Olympic Analysis by Pukhraj Garasiya

Problem Statement:

1. Medals won by gender.

2. Male and female participation ratio by year.

3. Top 10 countries by total medals won.

4. Top Olympic medalists by total medals.

5. Total number of participating countries.

6. Total number of sports.

1) Data Transformation:

1.1 **Remove Unnecessary Columns:**

- **athlete_events table:**

    o **Games:** This column contains a combination of years and seasons, which are already available in separate columns. Therefore, we need to remove this column using the "Remove Columns" option.

    o **Age:** The age column is not required, so it should be removed using the "Remove Columns" option.

    o **Height:** Similarly, the height column is not needed and should be removed using the "Remove Columns" option.

- **country_definitions table:**

    o **Notes:** The notes column is not needed for the analysis, so it should be removed using the "Remove Columns" option.

1.2 **Make First Row as Header:**
Set the first row as headers in the country_definitions table using the "Use First Row as Headers" option.

1.3 **Merge Tables:**
There is an NOC column in both athlete_events and country_definitions tables. To avoid redundancy, we merge both tables using a full outer join on the NOC column to consolidate all data into the athlete_events table. This can be done by navigating to "Merge Queries," selecting the NOC column, and choosing the "Full Outer Join" option.

1.4 **Remove Duplicate NOC Column:**
After merging, there will be two NOC columns in the athlete_events table. Remove the duplicate column using the "Remove Columns" option.

1.5 **Rename Columns:**
Rename the columns for better clarity:

- Change country_definition.NOC to Country.

- Change Sex to Gender.
  To rename, right-click the column header and select the "Rename" option.

1.6 **Standardize Data in the Gender Column:**
Replace values in the Gender column:

- Replace "M" with "Male."

- Replace "F" with "Female."
  Go to the "Transform" option and use the "Replace Values" feature.

1.7 **Check for Null Values:**
Check for null values in the Country column. Exclude rows with null values as they are not required for analysis.

Solution:

1. **Medals Won by Gender:** Use a pie chart to visualize this data.

2. **Male and Female Participation Ratio by Year:** Use a stacked area chart to display the ratio over time.

3. **Top 10 Countries by Medals:** Use a matrix to present the top 10 countries by total medals.

4. **Top Olympic Medalists by Total Medals:** Use a stacked column chart to show the medalists with the highest totals.

5. **Total Number of Participating Countries:** Display this information using a card visual.

6. **Total Number of Sports:** Use a card visual to show the total number of sports.