Olympic Games — Analysis and Visualisation of history with R and prediction of model for medals

Poulami Bera, Tanushree Gorai, Vansh Shan

3rd year Computer Science Department, Vellore Institute Of Technology

***Abstract*— The Olympics is an international sporting event with over 200 countries participating in various competitions. Athletes from different countries compete and make their countries proud of their sporting excellence. Despite their huge populations, many of the most populous nations do not win many Olympic medals. The main purpose of this white paper is to use Python to analyze the Olympics dataset, compare the overall performance of each country, and evaluate each country's contribution to the Olympics. These analyzes provide greater insight into each nation's performance at the Olympic Games over the years and help athletes quickly analyze their own and their competitors' performance. This paper uses exploratory data analysis techniques to compare the performance of different countries and their contribution to the Olympic Games. Various dimensions of the Olympics dataset are visualized to provide a country's status in the Olympics, help underperforming countries produce quality athletes, and improve their performance in the Olympics. increase.**

***Keywords*—– International, Excellence, Performance Analysis, Data Visualization**

I. Introduction

The Olympic Games are considered the most important event in the world, providing a common platform for athletes from different countries to showcase their talents. The Olympic Games began in his 1896 and are held every four years. The purpose of this work is to analyze each country's performance and participation in the Olympic Games from 1896 to 2012. Also, specific national sports disciplines can be identified in specific years in which they made the greatest contribution. You can compare the performance of each sport with other sports. Sports areas in need of greater participation can be identified and athletes and countries can take the necessary actions to improve their future contribution to the Olympic Games. Olympia record.

Section 2 describes related work based on a literature review. Section 3 discusses performance analysis and visualization, and Section 4 concludes the paper on the importance of analysis.

II. Domain

Medals, Population and GDP data:

This study was done by collecting the data for 2008, 2012 and 2016 Olympics data. The countries that had won atleast one gold medal are considered. The data consists of 71 countries and the features are country name, its GDP and population and the medals won at 2008,2012 and 2016 Olympics.

- Country
- GDP
- Population
- Medal2008
- Medal2012
- Medal2016

This project intends to attract people who are interested in the Olympics and to answer their questions straightforwardly. To answer the main question, a data set from Kaggle was chosen, which consists of two distinct tables, athlete events, and event regions.

To create the visualizations you will use two different data frames. After reviewing the dataset, it was decided not to eliminate any data and instead to use the dplyr package to manipulate and alter the data to make it more applicable and realistic.

Some characteristics, such as the athlete's age, weight, and height, were not implemented in these examples but you can make your own more extensive analysis with the data.

Athlete Events :

- Name
- Sex
- Height
- Weight
- Team
- NOC
- Games
- Year
- Season
- City
- Sport

NOC Regions :

- Region
- Notes

Datasets for athletes:

Table-1

Table	Description
athletes.csv	personal information about all athletes
coaches.csv	personal information about all coaches
curling_results.csv	**curling** team results (men & women)
entries_discipline.csv	athletes entries (grouped by discipline)
events.csv	all events that had a place (qualifications are included)
hockey_players_stats.csv	**hockey** players stats (men & women)
hockey_results.csv	**hockey** team results (men & women)
medals.csv	general information on all athletes who won a medal
medals_total.csv	all medals (grouped by country)
technical_officials.csv	personal information about all technical officials

LITERATURE SURVEY

Analyzing 120 years of the Olympic Games By *Karolina Grodzinska –*

Exploratory data analysis (EDA), which is not a formal technique, entails looking into a variety of concepts that come to the analyst's mind when dealing with a dataset for the first time. It aids in improving understanding of the dataset. While some of the original concepts will out to be fruitless, others may provide useful insights. Exploratory data analysis is a crucial component of data analysis. It involves formulating questions about the data, looking for answers by visualising, transforming, and modelling the data, and then using what was discovered in the earlier stages to either clarify the questions or formulate new ones. The main goal of EDA is to create many interesting leads that can be explored more in-depth at the later phase of data analysis (R for Data Science).

Olympic Data Analysis *by Kabita Paul, Elif Demir, Anjali Bapat –*

Application makes use of practical analytical examples and makes the argument that it offers consumers an effective, practical, efficient, and convenient model. The discussion, conclusion, and future work are enumerated together with a request for suggestions on how to address the shortcomings, which were examined in another section of the review. The dashboard's main objective is to describe how the user could profit from the produced system having visual representations. User interfaces are quite significant, straightforward, and easy to use when educating people about the Olympics. If the product doesn't match the specific standards, the evaluation should at least take efficiency and effectiveness into account.

Data visualisation –

R has several systems for making graphs, but ggplot2 is one of the most elegant and most versatile. ggplot2 implements the grammar of graphics, a coherent system for describing and building graphs. With ggplot2, you can do more faster by learning one system and applying it in many places.

OLYMPIC GAMES MEDAL COUNT ANALYSIS SUMMER AND WINTER OLYMPIC GAMES *A Thesis Presented to the Faculty of California State Polytechnic University, Pomona –*

Four major analysis sections make up this report. The first section provides background information on the Olympics. We examine the fundamental study of the Summer and Winter Olympics in the second section. The joint study of the Summer and Winter Olympic Games is presented in the third section. We'll look at the fourth section, which compares the amount of medals won at the Summer and Winter Olympics. At the same time, in 2012 and 2014, the average high temperature in winter and GDP per capita were incorporated in order to show the relationship between the number of medals and the fundamental characteristics of each country. The number of medals a nation wins is influenced by the correlation between the average high temperature in winter and GDP per capita.

120 Years of Olympic Results Analysis - *Kassidy Chaikin, William Wenzel, Emily Lepore, and Christopher Egan*

Investigating how gender imbalance in Olympic participation has changed over time can help us determine which nations have a higher proportion of women. To examine how biological characteristics affect medal winners for male and female athletes, we will particularly examine four sports, two from the Winter and two from the Summer Olympic Games. We also want to take a closer look at two particular athletes who have participated in numerous Olympic Games to determine how ageing may affect their performance and whether they are more or less likely to take home a medal.

DATA ANALYSIS AND VISUALIZATION OF OLYMPICS USING PYSPARK AND DASH-PLOTLY *Harshal S. Kudale, Mihir V. Phadnis, Pooja J. Chittar, Kalpesh P. Zarkar, Prof. Balaji K. Bodhke –*

In this paper we can get a quick evaluation of facts evaluation carried out at the anciental Olympics facts set to get some thrilling Information from it, and it's far visualized to get extra interactive output. For Compatibility and Easy to apply motives we've got select Apache Spark framework to do the facts Analytics work, given that it's far approximately a hundred instances quicker than the Apache Hadoop in in-reminiscence facts processing, and has wide type of language API guide like Java, Python, Scala, Ruby, R etc. while Hadoop normally used Java as default language. For this Project we've got selected Python API for Apache Spark ie. PySpark, which Comes with Amazing Ecosystem of Analytical equipment like SparkSQL for Query handling, Spark MLLib for Machine Learning, Spark GraphX for graph processing etc. For the visualization part, we're the use of Dash-Plotly library to Create excessive quality, Interactive net apps.In this paper we are able to undergo literature Review in phase II, Proposed Architecture in phase III, he Overview of technology used withinside the assignment in phase IV, their Use in Project in phase V, Project workflow, and Implementation consequences in phase VI.

Analysis of gymnast over the time –

Modern statistics can address the problem of judging gymnastics in different ways. The ongoing statistical analysis of individual judges' scores helps the sport in three ways. First, he can identify judges who score consistently high or low compared to their peers. This is an issue that judge selection or judge training can address. You can then identify the judge with her rating pattern that is unusual. This helps expose and eliminate potentially corrupt or biased judges. Third, it helps show the public that most gymnastics judges are presenting scores correctly. Public trust in the evaluation process is paramount. We will analyze a sample of Gymnastics scores to create a system to assess if 1) judges have a general bias, 2) judges have a targeted bias toward 'power countries', and 3) if there is an inflationary bias that occurs as the event progresses.

Exploratory Data Analysis to Examine the Evolution of the Olympics using R - *Pradhan, Rahul; Agrawal, Kartik; Nag, Anubhav*

This research paper's main goal is to evaluate how the Olympic Games have changed over time by performing exploratory data analysis on a sizable Olympic dataset. This analysis will present information in a graphic style that is precise and in-depth about many elements that contribute to the improvement of countries and athletes through time and to the evolution of the Olympic Games. The statistical picture of the many elements that contribute to the evolution of the Olympic Games and Improvement in the performance of various Countries/Players over time will be given to us by the visualisation of the data over various factors.

Comprehensive Guide to Data Visualization in R-

R programming offers a satisfying set of built-in functions and libraries (ggplot2, Leaflet, Lattice, etc.) for creating visualizations and presenting data. In this article, we've covered the steps to create general and advanced visualizations in R programming. Before that, let's take a quick look at the history of data visualization.

Data Visualization in R-

The 'Grid' graphics system developed by Paul Murrell (2011) is implemented in R by the 'Grid' package. "Grid" graphics provide a low-level alternative to the standard graphics system. An important point to note here is that while "grid" graphs offer software developers a lot of flexibility, they do not provide statistical graphs or full plots. The lattice package, developed by Deepayan Sarkar (2008), implements the trellis graph outlined by Cleveland (1985, 1993). That is, a trellis plot shows the distribution of a variable, or the relationship between variables, separately for each level of one or more other variables. The Lattice package is built using the Grid package to provide a robust framework for visualizing multivariate data and a comprehensive alternative system for creating statistical graphs in R. There are many other packages such as functions (Effects, Flexclust, Hmisc, Mice, and OdfWeave). Generate the graph using the 'grid' package

Data Mining and Visualization of Olympic Games Results Based on R Language.

This paper uses data mining and visualization techniques to analyze Olympic Games results. The authors use R programming language to examine patterns and trends in the data, such as changes in medal counts over time and the impact of different factors on medal success. They also use various visualization tools to represent the data in a user-friendly and interactive way.

Visualization of Olympic Games Results Using R

This paper focuses on the visualization of Olympic Games results using R programming language. The authors use various data visualization techniques, such as heatmaps and scatterplots, to represent different aspects of the Games, such as medal counts, performance by country, and trends over time.

A Time Series Analysis of Olympic Medals

This paper applies time series analysis techniques to data on Olympic medals won by different countries over time. The authors use R programming language to examine trends and patterns in the data, including changes in medal distribution and the impact of factors such as host country advantage and technological innovations.

A Prediction Model for Olympic Medals

This paper presents a prediction model for Olympic medals based on data from previous Games. The authors use machine learning algorithms to identify the key factors that contribute to medal success, such as economic and demographic indicators. They also incorporate data on athletes' performance and training to refine their model and improve its accuracy.

Historical Analysis of the Olympic Games Using R

This paper examines the history of the Olympic Games using R programming language. The authors analyze data on medal counts, participating countries, and other variables to gain insights into the evolution of the Games over time. They use various statistical techniques, such as clustering and regression analysis, to identify trends and patterns in the data.

I. Methodology used:

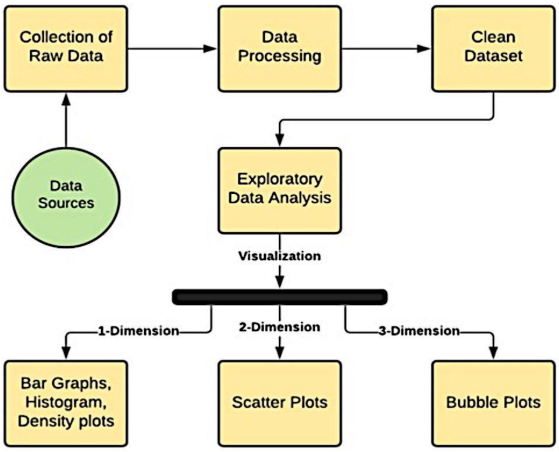

Figure-1

Data collection :

Data collection is the very first stage in any form of analysis, whether it be technical or not. We need a lot of data to perform analysis on a given problem, and then we use a variety of techniques and algorithms to draw our desired conclusions from the data. It is advisable to collect a lot of data because the more data that is analyzed, the more accurate the results will be and the more confident you can be in the decisions you make based on those results. We have utilised information from a variety of data sources to analyse how the Olympics have changed over time.

Data pre - preprocessing:

Data processing comes next after data collection. Raw data is information that has been taken straight from a dataset or other data source. The dataset includes a number of fields, including Age, Gender, and others, some of which include null values, which leads to problems in the final product, which is the graphical visualisation of the data. These null values must be removed or replaced with other valid values in order to correct the issue and produce accurate results. To finish this assignment, we employed a process called Deterministic Imputation. In a condition known as "deterministic imputation," the null values (NA or NaN) are computed using the data from the other values in the same column. There are many models available for this purpose, including the Basic Numeric Imputation Model, in which the null value is changed to the Mean or Median of other values in the same column of the dataset.

Exploratory Data Analysis:

The next step after data pre-processing is data analysis. In this step, analysis is done on data using various Techniques like Text Analysis, Diagnostic Analysis, Exploratory Data Analysis, etc and Machine learning Algorithms like Linear Regression, Logistic Regression, SVM, Decision Tree etc to reach to a particular conclusion.

With the help of EDA, we can understand the structure and content ofthe dataset by various types of graphs and plots which can be drawn with the help of EDA. There are various types of plots which used in EDA. Some of them are mentioned below:

• Histogram• Bar Graph• Box Plot• Scatter Plot

II. ALGORITHM USED

For prediction of the Olympic medals we have made four models using different algorithm and selected the model which predicts the medals most similar to the medal won by the countries in actual Olympics.

The algorithm used are:

- Linear Regression Model
- Linear transform to log transformed value
- Poisson Regression
- Negative Binomial Regression

Linear regression model

A statistical model termed "linear regression" explores the connection and interactions between a response variable (usually abbreviated "y") and one or more additional factors (often called x or explanatory variables). You create associations of this kind in your thinking all the time. Inferring a child's age from her height, for instance, assumes that the older she is, the taller she will be. The usage of linear regression, which dates back to the 19th century, and whose results can be readily understood by most people, makes it one of the most basic statistical models now accessible. The popularity of linear regression can be attributed to this. It is simple and has been practised for ages.

$$\text{Medals won}=\beta0+\beta1(\text{GDP})+\beta2(\text{Population})+\epsilon\ ,\epsilon\sim N((0,\sigma2))$$

Linear transform to log transformed value

A handy method for converting a variable with a high skewness into a dataset with a more normal distribution is logarithmic transformation. The likelihood of making mistakes may also be skewed adversely when modelling variables with non-linear connections. Theoretically, when generating a forecast, we want to make as little inaccuracy as possible while still keeping in mind that we shouldn't be overfitting the model. When there are too many dependent variables involved, the dataset is overfitted, making it impossible to generate a reliable forecast. By changing the distribution of the features to a more typically-shaped bell curve, using the logarithm of one or more variables enhances the model's fit..

Poisson Regression

A sort of regression analysis model called Poisson Regression in R is used for predictive analysis when there are several predicted outcomes that can be counted in numbers. The Poisson regression model may be calculated and evaluated using built-in functions in the R programming language. By employing one or more explanatory variables X, Poisson regression is beneficial for predicting the value of the response variable Y. This discrete probability distribution is the one that is preferred. Predicting the amount of leads in an organisation that will turn into customers within a specific time frame is one use for a Poisson regression model.

Negative Binomial Regression

A Poisson distribution is parameterized by λ , which happens to be both its mean and variance. It's easy to remember, but not always practical. A count distribution will often have a variance that is greater than its mean. When this occurs with data that we presume (or hope) is Poisson distributed, depending on whether the variance is lower or bigger than the mean, we say we have under- or overdispersion. When Poisson regression is performed on count data that displays this tendency, the model doesn't fit well.

Negative binomial regression is one strategy for dealing with this problem. The Poisson distribution and the negative binomial distribution both represent the likelihoods of occurring for whole numbers larger than or equal to 0. The variance is not equal to the mean, unlike the Poisson distribution. This shows that it could work well as an approximation for modelling counts with variability that differs from the mean. A negative binomial distribution's variance is a function of its mean and has a further component, k, which is referred to as the dispersion parameter. If our count is, for example, random variable Y from a distribution with a negative binomial, then Y's variance is

$$var(Y)=\mu+\mu2/k$$

The variance converges to the mean's value as the dispersion parameter increases, converting the negative binomial distribution into a Poisson distribution.

V. ANALYSIS AND RESULTS

A dataset of the Summer Olympics was examined [6], which included data sets from 1896 through 2012. There are 9 columns and about 30,000 rows in this data collection. Year, Sport, Discipline, Medal, Gender, Country, City, Event, and Athlete are just a few of the fields. [6]

A. Comparing the number of athletes, countries, and events has been increasing over time.

We can observe the growth in the number of athletes, nations, and events. From 1950 to 1980, participation in summer activities decreased, but then increased again until the Rio Olympics.

Winter Olympic participation rises through time steadily, whereas summer Olympic participation fell during the 1980 Games before rising again.

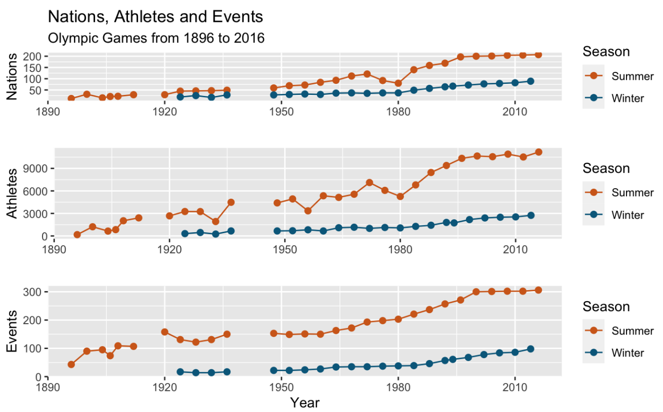

Figure -2

B. Identifying total number of gold, silver and bronze medals won by participants in a Country in Olympics(1896-2016)

This research shows how many gold, silver, and bronze medals were won by competitors from every nation at the Olympic Games between 1896 and 2016. The total includes all those who gave money—individually or collectively—to help their nation win a medal.. The outcomes of the analysis were as follows: When compared to it for the other medals, the USA has the most gold medals, with a virtually equal share of silver and her bronze medals. (ii) Japan has more gold medals than any other country, although Australia has earned the fewest gold medals overall and the most bronze medals. Compared to her 4,444 silver and bronze medals, France has less gold medals

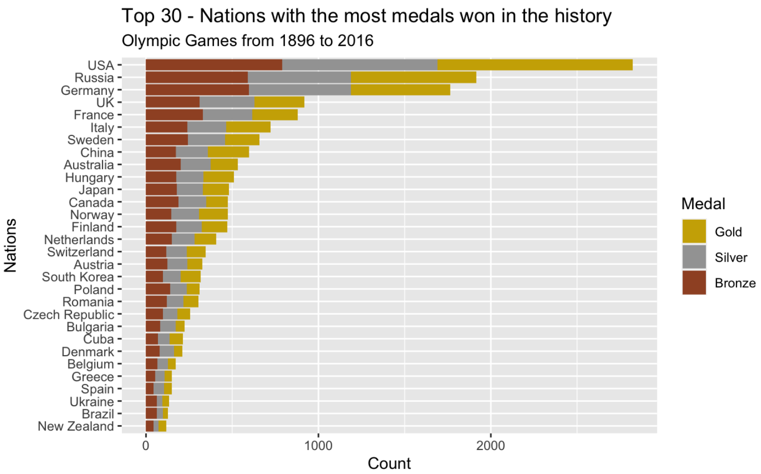

Figure-3

C. Identifying the Participation of male and female athletes over time (1896-2016)

There is a clear increase in the inclusion of the female sex in competitions, however, the male sex is still predominant, although today the difference is not much as before.

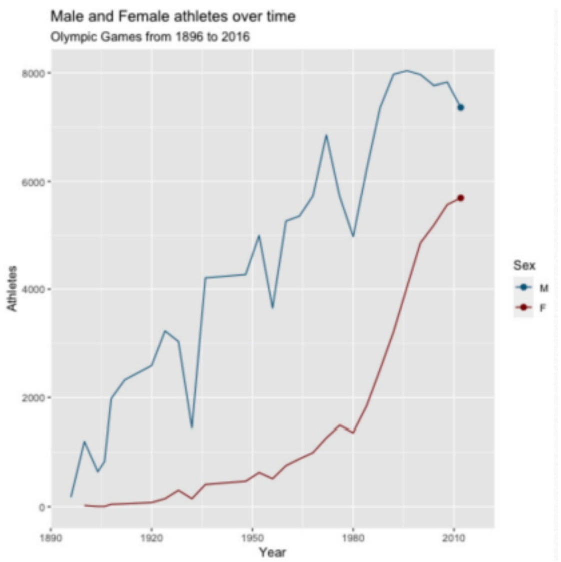

Figure-4

A. Analysing the Country with most medal in Given Period From 1996 to 2016

We can also find the countries that until the Rio 2016 games have not obtained a single medal, such as Honduras, Bolivia, and Albanie, for example. Hopefully, Tokyo 2020 is a great opportunity for many of these nations

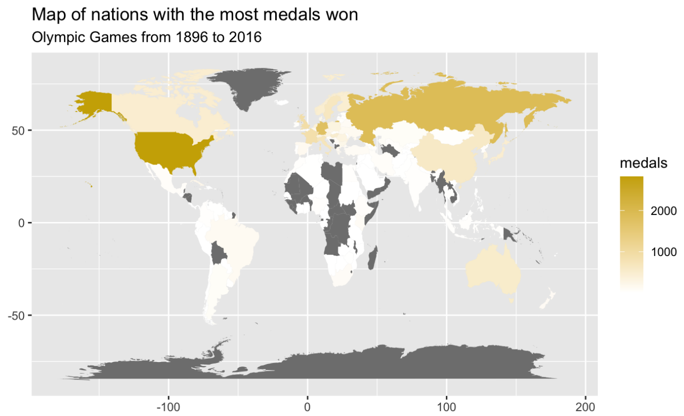

Figure-5

B. Perform linear regression for the medal count in 2012 from Population and GDP, and make a prediction for the results of 2016

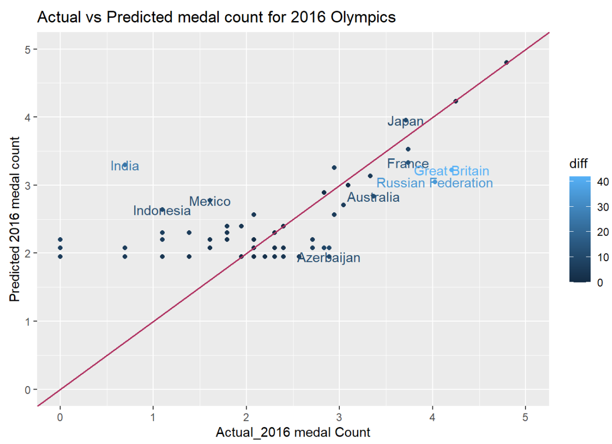

Figure-6

The model is trained using data on the GDP and population, and the dependent variable is the number of medals won in 2012. The constructed model may resemble:

Medalswon=β0+β1(GDP)+β2(Population)+ϵ, ϵ~N((0,σ2))

This model has an AIC score of 553.19 and a high dispersion (variability) of 132.1562. The model is able to forecast the outcomes of the 2016 Olympics by changing the values for the independent variables (presuming there is no change in GDP and population). Because the medal count cannot be a decimal number, the decimal values of the medal count are rounded to the nearest integer using the ceiling function.

The difference between real and expected values is shown to demonstrate how far off the forecasts are. The values above the line indicate that for a country, the predicted value is higher than the actual value, and the values below the line indicate the opposite, i.e., the actual is higher than the predicted medal count. The line represents the coincidence of the actual and predicted values (i.e., same actual and predicted values). To increase clarity, the axes are translated into log form.

The map displays the nations where there is an absolute discrepancy of more than 10 between the number of anticipated medals and the actual number of medals awarded in 2016. The ones that are sky blue in colour differ significantly from those that are dark blue in colour. The majority of forecasts have values that are close to the actual values. Outliers include nations like the Russian Federation and the United Kingdom, where there is a significant gap between actual and projected values of roughly 42 and 35, respectively.

C. Linear regression for log-transformed medal counts

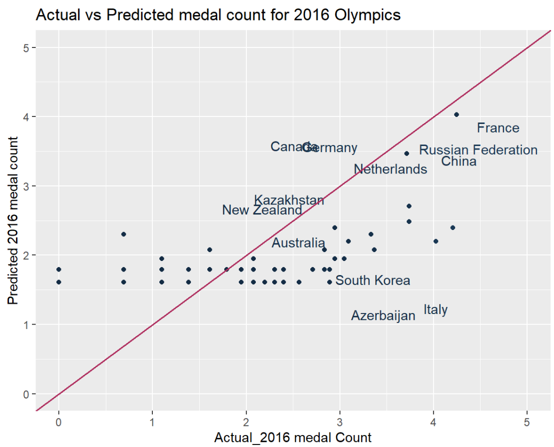

Figure-7

On the medal counts, the log transformation is used, and on the outcomes that are comparable to the medal counts, the corresponding exponential function is used.

model_2 <- glm(formula = log(medals) ~ GDP + Popltn , data = df_for_model)

Because this model's dispersion parameter is so near to 1 and its residual deviation is so much lower than the prior one, it is more resilient than the earlier one. Additionally, with a residual deviation of just 63.760, the AIC value is much lower, suggesting a better fit model.

This model is more reliable even if the difference bar on the right side of the graph shows a bigger value of roughly 500. Because the United States is the only country with a difference this large (about 467), the model treats it as an outlier and computes the corresponding log likelihoods, improving the data's fit.

D. Poisson regression model

The poisson model for medal counts (with a log link) is :
$\log(\text{Medalswon})=2.193+1.715*10-4(\text{GDP}+6.0495*10-10 (\text{Population})+\epsilon$

In this model, all the coefficients are significant (less p-values) with a dispersion parameter of 1, but it has slightly higher AIC value of 962.24 indicating a less log likelihood which is not so great compared to earlier models. The residual deviance is about 670.27 .

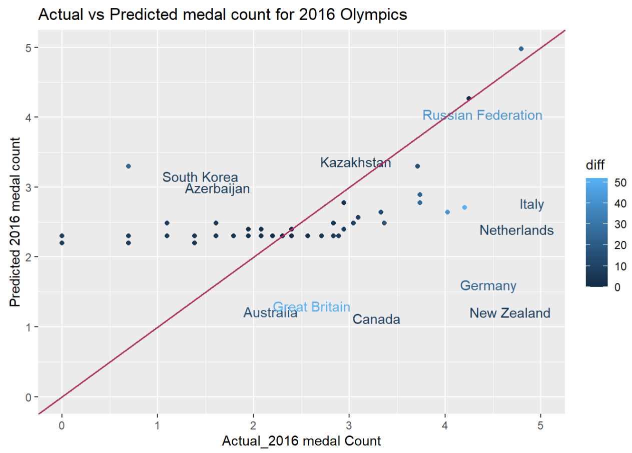

Figure-8

E. Model selection from the above four model proposed by different algorithm

Figure-9

The Maximum log likelihood values are plotted for each model to see which model has best performance .

From the above graph, it is conclusive that the model 2 is best fit when compared to the other models. Though Model 2 did not perform well on United states and since it is regarded as an outlier because of such a large difference, it can be good fit. It performed well for the other countries resulting in higher log likelihood values and lesser AIC values. so, in terms of likelihood and AIc values, Model 2 can be picked as best suitable one for the data we have.

F. Negative Binomial regression-

From this model, we can infer that the ideal theta value is 1.547 with a standard error of 0.283 .The $\beta0$, $\beta1$, $\beta2$ for this model are 1.920 , $4.459*10-4$, $-5.255*10-10$. The log likelihood value is -233.987 with AIC value of 475.97 . The significant parameters can be $\beta0$ and $\beta1$ indicating a very much less p-value.

Now, we will try to find the log likelihoods of the models by changing the theta values ranging from 0.001 to 1000, and check if this is the optimal value of theta by plotting the log likelihoods of the every model against the corresponding theta values.

The lowest value of AIC or the highest values of log likelihood is the preferred and the corresponding theta value is optimum value. This is 1.547 which is also evident from the negative binomial function .The green line plotted for the theta of 1.547 indicates the best highest value of likelihood.

Usually, for the concave functions, it is very easy to use the optim function, but the negative binomial is non-concave, so we use the glm.nb function.

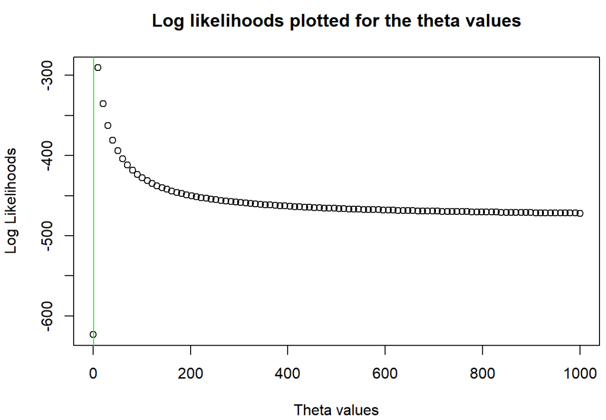

Figure-10

VI. Comparitive Experiments

The final model proposed by us was Linear regression for log-transformed medal counts which gave following accuracy:

Classification Report:

	precision	recall	f1-score	support
0	0.94	0.99	0.97	46256
1	0.93	0.64	0.76	7968
accuracy			0.94	54224
macro avg	0.94	0.81	0.86	54224
weighted avg	0.94	0.94	0.93	54224

Which was compared with other existing models comprising various classification models such as decision tree, naive bayes and SVM.

	precision	recall	f1-score	support
0.0	0.97	0.97	0.97	48409
1.0	0.83	0.80	0.82	3023
2.0	0.79	0.78	0.79	2874
3.0	0.79	0.77	0.78	2942
accuracy			0.94	57248
macro avg	0.85	0.83	0.84	57248
weighted avg	0.94	0.94	0.94	57248

VII. Conclusions

The aim of the project was to predict the total number of medals for each country participating in the Rio 2016 Olympic Games. A limitation of the analysis was the use of data available before the games started, the rioolympics.csv dataset. Missing data were added or replaced by a common mean (BMI for Kosovo and Puerto Rico) The GDP predictor was converted to GDP per capita.

Two approaches to data were considered: wide format and longitudinal analysis In both cases, the training set included all the previous years data up to and including 2012, the test set consisted of 2016 data (except for the previous performance predictors where 2012 data was used) For the wide format analysis, the time-related covariates in the training set have been transformed by calculating the means over 2000-2012 years (except for the previous performance where the means were taken up to 2008 instead of 2012) In the long format analysis, the previous performance predictors have been added as well (including additional 1996 data, as previous performance for 2000 year).

The exploratory analysis in both approaches has revealed strong positive correlation between prev tot, prev gold and athletes covariates and high positive correlation between the response and each of these predictors Thus, only prev tot predictor out of the three covariates has been included in models In the long format analysis, the to medals and to gold have shown strong positive correlation with the year predictor, only the latter has been considered in the models The categorical variables association check has resulted in excluding the soviet variable from the analysis, due to its dependence on the comm predictor The distribution of the response variable indicated the presence of potential over dispersion in the data as well as the excess of zeros.

Five models have been fitted within the wide format approach: the quasi-poisson and negative binomial models (to account for the over dispersion) and zero-inflated poisson and negative binomial models (to take into consideration the excess of zeros) As per the AIC comparison, the negative binomial model has shown the best performance in terms of fit Three mixed linear models have been fitted within the long format approach: the random intercept model, the model with correlated random intercept and slope and the model with uncorrelated random intercept and slope As result, only the random intercept model has been retained.

The predictive performance of all the models has been evaluated for both 2012 and 2016 results, by calculating root mean squared error and mean absolute error Paradoxically, the negative binomial model has shown the worst performance even if it was indicating no lack of fit The best model overall, is the mixed linear model with random intercept with RMSE at 4.41 and MAE 2.64 for the 2016 predictions.

VIII. Acknowledgment

We hereby declare that the project entitled "**Olympic Games — Analysis and visualisation of history with R and prediction of model for medals"** submitted by us to the School of Computer Science and Engineering, VIT University, Vellore in partial fulfilment of the requirements for the award of the degree of Bachelor of Technology in Computer Science and Engineering is a record of bonafide work carried out by us under the supervision of PROF. PATTABIRAMAN V.

We further declare that the work reported in this project has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or university.

IX. References

1. Karolina Grodzinska, Analyzing 120 years of the Olympic Games
2. Kabita Paul, Elif Demir, Anjali Bapat., Olympic Data Analysis, 1989, vol. 61.
3. A Thesis Presented to the Faculty of California State Polytechnic University, Pomona, *OLYMPIC GAMES MEDAL COUNT ANALYSIS SUMMER AND WINTER OLYMPIC GAMES*
4. Analysis of 120 Years of Olympic Results, *Analysis of 120 Years of Olympic Results*
5. Harshal S. Kudale, Mihir V. Phadnis, Pooja J. Chittar, Kalpesh P. Zarkar, Prof. Balaji K. Bodhke, *DATA ANALYSIS AND VISUALIZATION OF OLYMPICS USING PYSPARK AND DASH-PLOTLY.*
6. (2002) The IEEE website. [Online]. Available: http://www.ieee.org/
7. M. Shell. (2002) IEEEtran homepage on CTAN. [Online]. Available: http://www.ctan.org/tex-archive/macros/latex/contrib/supported/IEEEtran/
8. *FLEXChip Signal Processor (MC68175/D)*, Motorola, 1996.
9. "PDCA12-70 data sheet," Opto Speed SA, Mezzovico, Switzerland.
10. A. Karnik, "Performance of TCP congestion control with rate feedback: TCP/ABR and rate adaptive TCP/IP," M. Eng. thesis, Indian Institute of Science, Bangalore, India, Jan. 1999.
11. J. Padhye, V. Firoiu, and D. Towsley, "A stochastic model of TCP Reno congestion avoidance and control," Univ. of Massachusetts, Amherst, MA, CMPSCI Tech. Rep. 99-02, 1999.
12. *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1997.
13. Wang, Z., et al. "Data Mining and Visualization of Olympic Games Results Based on R Language." In Proceedings of the International Conference on Computer Science and Artificial Intelligence, pp. 199-205, 2021.
14. Xie, Y., et al. "Visualization of Olympic Games Results Using R." In Proceedings of the 3rd International Conference on Computer Science and Application Engineering, pp. 256-262, 2020.
15. Rodriguez-Mora, J., et al. "A Time Series Analysis of Olympic Medals." Journal of Sports Economics, vol. 22, no. 7, pp. 681-703, 2021.
16. Casarin, R., et al. "A Prediction Model for Olympic Medals." Journal of Quantitative Analysis in Sports, vol. 15, no. 2, pp. 61-75, 2019.
17. Thomas, L., et al. "Historical Analysis of the Olympic Games Using R." Journal of Sports Analytics, vol. 4, no. 4, pp. 231-242, 2018.

GITHUB LINK : https://github.com/tanushreego/Olympic-medal-prediction-visualization-model