General Sri John Kotelawala Defence University

Faculty Of Management, Social Sciences and Humanities

Department of Languages

BSc in Applied Data Science Communication

Intake 40

A.G.A.U.S.Gunasekara - D/ADC/23/0034

Y.G.G.D. Ranasinghe - D/ADC/23/0025

M.L.B.T.S.Perera - D/ADC/23/0047

Year 02: Semester 04

26/08/2024

Advanced SQL and Cloud Databases

Semester 04 - Assignment 1

Assessment Title: Developing a Power BI Dashboard and Microsoft Report Builder reports to Analyze Historic Olympic Games.

---

1) Introduction

Olympic Games in the 21st Century

The early 21st century marked an important era for the Olympic Games, characterized by a blend of tradition, innovation, and global sportsmanship. These competitions, which span the years from the 2000 Sydney Olympics to the most recent Olympics in Paris in 2024, demonstrated the height of athletic success and the determination of international competition.

The Olympic movement had a significant growth during this time, with a greater focus on participation and the addition of new countries. Athletes used the Games as a platform to smash records, motivate future generations, and foster international friendship.

We can identify many of the iconic moments, renowned athletes, and enduring legacies that have formed the modern Olympic era as we focus the Olympic Games from 2000 to 2024. Accordingly, based on the official information recorded between the year 2000 and the year 2024, here is a report of an in-depth analysis that we have done on various variables that are different from each other.

About Dashboard:

This Power BI dashboard provides a comprehensive analysis of the Summer Olympics held between the years 2000 and 2016. It offers insights into various aspects of the Games, including:

- Year Representation
- Country Representation
- Gender Representation
- Sports Evolution
- Medal Distribution

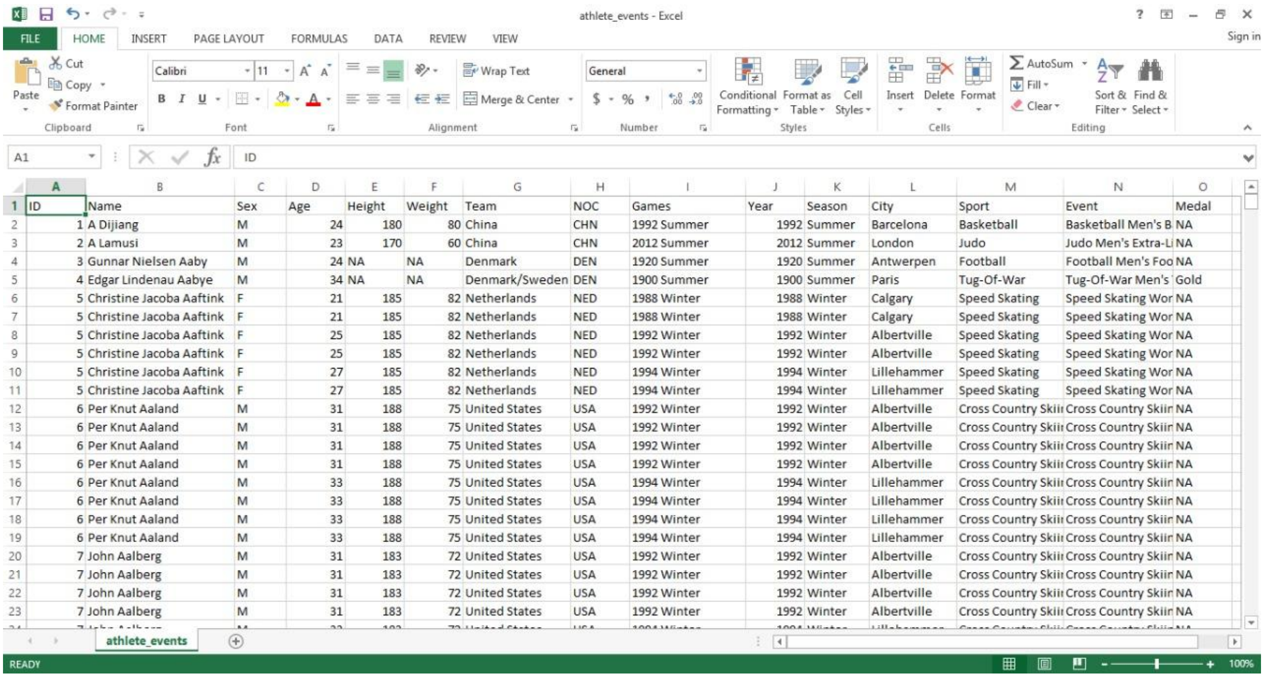
By exploring this dashboard, users can gain valuable insights into the trends and patterns of the modern Summer Olympics. Some potential areas of analysis include:

- Identifying countries that have experienced the most significant growth in Olympic participation and performance.
- Analyzing the impact of gender equality initiatives on the representation of female athletes.
- Identify the evolution of popular sports and the emergence of new Olympic events.
- To understand the factors affecting the success of certain countries in the Olympic Games.

This dashboard serves as a powerful tool for sports analysts, researchers, and enthusiasts to delve deeper into the fascinating world of the Olympic Games and gain valuable insights into the global sports landscape.

2) Exploration of data

• Explanation of the data set



ID	Name	Sex	Age	Height	Weight	Team	NOC	Games	Year	Season	City	Sport	Event	Medal
1	A Dijiang	M	24	180	80	China	CHN	1992 Summer	1992	Summer	Barcelona	Basketball	Basketball Men's B	NA
2	A Lamusi	M	23	170	60	China	CHN	2012 Summer	2012	Summer	London	Judo	Judo Men's Extra-Li	NA
3	Gunnar Nielsen Aaby	M	24	NA	NA	Denmark	DEN	1920 Summer	1920	Summer	Antwerpen	Football	Football Men's Foo	NA
4	Edgar Lindenau Aabye	M	34	NA	NA	Denmark/Sweden	DEN	1900 Summer	1900	Summer	Paris	Tug-Of-War	Tug-Of-War Men's	Gold
5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Wor	NA
5	Christine Jacoba Aaftink	F	21	185	82	Netherlands	NED	1988 Winter	1988	Winter	Calgary	Speed Skating	Speed Skating Wor	NA
5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Wor	NA
5	Christine Jacoba Aaftink	F	25	185	82	Netherlands	NED	1992 Winter	1992	Winter	Albertville	Speed Skating	Speed Skating Wor	NA
5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Wor	NA
5	Christine Jacoba Aaftink	F	27	185	82	Netherlands	NED	1994 Winter	1994	Winter	Lillehammer	Speed Skating	Speed Skating Wor	NA
6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA
6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA
6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA
6	Per Knut Aaland	M	31	188	75	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA
6	Per Knut Aaland	M	33	188	75	United States	USA	1994 Winter	1994	Winter	Lillehammer	Cross Country Skiir	Cross Country Skiir	NA
6	Per Knut Aaland	M	33	188	75	United States	USA	1994 Winter	1994	Winter	Lillehammer	Cross Country Skiir	Cross Country Skiir	NA
6	Per Knut Aaland	M	33	188	75	United States	USA	1994 Winter	1994	Winter	Lillehammer	Cross Country Skiir	Cross Country Skiir	NA
6	Per Knut Aaland	M	33	188	75	United States	USA	1994 Winter	1994	Winter	Lillehammer	Cross Country Skiir	Cross Country Skiir	NA
7	John Aalberg	M	31	183	72	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA
7	John Aalberg	M	31	183	72	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA
7	John Aalberg	M	31	183	72	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA
7	John Aalberg	M	31	183	72	United States	USA	1992 Winter	1992	Winter	Albertville	Cross Country Skiir	Cross Country Skiir	NA

This file we used for analysis contains 271116 rows and 15 columns. Each row corresponds to a single athlete competing in a single Olympic event (athlete event). The columns are:

1. ID - Row Number
2. Name Athlete's Name
3. Gender M or F
4. Age - Integer
5. Height - In Centimeters
6. Weight - In Kilograms
7. Team – Name of the team
8. NOC - National Olympic Committee Letters 3 Code
9. Sports - Year and Season

10. Year Integer
11. Season - Summer or Winter
12. City Host City
13. Sports-Sports
14. Event - Event
15. Medal - Gold, Silver, Bronze, or NA

- Preparation of the data set

Then, this data file was cleaned and processed in a way that is appropriate to our analysis. The basic steps used there were as follows:

- Handle missing values: impute missing data points using methods like imputation or deletion.
- Correct and remove inconsistent data: Find and fix errors, inconsistencies, or outliers in the data.
- Data Formatting and Standardization: Make sure the data is formatted consistently by converting dates or standardizing units.
- Feature Engineering: Develop new features or modify existing ones to enhance the performance of the model.

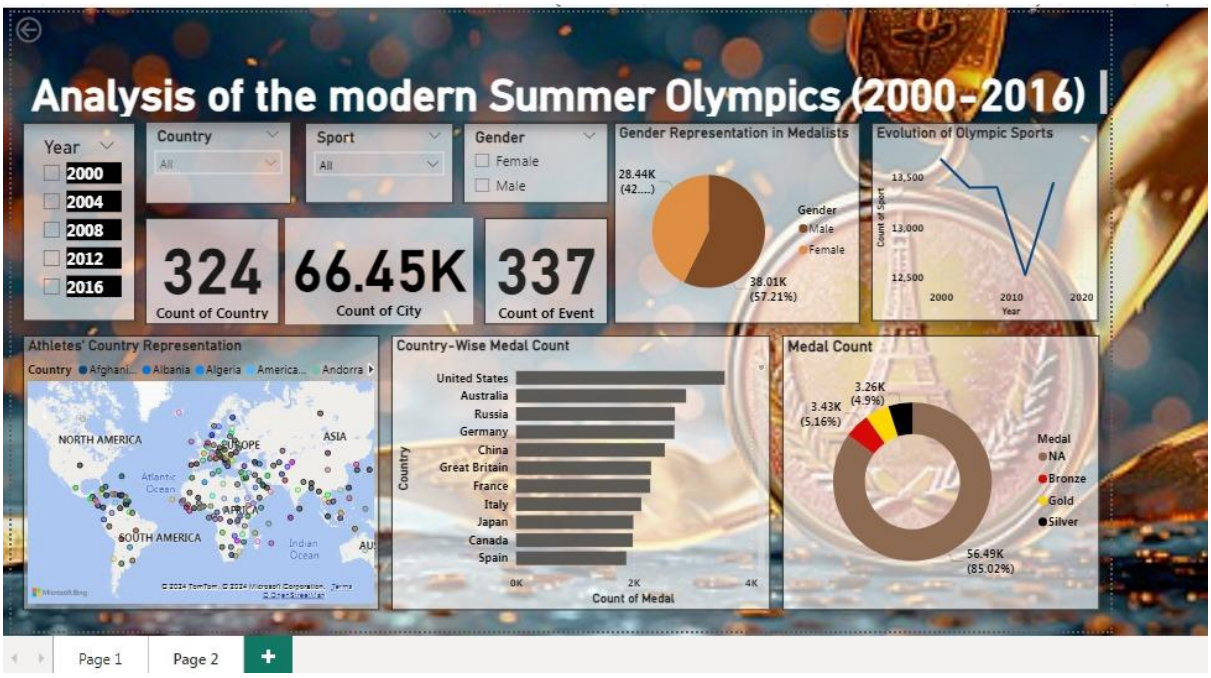
Accordingly, in the Data cleaning and processing step, Data cleaning was done as follows.

1. **Removing the column named "Name".**
2. **Changed the name of the column named "sex" to "Gender" and changed the value of "M" to "Male" and the value of "F" to "Female".**
3. **Changing the name of the column named "Team" to "Country".**
4. **Filtering the years between 2000 and 2024 into only 05 categories in the column named "Year".**
5. **Removal of columns named "NOC" and "Games".**
6. **Filtering only the section labeled "Summer" in the column named "Season".**
7. **Removed "NA" (not available) values in all columns except column named "Medal".**
8. **The first column of the dataset named "ID" is arranged in ascending order.**

ID	Gender	Age	Height	Weight	Country	Year	Season	City	Sport	Event	Medal
707	Male	24	180	83	Slovenia	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
1855	Male	28	172	65	Oman	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
1959	Male	28	172	69	Oman	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2041	Male	26	164	60	Oman	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2127	Male	28	175	70	Saudi Arabia	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2165	Male	28	174	68	Oman	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2193	Male	20	170	65	Saudi Arabia	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2194	Male	18	170	65	Saudi Arabia	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2579	Male	23	170	69	Trinidad and Tobago	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2601	Male	29	185	84	Greece	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
2760	Male	24	187	75	Nigeria	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
3663	Male	24	172	79	Ghana	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
4618	Male	28	175	71	Nigeria	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
5456	Male	28	179	76	Japan	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA
5962	Male	20	180	75	Mauritius	2000	Summer	Sydney	Athletics	Athletics Men's 4 x 100 metres Relay	NA

3) Dashboard design and Implementation

- Dashboard draw up

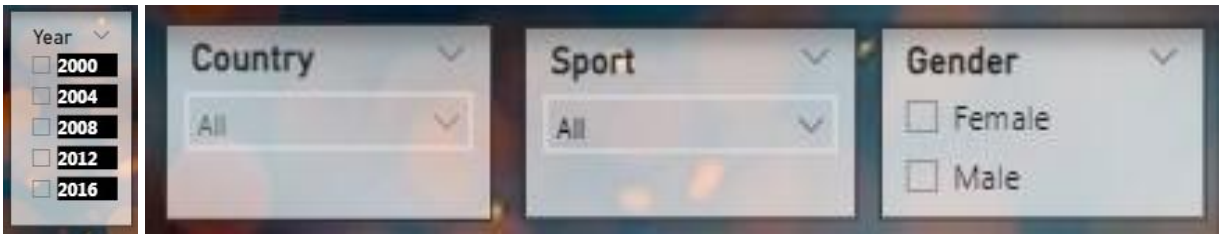


- Above is the final power BI dashboard we got as a result.

The dashboard is divided into two pages and as shown here, the first page includes a simple description of what's inside with a welcome.

Let us see a detailed explanation about the analysis part of the dataset.

This five-year filterable Slicer allows the entire dashboard to be filtered by the same relevant year. Assuming that the year 2012 is chosen as an example from the years 2000, 2004, 2004, 2008, 2012, 2016, then the number of women and men who have participated from each country in relation to that year will be represented in percentage and quantity, and the number of events in the festival will be represented in relation to the year. , showing the number of countries filtered for analysis and the size of cities, showing the evaluation of each other's games in the time period from 2000 to 2020, depicting the representation of each other's countries with the help of a world map, as well as showing the form of victories obtained in comparison to each other's countries in a three-dimensional graph and as a whole The percentage of medals won in the respective year is also displayed on this dashboard in a colorful way based on a pie chart.



When the dashboard is taken as a single unit, the following Slicers fulfill the following requirements.

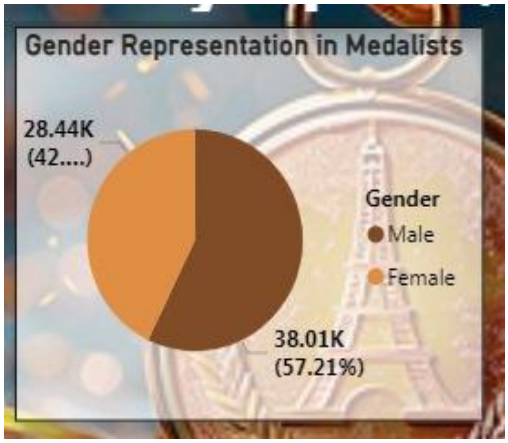
Year Slicer: Allows users to filter by year to focus on specific time periods.

Country Slicer: Allows users to filter the map by region (North America, Asia, etc.) to focus on specific geographical areas.

Sport Slicer: Allows users to filter by sports to focus on specific sports.

Gender Slicer: Enables filtering by year to analyze changes in gender representation among athletes.

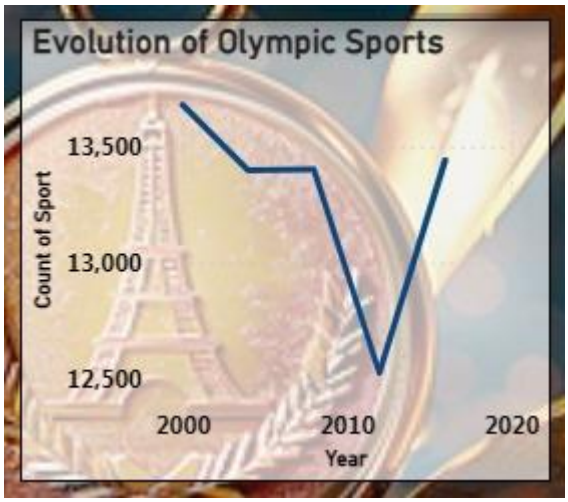
Here are some cards to represents the number of countries, number of cities and number of events in relation to the values applied to the filters according to the requirements above.



This is a pie chart illustrating the gender representation among medalists.

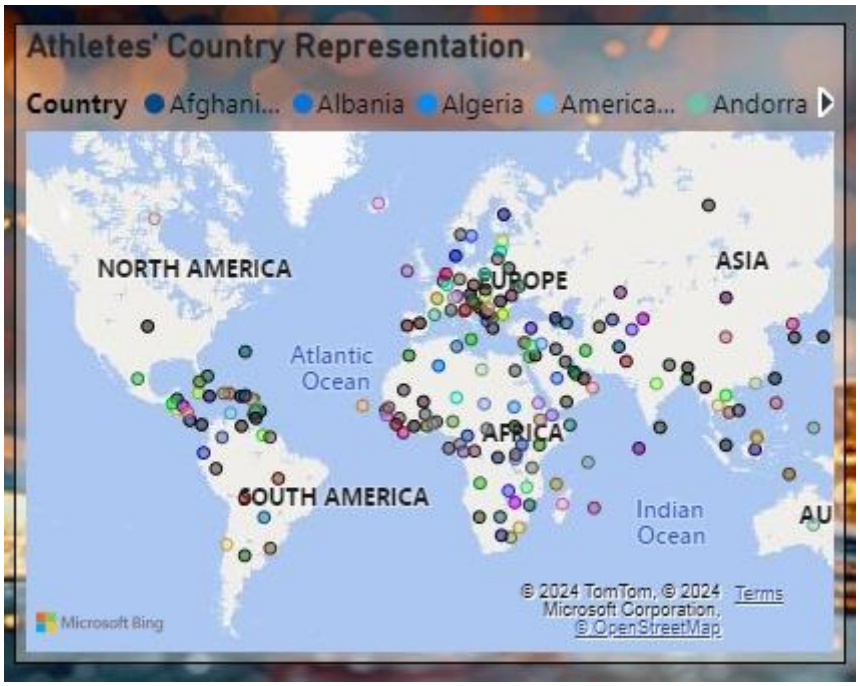
- The chart is divided into two segments, one for female medalists and one for male medalists.
- The female segment is represented in a lighter orange color, while the male segment is in a darker brown color.
- Each segment is labeled with the corresponding gender, the number of medalists, and the percentage of the total medalists.

This value will vary analytically depending on the requirements related to the filters above.



This line graph "Evolution of Olympic Sports."

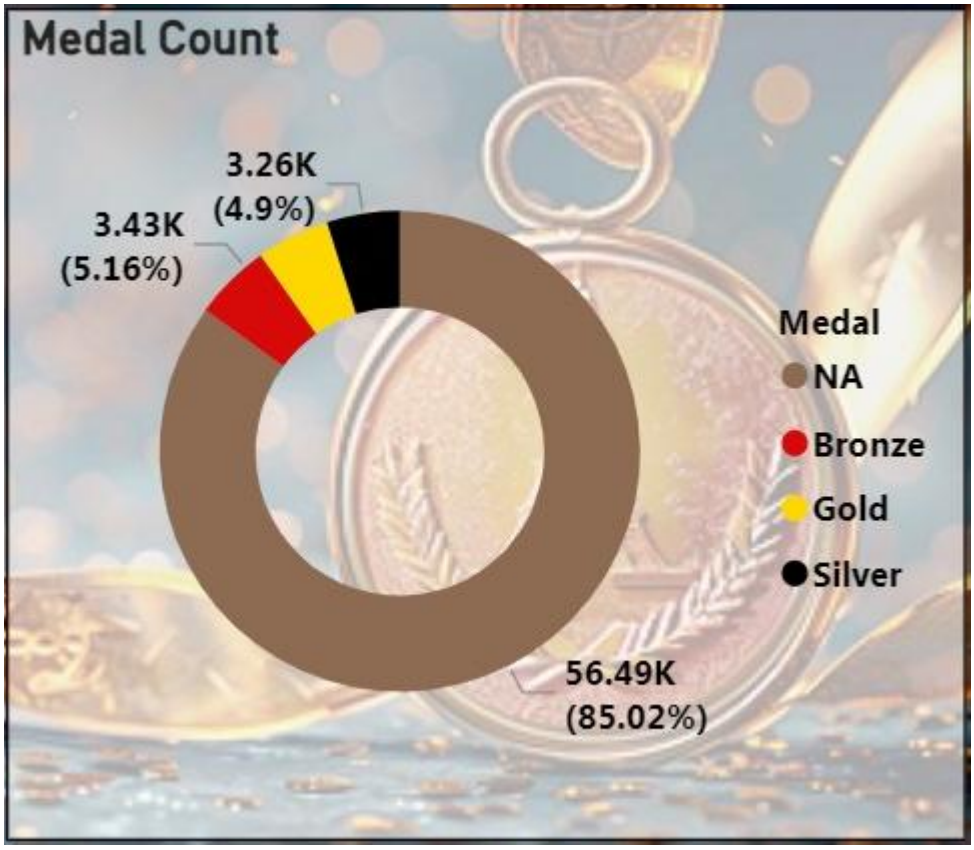
- The x-axis represents the years from 2000 to 2020, and the y-axis represents the count of sports. The graph depicts a fluctuating trend in the number of Olympic sports over time.



This is a world map that visually represents the countries participating in the Olympic Games. Each country is marked with a colored dot, and the size of the dot may indicate the number of athletes representing that country.

- **Global Participation:** The map shows that athletes from a wide range of countries participate in the Olympics, spanning across continents.
- **Regional Clusters:** There appear to be clusters of dots in certain regions, such as Europe, North America, and Asia, suggesting that these regions have a higher concentration of participating countries.
- **Variations in Dot Size:** The size of the dots may indicate differences in the number of athletes sent by each country. Larger dots could represent countries with higher athlete delegations.
- **Emerging Nations:** The map might reveal the participation of emerging nations in the Olympics, indicating a growing global reach for the event.
- **Historical Trends:** By comparing this map with data from previous Olympics, it would be possible to identify changes in participation patterns over time.

This is a horizontal bar chart showing the medal count for different countries. The x-axis represents the count of medals, while the y-axis lists the countries.



This shows a donut chart representing the distribution of medals in an Olympic competition. The chart is divided into four segments, each representing a different type of medal.

---

4) Implications & Limitations

Implications:

• Gaining insight into Global patterns: The dashboard offers insightful information about the global patterns surrounding the Olympic Games, including the rise in the number of participating nations and the increasing number of female participants.

• Identifying Dominant Nations: For analytical and comparison purposes, the dashboard shows the nations that consistently placed highly in the Olympics.

• Evaluating Gender Equality: The gender representation visualizations provide insightful data that can be used to gauge the Olympic movement's progress toward gender equality.

• Inform Policy Decisions: Decisions about international cooperation and sports development can be influenced by the statistics on the dashboard.

Limitations:

• Data Restrictions: Only data from the years 2000–2016 is available on the dashboard. A more current knowledge of the trends could be obtained by analyzing more recent data.

• Data Granularity: While offering a broad summary of the Olympics, the dashboard might not go into detail about particular sports, occasions, or athletes.

• Context and Interpretation: A thorough grasp of the variables affecting Olympic success, such as the political and economic environments as well as the infrastructure supporting sports, is necessary for the interpretation of the data.

• Limitations of Power BI: The extent of analysis and the kinds of visualizations that can be produced may be restricted by Power BI's features.

5) Recommendations

Recommendations:

- **Expand the Data Range:** Incorporate data from more recent Olympic Games to provide a more comprehensive analysis.
- **Increase Data Granularity:** Include data at a more granular level, such as individual athlete performance or specific sports.
- **Explore Advanced Analytics:** Utilize advanced data analysis techniques, such as machine learning or statistical modeling, to uncover deeper insights.

The Power BI dashboard can be even more helpful for comprehending the current Olympic Games and guiding future research and decision-making by resolving these issues and broadening the analysis's scope.

6) Conclusion

This analysis's Power BI dashboard provides a thorough rundown of the Summer Olympics from 2000 to 2016. Through the analysis of significant indicators including national participation, gender parity, medal allocation, and sports development, the dashboard offers insightful information on the tendencies and patterns influencing the contemporary Olympic movement. Lastly, we can state that the rising worldwide reach of the Olympics is shown in the consistent rise in the number of participating nations throughout time. Gender equality is still a work in progress, as female athletes are still underrepresented in several sports.

The medal count is dominated by some nations, which emphasizes the significance of elements like government funding, sports infrastructure, and athletic development initiatives. New sports that represent changing cultural trends and societal preferences have been included for the Olympic Games.

7) References

https://figshare.com/articles/dataset/Olympic_history_longitudinal_data_scraped_from_www_sports-reference_com/6121274