PROJECT WORK

Project Title: DATA-DRIVEN INSIGHTS ON OLYMPIC SPORTS PARTICIPATION AND PERFORMANCE

Faculty Mentor: P. SIVA KIRAN KUMAR

Team ID: LTVIP2023TMID00417

Team Size: 5

Team Leader: GUNAPU SHARON

Team Members: SANCHANA MAMATHA

PONNADA NEELIMA

KATTAMURI APARNA

POTIPIREDDI JHANSI

Data-Driven Insights on Olympic Sports Participation and Performance

1 INTRODUCTION:

1.1 Overview:

The modern Olympic Games or Olympics are leading international sporting events featuring summer and winter sports competitions in which thousands of athletes from around the world participate in a variety of competitions. The Olympic Games are considered the world's foremost sports competition with more than 200 nations participating. The Olympic Games are held every four years, with the Summer and Winter Games alternating by occurring every four years but two years apart.

Olympic Games are one of the main international events and also a matter of prestige for countries and therefore each country tries to give their best performance during the event. An analysis needs to be done by each country to evaluate the previous statistics which will detect the mistakes which they have done previously and will also help them in future development. Visualization of the data over various factors will provide us with the statistical view of the various factors which lead to the evolution of the Olympic Games and Improvement in the performance of various Countries/Players.

1.2 Purpose:

The purpose of this project is to analyse the large Olympic dataset using Exploratory Data Analysis to evaluate the evolution of the Olympic Games over the years. An analysis can also be done by the host country to find out the mistakes in the arrangements of the event which will help them in overcoming these mistakes and host the event accurately. This analysis will provide detailed and accurate information regarding various factors which lead to the evolution of the Olympic Games and the improvement of Countries/Players over time in a visual format.

The Analysis will include the visualisation and explanation of the change in trends of the various factors over the years which will help to predict the information of future Olympic Games. As the Olympic Games are one of the most important sporting events across the world, each country and each player tries to give their best performance in the event. To improve their performance, every country should perform such an Analysis which would help them in the improvement of their policies and strategies by providing current statistics to them.

2 LITERATURE SURVEY:

A literature survey for Olympic sports involves reviewing academic articles, books, and other sources related to the history, governance, economics, athlete development, social and cultural impact, and technology and innovation in Olympic sports. The survey can provide a comprehensive understanding of the significance, challenges, and opportunities associated with Olympic sports.

2.1 Existing Problem:

Lack of analysis of Olympic sports can result in limited information about previous statistics which will detect the mistakes which they have done previously and will also effects for future development. And also lack of insights which results in statistical view of various factors to improves the performance of players. Doing data analysis will solve this problem.

2.2 Proposed Solution:

Data analysis is the solution for analysis of Olympic sports. There has been a lot of analysis on the Olympic Games like statistics visualisation, performance analysis of players, improvement in the performance of various countries, and many more. To accomplish this, we have to complete all the activities listed below,

- Data Collection & Extraction from Database

  Here, we collect dataset, Storing Data in DB2 & Perform SQL Operations and connect DB2 with Cognos.
- Data preparation

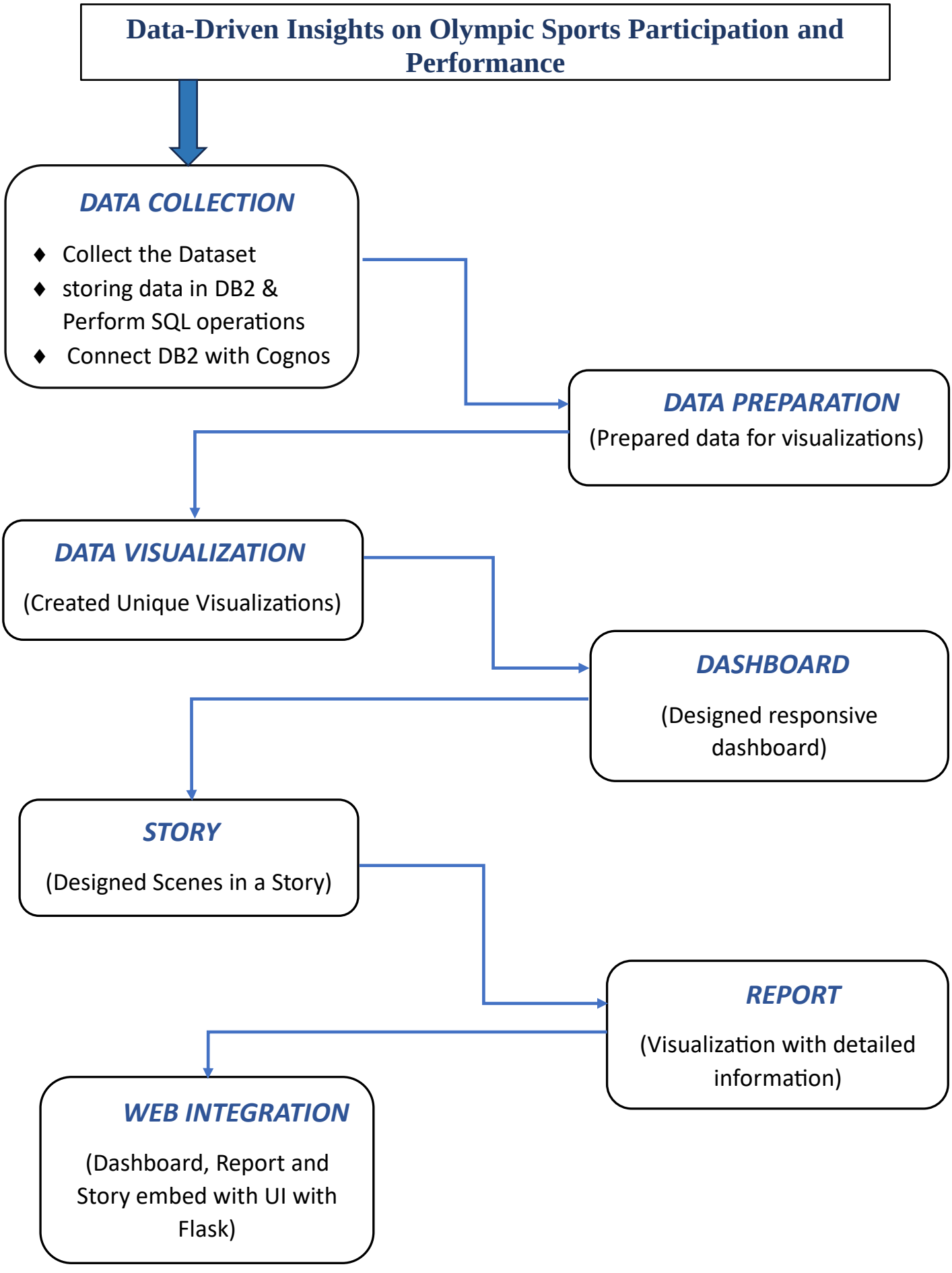
  Here, we prepare data module for visualizations. Preparing the data for visualization involves cleaning the data to remove irrelevant or missing data, transforming the data into a format that can be easily visualized, exploring the data to identify patterns and trends, filtering the data to focus on specific subsets of data.
- Data visualization

  Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information. The goal of data visualization is to make complex data sets more accessible.
- Creating Dashboard, Report and Story

  A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. A data story is a way of presenting data and analysis in a narrative format, with the goal of making the information more engaging and easier to understand. Report in data analytics typically involves analysing and interpreting data to draw insights. The report should begin with executive summary, which provides a brief overview of the main findings and recommendations.
- Web Integration, here we embed Dashboard, Report and Story with UI with Flask to make better decisions, and communicate their performance to others.

3 THEORITICAL ANALYSIS:

3.1 Block Diagram:



3.2 HARDWARE/SOFTWARE DESIGNING:

As the Olympic Games are one of the most important sporting events across the world, each country and each player tries to give their best performance in the event. To improve their performance, every country should perform such an Analysis which would help them in the improvement of their policies and strategies by providing current statistics to them. In this project Data-Driven Insights on Olympic Sports Participation and Performance we use some hardware/software systems.

Hardware requirements of this project:

- PC/Laptop with latest version of Windows, Mac, Linux ect .

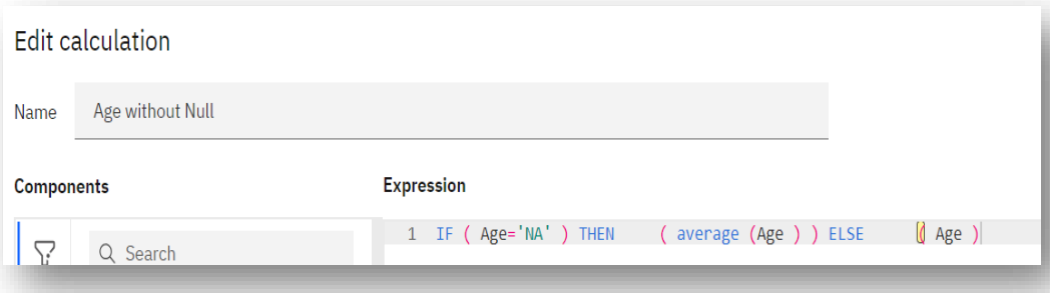
Software requirements of this project:

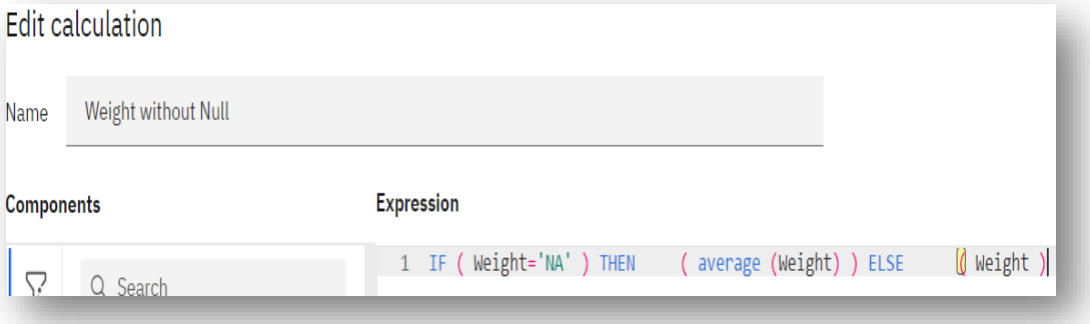
- IBM Cognos Analytics, Anaconda Navigator, Python, Jupyter Notebook & Spyder.

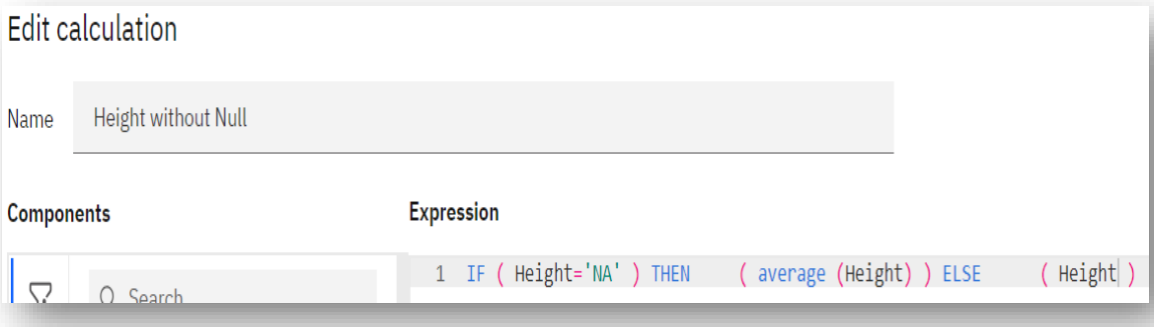
4 RESULT:

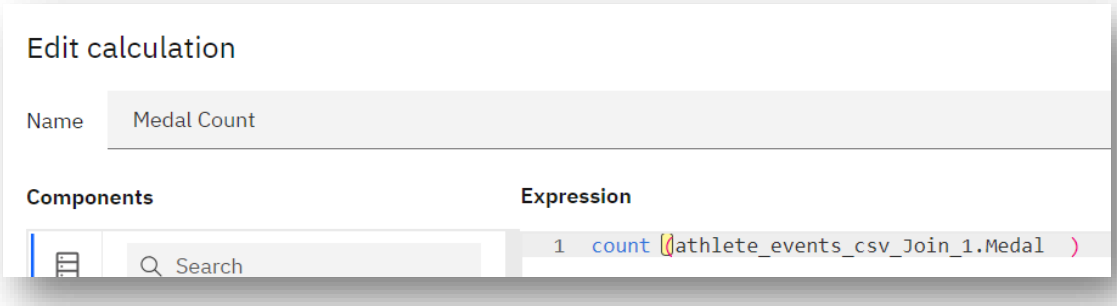
Following are the outputs of this project Data-Driven Insights on Olympic Sports Participation and Performance.

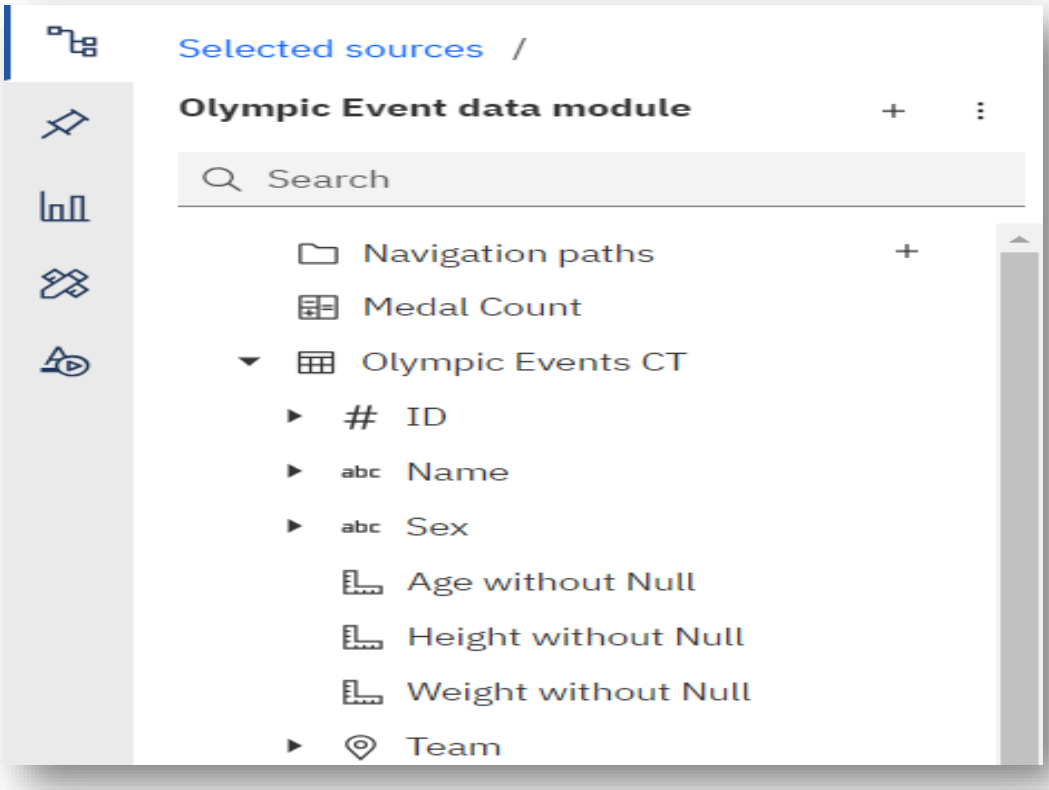
No OF Calculation Fields Created:



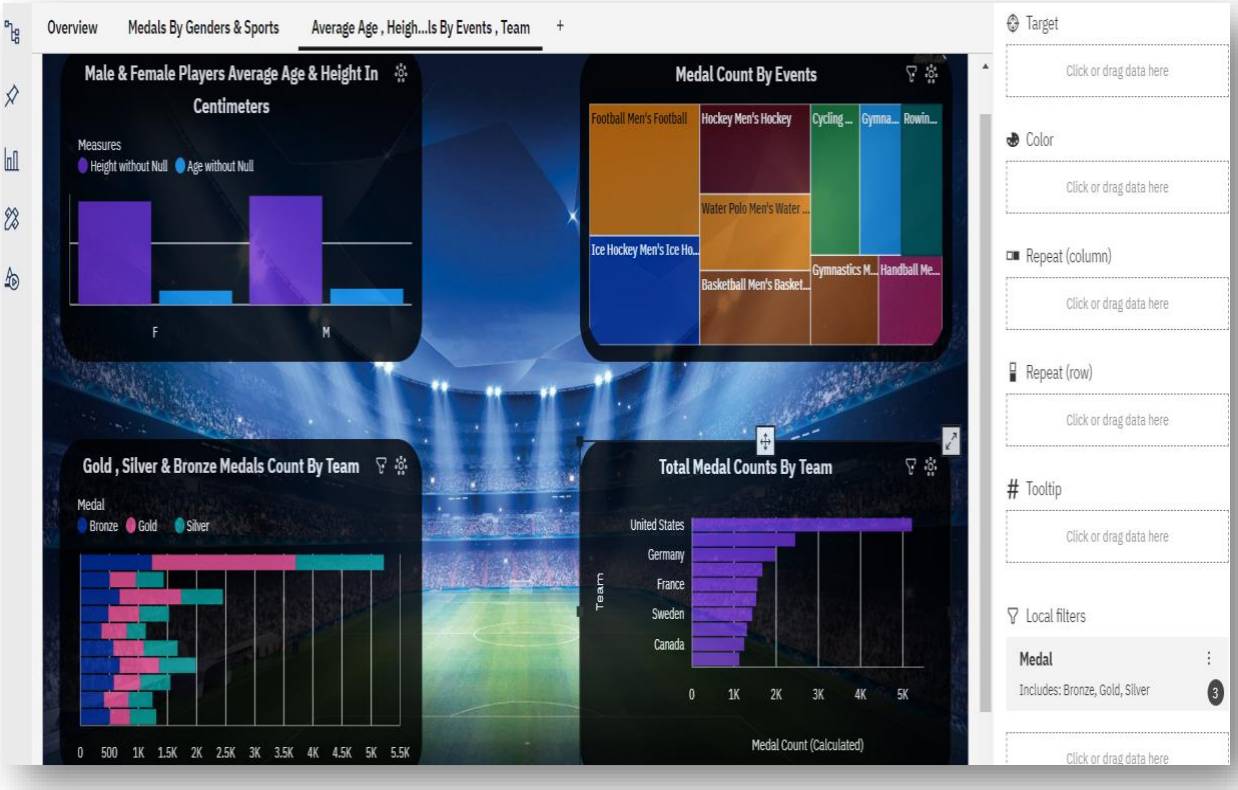


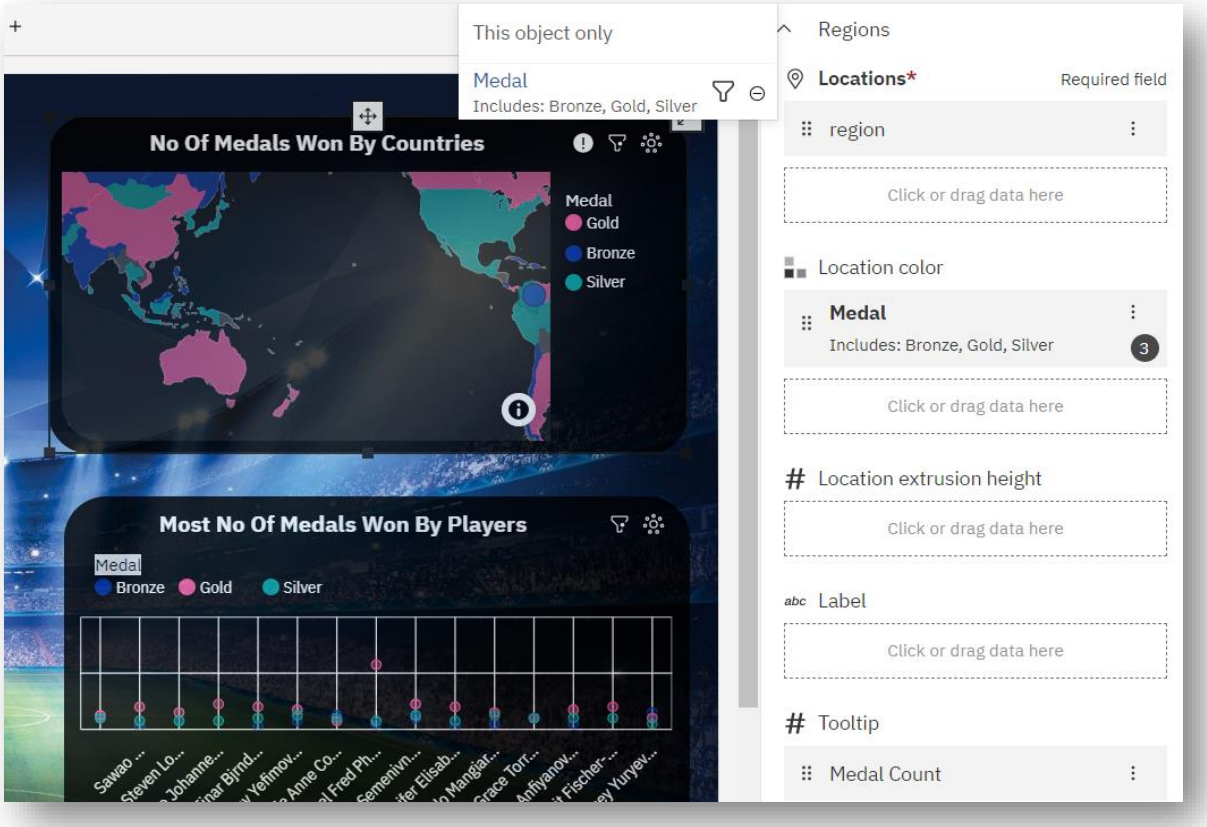

Edit calculation
Name
Height without Null
Components
Expression
Search
1 IF (Height='NA') THEN (average (Height)) ELSE (Height)


Edit calculation
Name
Medal Count
Components
Expression
Search
1 count (athlete_events_csv_Join_1.Medal)


Selected sources /
Olympic Event data module
Search
Navigation paths
Medal Count
Olympic Events CT
ID
Name
Sex
Age without Null
Height without Null
Weight without Null
Team

Utilization Of Data Filters:

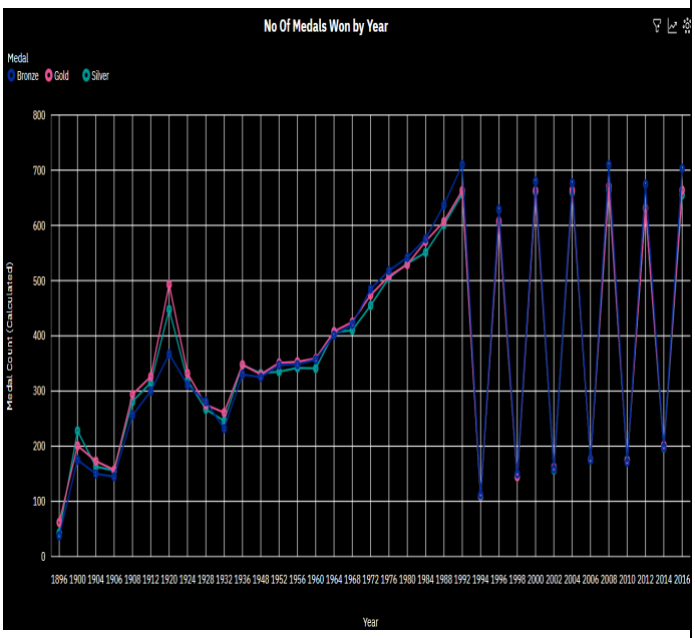




No Of Visualizations/ Graphs:

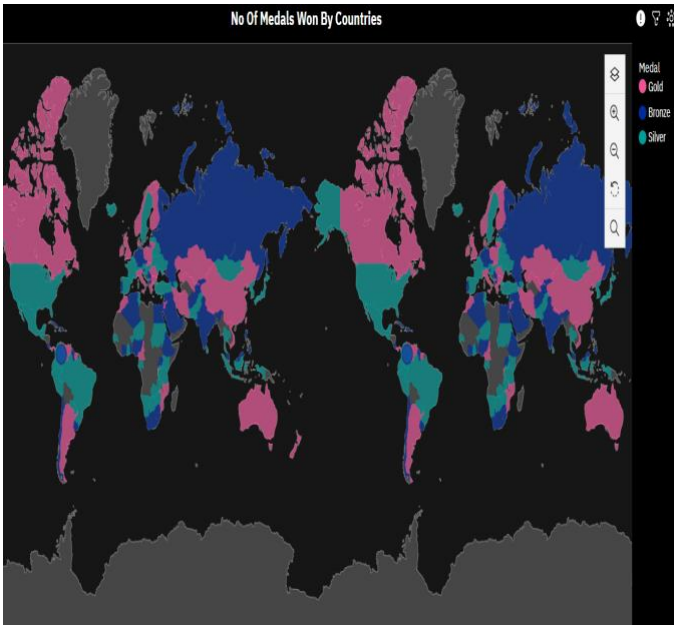
Data visualization is the process of creating graphical representations of data in order to help people understand and explore the information.

No Of Medals Won by Year



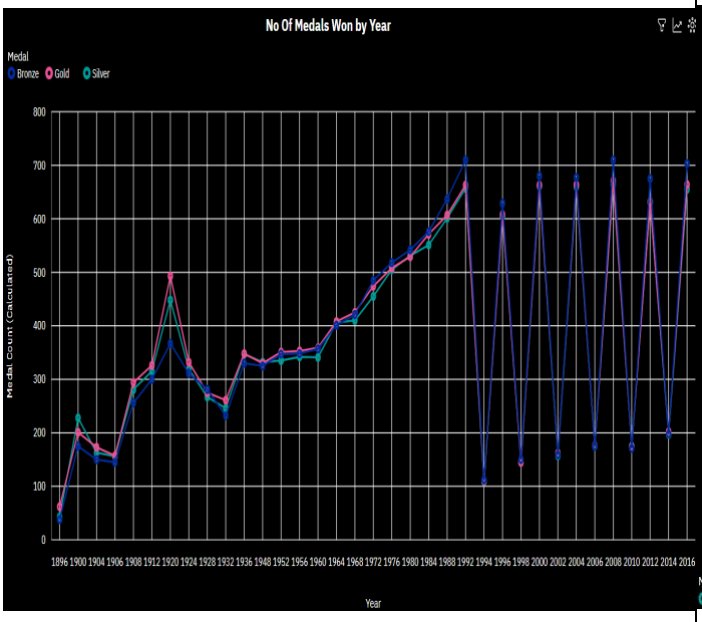
This visualization shows the total number of results for **Medal Count**, across all **years**, is **105**.

No of Medals Won by Countries



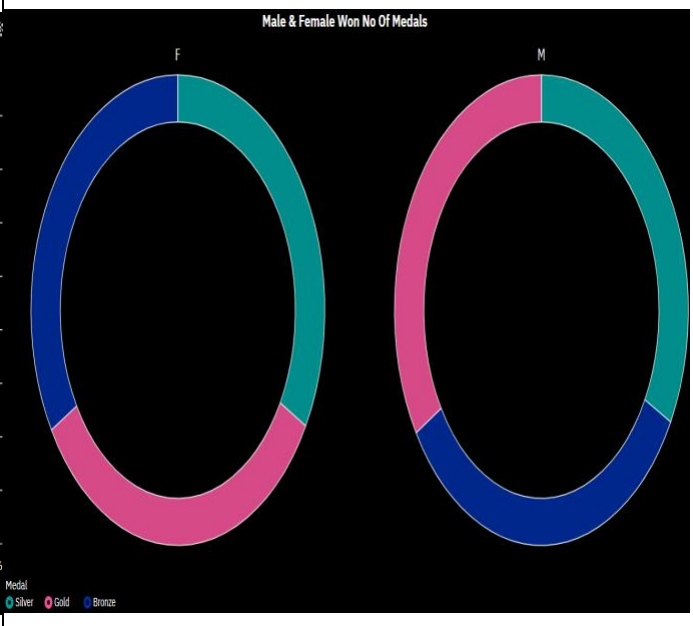
This visualization shows total number of results for **Medal Count**, across all **regions**, is **327**. The total number of results for **Medal**, across all **regions**, is **nearly 40 thousand**

Male & Female Won No of Medals



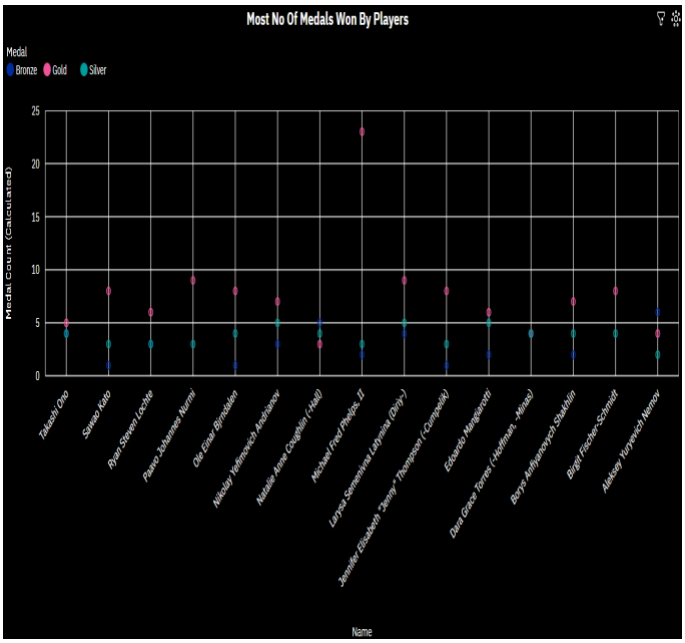
This visualization shows the total number of results for **Medal Count**, across all **medals**, is **6**.

Most No of Medals Won By Player



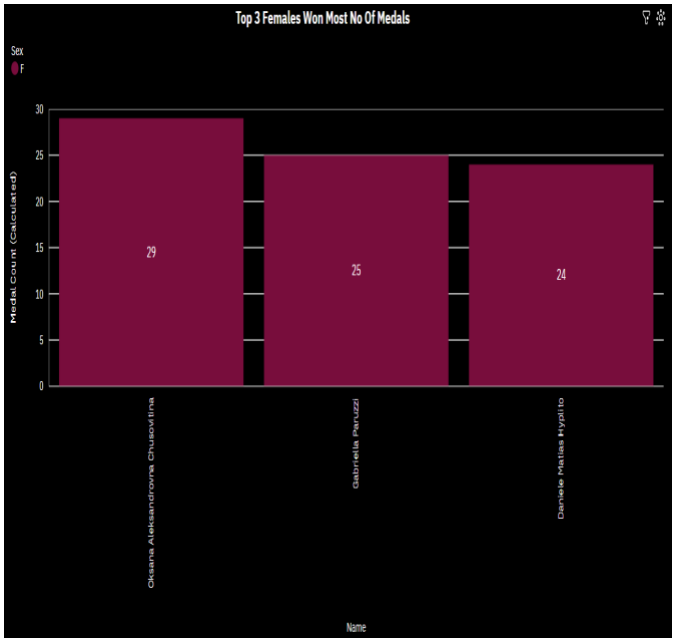
This visualization shows the total number of results for **Medal Count**, across all names, is 43. **Gold** (**34.9** %) and **Silver** (**34.9** %) are the most frequently occurring categories of **Medal** with a combined count of **30** items with **Medal Count** values (**69.8** % of the total).

Top 3 Females Won Most No of Medals



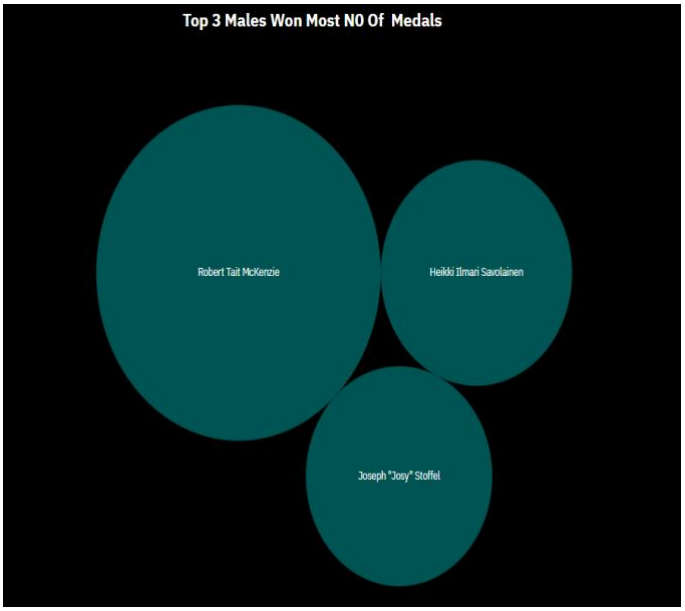
This visualization shows the total number of results for **Medal Count**, across all names, is 3. City Beijing has the highest Medal Count at 11, out of which name Oksana contributed the most at 6

Top 3 Males Won Most No of Medals



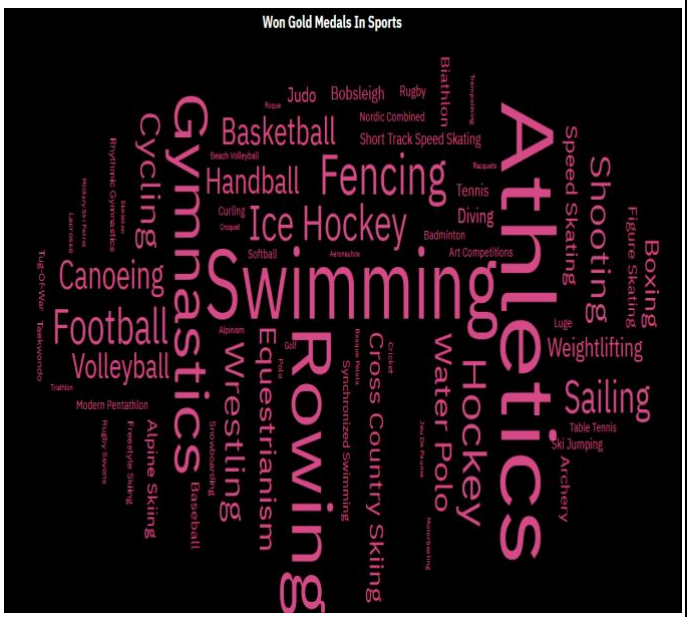
This visualization shows the total number of results for **Medal Count**, across all **names**, is **3**. **City Los Angeles** has the highest **Medal Count** at **52**, out of which **Name Robert Tait McKenzie** contributed the most at **44**.

Won Gold Medals in Sports



This visualization shows the total number of results for **Medal Count**, across all **sports**, is **66**. **City London** has the highest **Medal Count** at **1215**, out of which **Sport Athletics** contributed the most at **140**.

Won Silver Medals in Sports

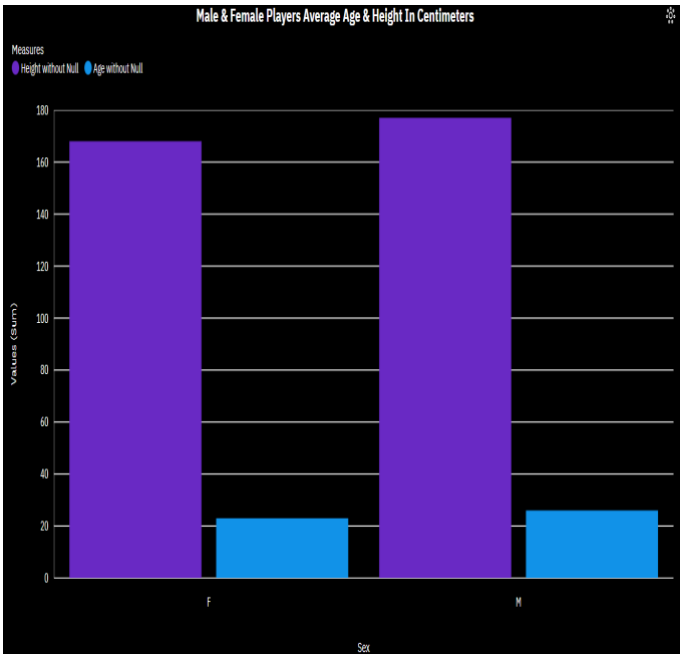


This visualization shows the total number of results for **Medal Count**, across all **sports**, is **62**. **City London** has the highest **Medal Count** at **1195**, out of which **Sport Athletics** contributed the most at **140**.

Male & Female Players Average Age & Height

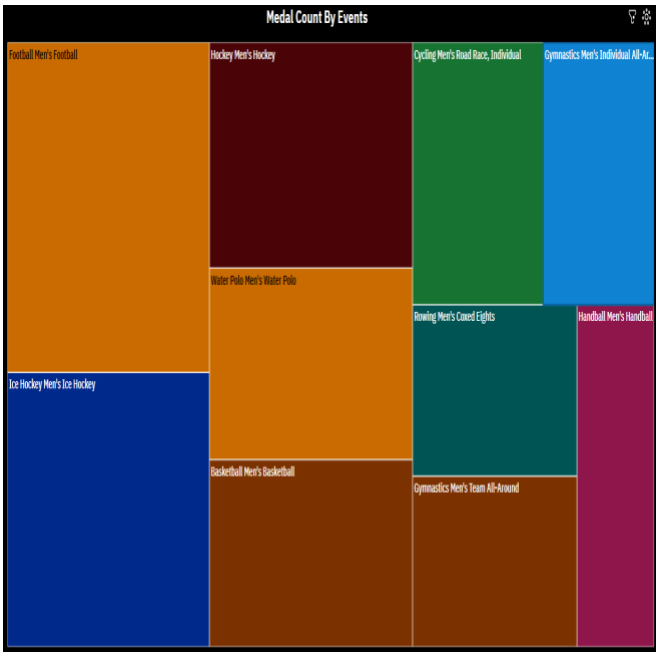
This visualization shows the total number of results for **Age without Null**, across all **sexes**, is **almost 2500**. The total number of results for **Height without Null**, across all **sexes**, is **almost 2500**.

Medal Count by Events



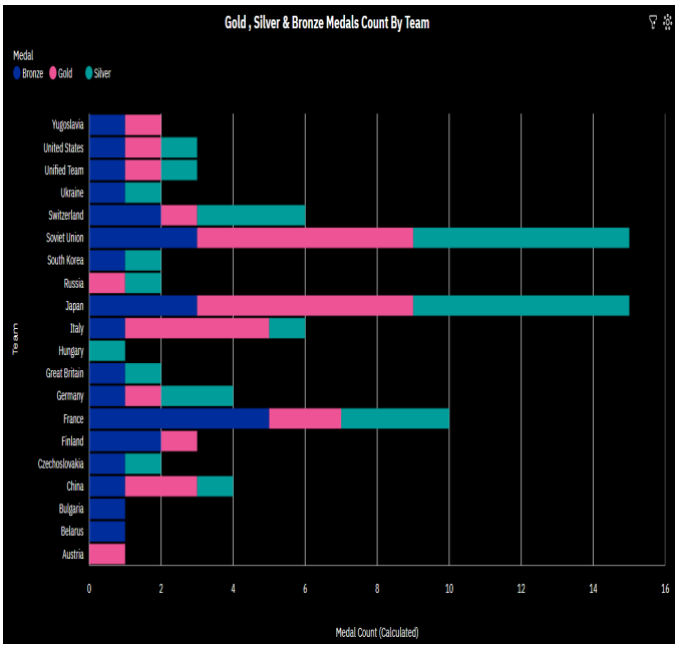
This visualization shows the total number of results for **Medal Count**, across all **events**, is **10**. **Event Football Men's Football** has the highest **Medal Count** due to **City London**.

Gold, Silver & Bronze Medals Count by Team



This visualization shows the total number of results for **Medal Count**, across all **teams**, is **30**.

Total Medal Counts by Team



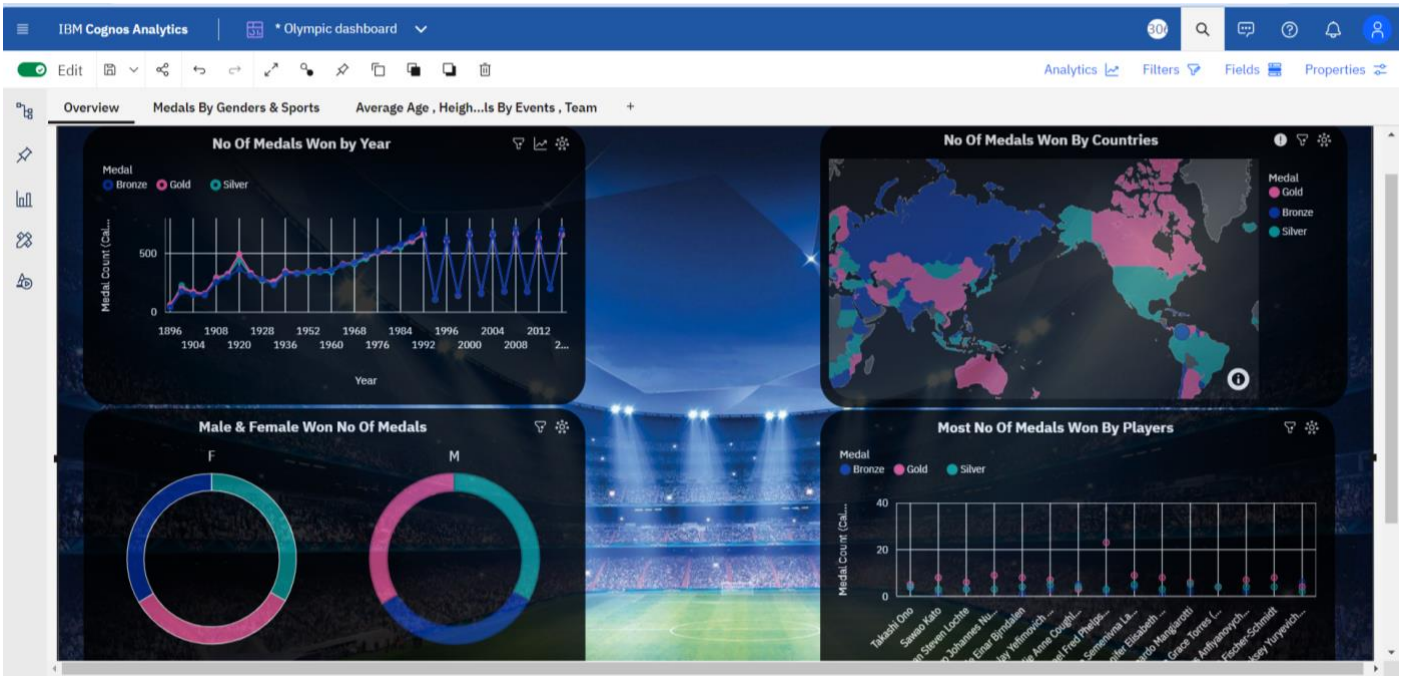
This visualization shows the total number of results for **Medal Count**, across all **teams**, is **10**. **Team United States** has the highest **Medal Count** due to **City Los Angeles**.

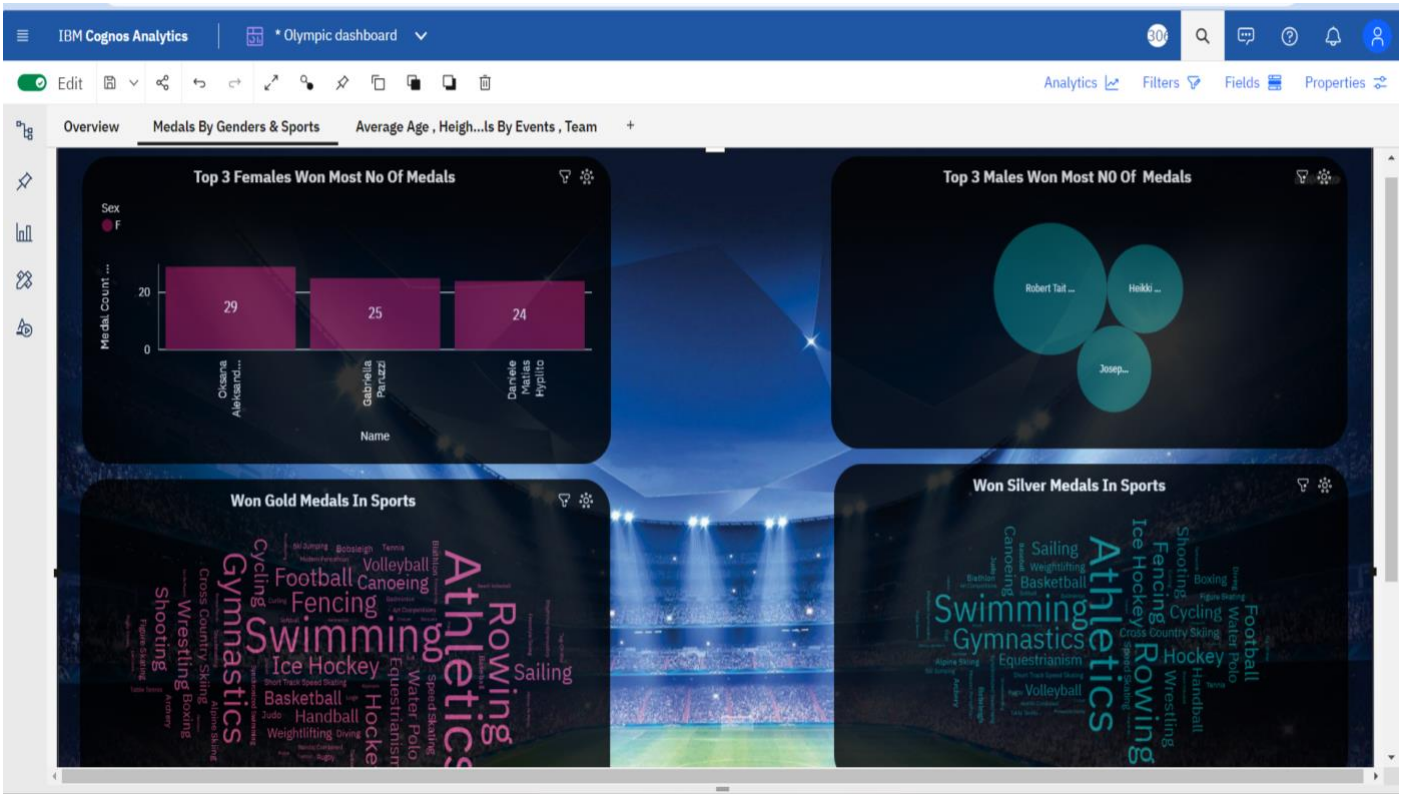
DASHBOARD:

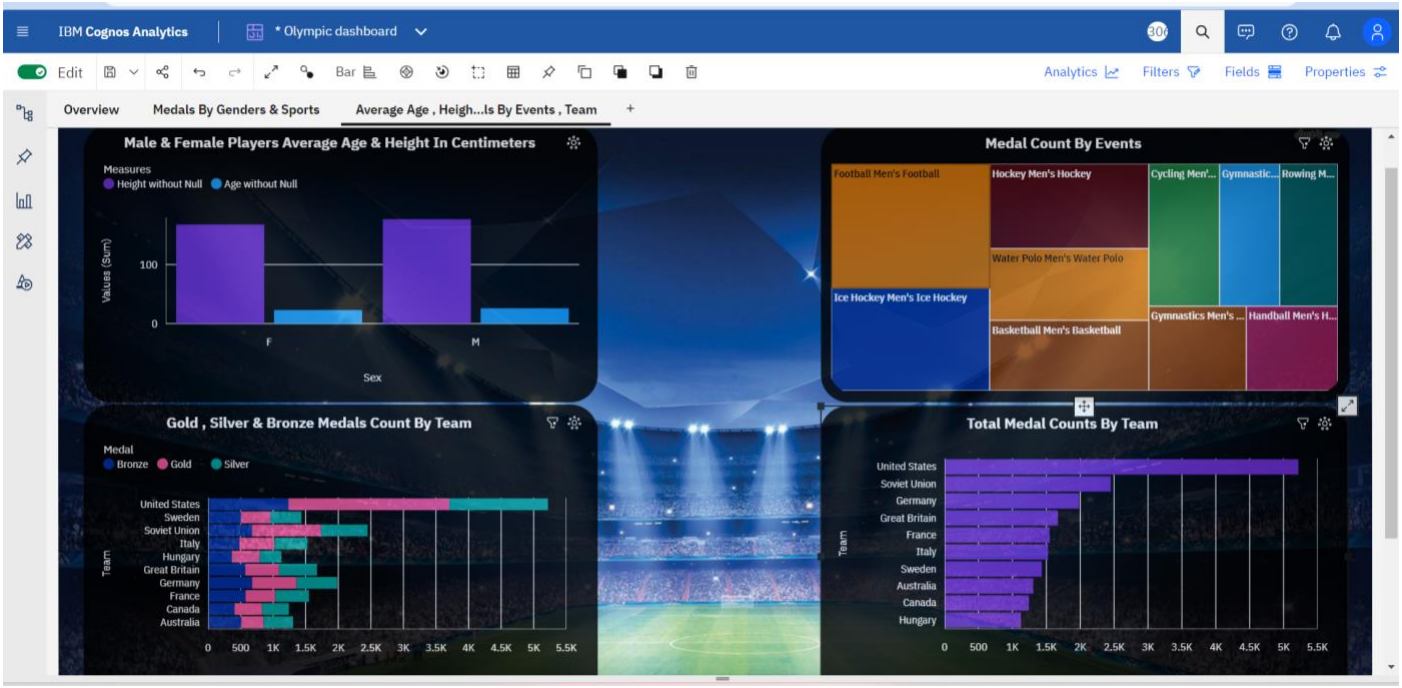
A dashboard is a graphical user interface (GUI) that displays information and data in an organized, easy-to-read format. Dashboards are often used to provide real-time monitoring and analysis of data, and are typically designed for a specific purpose or use case. Dashboards can be used in a variety of settings, such as business, finance, manufacturing, healthcare, and many other industries. They can be used to track key performance indicators (KPIs), monitor performance metrics, and display data in the form of charts, graphs, and tables.

Olympic sports dashboard:

We created Olympic sports dashboard by using visualizations which are shown in above for better information.



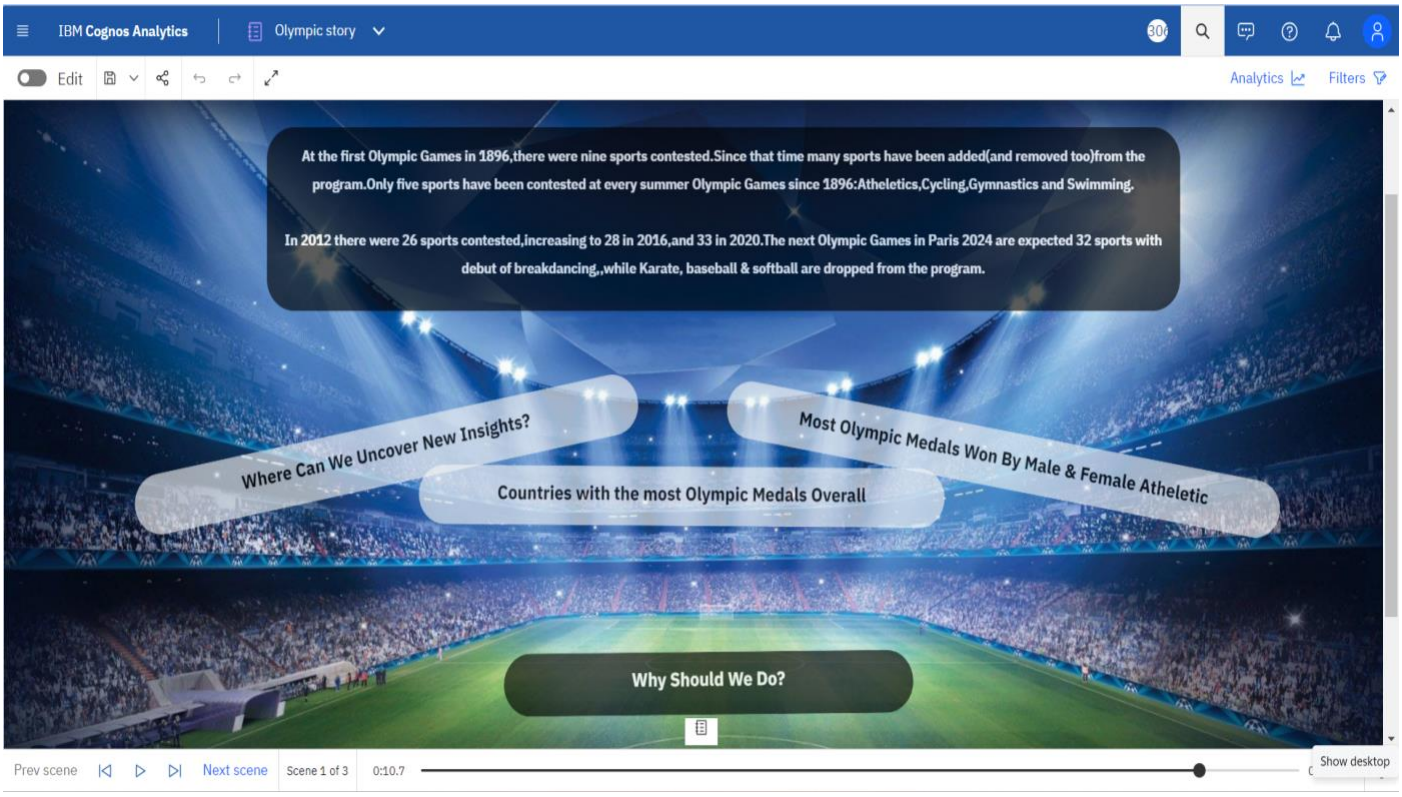




STORY:

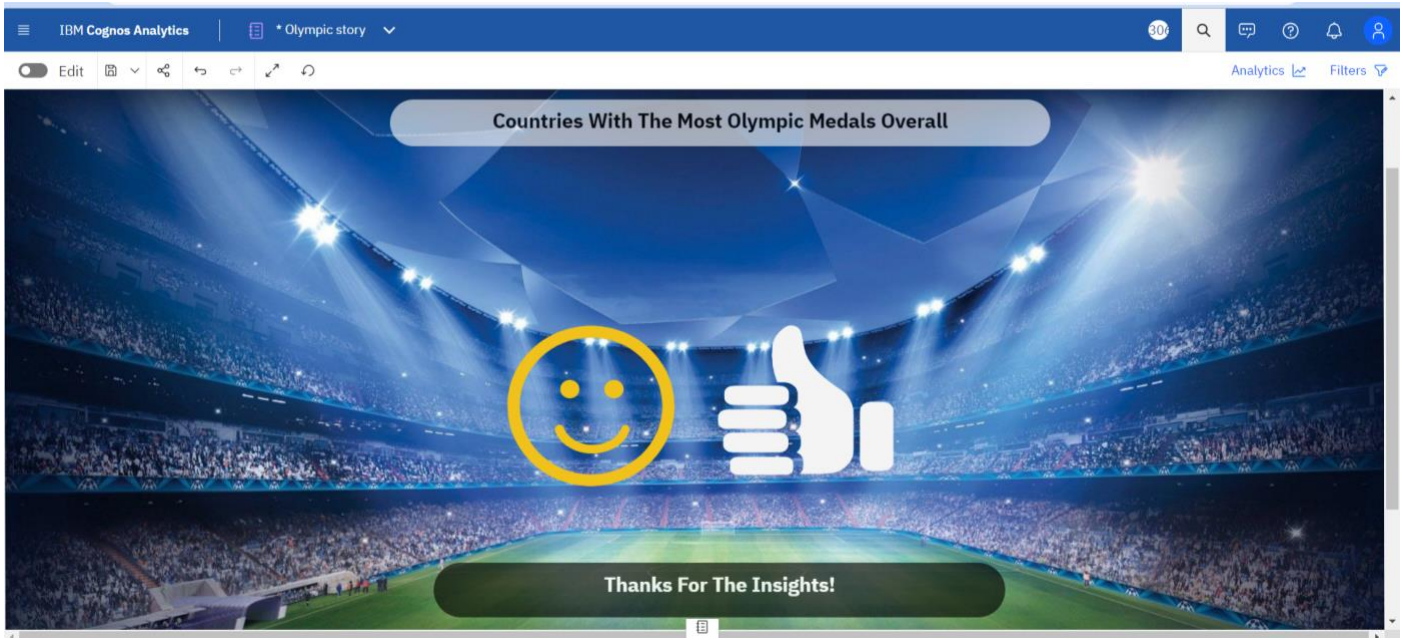
We created Olympic story in the form of scenes. The number of scenes in a storyboard for Data-Driven insights on Olympic Sports Participation and Performance will depend on the complexity of the analysis and the specific insights that are trying to be conveyed. A storyboard is a visual representation of the data analysis process and it breaks down the analysis into a series of steps or scenes.

SCENE-1:



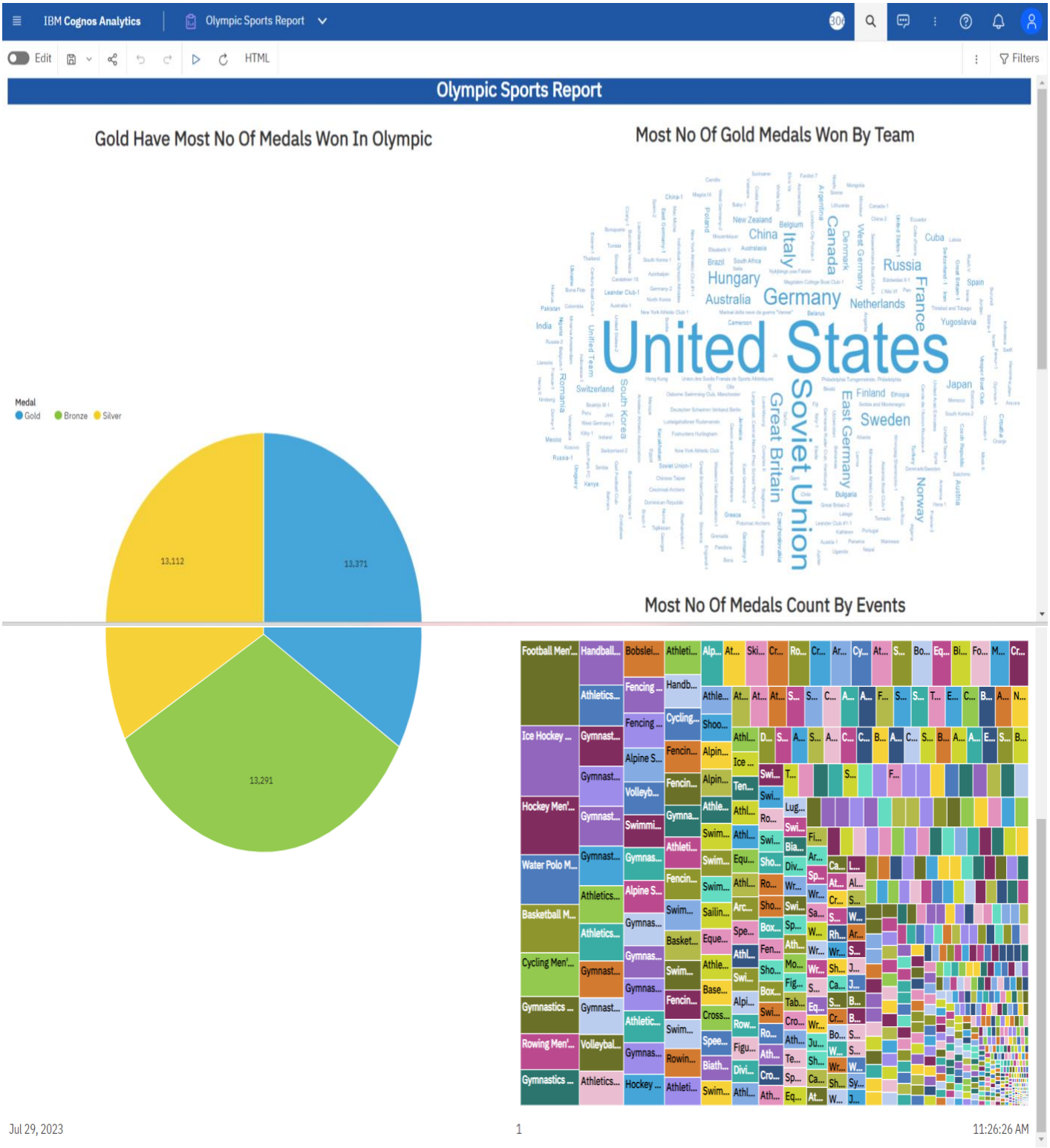
SCENE-2:

SCENE-3:



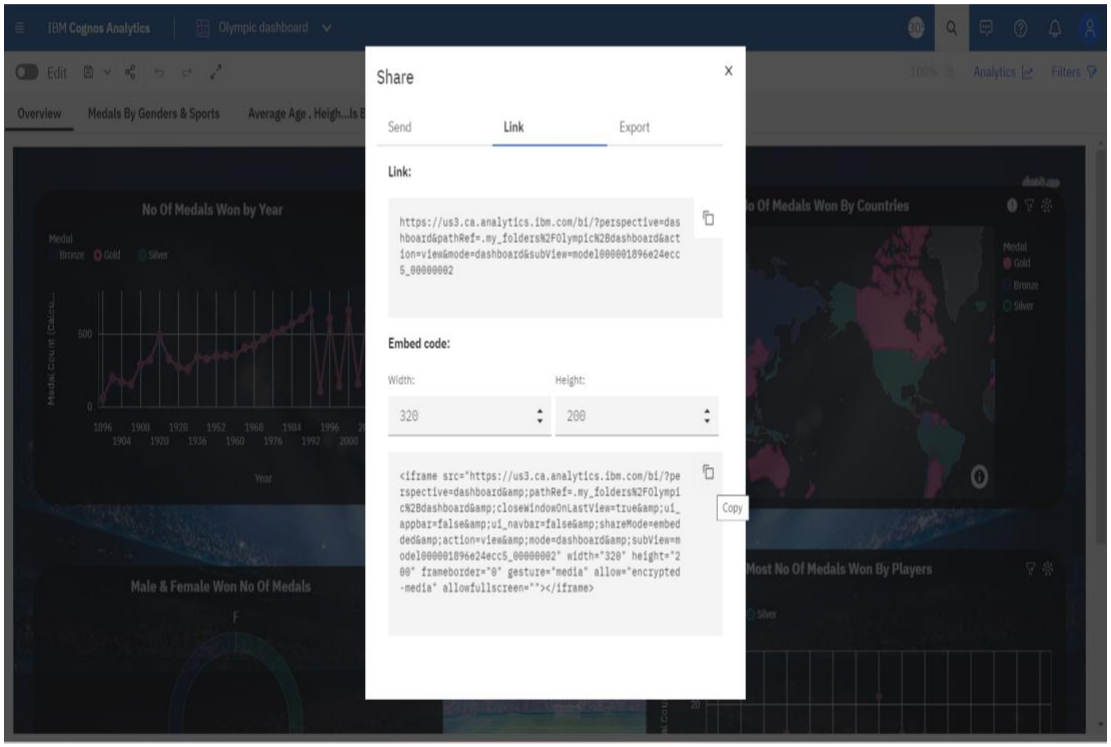
REPORT:

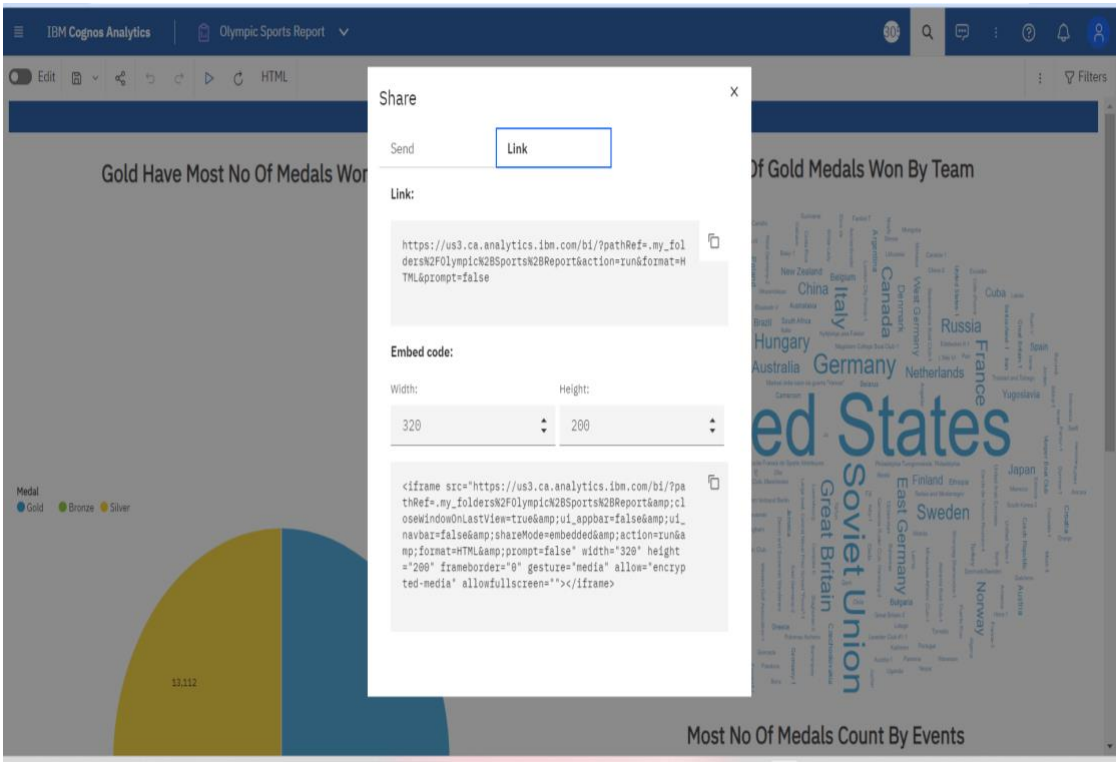
When creating a report in Cognos, it is often helpful to include visualizations to help communicate the findings of the analysis. Report for Data-Driven insights on Olympic Sports Participation and Performance shown below,

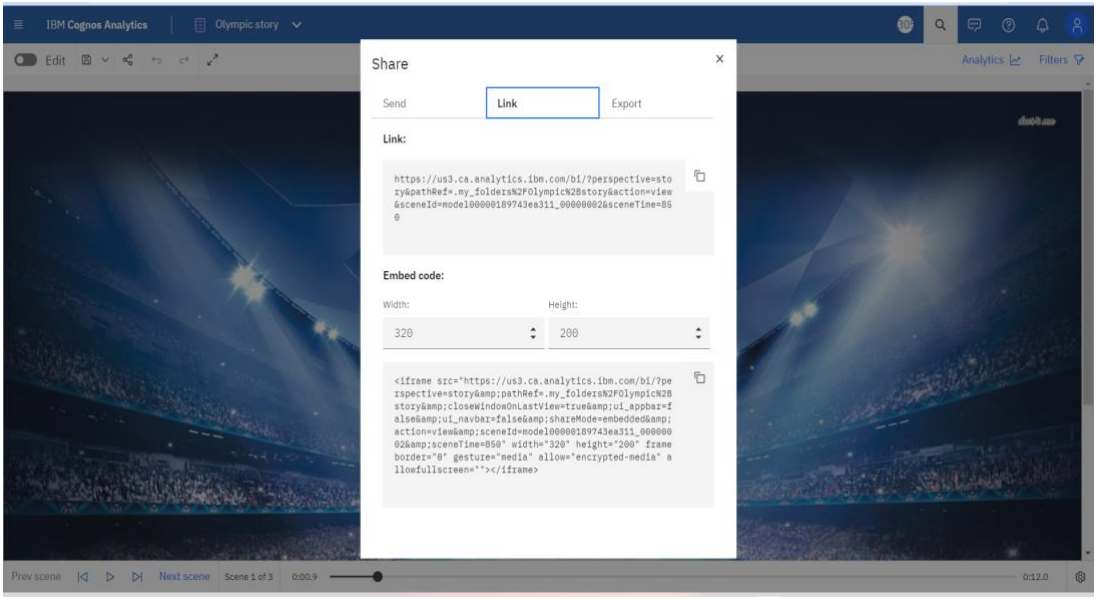


Web Integration:

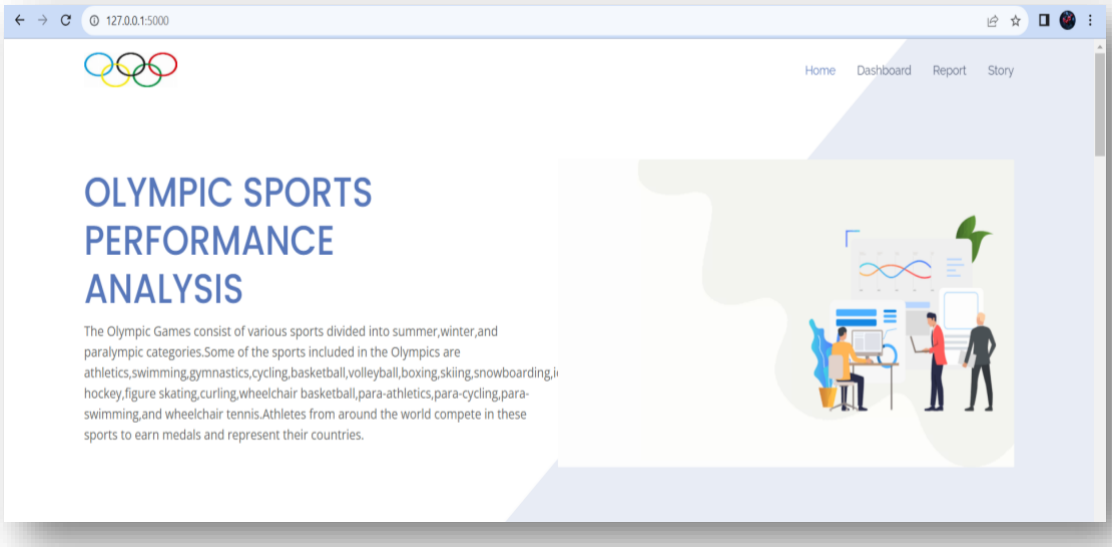
Here we embed Dashboard, Report and Story with UI with Flask to make better decisions, and communicate their performance to others. Here first go to Dashboard, Report & Story, click on share button on the top. Next copied URL links of Dashboard, Report & Story and paste it in html code in flask then we run app.py code we get a URL address which we run at browser then we get our web page.

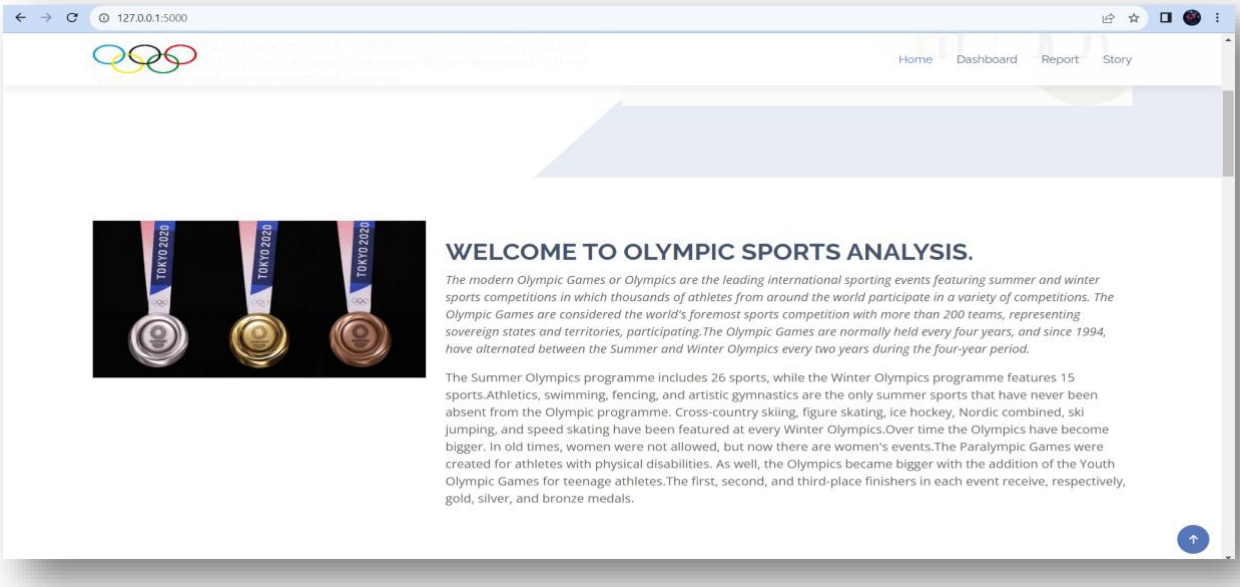


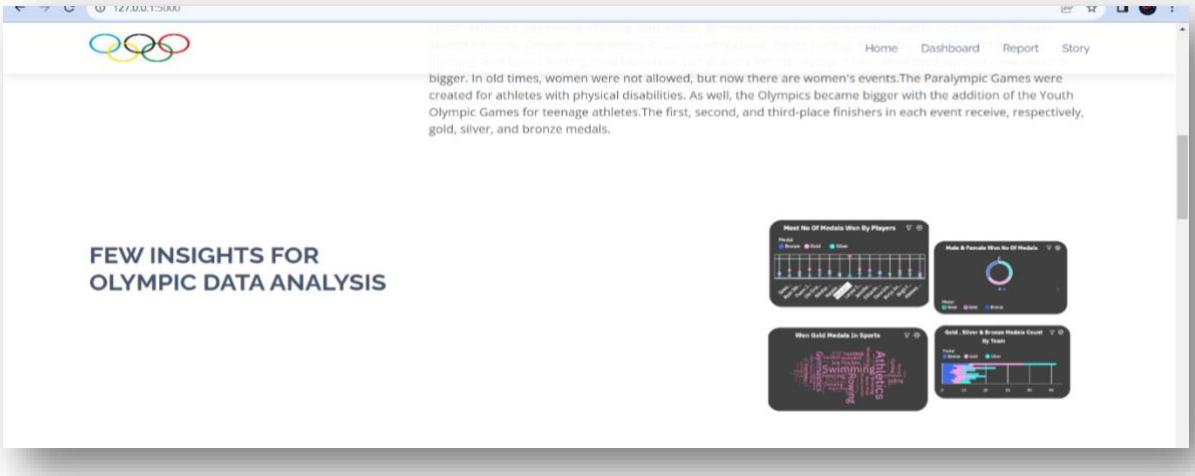


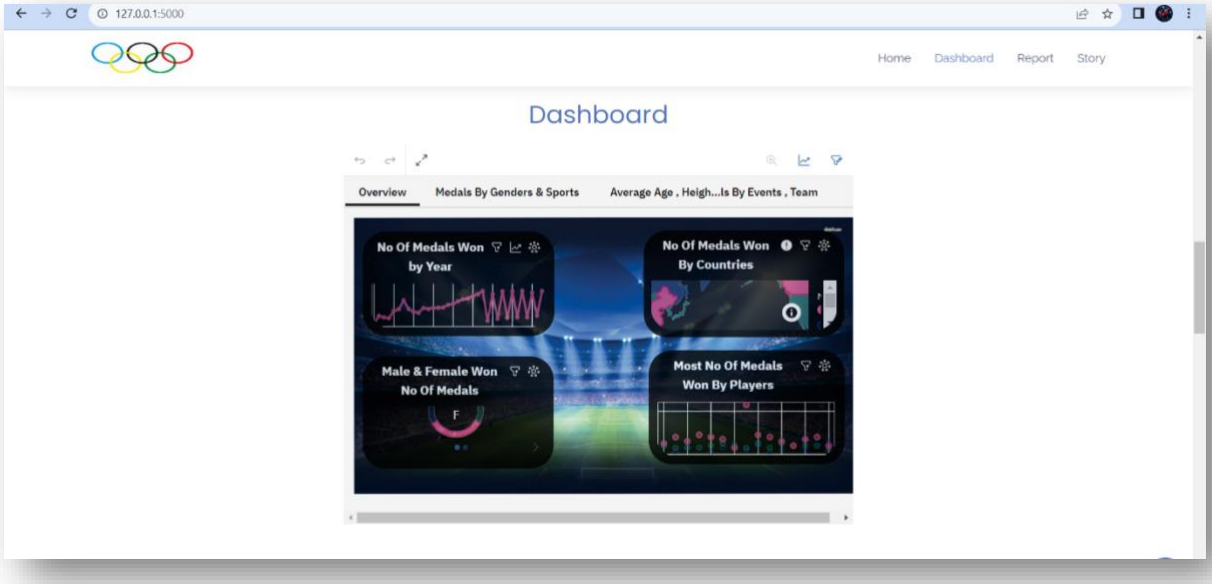


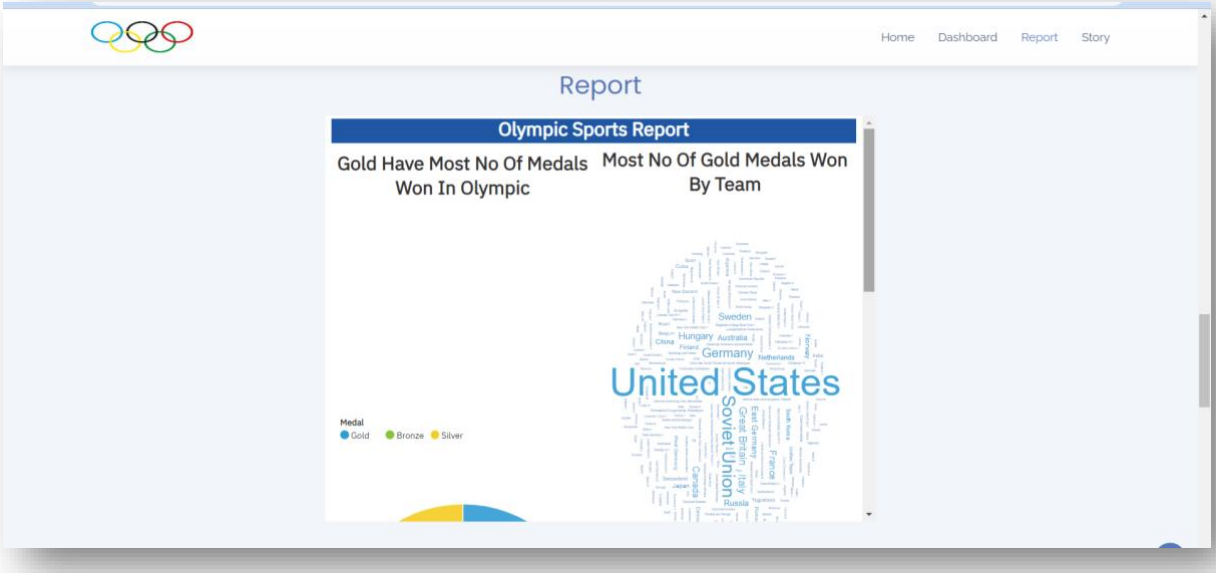
Publishing Dashboard, Report & Story on Web Page

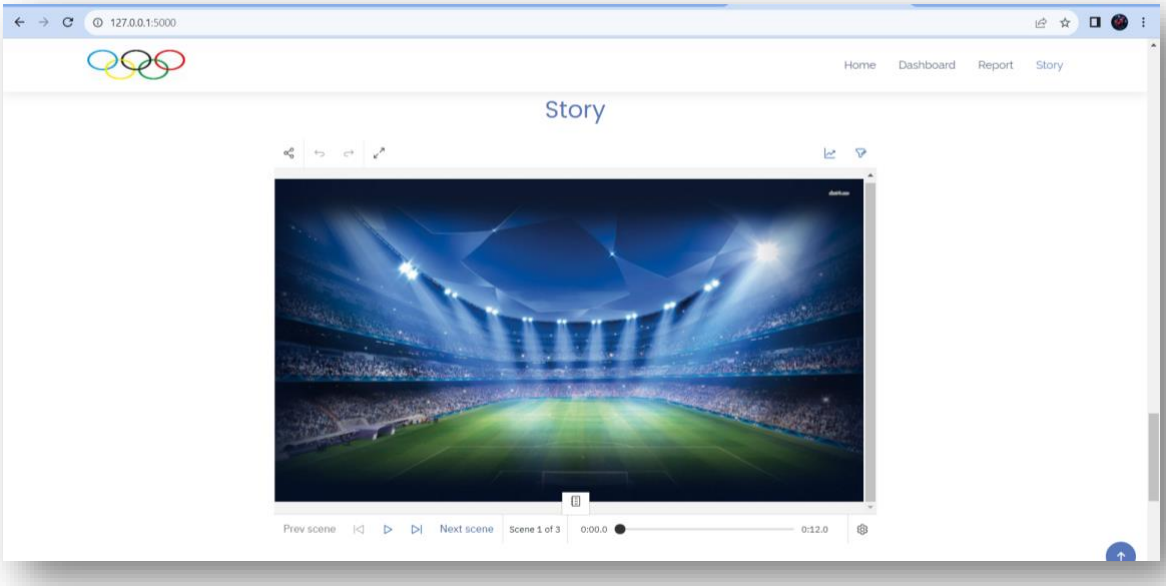


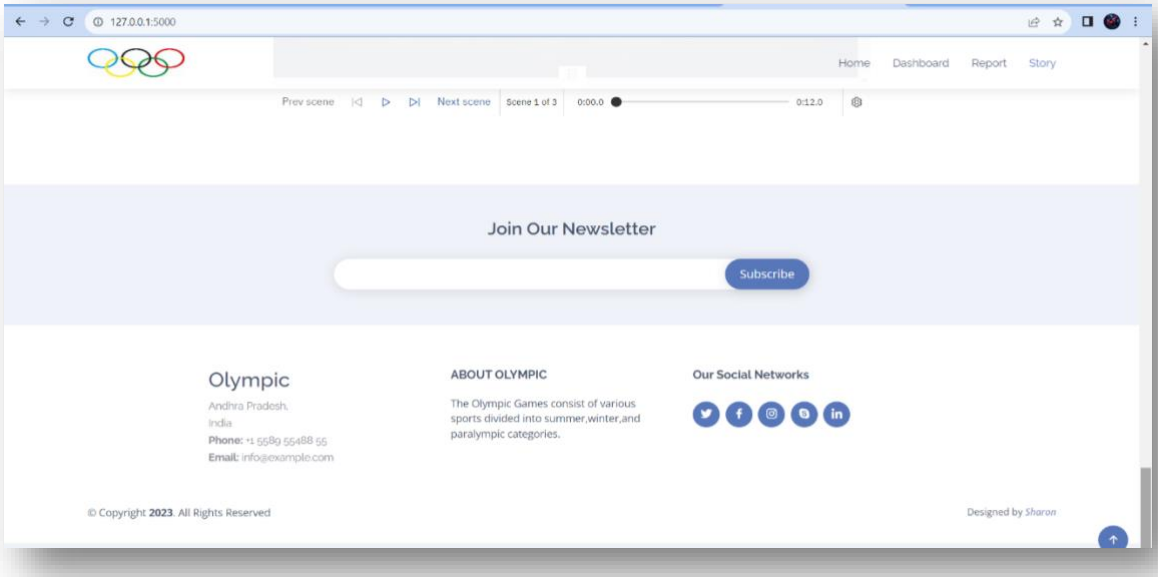



Home
Dashboard
Report
Story
bigger. In old times, women were not allowed, but now there are women's events.The Paralympic Games were created for athletes with physical disabilities. As well, the Olympics became bigger with the addition of the Youth Olympic Games for teenage athletes.The first, second, and third-place finishers in each event receive, respectively, gold, silver, and bronze medals.
FEW INSIGHTS FOR
OLYMPIC DATA ANALYSIS


127.0.0.1:5000
Home
Dashboard
Report
Story
Dashboard
Overview
Medals By Genders & Sports
Average Age , Heigh...ls By Events , Team
No Of Medals Won by Year
No Of Medals Won By Countries
Male & Female Won No Of Medals
Most No Of Medals Won By Players


Home
Dashboard
Report
Story
Report
Olympic Sports Report
Gold Have Most No Of Medals Won In Olympic
Most No Of Gold Medals Won By Team
United States
Soviet Union
Germany
Sweden
Hungary
Australia
Netherlands
Great Britain
France
Italy
Russia
Medal
Gold
Bronze
Silver





5 ADVANTAGES & DISADVANTAGES:

Advantages:

- Detects and correct the errors from data sets with the help of data cleansing. This helps in improving quality of data.
- Deeper insight into the performance of countries in the Olympics over the years and helps athletes to quickly analysis their own and competitors' performance.
- Helps in displaying relevant advertisements on the online websites based on historic data and purchase behaviour of the users.
- Helps in preventing any wrongdoings and/or calamities.

- Deeper insight into the performance of countries in the Olympics over the years and helps athletes to quickly analysis their own and competitors' performance.

Disadvantages:

- Due to former geographical or historical changes analysis may vary
- This may breach privacy of the customers as their information such as purchases, online transactions, subscriptions are visible to their parent companies.
- The cost of data analytics tools based on applications and features supported. Mostly some of the data analytics tools are complex to use and require training. This increases cost to the company willing to adopt data analytics tools or software.
- It is very difficult to select the right data analytics tools. This is due to the fact that it requires knowledge of the tools and their accuracy in analysing the relevant data as per applications.
- The information obtained using data analytics can also be misused against group of people of certain country or community.

6 APPLICATIONS:

- Data analysis helps sports entities evaluate the performance of their athletes and assess the recruitment necessary to improve the team performance.
- Help coaches and athletes to analysis the previous training and competition sports behaviour to improve their competition performance.
- Data analysis technology can also help coaches and athletes understand the strengths and weaknesses of their opponents to achieve excellent results in large-scale events.
- By using Data analysis, the number of medals received by countries can be classified and can be displayed. And also determine which country hosted the Olympics in which year.
- Data analytics can play a major role in revolutionizing the transportation industry. It is especially useful for transporting a large number of people to a certain location that requires seamless transportation. This data analysis technique was used in the 2012 London Olympics. Approximately 18 million people required for this event.

7 CONCLUSION:

The main objective of this study was to analyse and visualise the various factors which have contributed to the Evolution of the Olympic Games over the years. This type of analysis is very helpful as this type of analysis can be performed by any Country or Player which can help them in analysing their performance so that they can improve their performance by changing their strategies. We have used a technique named Exploratory Data Analysis which enables you to encapsulate the primary factors of a dataset into a visual format. We selected Python language to implement our work because it is one of the best languages suitable for Data Analysis and is the platform where we have performed this Analysis.

As a result of the Analysis, we can conclude that it is true that Olympic Games have evolved considerably over time from the 1896 Olympic Games till the 2016 Rio Olympics. Various factors provide valid evidence that the Olympics have changed a lot. some of these factors are the launch of the Winter Olympic Games apart from the Summer Olympic Games in 1924, an increase in the number of participating countries in both Summer and Winter Olympics, the Average age of players in the Olympic Games, the increase in the participation of the females in both Summer and Winter Olympics over the time, Total number of medals won by various participating countries over the years, Average height and the weight of Players who contributes to victory of Games in the event. Apart from these, there are many more factors that depict the Evolution of the Olympic Games over time. Visualisation of these factors has been done to explain and validate the Analysis in various Graphical formats like a Line graph, Scatter Plots, Bar, Graphs, etc.

8 FUTURE SCOPE:

We all know that any Analysis is not perfect and it consists of some limitations which define the Future scope of the Research Work. This project work also contains some limitations which we are considering as the Future Scope of the Project. We have visualised our data only in Graphical format. We can also describe the data in other formats like Geographical format where we can depict the countries on the World map. Till now we have only performed Data Analysis using Exploratory Data Analysis. We can also apply various Machine Learning Algorithms to the data set after Analysis and can create a Predictive Model which can predict the statistics of the Future Olympic Games. We can also perform Correlation Analysis on the data set and analyse the relationship between two continuous variables.