# Analysis of Olympic Medal Data using Machine Learning Techniques

Introduction:

The project work given below focuses on the investigation of the relationship between different metrics of Olympic medals using data visualization and machine learning techniques.

Data Preprocessing:

Import necessary libraries: NumPy, Pandas, Matplotlib, Seaborn, TensorFlow, Keras

Warnings ignored through warnings.simplefilter("ignore")

Data Visualization:

Scatter plots drawn for visualizing relationships between:

Gold medal vs. Total medals

Country Code vs. Gold medals

Silver medals vs. Total medals

Country Code vs. Silver medals

Gold medal vs. Total medals

Country Code and Bronze medals

Country Code and Total medals

Outlier Detection and Removal:

Wrote a python function to automatically detect outliers

Dropped columns "Country Code" and "Country"

Feature Selection and Target Variable:

Selected features X by dropping "Total" column from the dataset

Designated column "Total" as target variable y

## Machine Learning Implementation

Implementation of regression models as below:

Linear Regression

Polynomial Regression

Ridge Regression

Elastic Net Regression

Lasso Regression

Artificial Neural Network

# Evaluated the performance of various models using the train-test and cross-validation scores. Outcome: Linear Regression and Polynomial Regression have the best scores of 1.0 for both train-test and cross-validation. That will mean both are the best amongst the implemented models in terms of performance. Conclusion: This analysis has shown how effective linear and polynomial regression models are at solving prediction problems using medal totals at the Olympic Games. The further investigation of feature engineering and hyperparameter tuning might improve the performance of such models.