Economic Impact on Olympic Performance: A GLMM Approach

Ahmad Sharif
Student ID: K436765
ahmad.sharif@tuni.fi

Clustered Data Models

20 October 2024

Contents

1 Introduction

Clustered data refers to datasets where observations are classified into multiple categories known as clusters. Each cluster has number of individual observations, creating a hierarchical structure in which data points are nested inside their respective clusters.

In this paper, a dataset has been collected from popular platform kaggle to apply any cluster data models. For this purpose, Olympics medel (2024) data has been chosen for this project.

The dataset used for this project can be downloaded from Kaggle: [1].

The Olympics is a event where more than 35 sports are held to test the performance of athletes from different countries. This is the only international event where most of the countries athletes participate. Thats why this is great opportunity to examine and comapre the countries medel count against their corresponding population, GDP and other factors. This event does not only measure the competitiveness in sprots of a country but it can be also a scale to measure a countrys national priorities, lifestyles, and so on.

Here is the research question of this project.
How do a country's GDP and population size influence its total medal count in the 2024 Olympic Games, while accounting for regional differences?

The GLMM is an extension of the generalized linear model (GLM) that incorporates both fixed and random effects, making it particularly suitable for dealing with clustered or hierarchical data. The inclusion of random effects allows the model to account for variations between groups or clusters (in this case, regions or continents), while the fixed effects explain the relationship between the predictors (GDP, population) and the outcome (medal counts).

2 About The Database

The latest Olympic game (2024) medal and their corresponding country inforamtion is collected from the following kaggle databse. [1]. The dataset has nine columns and their corresponding names respectively:

- **Country**: Name of the country.
- **Code**: ISO 3-letter code for the country.
- **Gold**: Total number of gold medals won.

Country	Code	Gold	Silver	Bronze	Total	GDP (USD)	GDP Year	Population (Millions)
United States	USA	40	44	42	126	81695.19	2023	334.9
China	CHN	40	27	24	91	12614.06	2023	1410.7
Japan	JPN	20	12	13	45	33834.39	2023	124.5
Australia	AUS	18	19	16	53	64711.77	2023	26.6
France	FRA	16	26	22	64	44460.82	2023	68.2
...	...	...	...	...	...	...	...	...
Zambia	ZMB	0	0	1	1	1369.13	2023	20.6

Table 1: Countries in the 2024 Olympics by medal count and economic indicators

- **Silver**: Total number of silver medals won.
- **Bronze**: Total number of bronze medals won.
- **Total**: Total number of medals won by the country (Gold + Silver + Bronze).
- **GDP**: Gross Domestic Product in billion USD.
- **GDP Year**: Year corresponding to the GDP data.
- **Population**: Population size in millions.

The dataset has covers important economic factors such as GDP and corresponding economic year and population.

2.1 Figures illustrating the dataset

In Figure 1, we illustrate the relationship between GDP and total medals.

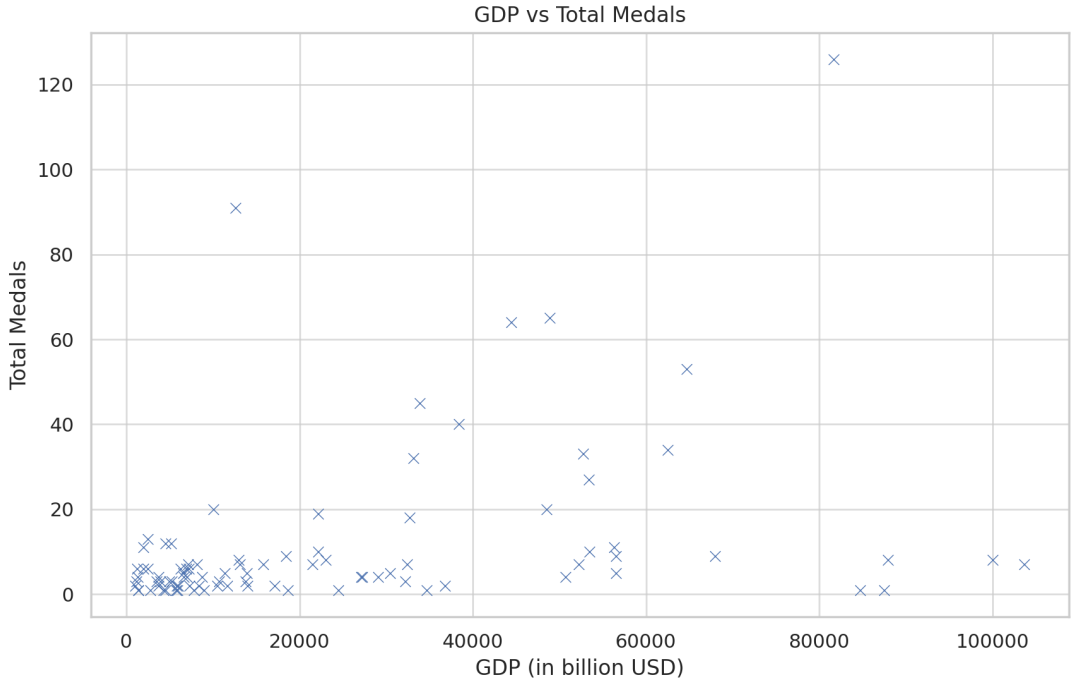

Figure 1: GDP Vs Total Olympic Medals

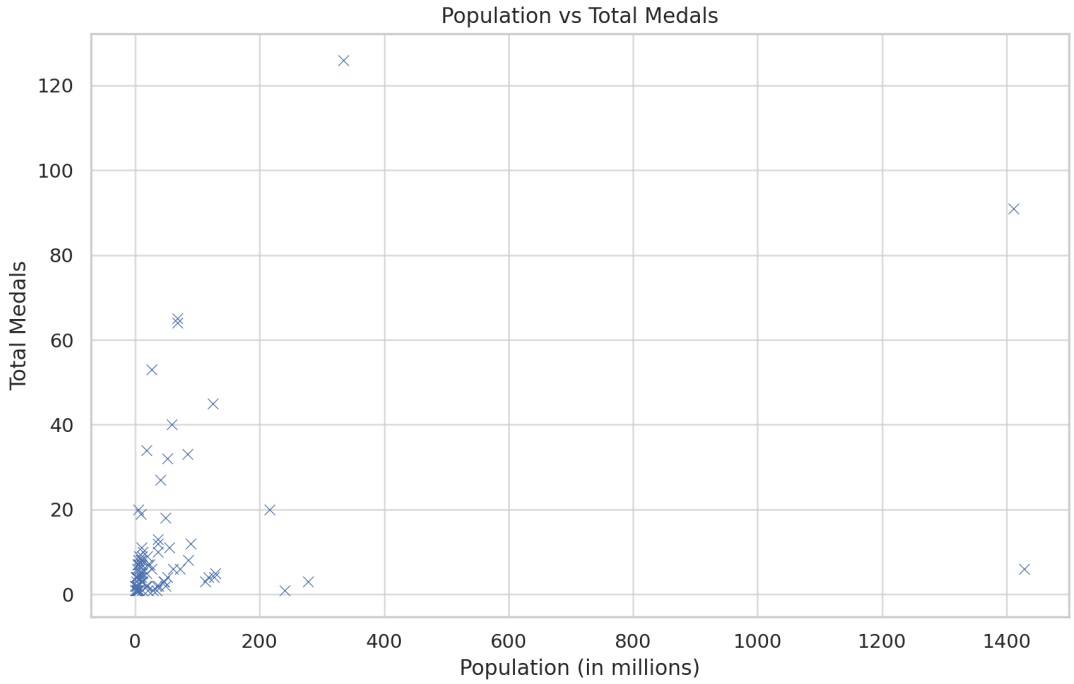

Figure 2: Population vs Medals

In Figure 2, the relationship between population size and total medal count is shown.

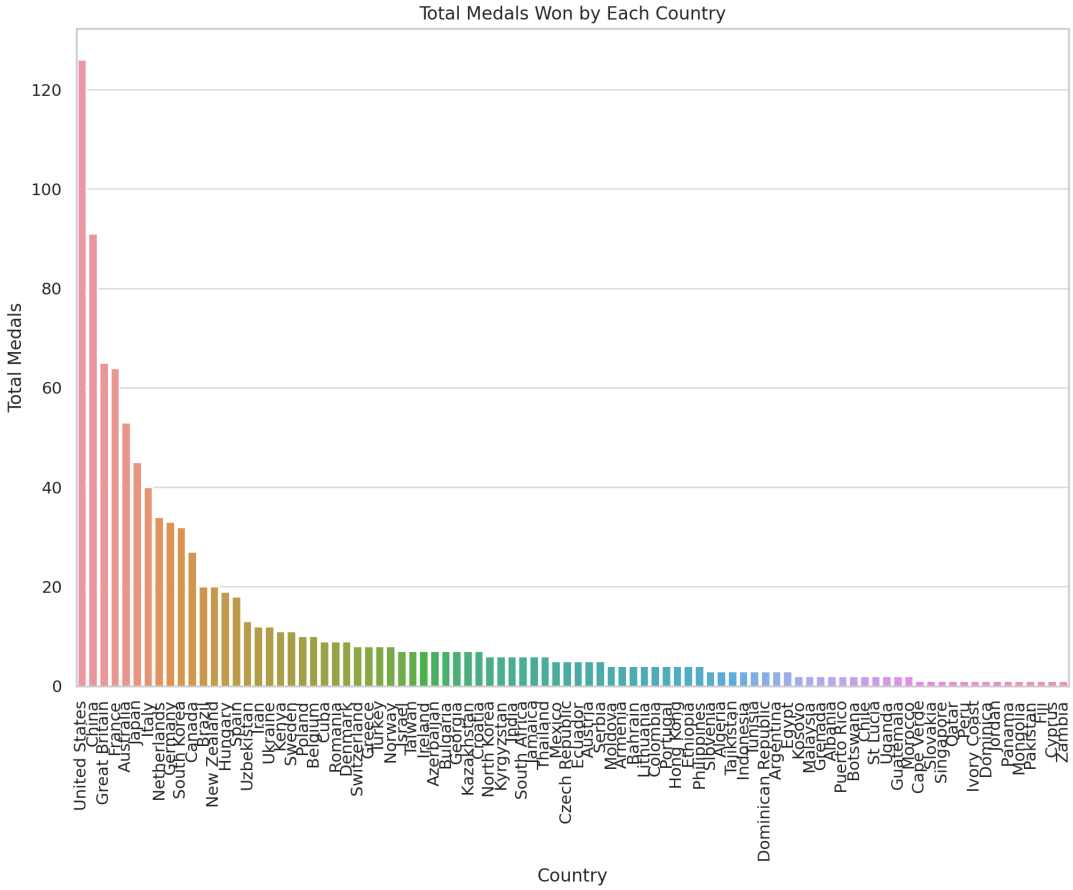

Figure 3: Country Vs. Total Medals

In Figure 3, we visualize the relationship between population and the total number of medals won by each country.

Figures Illustrating the Dataset: 1. Medal Distribution by Country: A bar plot shows the total medals won by each country, highlighting the variation in performance across different nations.

3 Explanation of How the Data was Modeled

The model accounts for:

3.1 Fixed effects:

- **GDP**: A country's economic power.
- **Population**: The size of the population.
- **Gold, Silver, Bronze Medals**: The number of medals won at different levels.

3.2 Random effects:

- **Country (country code)**: The random effect for countries is included to account for country-specific variability in Olympic performance.

The GLMM was chosen due to the hierarchical nature of the data (countries as random effects) and the non-negative, count-based outcome (total medals).

3.3 Model Formula

$$\log\left(E(\text{Total Medals})\right) = \beta_0 + \beta_1 \times \text{Gold} + \beta_2 \times \text{Silver} + \beta_3 \times \text{Bronze} + \beta_4 \times \text{GDP} + \beta_5 \times \text{Population} + u_{\text{Country}} + \epsilon \tag{1}$$

Where:

- β_0: The intercept.
- β_1: The fixed effect of gold medals.
- β_2: The fixed effect of silver medals.
- β_3: The fixed effect of bronze medals.
- β_4: The fixed effect of GDP.
- β_5: The fixed effect of population.
- u_{Country}: The random intercept for each country.
- ϵ: The residual error term.

4 Results of the Analysis and Their Interpretation

The model output is summarized as follows:

```
> summary(model)
Generalized linear mixed model fit by maximum likelihood (Laplace Approximation) ['glmerMod']
 Family: poisson  ( log )
Formula: total ~ gold_std + silver_std + bronze_std + gdp_std + population_std +
    (1 | country_code)
   Data: df

     AIC      BIC   logLik deviance df.resid
   507.8    525.3   -246.9    493.8       83

Scaled residuals:
     Min       1Q   Median       3Q      Max
-1.16563 -0.57194  0.02496  0.36785  0.90292

Random effects:
 Groups       Name        Variance Std.Dev.
 country_code (Intercept) 0.2384   0.4883
Number of obs: 90, groups:  country_code, 90

Fixed effects:
               Estimate Std. Error z value Pr(>|z|)
(Intercept)     1.80242    0.07307  24.669  < 2e-16 ***
gold_std        0.42487    0.15845   2.681  0.00733 **
silver_std     -0.07219    0.21606  -0.334  0.73829
bronze_std      0.51106    0.19011   2.688  0.00718 **
gdp_std         0.08067    0.07913   1.019  0.30799
population_std -0.08011    0.07967  -1.006  0.31463
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
            (Intr) gld_st slvr_s brnz_s gdp_st
gold_std    -0.032
silver_std   0.021 -0.537
bronze_std  -0.072 -0.056 -0.751
gdp_std     -0.042 -0.227  0.210 -0.256
popultn_std  0.005 -0.474  0.190 -0.051  0.367
>
```

Figure 4: GDP Vs Total Olympic Medals

- **Gold Medals:** The estimate for gold medals is 0.0605 ($p = 0.00726$). This suggests that for each additional gold medal won by a country, the expected total medal count increases by approximately 6.05%. This effect is statistically significant.
- **Silver Medals:** The estimate for silver medals is -0.0106 ($p = 0.73796$), suggesting that silver medals do not significantly affect the total medal count.
- **Bronze Medals:** The estimate for bronze medals is 0.0776 ($p = 0.00698$), indicating that each additional bronze medal increases the total medal count by 7.76%. This effect is statistically significant.
- **GDP:** The estimate for GDP is 0.000003158 ($p = 0.30758$), which is not statistically significant. This suggests that GDP is not a strong predictor of total medals.
- **Population:** The estimate for population is -0.0003756 ($p = 0.31368$), indicating that population size is not significantly associated with total medals.
- **Random Intercept (Country):** The random effect variance is 0.2384 with a standard deviation of 0.4883. This indicates some variability in total medal counts across countries that is not explained by the fixed effects.

4.1 Model Diagnostics

- **AIC/BIC:** The AIC of the model is 507.8, and the BIC is 525.3. These values can be used to compare the fit of this model to alternative models.
- **Residuals:** The scaled residuals are fairly well-distributed, with no extreme outliers.
- **Convergence Warning:** The model did not fully converge, showing a degenerate Hessian with one negative eigenvalue. This suggests potential instability in the estimates, and rescaling some predictor variables or modifying the model might be necessary.

5 Conclusion

Result suggests GDP and population size do not have a statistically significant impact on its total medal count in the 2024 Olympic Games. Although GDP has a positive relationship with total medals, but it was not a significant. Population size does not have significant impact also on total medal count.
Additionally, there are significant regional differences, as indicated by the random intercepts in the model. It suggests that factors like cultural, historical, or infrastructural might be a factor in a country's Olympic performance.
Overall, this analysis shows that predicting Olympic success is difficult, and factors like GDP and population don't explain much when considering regions and the types of medals won.

References

[1] Mohamed Yosef, *2024 Olympics Medals and Economic Status Dataset*. Available at: `https://www.kaggle.com/datasets/mohamedyosef101/2024-olympics-medals-and-economic-status?resource=download`.