iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** VIII **Month of publication:** August 2021

DOI: https://doi.org/10.22214/ijraset.2021.37688

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Insights of the Olympics: Athens to Rio

Rakshit Patel[1], Prince Patel[2], Abhikumar Patel[3], Deepkumar Patel[4]
[1, 2]*Department of Computer Engineering, G.H Patel College of Engineering & Technology, Anand, Gujarat*
[3, 4]*Department of Information & Technology, Government Engineering College, Gandhinagar, Gujarat*

Abstract: ***The Olympics is one of the eye-catching platforms for the sports community as well as sports' lovers in the 21st Century. Sports also adopted the marvelous advancements that took place in vivid fields of technologies. In this modern era of analytics, every category demands prior research before indulging in the actual event. The sports analysis is the primary focus of this paper which mainly displays analysis of comprehensive data about different sport events involved in Olympics taking place from its beginning to the year 2016. The changing time brings new challenges and understanding the tactics along with the playing style of the opponent has become a necessity which basically constitutes to sports analytics. The analysis of the Olympics revealed certain riveting and fascinating facts referring to different aspects and areas. Numerous captivating results can be observed after carrying out research and analysis over the dataset of the Olympics. The analysis implies survey of age wise distribution of the competitors along with the winners' age classification. The medals tally list of countries with respective medals won till Rio Olympics is displayed and studied. In addition, the different venues which hosted the Olympics along with their respective year and number of participants is listed. The competitors with highest numbers of medals also contribute to our research's aim. The Indian Olympic medalists are also focused keeping their participation and winning ratio in consideration. With our research, we would like to light up unexposed and unexplored gospel.***
Keywords: Statistical Data, Sports Analytics, Data visualization, Factors, The Olympics

I. INTRODUCTION

Taking the history of the Olympics into consideration, the ancient Olympics were held in honor of Zeus at Olympia. During ancient times, Olympics gained massive fame such that the time period of four years was labelled as Olympiad and such references were used in common conversation taking place among local people. The ancient Olympics commenced in Athens which later on advanced at global level. The first ancient Olympic champion listed in the records was considered as Coroebus of Elis. There are several religious strings linked to native Greeks regarding the Olympics. Foot Races were considered as one of the earliest Olympic events. With the passage of time, events like wrestling, boxing, javelin throw, pentathlon, equestrian and chariot racing have also been included. In the early phases of Olympic competition, all the contests took place on a single day which later on expanded to four days, reserving fifth day for the closing ceremony. During the 19th century, Greece suffered massive losses due to invasions carried out by various empires, resulting in diminishing popularity of the Olympics. A revolutionary movement began to elevate the downfall of fame and dignity of Olympics. A successful outcome resulted in reinstatement of the Olympics.

With the aim to revive the glory of the Olympics at World level, the International Olympic Committee was established in the year 1894. Moving back to history, it all started in 1896 at Athens where the first modern Olympics was organized at international level. The Olympics observed multiple ups and downs in its popularity considering participation as the major aspect. During its initial phase, the Olympics was organized once every four years which then divulged in two phases including Winter and Summer Olympics held every four years each. The craze of Summer Olympics overcomes winter Olympics as a higher rate of participation is observed in summer Olympics with more than 200 nations taking part in it while just about 92 countries are part of Winter Olympics. To everyone's surprise, the Paralympics started in the year 1989 which provided a suitable International platform to differently abled humans in order to showcase their talent in the respective sport events. Few considerable cancellations were made in the Olympics considering the International plight at that time. It consists of three halts of Olympics all due to war going on among certain nations. The remarkable support and enthusiasm towards sports led to involvement of various new categories at every Olympics. The rate of participants has been drastically elevated in recent years and keeps on increasing at every event.

II. PROBLEM STATEMENT

The Olympics have a history of more than 120 years which makes it extremely hard to get a general overview about several events which took place till date. Assessing such mighty data can be tedious and difficult. So, in order to simplify the process and get a better understanding of the insights of the Olympics, a plain and straightforward analysis will be required. To get an understanding about the players as well as various countries' performance in different categories of sports. Moreover, skimming huge amounts of data can be a time consuming and monotonous task which can result in leaving the task unfinished.

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

III. OBJECTIVE

The main goal of this project is to analyze the previously available dataset of the Olympics from its initiation in 1896 at Athens to 2016 held at Rio. The primary focus was to understand and unveil the unrevealed facts and figures about the historic Olympics events. Along with this, it also aimed at differentiating statistics involved at Summer and Winter Olympics considering various parameters like age and gender distribution of the contestants. Besides this, we also targeted the physical factors such as height and weight of the players to discover their average mass and highness. To lighten up the list of countries which won the maximum number of medals along with candidates possessing the highest individual medals. To get an idea about the participation rate of male and female competitors in the Olympics along with the cities that hosted it.

IV. TOOLS AND METHODOLOGY

The Olympic games is an international sports festival for sportsmen which commenced in 1896 and is still in action with the agenda to widespread the popularity of sports events and to motivate youngsters to indulge in sports activities. In this case, in order to get insights about the Olympics we have used an immense dataset from the time span of 1896 to 2016. To initiate with, the dataset needs to get cleaned and analyzed. For the respective task of exploring the data, we have used a Jupyter notebook which is basically an open-source web application. The necessary code is required to clean and explore the data using Python programming language. Various distinct packages such as Numpy and Pandas have been used to smoothen our process of analyzing datasets. Visuals provide a mesmerizing effect to any analysis as well as provide a supporting hand in easy and proper understanding of the results. Few visuals have been created with the use of Matplotlib and Seaborn. In order to create stunning and fascinating visuals we took the help of PowerBI which is a business intelligence tool. It enables us to create eye-catching and custom visuals with ease and far more simplicity.

V. DATA COLLECTION

The data has been gathered from www.kaggle.com. It has been published by Randi H Griffin on the former platform. The name of the dataset on the platform is referred to as “120 years of Olympic history: athletes and results”. It is a historical dataset on the modern Olympic Games, including all Games held at Athens in 1896 to Rio in 2016. It consists of two separate files. a) athlete_events.csv b) noc_regions.csv.

The file named ahtlete_events.csv comprises 15 columns and 2,71,116 records of individual athletes who have participated in Olympics from the time span of 1896 to 2016. Table 1 displays the description of 15 attributes of the dataset. Mainly, it specifies the names of the athletes who took part in the respective Olympics along with their sex, age, weight and height.

The data also illustrates information regarding the year of their participation as well the corresponding season in which they participated. Apart from this, the category of sports they compete in is also mentioned along with other details. In addition to this, it also states whether the competitor was successful in achieving the medal or not.

The file “country_definitions.csv” consists of 3 columns and 230 records. Table 2 portrays the attributes of the former file. Basically, each record has a 3-letter unique code representing the name of the country/region. The attribute “NOC” is common in both the data files.

Table 1: Information regarding athletes

Athlete Events	
Attributes	**Description**
ID	Unique for each athlete
Name	Athlete's Name
Sex	M / F / Unspecified
Age	Age of athlete
Height	In centimeter's
Weight	In kilograms
Team	Team Name
NOC	3-letter code
Games	Year and season
Year	Year in which game was played
Season	Summer or Winter
City	City in which event was hosted
Sport	Sport category
Event	Event details
Medal	Medal type

©IJRASET: All Rights are Reserved

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue VIII Aug 2021- Available at www.ijraset.com

Table 2: Information regarding Country Code and Region

Country Definitions	
Attributes	**Description**
NOC	National Olympic Committee 3 letter code
Region	Country name used for geospatial mapping
Notes	Real country name if Region isn't an exact match

VI. DATA PREPROCESSING

While working on a huge and real-world dataset, there might be incomplete, noisy or inconsistent data. So, in order to make it properly usable data, it should be cleaned, transformed and standardized as per the requirements.

In this scenario, there are two tables available. We have merged both the tables by performing an inner join on "NOC" attribute. Our final dataset has 17 columns and 2,71,116 records. We didn't find any outliers or missing data for the attributes which have been used for analyzing and finding the insights about the athletes who have participated in various sports events. As the data was normalized and cleaned, there were no major preprocessing steps required for this dataset.

VII.DATA ANALYSIS AND VISUALIZATION

To get insights from the data, one of the necessary steps is to create lucrative visuals in the form of charts and graphs. The visuals are easy to understand and can express a lot about the data in simplified form. We have analyzed 2,71,116 records of athletes from 230 different countries starting from the 1896 Athens Olympics to 2016 Rio Olympics. This data collectively represents information about 120 years of the Olympics events. Till date, it has been hosted in 42 different cities and 66 different varieties of sports events have been organized. By the year 2016, 765 events have taken place and 1184 teams across the globe have participated in the modern Olympics.

Before deep diving into major findings, let's take a quick overview about when modern Olympics started in the year 1896 it started in Athens with just 380 participants representing 9 different segments of sports. The Modern Olympics take place every four year. Over a period of time, the number of participants and sport events surged to a greater height. From the year 1924, winter Olympics started separately and that too take place every four years. There are less number of events and participants in winter Olympics compared to summer Olympics. The visual in the Fig-1 illustrates the rise in the number of participants over the period of time.

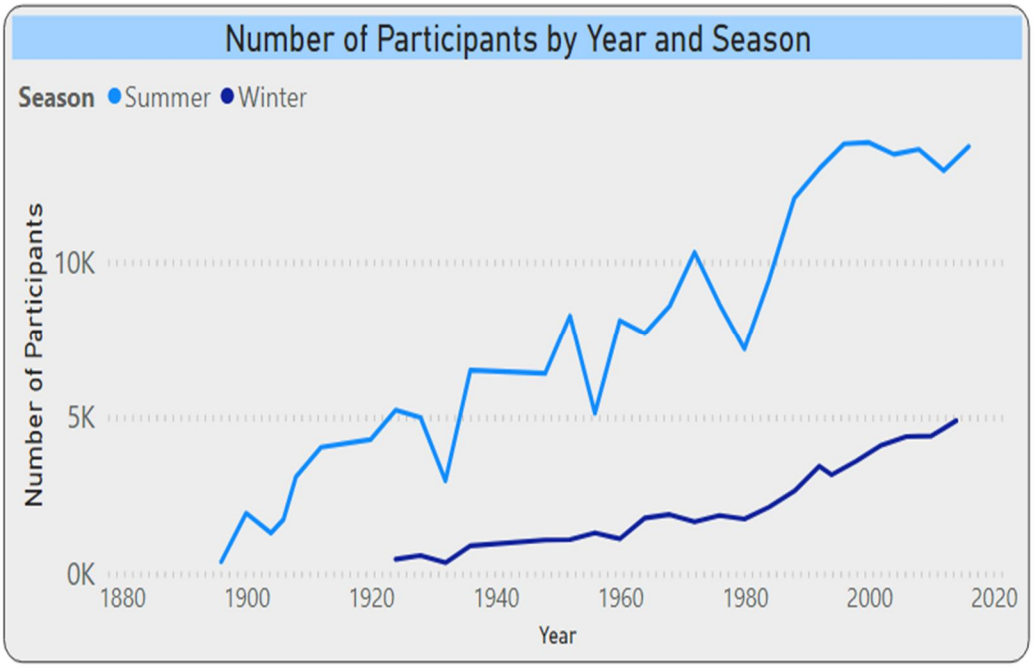


Fig-1: Number of Participants by Year and Season

©IJRASET: All Rights are Reserved

Fig-2 indicates several categories of sports played from 1896 to 2016 during summer and winter. Fig-2 provides a general overview about the number of categories of sports played from its commencement in 1896 to 2016 in the summer as well as winter season. Performing an overall assessment, it can be clearly spotted that there is a significant ascend in the number of categories of sports in both seasons. After 1980, it soared sharply with various new inclusions in different categories. It included more than 30 different sports categories in the summer season while 15 new sports categories in the winter season.

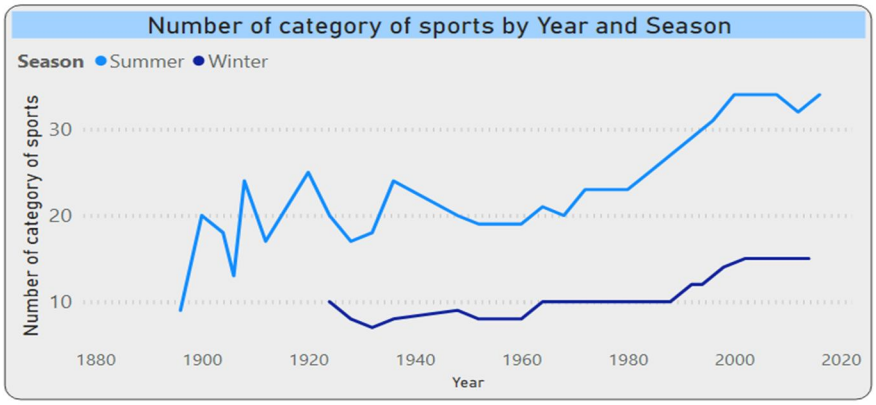


Fig-2: Number of category of sports by Year and Season

Fig 3 and Fig 4 showcase the categories of sports played in both summer and winter seasons respectively. It is a word cloud based on the number of participants. It illustrates that maximum participation can be noted in Athletics and Gymnastics categories during the summer season whereas skiing and skating contribute to highest participation in the winter season.



Fig- 3: Sports category in Summer Olympics



Fig-4: Sports category in Winter Olympics

©IJRASET: All Rights are Reserved

In modern Olympics, contestants have taken part from 230 different countries having an average age of 25.56 years in which average age of male participants is 26.67 years while it is 23.73 years in case of female participants. The average weight of male competitors and female competitors are 75.74 Kg and 60.02 Kg respectively while overall average considering both the genders is 70.70 Kg. Similarly, the average height of men competitors and women competitors are 178.85 cm and 167.83 cm respectively with an overall average of 175.33 cm.

Fig. 5 represents the overall age wise distribution of competitors from all the Olympics events held till 2016. The age distribution of the participants follows the normal distribution. Majority of the participants belong to the age group of 20-25 years which is followed by the age group of 25-30 years. Youngsters contribute to maximum participation in major Olympic events with around 150000 participants ranging from 20 to 30 years of age. Similar number of participants belong to the age group 15-20 and 30-35. A significant drop can be noticed among the age group beyond 35 years in terms of rate of participants.

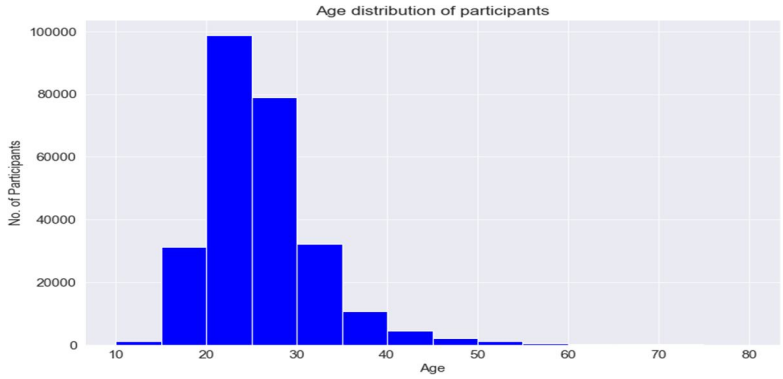


Fig-5: Age distribution of participants in Olympics.

In an overview, approximately 72% of total players were male participants while only 27% belong to the female category considering both Summer and Winter Olympics. Fig 6 and Fig 7 display gender wise participation ratio in Summer and Winter Olympics with the respective year. Fig 6 and Fig 7 also show the top 5 nations in terms of participants during both Summer and Winter Olympics. The year wise ups and downs trend can also be noticed ranging from the year 1896 to 2016 in Fig 6 and Fig 7.

When comparing data of Summer and Winter Olympics, it is observed that more female competitors take part in Winter Olympics compared to Summer Olympics. Along with that, The USA, Germany and Russia can be found in the list of countries with maximum participation in both the seasons of Olympics. Canada and Italy greatly focus on the Winter Olympics while UK and France contestants highly take part in the Summer Olympics.

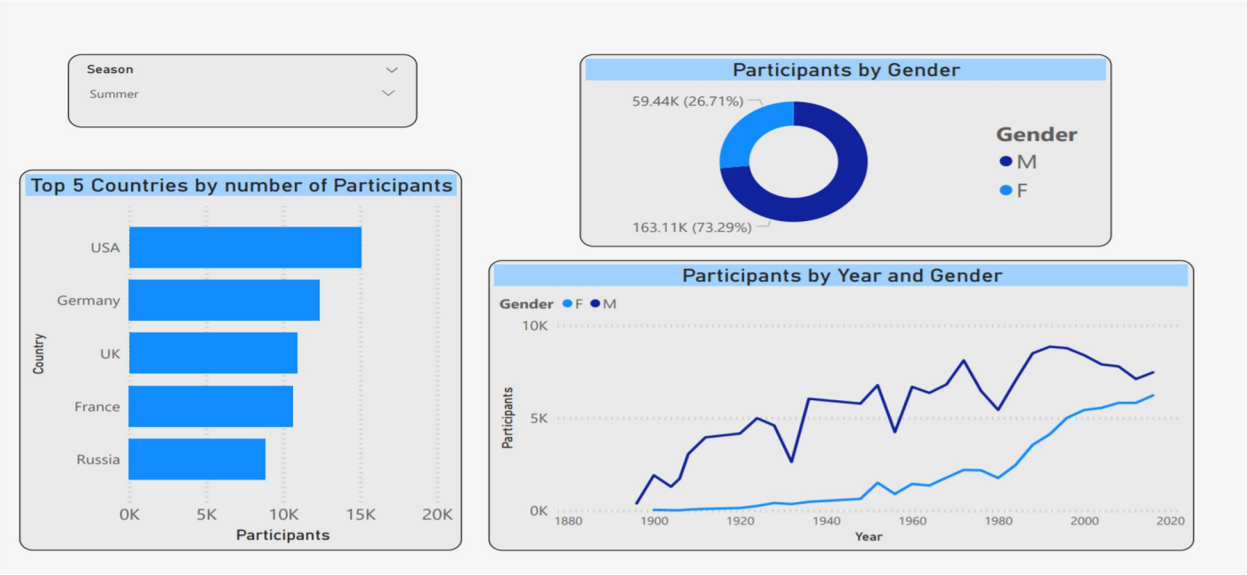


Fig- 6: Participation by Gender, Year and by Country in Summer

©IJRASET: All Rights are Reserved

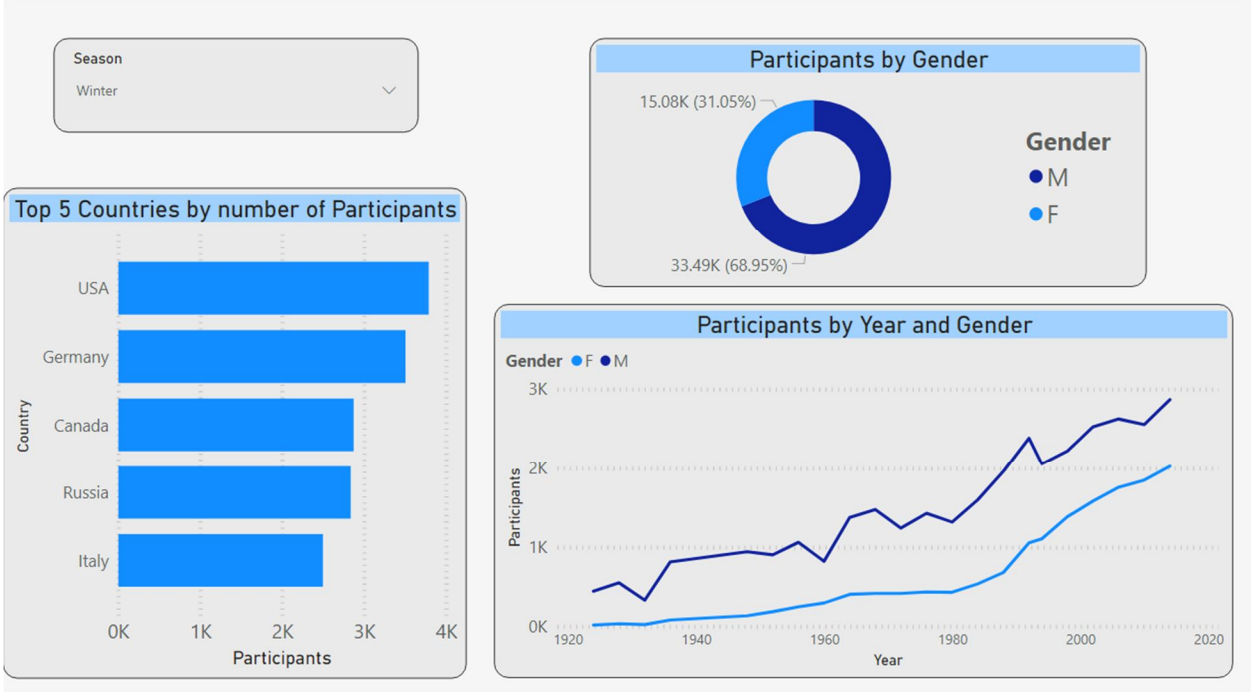


Fig-7: Sports category in Summer Olympics in Winter

In the modern Olympics, in total 39,783 medals have been awarded to the contestants in the past 120 years (1896 -2016) in various categories of sports. To break it down further, 13,372 -Gold medals, 13116 -Silver medals whereas 13,295 -Bronze medals were awarded to winning participants. Out of all, 28.29% medals have been attained by female candidates whereas 71.71% medals have been won by men contestants. Figure 8 showcases medals won by men as well as women candidates.

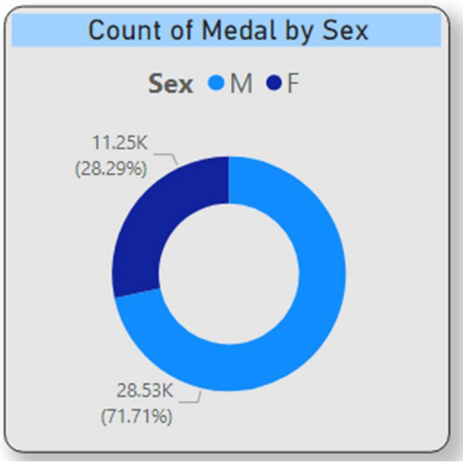


Fig-8: Count of Medal by Sex

Figure 9 illustrates that the USA has won 5637 medals till 2016 which is the highest record in terms of number of medals won. Russia and Germany have almost won the same number of medals. Russia possesses 3947 medals whereas Germany has 3756 medals. UK holds fourth position in the list with 2068 medals. France, Italy, Sweden, Canada and other countries have seized less than 2000 medals. The visuals also unveil the number of gold, silver and bronze medals won by the respective nation.

The second visual in Fig. 9 shows the numbers of medals won in different categories of sports. It can be noticed that nearly 40000 medals have been won in the athletics category which is the highest in any other sport category. Gymnastics and Swimming achieved second and third position respectively.

©IJRASET: All Rights are Reserved

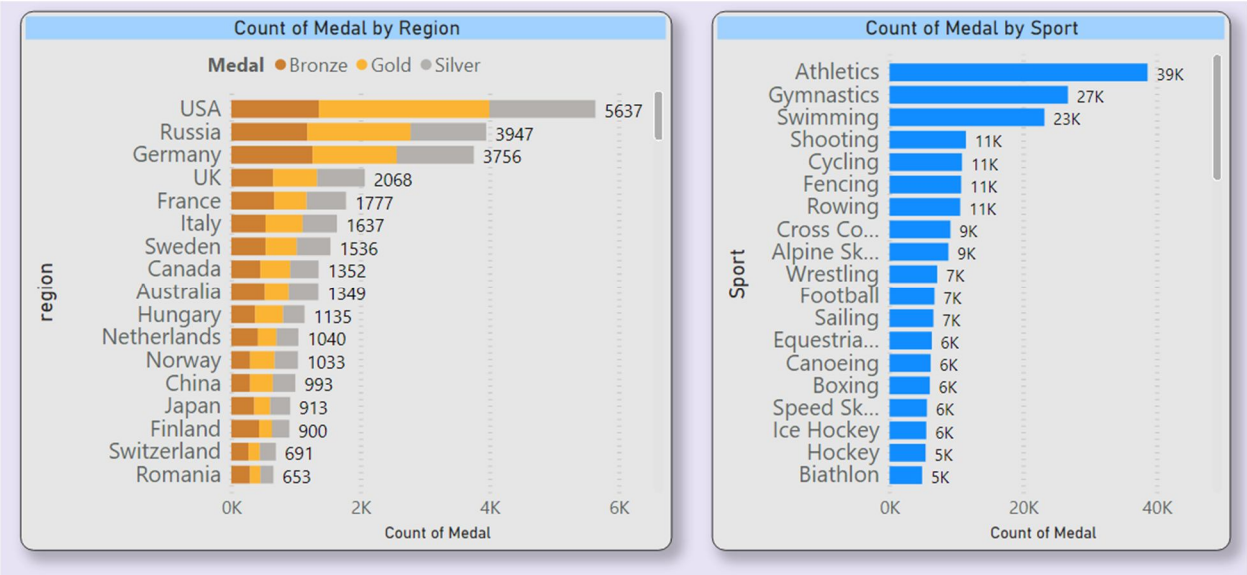


Fig-9: Count of Medal by Region and sports category

Fig. 10 illustrates the age wise distribution of the participants who were successful in winning medals. As an overall assessment, it is clearly observed that most of the winners belong to the age group of 20-25 years which is followed by the age group of 25-30 years. Rest of all age groups have insignificant winners. The most surprising fact unveiled is that the youngest participant won the medal at age of 10 while the oldest contestant won at age of 73 years. After the age of 30 years, there is a significant drop in the number of winners.

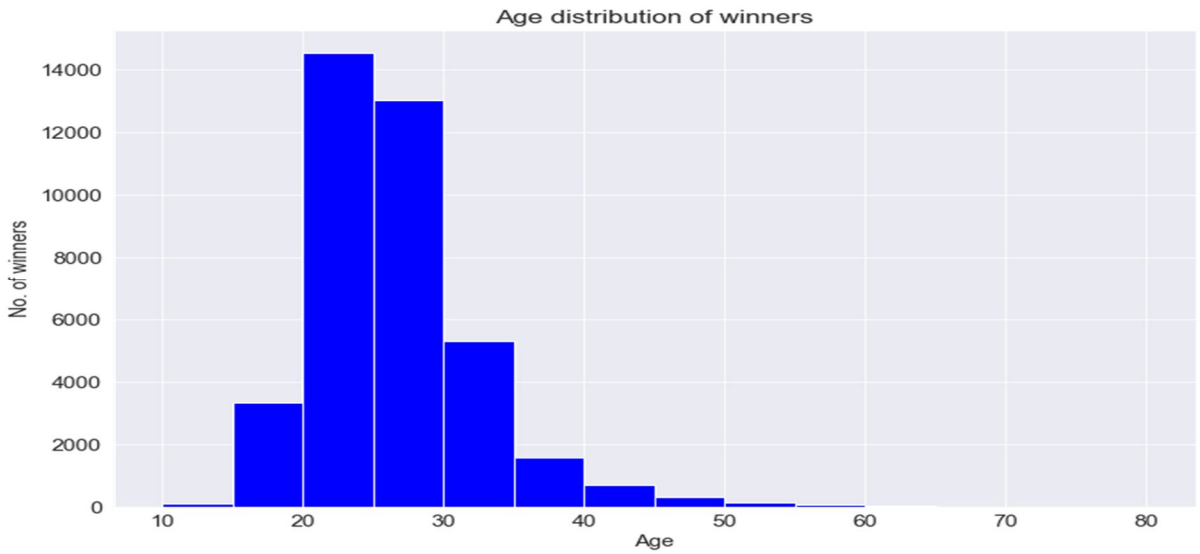


Fig.-10: Age distribution of winners

The data shows that till 2016, Michael Fred Phelps, an American swimmer, possessed the highest number of medals. He achieved 23 gold medals, 3 silver medals and 2 bronze medals which is quite impressive. Also, he holds the record of winning maximum gold medals by any individual. Larisa Latynina, a Ukrainian female gymnast, has obtained 18 medals out of which 9 are gold, 5 are silver and 4 are bronze medals. Figure 11 is a bar chart that portrays name and number of medals won by top 7 winners in maximum medal tally.

©IJRASET: All Rights are Reserved

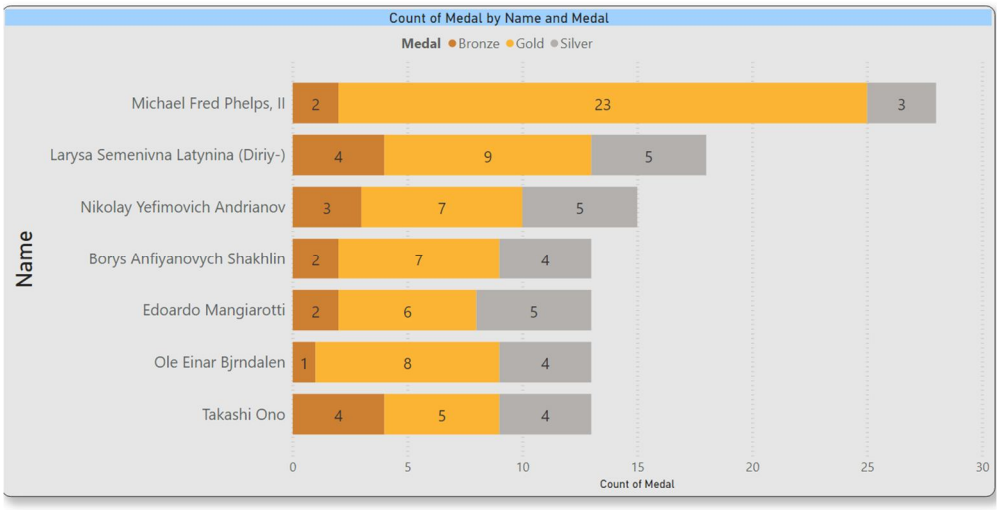


Fig. 11: Highest number of Medal by Name

In the modern Olympics, participants from across 230 different regions took part. In total, 1408 participants have taken part from India in various categories of sports like Hockey (which is national sports of India), Athletics, Shooting, wrestling, Gymnastics and Archery. Figure 12 shows the number of participants by their sports category. From India, 1167 male contestants and 241 female participants have participated till date.

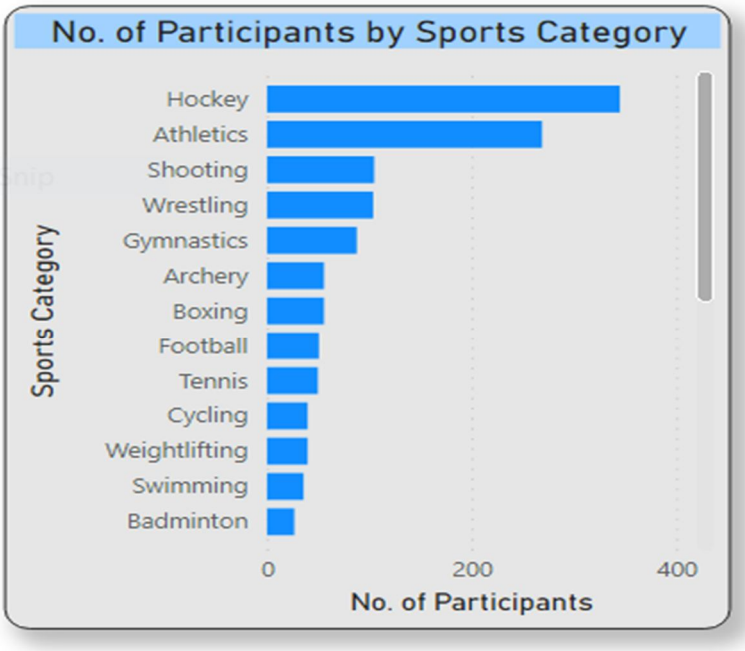


Fig. 12: No. of Participants by Sports Category

Till now, India has successfully achieved 197 medals in the Olympics till 2016. Out of which, 138 are golden medals, 19 are silver medals and 40 are bronze medals. Fig. 13 is a Pie chart describing the number of gold, silver and bronze medals won by Indian contestants. Data reveals that the average age of Indian participants is almost the same compared to players across the globe. But the average weight of Indian participants is about 5% less compared to other international players. The height of Indian participants is 171.64 whereas average height of other international players is 175.33. Overall, India has won 173 medals in Hockey, 7 medals in Alpinism and 5 medals in wrestling till 2016 Rio Olympics. Fig. 14 illustrates the medals won by India in various sports categories.

©IJRASET: All Rights are Reserved

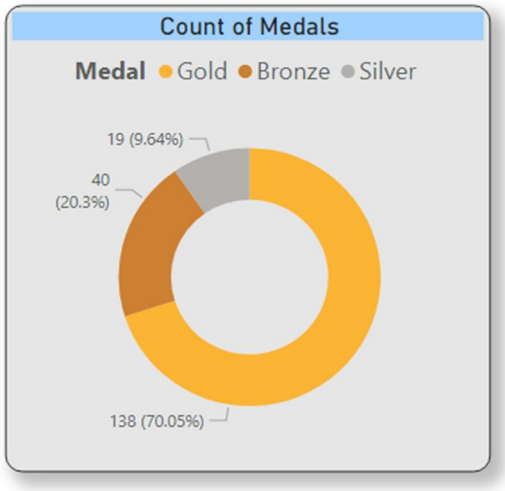


Fig. 13: Medals won by India

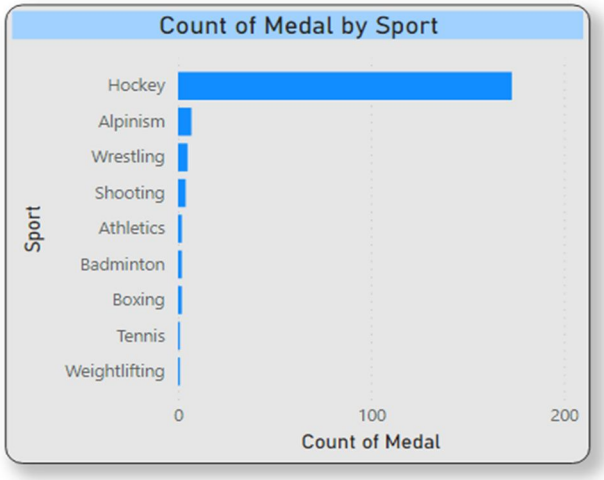


Fig. 14: Medals won by sports category.

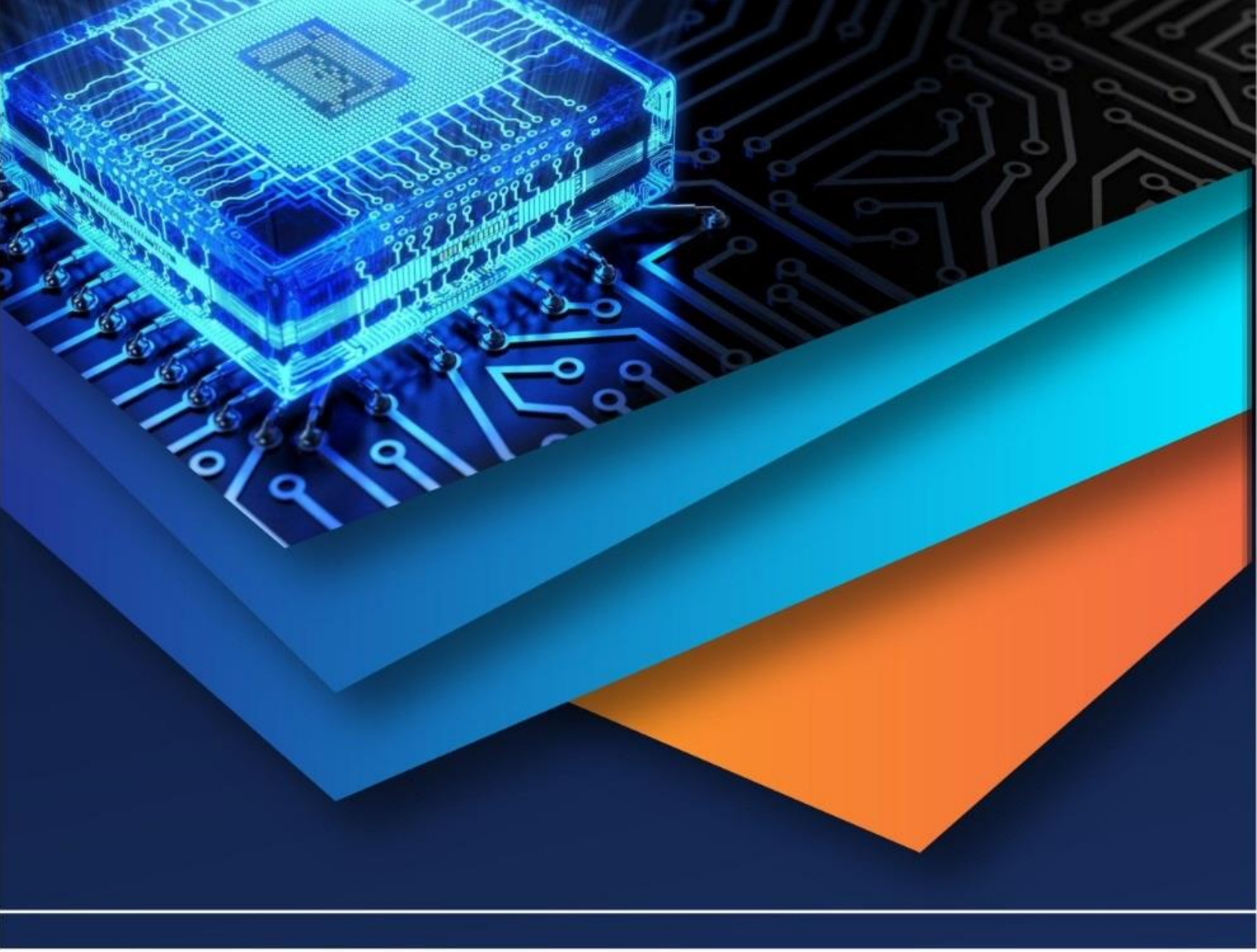
VIII. CONCLUSION

To conclude, it has been revealed that some interesting figures such as the age of the youngest and eldest player who are successful in securing a medal. The list of regions with maximum medals have also been determined along with contestants securing highest numbers of medals in various sports categories. In addition to this, numerous overall information about distinct events have been determined. The general ratio of men contestants to female competitors in both Summer as well as Winter Olympics have been obtained. The information regarding the hosting cities along with their respective year can be observed in association with the rate of participation in that corresponding Olympics.

REFERENCES

[1] https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results
[2] D., Yamunathangam & Kirthicka, G. & Parveen, S.. (2019). Performance analysis in Olympic games using exploratory data analysis techniques. International Journal of Recent Technology and Engineering. 7. 251-253.
[3] Rahul Pradhan et al 2021 IOP Conf. Ser.: Mater. Sci. Eng. 1099 012058
[4] https://olympics.com/en/
[5] Marchi, Leonardo. (2021). Data mining of sports performance data.
[6] https://en.wikipedia.org/wiki/Olympic_Games

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)