

GDP based Medal Count Analysis in Summer Olympics Games for Two Decades - An Exploratory Analysis

Sumathi VP, Vanitha V, Divyadarshini M

Abstract: *The Olympics games started way long back with many participants and winners from all over the world. The game involved in many disciplines and made a bigger impact on the participants and the audience as well. A big data boom is on the horizon, so it's more important than ever to take control of this data. Instinctively this analysis recognise that to perform better than the competitors, this need accurate evidence and data to base the decisions on. The game had its debut in the year 1896 and the progress till now is recorded with the athlete's respective years, disciplines, total medal counts. The goal of this thesis included improved understanding of the competing countries and to develop the players' skills more efficiently for both the extremes (First 10 and Last 10 countries). The analysis is taken by the data of last 5 summer Olympics Games using statistical methods such as correlation factor. Performance analysis is based on the correlation factor with respect to country's GDP (Gross Domestic Product), total medal counts and gold medal counts. This analysis results in an outcome for both extremes meant to amplify the information, which can make the users get higher knowledge about their competitors and country to proceed. There are attributes (year, GDP in million) taken from the dataset and derived attributes (country wise total medal count and country wise men and women athletes count and distinct medal counts for men and women) obtained and analysed to give the knowledge of both extreme countries' (First 10 and Last 10 countries) performance in each year. Finally, the analysed data is plotted in graphs, which can help to find the successes as well as disappointments.*

Keywords: *exploratory data analysis; olympics analytics ; performance; medal count analysis; gross domestic product (GDP); analysis; competitors skillsets; disciplines; regular expository; statistical methods; sample variance; graphs;*

I. INTRODUCTION

The Olympics Games is the most important sports event in the world. There are two categories in Olympics, Summer Olympics and Winter Olympics. The success of a country is marked by the number of medals they obtained, GDP of the country and their overall population. This paper mainly focuses on these factors that are used to analyse the performance of the nation. The Olympics games is an International sports competition. In the Olympics, athletes play in many types of games. Some athletes participate in Summer Olympics

while the others participate in Winter Olympics from same country. Analysis is applied on data discussion is only about the Summer Olympics from [2000-2016] that is two decades. Data about Gross Domestic Product [GDP] and medal count is collected for first ten countries and last ten countries and Exploratory data analysis [EDA] is made with this data. Interesting facts about first ten and last ten winning countries are acquired by this analysis. Each information is depicted as charts to give clearer knowledge about the analysis made with two decades of Summer Olympics data.

Several analyses have already been carried out using the Olympics dataset. This [1] paper infers about the performance of each nation participating in Olympics is analysed based on their medal count. This [2] paper is based on the fans' tweets where the popularity of the player has been identified using sentiment analysis. Similar approach in paper [3] has been done which is based on the GDP and population of the country as first stage process and identifying the medal count for the same as the second stage process. This is called two stage process approach. [4] A key article talks about political implications of Olympics to avoid wars across nations. In detail, when there is a war between countries, only those countries gets boycotted from the Olympics and other countries play to the traditional Olympic Spirit. This paper [5] helps to understand the different kinds of factors affecting the Olympics performance across countries. This narrows down to a significant factor, GDP. In the paper [7], exploratory data analysis was applied over the call data records.

In this paper, methods of Exploratory Data Analysis are applied on the data to obtain the results. The raw data set was obtained from Kaggle and the data related to GDP was extracted manually for each country. The dataset related to the events in the Olympics was obtained from 1896 – 2014. The graph is obtained for the first ten and the last ten countries in the world participated in Olympics from 1896 – 2014. For the ten countries, medal categories like Gold, Silver and Bronze for all countries for men and women are analysed from 2000 – 2016 to get accurate results. For each nation, analysis is carried out for number of athletes participated and the medal count in each category. With the combination of GDP, population of the country and the medal counts, a nation's success is analysed. That is, if a country with highest GDP, it will get the highest medal count. The dataset from Kaggle helps to identify the medal count for all the participating nations and is helpful in

Manuscript published on 30 November 2018.

*Correspondence Author(s)

Sumathi VP Department of Computer Science and Engineering
Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India

Vanitha v, Department of Computer Science and Engineering
Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India

Divyadarshini M, Department of Computer Science and Engineering
Kumaraguru College of Technology, Coimbatore, Tamil Nadu, India

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an [open access](https://creativecommons.org/licenses/by-nc-nd/4.0/) article under the CC-BY-NC-ND license <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

identifying their ranking in the world for each Olympics.

II.OBJECTIVE

Analyze the relation between number of athletes and medal count distribution

- Identify the correlation between GDP(in million) and total medal count year-wise
- Find the men and Women Gold medal count for first ten countries and last ten countries year-wise.

III.DATA COLLECTION

Data is collected manually, and a new dataset is created with 15 fields from 2000-2016 separately. They are country, country code, number of athletes, population (in Billion),

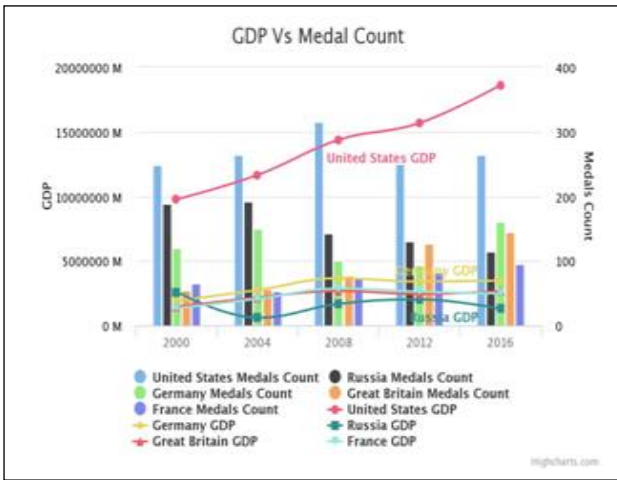


Fig 1.GDP and total medal count year-wise for the first five countries

GDP (in Million), Gold medal count, Silver medal Count, Bronze medal count, Women gold count, men gold count, women silver count, men silver count, women bronze count, men bronze count and total medal count of the country. All these data are not obtained from single dataset. The basic idea of what data to be collected is taken from Kaggle dataset then each fields are actually set to be derived separately from each website as a reference. Total medal count of each country is an derived attribute which is the sum of gold, silver and bronze medal count of that particular year. Gold, silver and bronze medal count is noted separately, further summing up all these the total medal count is obtained. each country’s total medal count is referred with different website to cross check whether it matches correctly with the medal division. Number of athletes participated in each country is searched and collected from different website for each country year wise. By keeping gold, silver and bronze medal count as reference men_gold, women_gold, men_silver, women_silver, men_bronze and women_bronze of each country with year-wise data is collected by surfing through different websites. Data collection is done manually for each column is obtained by surfing through the journals and websites related to olympics and other sport journals. Gross Domestic Product (GDP) of each country year wise data is collected, where the collected GDP of each is given in each way all of that are converted uniformly to million. The dataset contains the GDP in million. Population of each

country is collected in same way and converted uniformly in billion. The present dataset contains population in billion for each country year-wise. These fields are used for further analyses and the graphs are plotted.

IV.DATA ANALYSIS AND VISUALIZATION

Only limited fields from the dataset is taken for analysis. Fields like Medal, Gender, Year and Country are into consideration. For further analysis the data is extracted manually to identify the GDP of each country. At first, from year 1896 – 2016, the first ten countries and the last ten countries are identified and the GDP data set for those countries are extracted manually and the Graph is plotted. For each concept, two graphs are generated. One for the first ten countries and the other for the last ten countries.

Correlation= CORREL (X 1: Xn, Y1: Yn)

$$Correl(X, Y) = \frac{\sum (X - \bar{X}) (Y - \bar{Y})}{\sqrt{(\sum (X - \bar{X})^2) (\sum (Y - \bar{Y})^2)}}$$

This formula is used to find the correlation between any two relationships in the collected data, if it results in negative there is a negative correlation, if it gives the result in positive there is a positive correlation, if this results in zero then no correlation. When the correlation is above 0.5, it results in strong positive correlation and when it is below 0.5, it results in weak positive correlation.

A. GDP and Total Medal Count Year-Wise

First, the GDP (in Millions) for the first five and the last five countries are identified. Total medal count for each of those countries are calculated.

From the above graph, a country with highest GDP will get highest medal count. Therefore, USA has a very high GDP, thus the medal count of USA is higher in successive years. The other four countries have average increase in GDP, thus the medal count for those countries are increasing in the average manner.

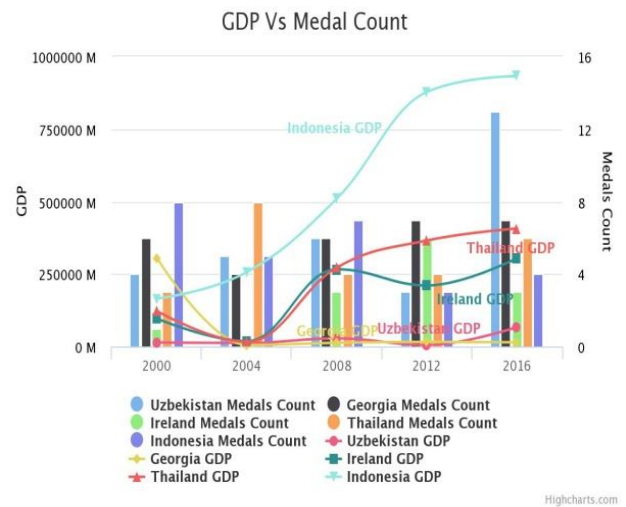


Fig2. GDP and total medal count year-wise for the last five countries



There is a rise and fall in the GDP for last five countries. Thus, there is an increase and decrease in the medal count in the subsequent year

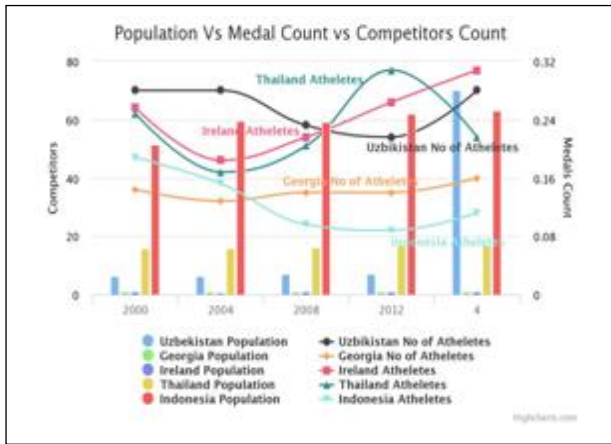


Fig 3. Population with number of athletes in top five countries

B. Athletes and Medal Count Distribution

At first, Number of Athletes and the medal count for each country is determined for 2016 is plotted as graph separately.

No of Athletes Vs Medal Count Distribution (Year 2016)

Source: Wikipedia.org

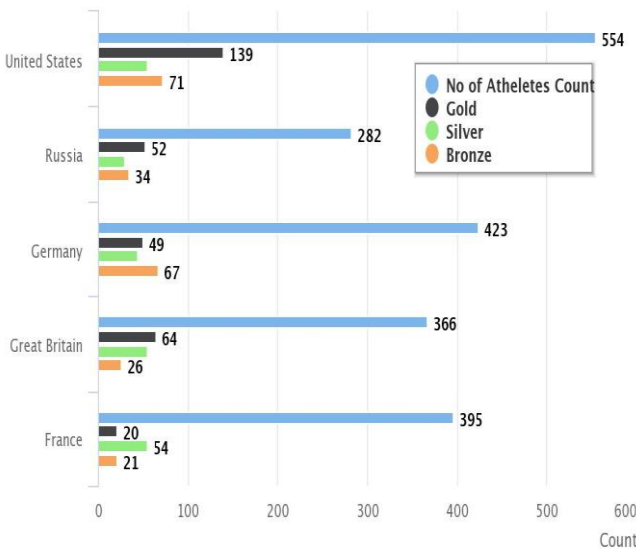


Fig 4. Number of Athletes Vs Medal count Distribution (Year 2016)

As and when the GDP of the country is increased, the number of athletes will also get increased. Based on these two factors, Gold medal count will gradually increases. By this an EDA is made that first five countries are maintaining the same even the number of athletes increase or decrease in number they perform good and get the highest medal count.

C. Population Versus Athletes Participated

A graph based on population of the country, number of athletes participated in each country and the total medal count for each year is plotted.

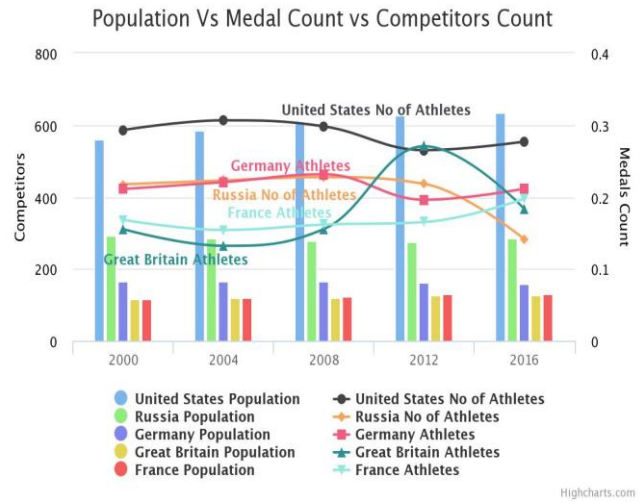


Fig 5. Population with number of athletes in top five countries

D. Correlation Between Gdp (In Million) And Total Medal Count Year Wise

The first ten countries have positive correlation for two decades. As there is a consistency in correlation of greater than 0.5, the graph infers that at this state, there is an increase in GDP and also in medal count.

GDP vs Total Medal Count (Co-Relation Factor)--> Top 10 Countries

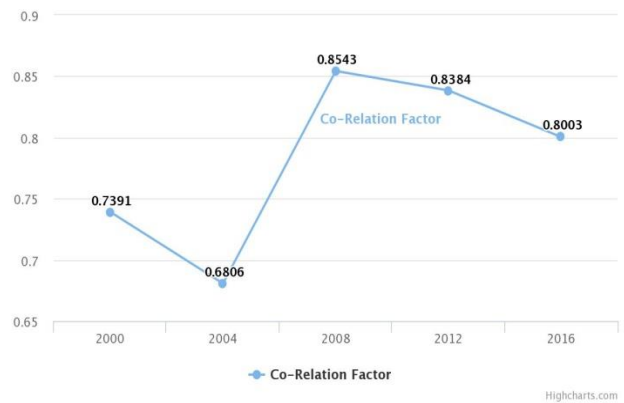


Fig 6. Correlation between GDP and total medal count for first ten countries

GDP vs Total Medal Count (Co-Relation Factor)--> Least 10 Countries

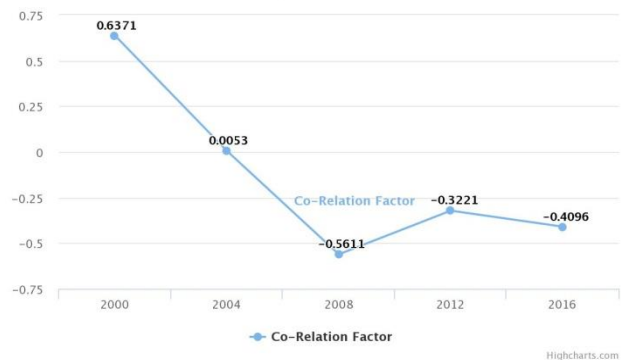


Fig 7. Correlation between GDP and total medal count for last ten countries

There is no consistency in correlation is maintained for the last ten countries. Thus, there is negative correlation factor. The graph infers that there is a rise and fall in the GDP and the medal count.

E. Correlation Between Gdp(In Million) And Gold Medal Count Year Wise

To get accurate correlation between GDP(in million) and Medal counts the analysis is based on the gold medal counts. Thus, it is inferred that GDP and the number of gold medals won by a country in the Olympics are directly related.

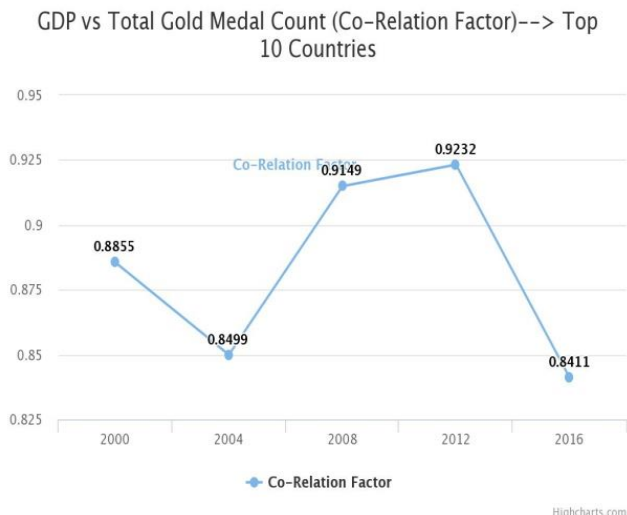


Fig 8. GDP and total Gold medal count year-wise for the top ten countries

The first ten countries have the positive correlation for two decades. Thus, it is concluded that the first ten countries will always have increase in both GDP and Gold medal counts. As their GDP increases, the gold medal count also increases

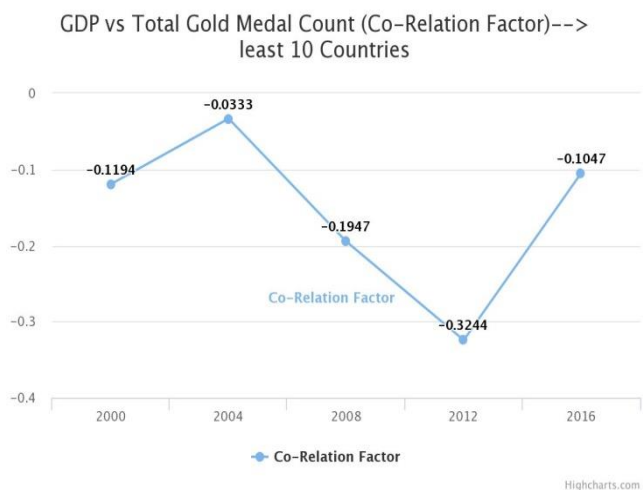


Fig 9. GDP and total Gold medal count year-wise for the top last countries

The last ten countries do not have consistency in their correlation like analysed from the correlation between GDP and total medal counts from above graph. Correlation between GDP and total gold medal count infers that winning of gold medal by these countries occur rarely. Hence this is negatively correlated, there is a fall and rise in GDP and gold medals count.

F. Men and Women Gold Medal Count

An interesting fact about winning gold medal counts by first five countries and last five countries either by men or women or by both is analysed.

a. First 5 Countries

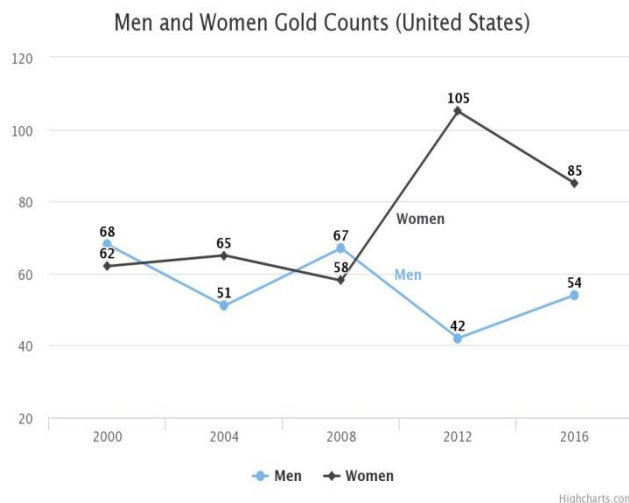


Fig 10. Men and Women Gold Medal Count for United States

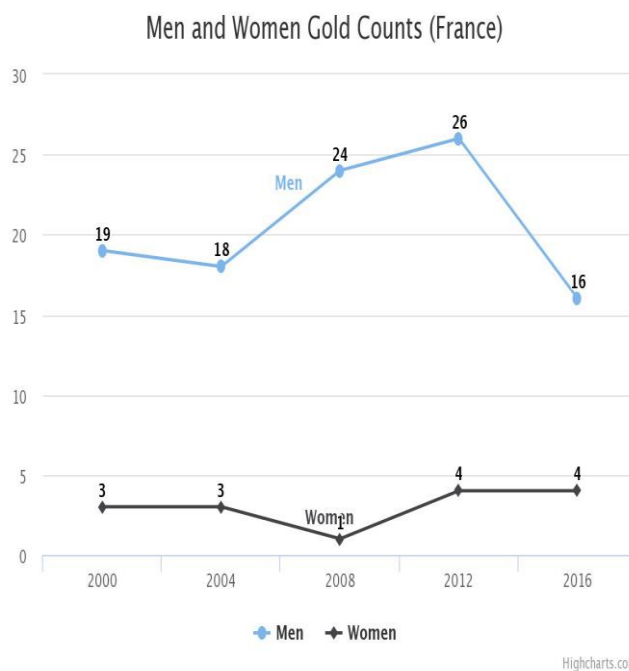


Fig 11. Men and Women Gold Medal Count for France

From the first Five countries the analysed data for two decades infers that in USA men got more number of gold medals than women, in Russia women performs well and got more number of gold medals than men, whereas in Germany and Great Britain both men and women performs equally and maintains the gold medal counts, and in France men got more number of gold medals.

b. Last 5 Countries



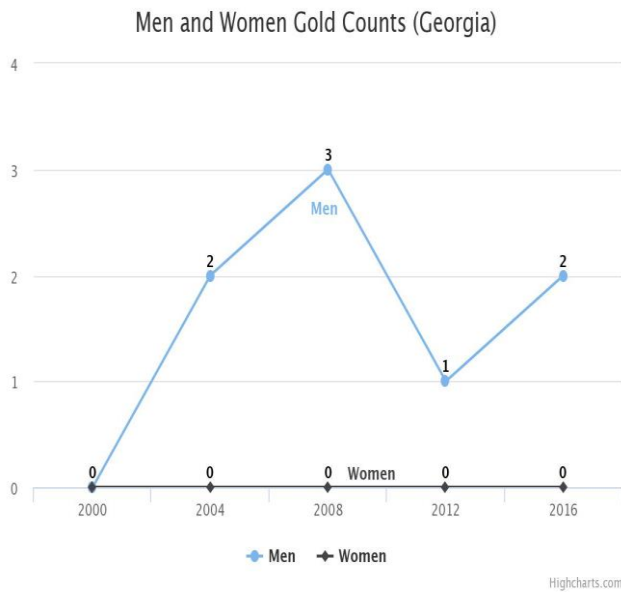


Fig 12. Men and Women Gold Medal Count for Georgia

GDP and total Gold medal count year-wise for the top last countries

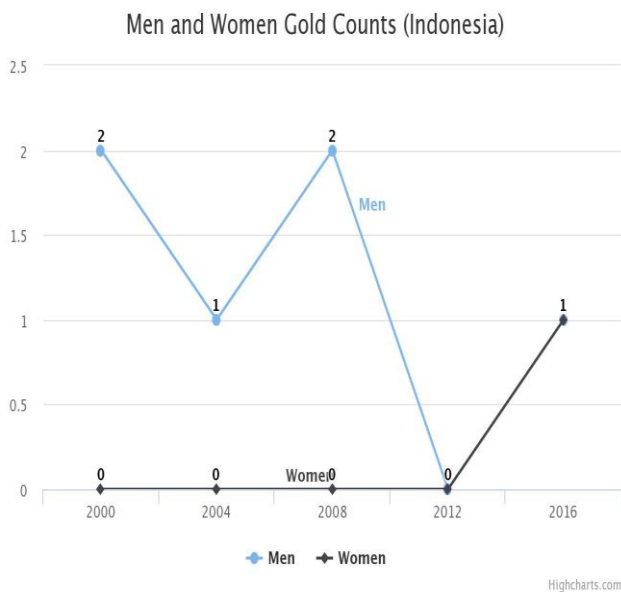


Fig 13. Men and Women Gold Medal Count for Indonesia

From the last Five countries the analysed data for two decades infers that in Georgia no women got gold medals, in Ireland both women and men participated but the gold medal count is only one for all the five summer Olympics. whereas in Uzbekistan and Thailand both men and women performs equally and maintains the medal counts in average, and in Indonesia only men scored gold medals.

V.CONCLUSION AND FUTURE WORK

Thus, the GDP of first ten and the last ten countries are determined and their performance rate has been analysed by finding the correlation factor with GDP and medal count year-wise. The correlation factor is also obtained for GDP with total gold medal count with each country year-wise. Thus a conclusion can be given that country with higher

GDP will have higher Gold medal count. Separate Analysis for identifying the number of medal counts for men and women is carried out. From this analysis, a nation can increase a number of male and female participants as needed to increase their success rate in olympics. This analysis will be more helpful for first ten winning countries as well as last ten winning countries, everyone can able to understand how they perform themselves and true comparison can be made with other countries. This will help in future to see their success and failure rate in olympics. Each country can separately look into their data and they decide how good or bad they performed. Based on the above work, participation rate and success rate of each country in olympics can be improved and each country can maintain the strong positive correlation with GDP and total medal counts. Future work can be done with the same dataset which are collected and referred through various website, to get the gender-wise participant percentage from the total population of each country, that will be more interesting fact for all countries that also make both men and women to get trained with sports skill, so the competitors count will get increased, by this winning a medal in olympics will add even more value and pride to the country.

REFERENCE

1. Lozano S et al., “ Measuring the performance of nations at the Summer Olympics using data envelopment analysis,” Journal of Operational Research Society. vol. 53, May 2002.
2. Yang Yu, Xiao Wang. World Cup 2014 in the Twitter World: “A big data analysis of sentiments in U.S. sport fans’ tweet.Computers in Human Behavior”,vol. 48, July 2015.
3. Yongjun Li et al., “ Performance evaluation of participating nations at the 2012 London Summer Olympics by a two-stage data envelopment analysis”,International Journal of Operational Research. December 2014.
4. Jayantha K, Ubayachandra E G.Going for Gold Medals:” Factors affecting Olympic Performance.International Journal of Scientific and Research Publications”,vol.5.6,June 2015.
5. Akalank Jayakumar ,Title “Is gdp of country and medal count are related”., In blog.September 2016.
6. Hartley Brody, Title:”Global Dominance: Olympics Vs GDP”.,In blog.August 2106.
7. Sumathi, VP, Kousalya, K, Vanitha, V, Cynthia, J, (2018), ‘ Crowd estimation at a social event using call data records’, Int. J. Business Information Systems, Vol 28, No. 2, pp 446-461.