OLYMPICS DATA

ANALYSIS

**TEAM** **: ALLIANCE**

Kakarla Lahari

Chinthapalli Mounika

Bikkina Mounika Priya

Rajanala Sai Vineela

Jerri Pavani

INTRODUCTION

Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset.

EDA involves generating summary statistics for numerical data in the dataset and creating various graphical representations to understand the data better. In this article, we will understand EDA with the help of an example dataset. We will use **Python** language (**Pandas** library) for this purpose.

Importing libraries

We will start by importing the libraries we will require for performing EDA. These include NumPy, Pandas, Matplotlib



Reading and Merging data



Data Analysis Steps

* Missing values
* Numerical variables
* Categorical variables
* Outliers

Feature Engineering

* Missing values
* Temporal variables
* Categorical variables: remove rare variables
* Standarise the values of the variables to the same range

Feature Selection

We specify the lasso regression model then we use the selectfrommodel object from sklearn

Data Visualization





