**OLYMPIC DATA SET ANALYSIS**

**POWER BI ANALYSIS (STEP BY STEP)**

In this project, we explore how to analyze a large dataset of Olympic data with PowerBI. We walk through the process of uploading the dataset, exploring its relationships, and analyzing the data for insights.

For this project, we have received the 11 csv files (city, events, games, games competitor, sports, medals etc). To do the analysis first of all I have uploaded these data to the PowerBI through get data tab. Then to see the relationship between the 11 table, I have used data model view. As there were so many tables hence there was no organized relationship build between them.

**Data Modeling in PowerBI**Data modeling is a critical step in analyzing large datasets. In this project, we use PowerBI's data model view to organize and visualize the relationships between all 11 csv files. This powerful tool allows us to see all the data in one place and to extract deeper insights.

**Streamlined Approach**

The data model view provides a streamlined approach to organizing complex data structures, making analysis more efficient and accurate**.**

**Visual Representation**

The data model is represented visually, which makes it easy to understand the relationships between the tables.  
  
**Easy To Navigate**

This tool is user-friendly and makes it easy to navigate through complex relationships within dataset.

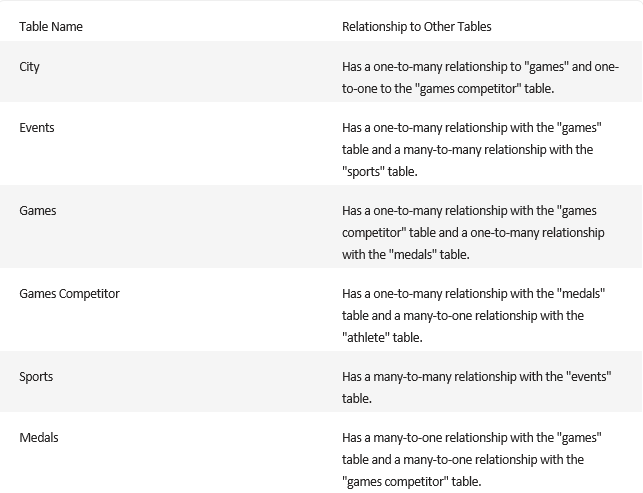
**Uploading Data in PowerBI**

Uploading data is an essential step in analyzing data in PowerBI. In this project, we explore how to upload 11 csv files to PowerBI.

---- To upload the data, navigate to the "Get Data" tab and select "Text/CSV". Then, select the csv file you want to upload.

**Exploring the Relationship Between the Tables**

Understanding the relationship between tables is the key to unlocking insights in complex data. In this project, we explore the relationships within the Olympic data tables.



**Identifying the Lack of Organized Relationship**

Before we can analyze data, we must identify the lack of organized relationships between tables. In this project, we identify and explore the lack of organized relationships within the Olympic data tables.

**No Clear Relationship**

**1**

The Olympic data tables had no clear, organized relationships, making it difficult to perform in-depth analysis.

**2**

**Identifying the Issue**

The data modeling view helped us to identify the problem and visualize the relationships between the tables.

**Resolution**

**3**

By understanding the relationships, we were able to build a data model and perform more advanced analysis.

Hence, I made three small tables by merging different tables (fuller outer join). Names of the table Games Table (Games competitor -> Games City -> City -> Games), Event Table (Competitor event -> Event -> Medal -> Sports) and Person Table (Noc region -> Person -> Person Region).  
  
**Games Table: A Nexus of Competitors, Cities, and Games**

A comprehensive table merging Competitor, Games City, City, and Games data.

**Benefits**

**Column Details**

Provides a holistic View of games and city associated, games year and id. Enables analysis of competitor participation across different games.

Competitors, Games, City, Games Id, Games Name, Games year etc.

**Event Table: Unveiling the Link between Competitor, Event, Medal, and Sports**

A cohesive table merging Competitor, Event, Medal, and Sports data.

**Benefits**

**Column Details**

Allows analysis of medal won by competitors in different events.  
Enable analysis of sports performance.

Eventid, Event Name, SportsId, Sports Name, MedalId, Medal Name etc.

**Person Table: Connecting Noc Region, Person, and Person Region**

An integrated table merging Noc Region, Person, and Person Region data.

**Benefits**

**Column Details**

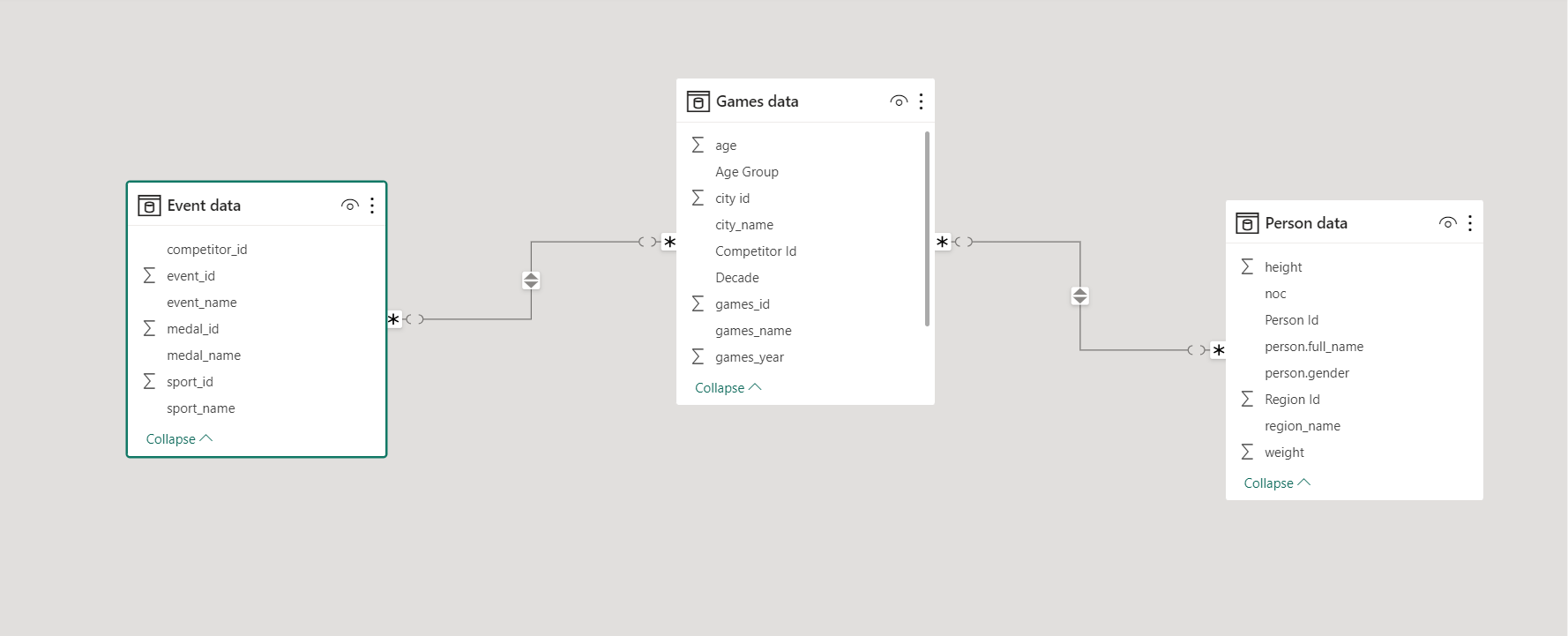
Facilitates analysis of individuals based on their gender, personid and regional affiliations.

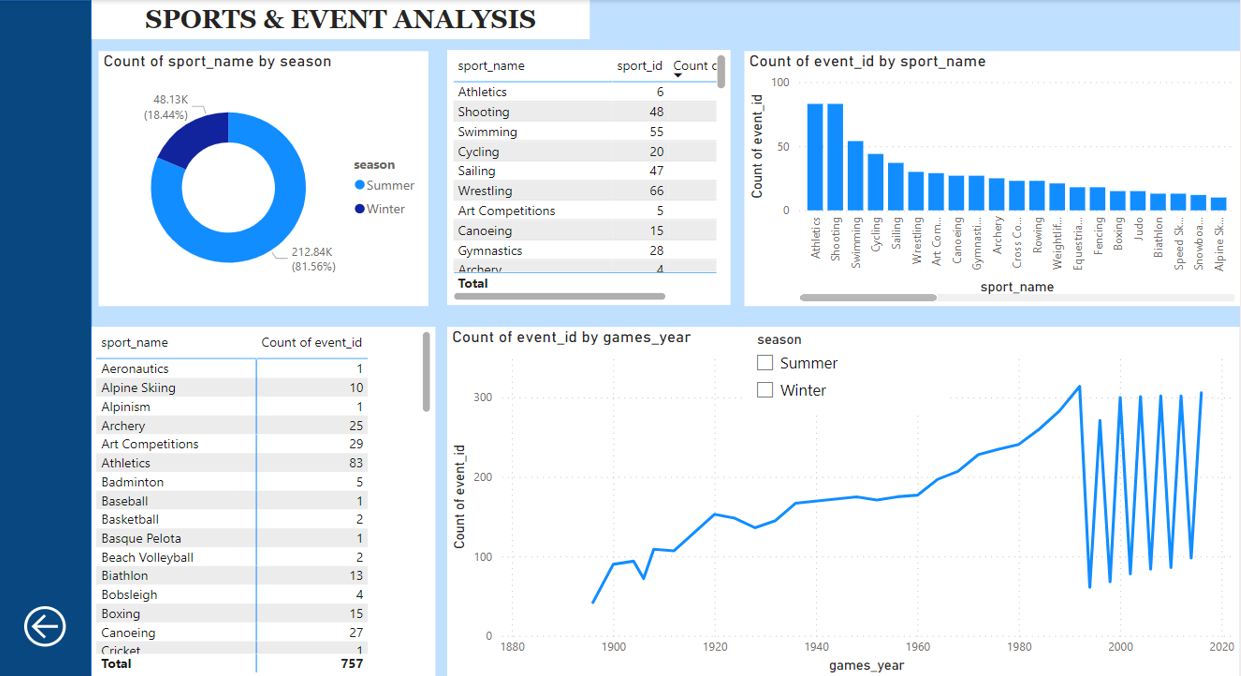
Person Name, Person Id, Gender, RegionId, Region Name etc.

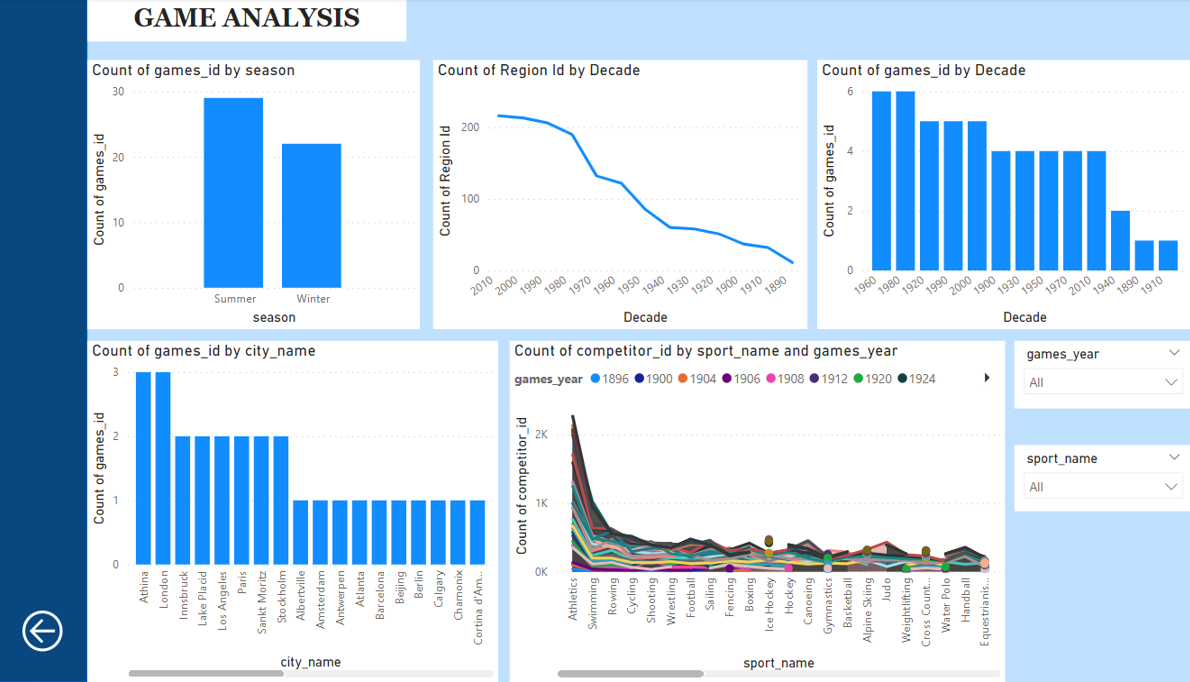
**Conclusion**

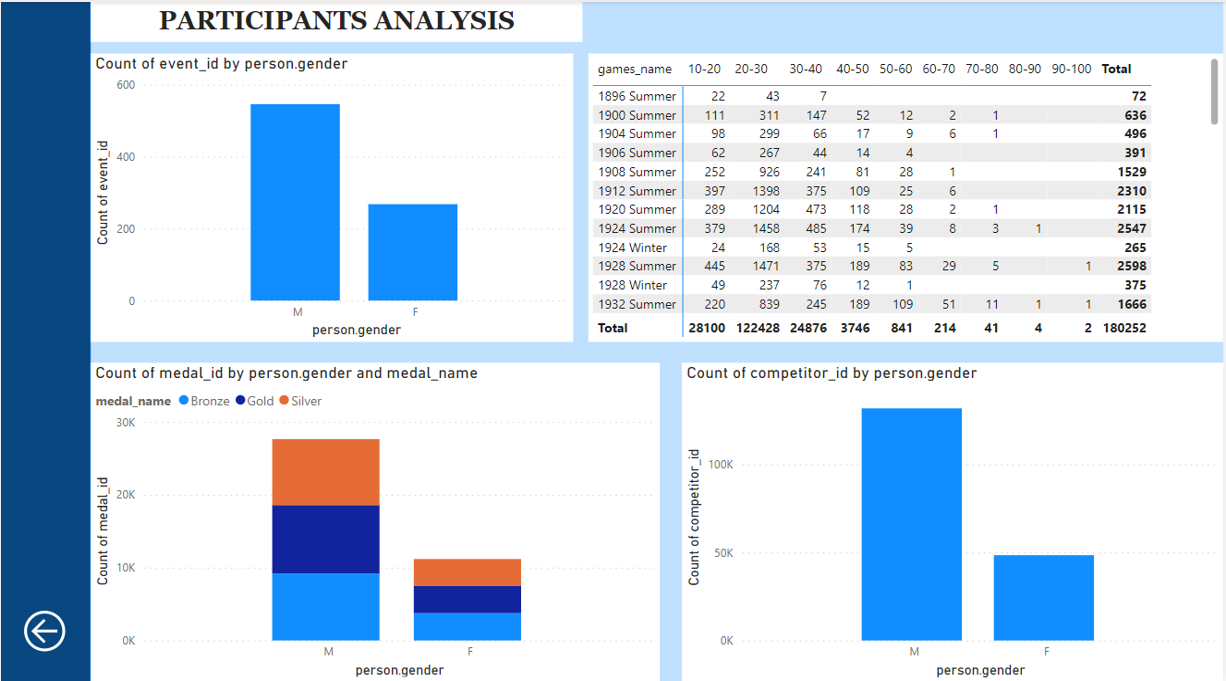
By merging tables using a Fuller Outer Join, we have unlocked valuable insights from the Games Table, Event Table, and Person Table. These merged tables provide a deeper understanding of the connections between competitors, events, medals, cities, regions, and sports. Whether it's analyzing competitors' performance or studying the geographical distribution of games, table merging enhances our ability to interpret and utilize data effectively.

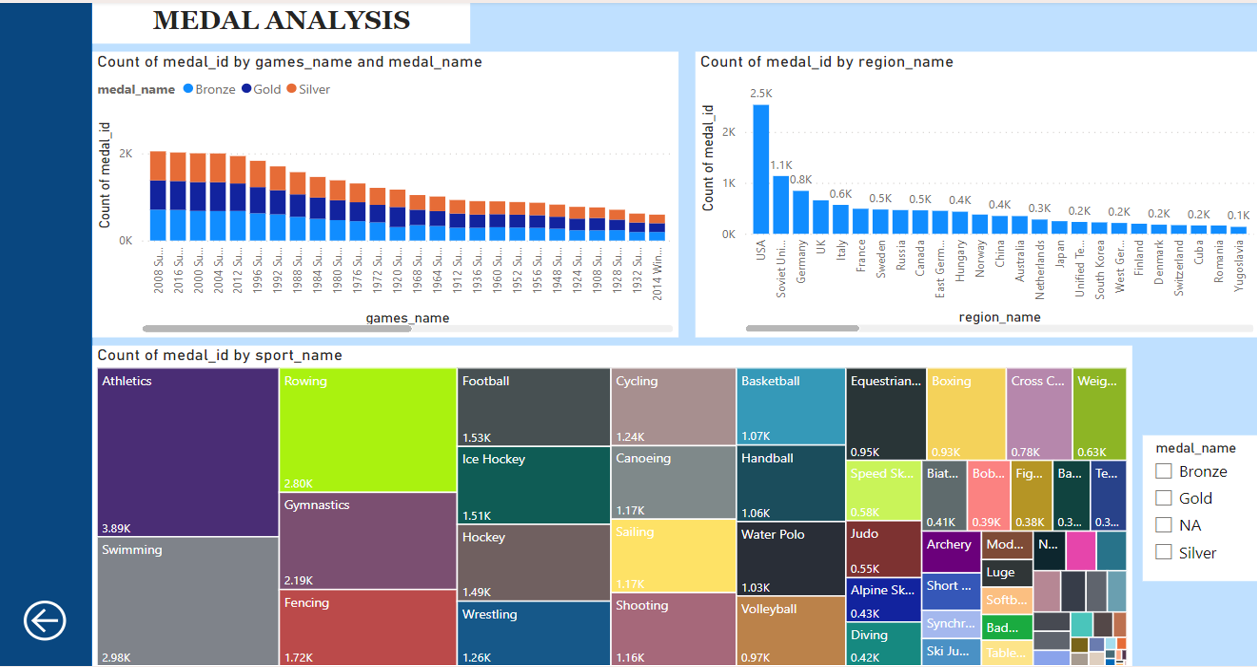
* **Showing the snapshots of PowerBI analysis dashboard for your reference.**

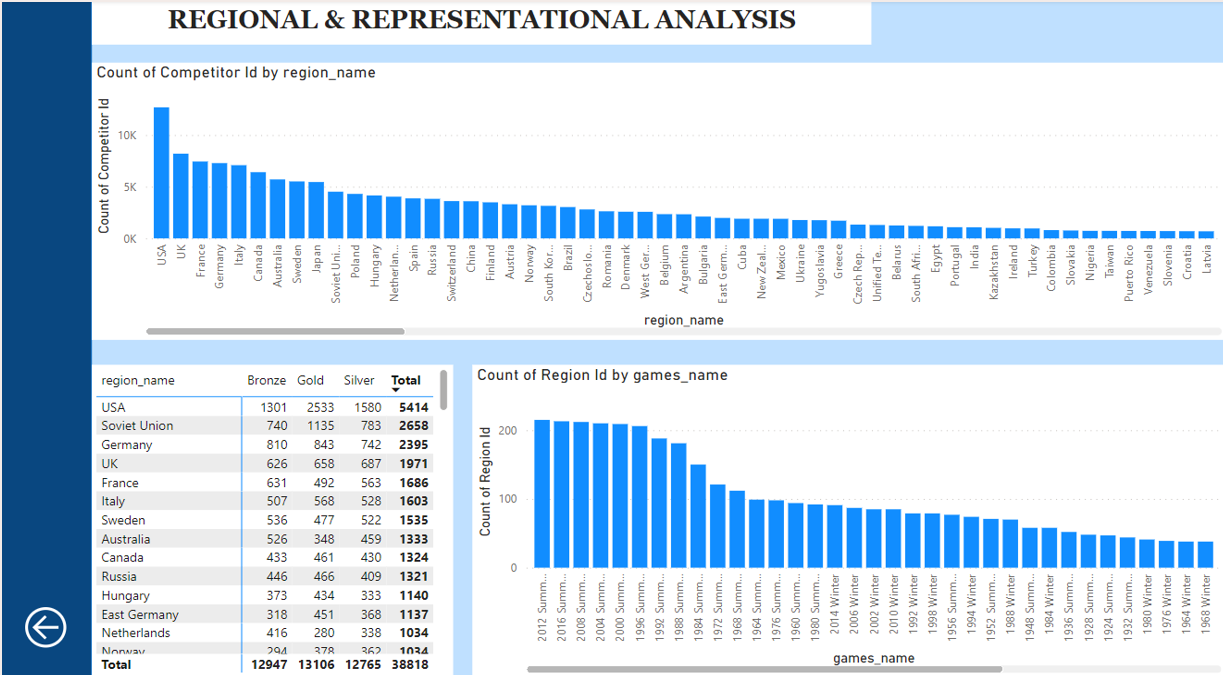




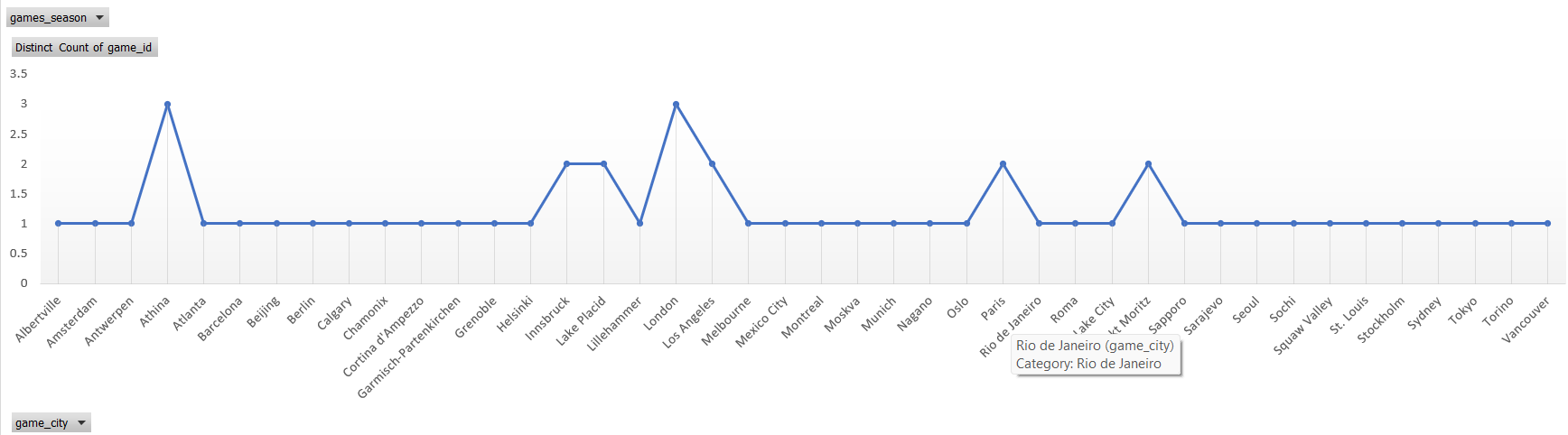




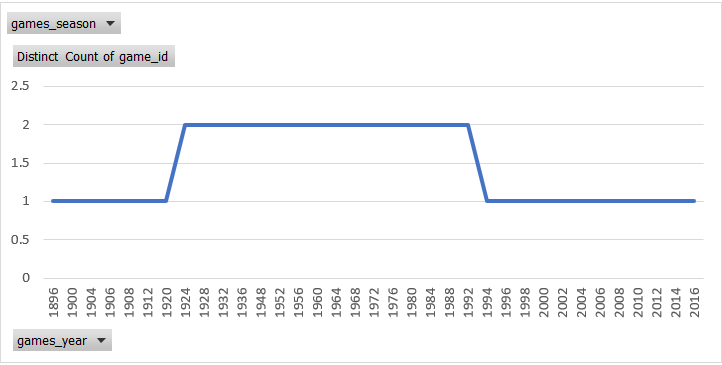




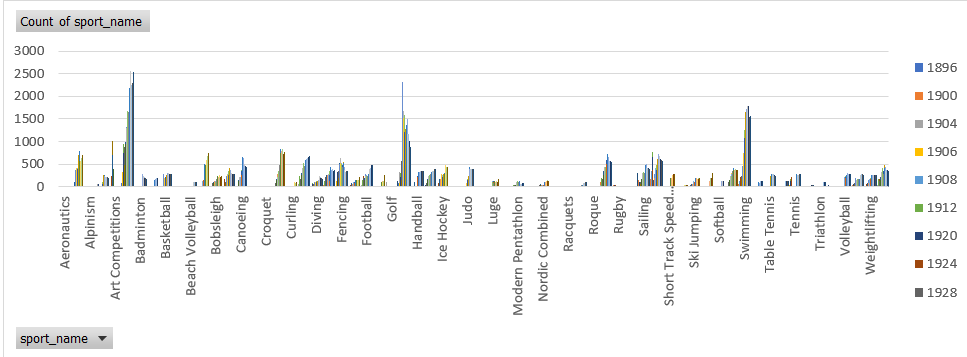
**EDA ANALYSIS (USING EXCEL)**  
Here, I have used the 3 tables made in PowerBI(Games, Event and Person) and further consolidated it using power query editor, and made a consolidated fact table.  
Using the consolidated fact table I have made different pivots and charts to do the eda analysis for the given question.



* Based on the dataset, the analysis reveals that regions like Athina and London have hosted Olympic games for more than 2 times.
* And regions like Innsbruck, Lake placid, Paris and Sankt Moritz have hosted Olympic games for 2 times.

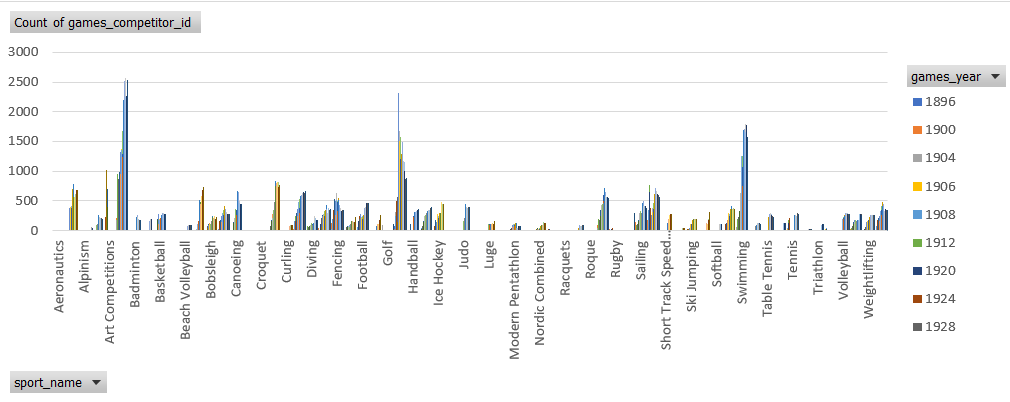


* Based on the dataset, the analysis reveals that from year 1896 to 1920 & year 1994 to 2016 there were only one Olympic game organized in an individual year.  
  Wherein, in the year 1924 to 1992 there were two Olympic games organized in an individual year.

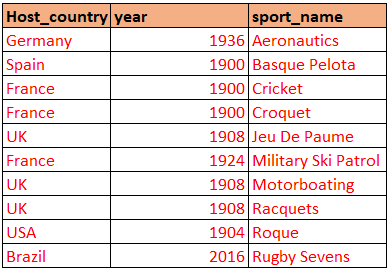


* Based on the dataset, the analysis reveals that the below mentioned sports are the emerging sports that have been recently added to the Olympics.

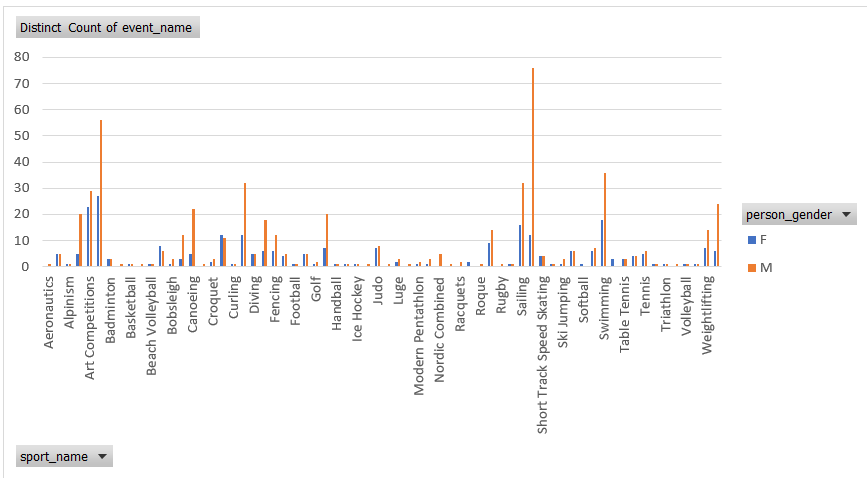
**Rugby Sevens** 2016 **Taekwondo** 2000 2004 2008 2012 2016 **Triathlon** 2000 2004 2008 2012 2016 **Tug-Of-War** 1900 1904 1906 1908 1912 1920



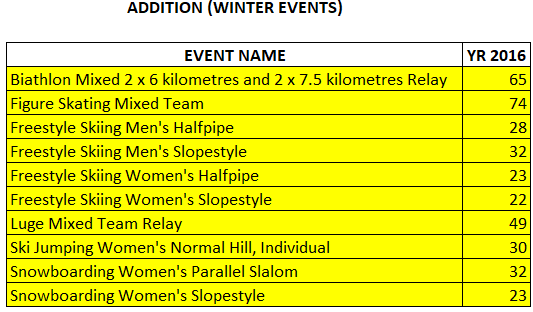
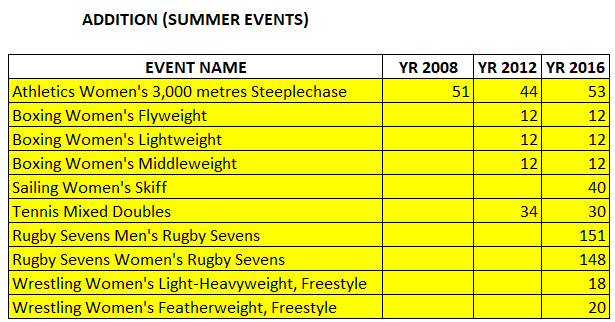
* Based on the dataset, the analysis provide that the popularity of certain sports have been increased tremendously. For e.g, Art Competition has 26 and 11 participants in the year 1912 and 1920, then we have Swimming which has 7, 1788 and 1777 in the year 1896, 2000 and 2008, and many other sports available in the dataset. Hence we can clearly see that the popularity for certain sports have increases year on year.

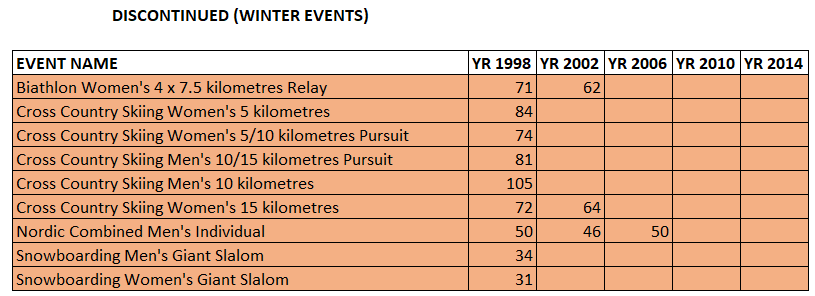
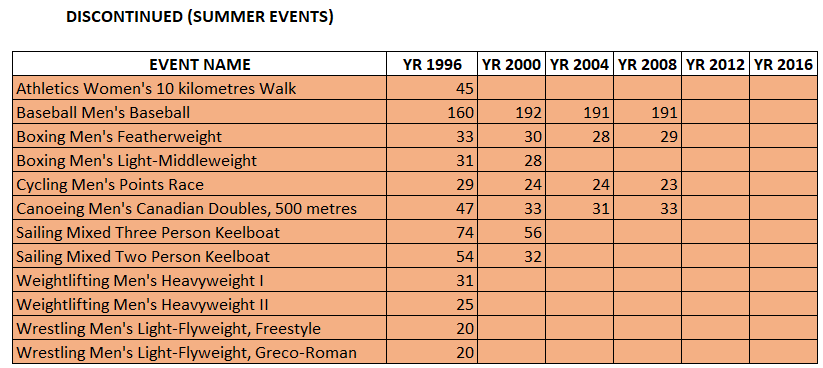


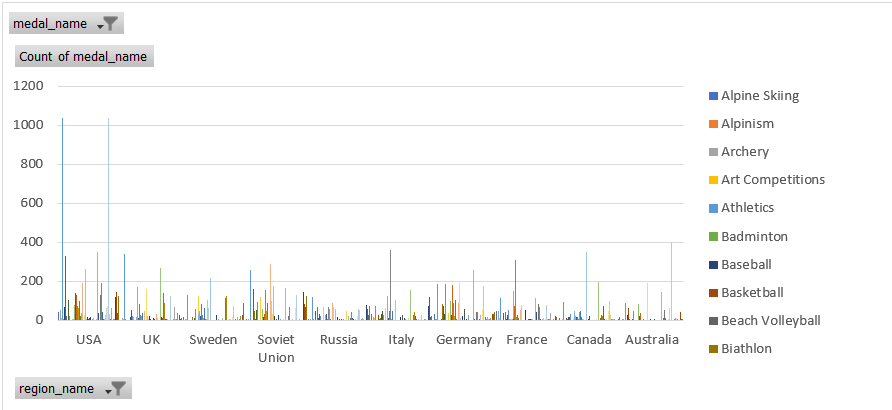
* Based on the dataset, the analysis reveals that the above mentioned sports are specific to these regions.



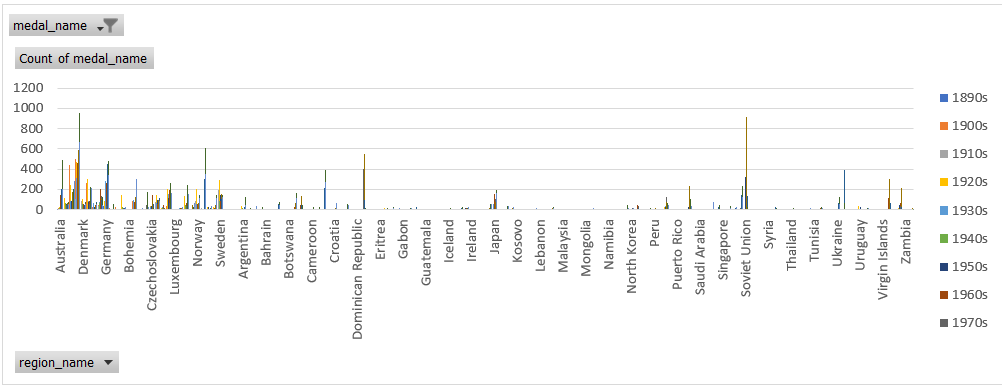
* Based on the dataset, the analysis reveals that almost all the sports have higher number of events for males in each Olympic game. Only in few of the sports number of events are equal for both male and female. For e.g, sports like Alpine Skiing, Badminton, Diving and Table tennis etc.



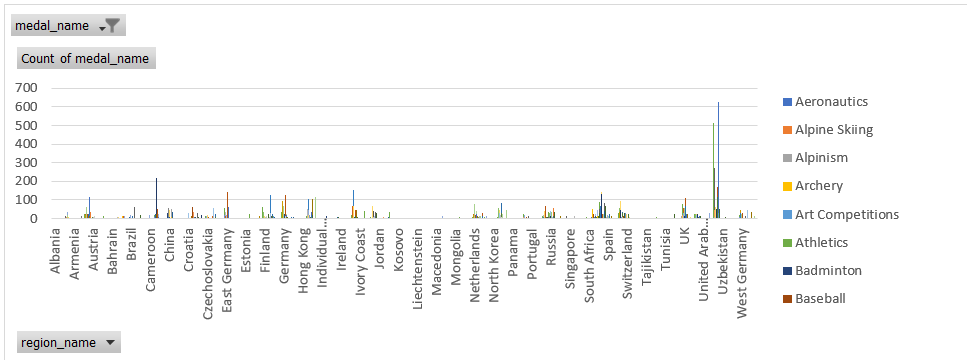




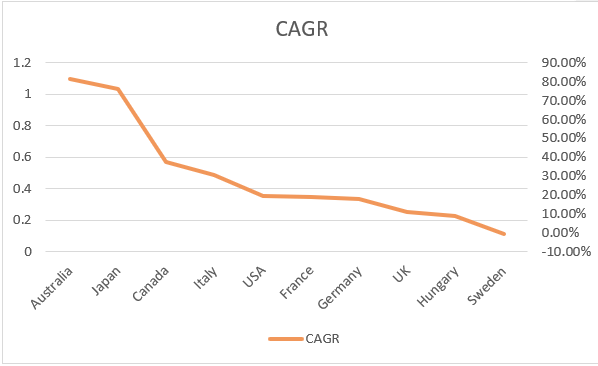
* Based on the dataset, this chart shows the number of medals won by every region in different sports. By using the slicer we can see the dominant countries in each sports and vice versa.



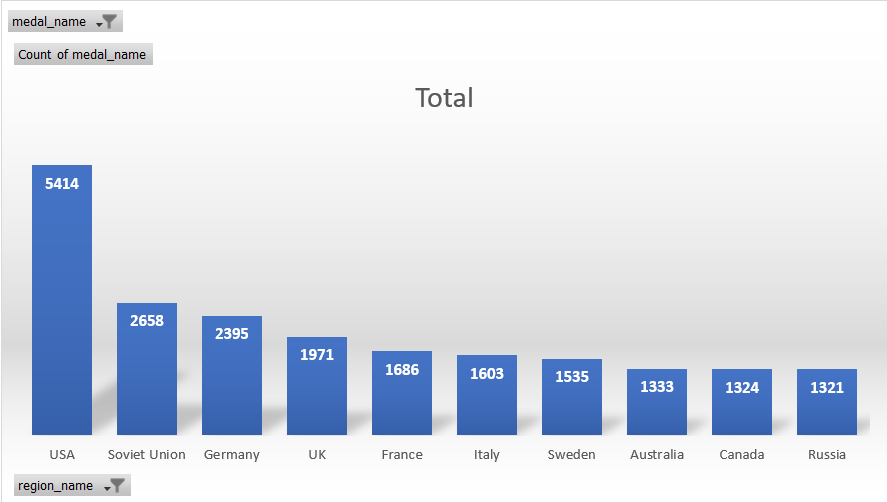
* + - Based on the dataset, the analysis reveals that USA is the region which has won the highest number of medals in multiple Olympic editions.



* Based on the dataset, the analysis reveals that there are many unexpected and surprising medals wins year on year.  
  I have analysed the surprising and unexpected wins for INDIA. India has won its first athletics medal in the year 1900 (SILVER- Athletics Men’s 200 metres), badminton meda in the year 2012 (BRONZE- Badminton Women’s single’s) and shooting medal in 2008 (GOLD- Shooting Men’s Air Rifle, 10 mtrs).



* Based on the dataset, the analysis reveals that some of the regions experience a significant growth in their participation number while other have seen the decline. Like Australia have seen a massive growth over the decades, from having 3 participants in 1910s to 1177 in 2010s. On the other hand Sweden have 713 participants in 1910s which is a decent number, but it declined to 294 in 1940s and then 676 in 2010s, which means that for Sweden seen growth and decline both. But if you will see through CAGR then Sweden has seen decline over the decades.



* Based on the dataset, the analysis reveals that USA has a notable impact on the overall medal tally as the highest number of medals have been won the USA.

**Conclusion: Exploratory Analysis of Olympic Games Dataset**

In this exploratory analysis of the Olympic Games dataset, we gained valuable insights into the historical aspects and general trends associated with the Olympics. The dataset provided a foundational understanding of the Olympic Games, including information about the years, names, seasons, host cities, and basic event details.

1. **Olympic Games Distribution:**
   * The dataset showcased a total of 51 Olympic Games, comprising 29 Summer Olympics and 22 Winter Olympics. This emphasizes the prevalence of Summer Olympics compared to the Winter counterpart in the dataset.
2. **Historical Decade-wise Distribution:**
   * We observed a varied distribution of Olympic Games across different decades, reflecting the evolution and growth of the Olympics over time.
3. **Seasonal Variation:**
   * The analysis highlighted a higher frequency of Summer Olympics compared to Winter Olympics, aligning with the historical pattern of Summer Games being more prevalent.

However, the dataset had limitations, restricting a deeper exploration into other aspects such as no proper data of height and weight of the participants, geographical and cultural conditions, or regional insights. Information regarding participant demographics, notable events, and sports trends over time was absent from the dataset, inhibiting a comprehensive analysis.

This analysis provides a foundation for further exploration, emphasizing the need for comprehensive data to uncover deeper insights into the Olympic Games and their dynamic evolution over the years