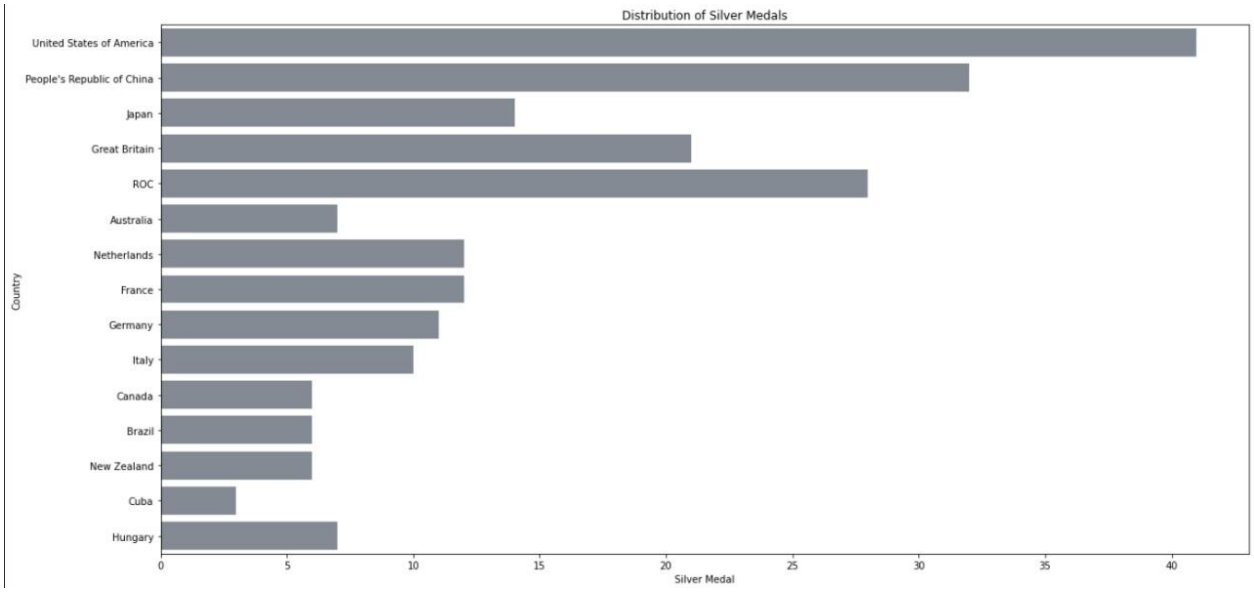
- Horizontal Bar plots of total medals :



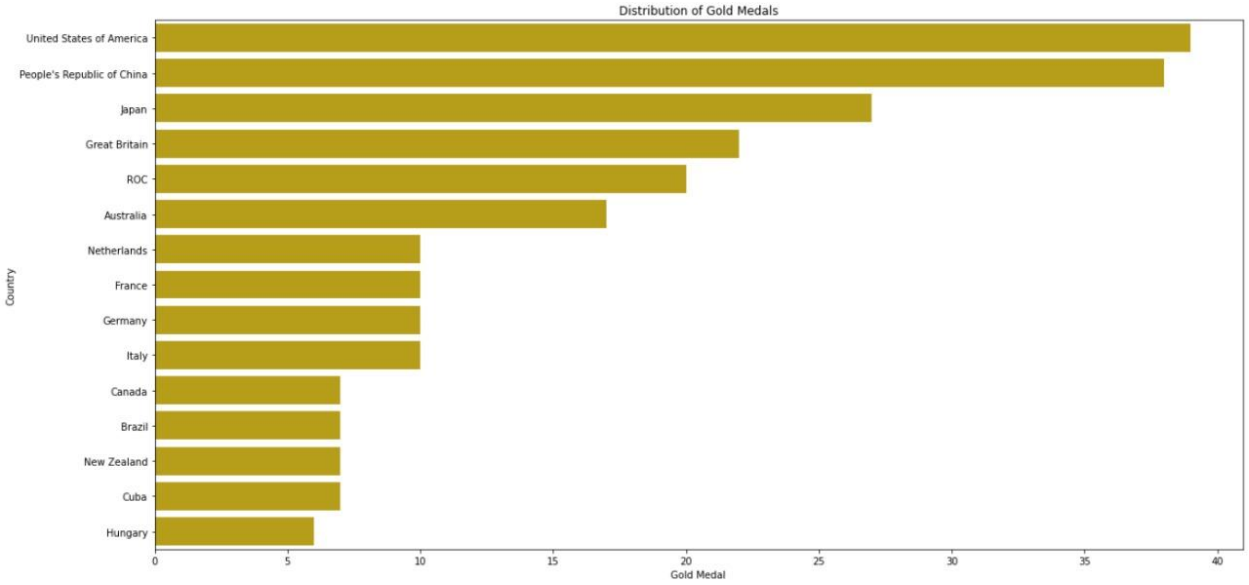
- Distribution of Bronze medal



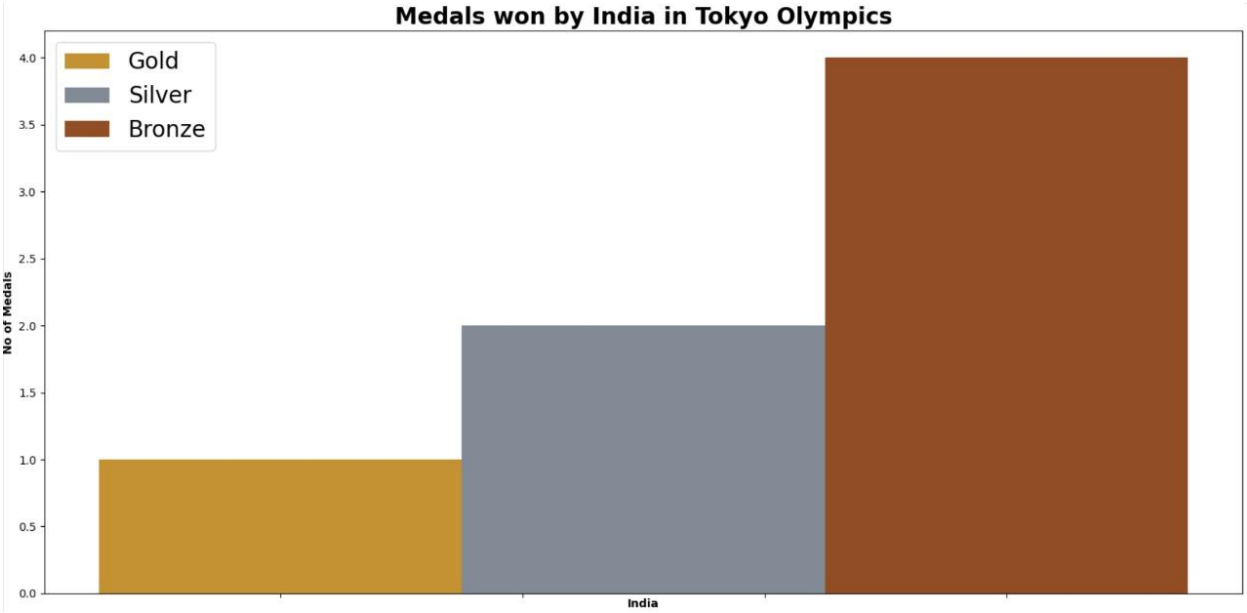
- Distribution of Silver medal



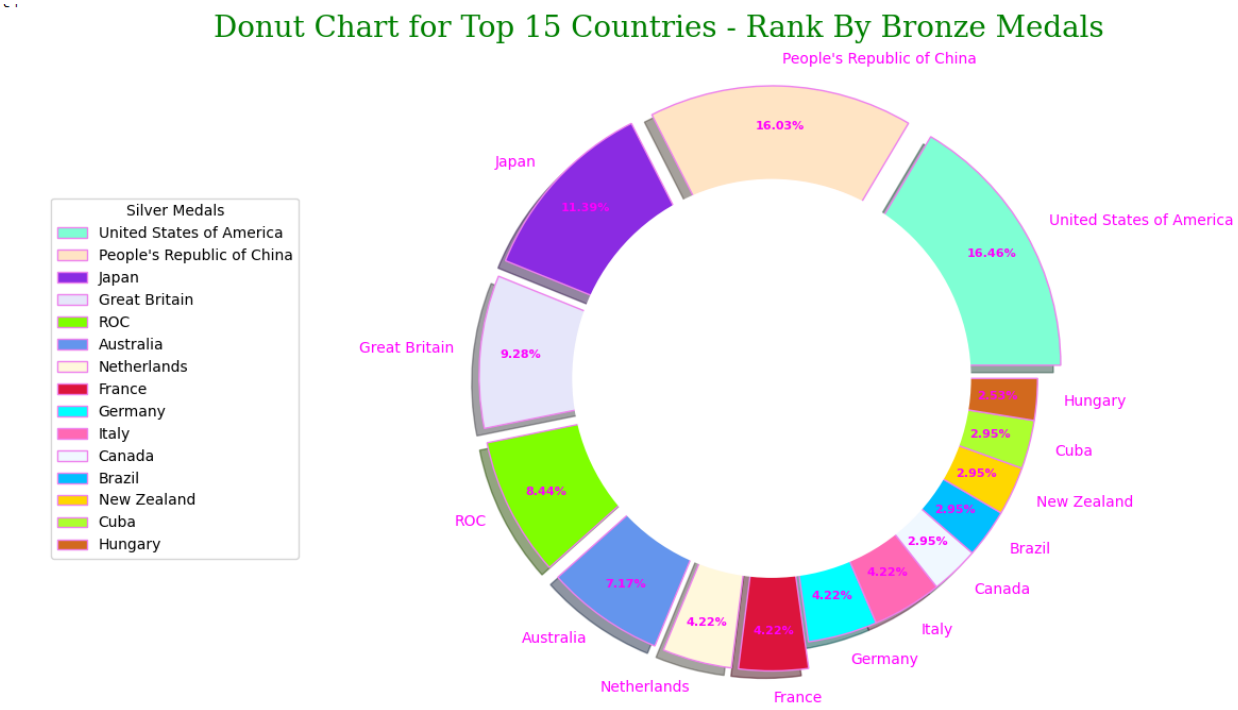
- Distribution of Gold medal



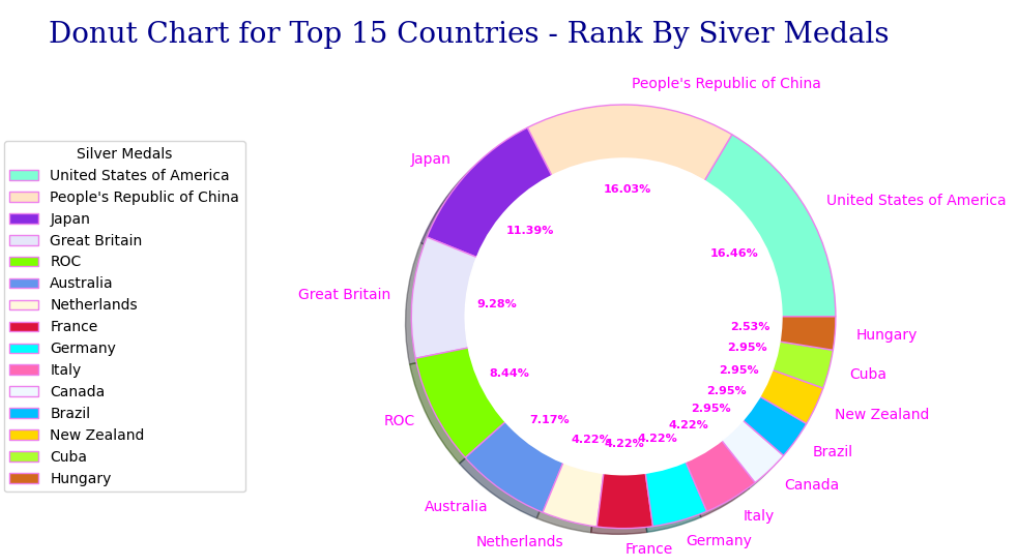
- Medals Won by India In TOKYO 2021



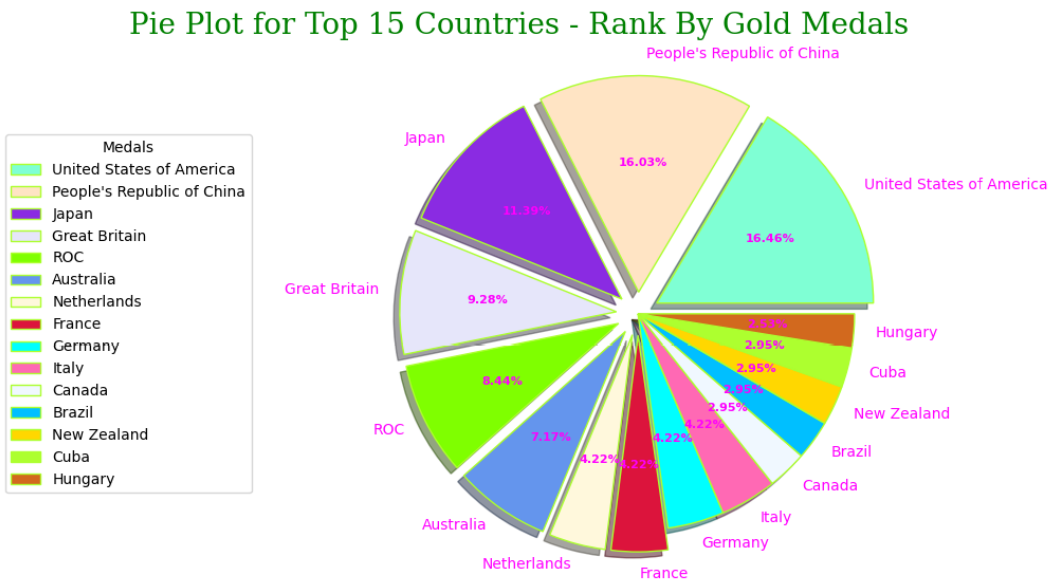
- Donut Chart for Top 15 Countries-Rank by Bronze Medals



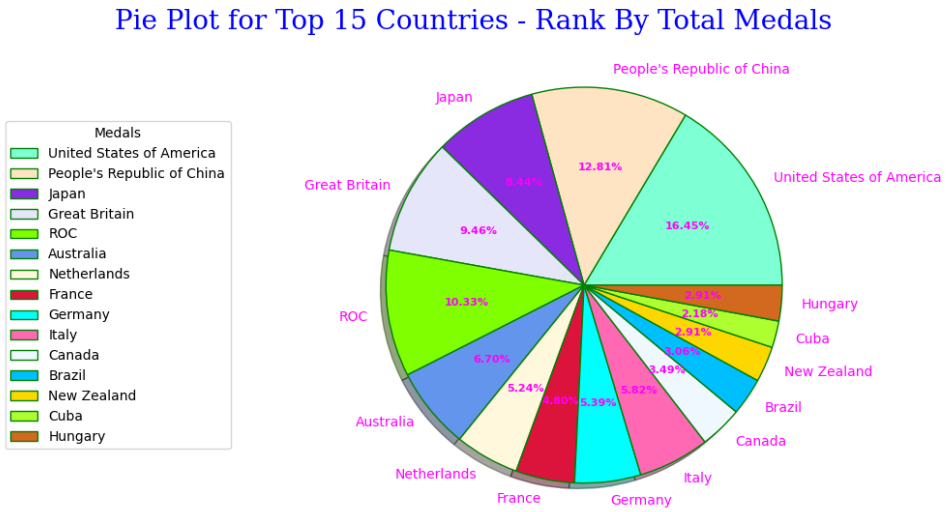
- Donut Chart for Top 15 Countries-Rank by Silver Medals



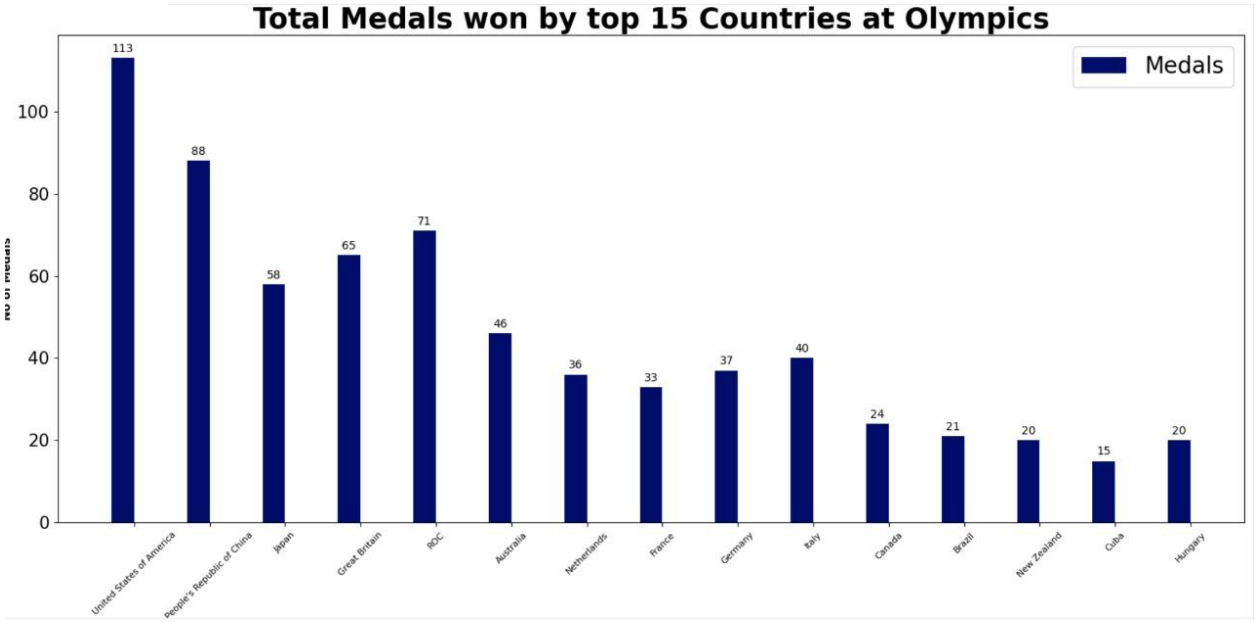
- Donut Chart for Top 15 Countries-Rank by Gold Medals



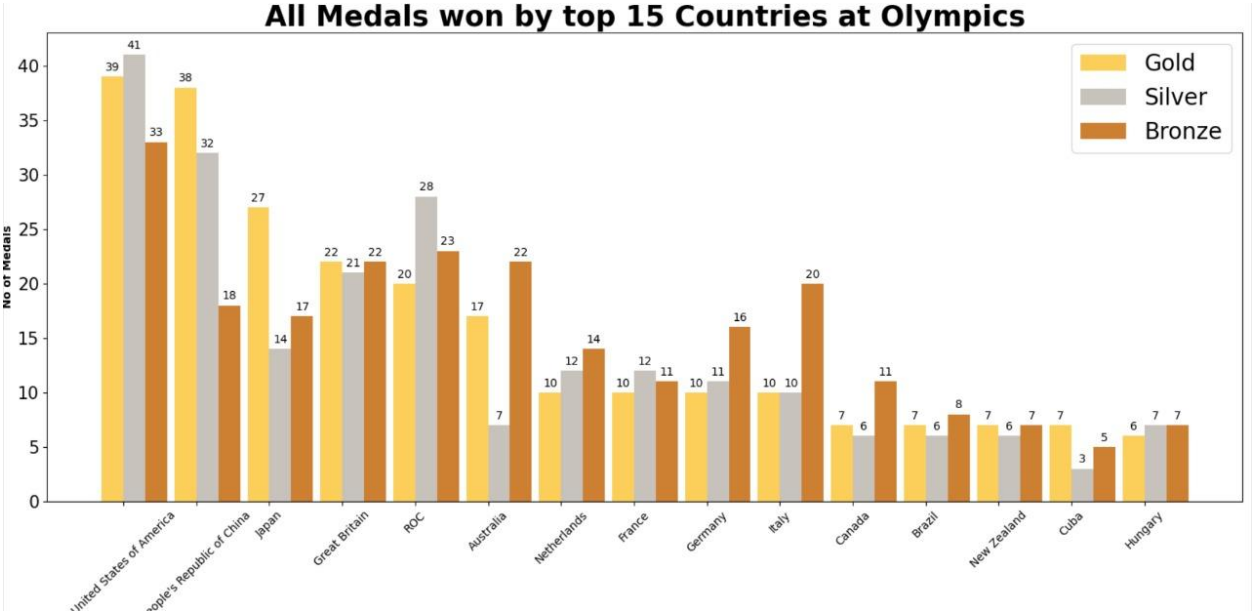
- Donut Chart for Top 15 Countries-Rank by Total Medals



- Total Medals won by top 15 countries at Olympic



- All Medals won by top 15 countries at Olympic



Conclusion

From the Analysis of the Data Set Received, It can be concluded that The Performance of Indian in the Olympics is improved as compared to its past performances in the previous olympics. This time India grabbed 1 Gold in the javelin Throw , 2 Silvers in the Weight Lifting and Wrestling , 4 Bronze in Women's Singles badminton, Women's Welterweight Boxing, Hockey and wrestling. From the data it can be analysed that the US is the most medal winning country, with the highest number of total medals and will do a comparable performance in the next olympics too. But surely as we can see that India is doing really well with an exponential growth in medals won, so we can conclude that India will do the same exponential growth in number of medals in upcoming olympics.

Group Members

- Anshuman Das - BT20CSE129
- Soumalya Roy - BT20CSE058
- Satyam Shukla - BT20CSE007
- Tanmayee Gosavi - BT20CSE011
- Ritika Sharma - BT20CSE026