

## Axel Ind GDP-Olympics Use Case

### Description of data

I have combined historical Olympic athlete data (<https://www.kaggle.com/heesoo37/120-years-of-olympic-history-athletes-and-results>) with information about the historical Gross Domestic Product changes of the world's countries (<https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG>) in order to discover insights about the relationship between Olympic participation and the economic state of a nation. This information could be related to the number of athletes, number of medals, or hosting nation of the games. To my knowledge, this represents a novel dataset. Compiling this data required a significant amount of data-preprocessing.

### Data Points of Interest for Predictions

I believe that the factors which make the Olympic games a successful economic venture are correlated to:

- The rate of increase of GDP after the games.
- The overall GDP increase after the games.
- The number of athletes sent by the host nation (both in total and as a proportion of all athletes present).
- The type of Olympic games (Summer or Winter).
- The historical growth or recession of the nation prior to hosting this Olympics.

### Other Notes

#### Data Preparation:

- The CSV format currently given does not represent my ideal version of the data and I have also included a database file which does.
- The data provided by the World Bank took some time to assemble as it did not properly comma delimit information and occasionally had entries like "Gambia, the".
- The Olympic dataset, on the other hand was well-developed and required minimal changes.

#### Intended Data Interpretation:

- I intend to use both classical statistics and ML to make predictions regarding the data.
- Classically, I believe that predictions can be made by treating each nation as a multi-variate normal distribution and determining if rate of GDP growth is statistically abnormal after hosting the games.
- In an ML context, I believe that predictions can be made by using classical Neural Networks to predict whether a majority of present who compete would benefit from hosting the next games.
- The statistical approach is backwards looking, and makes observations about what **has** already occurred, whereas the ML approach is predictive and seeks to answer if hosting **will** be a sound financial decision.