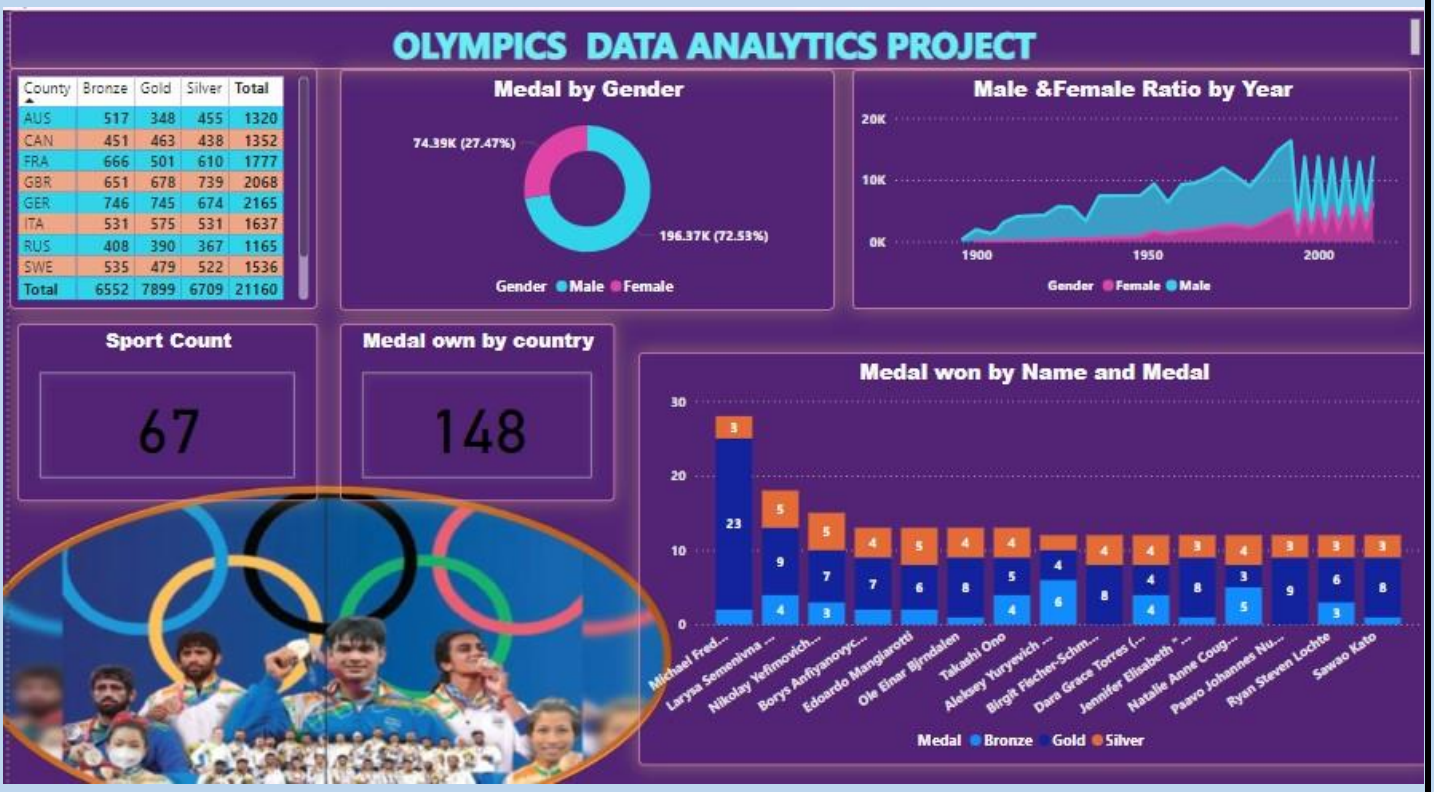
Olympics Dataset Analysis Project 2024 in Power BI



Problem Statement:

Analyse the Olympic 2024 dataset to determine the percentage distribution of medals won by gender (male vs. female) across all sports. The goal is to provide insights into gender representation and achievement in the 2024 Olympics .

1.Medals Won by Gender %

2. Male and Female Ratios by Year

3. Top 10 Countries by Medals

4. Top Olympic Medallists by Total Medals

5. Total Number of Participating Countries

6. Total Number of Sports

Key Performance Indicators (KPIs)

1. **Percentage of Medals Won by Male Athletes:** Measure the percentage of total medals won by male athletes across all sports.
2. **Percentage of Medals Won by Female Athletes:** Measure the percentage of total medals won by female athletes across all sports.
3. **Gender Disparity Ratio in Medals:** Calculate the ratio of male to female medals won to identify gender disparity.
4. **Top Sports for Each Gender:** Identify which sports had the highest medal counts for male and female athletes.
5. **Year-over-Year Gender Medal Comparison:** Compare the gender-based medal distribution in 2024 with previous Olympic years to identify trends.

Olympics Dataset Analysis2024

Understanding data: ***271117*15 columns***

- **ID**: Unique identifier for each athlete.
- **Name**: Athlete's name.
- **Sex**: Gender of the athlete (Male/Female).
- **Age**: Age of the athlete.
- **Height**: Athlete's height.
- **Weight**: Athlete's weight.
- **Team**: The team or country the athlete represents.
- **NOC**: National Olympic Committee code.
- **Games**: Edition of the Olympic Games.
- **Year**: Year the event took place.
- **Season**: Season of the Olympics (Summer/Winter).
- **City**: Host city.
- **Sport**: Sport category.
- **Event**: Specific event within the sport.
- **Medal**: Type of medal won (Gold/Silver/Bronze).

Data Preparation:

1.The Games column was removed as it included both year and season values, while the Height and Weight columns were deleted from the dataset from Athletes sheet

2. In the second sheet, the unnecessary `Notes` column was deleted, and the first row was set as the header to ensure proper data alignment and clarity.

3. To merge the `NOC` columns from both the `Athletes` and `Country Definitions` sheets, use a full outer join in Power Query Editor. This will integrate all records from both tables, preserving unmatched rows and combining relevant information.

4. After merging, delete the first `NOC` column from the `Athletes` sheet and retain the second `NOC` column to ensure consistency with the merged data.

5. Rename the Sex column to Gender, Rename the NOC column to Country. Rename the Region column to Region hen, use **Replace Values** to standardize the Gender column with 'Male' and 'Female'.

6. Check for missing values, in the `Country` column, and if a null value is found, handle it by removing the corresponding row from the dataset.

Olympics Dataset Analysis2024

Parimal A_Data Analyst