Group II

Project: Summer Olympics Insights for the new Millenia

Tunde Obatayo | Chisom Okere | Quanetta Caudle

Background

The first modern Olympics were held in Athens, Greece, in 1896 and since then men and women of all ages and race have come together to participate in the historic events' different editions. The allure of great cities which host these games, the acclaim that comes with getting to stand at the podium for your respective sport and the rewards they earn continues to draw sports men and women from all around the World.

With the advent of technology and better international relations between nations, the Olympic Games have replaced the vanity of war to prove national superiority. Now nations continue to emphasize their prominence and dominance through their medal hauls at these games. The individuals who have trained and sacrificed in more ways than one to attain peak body conditions to become medal winners at these events - bring prestige to their homelands, becoming national heroes and legends.
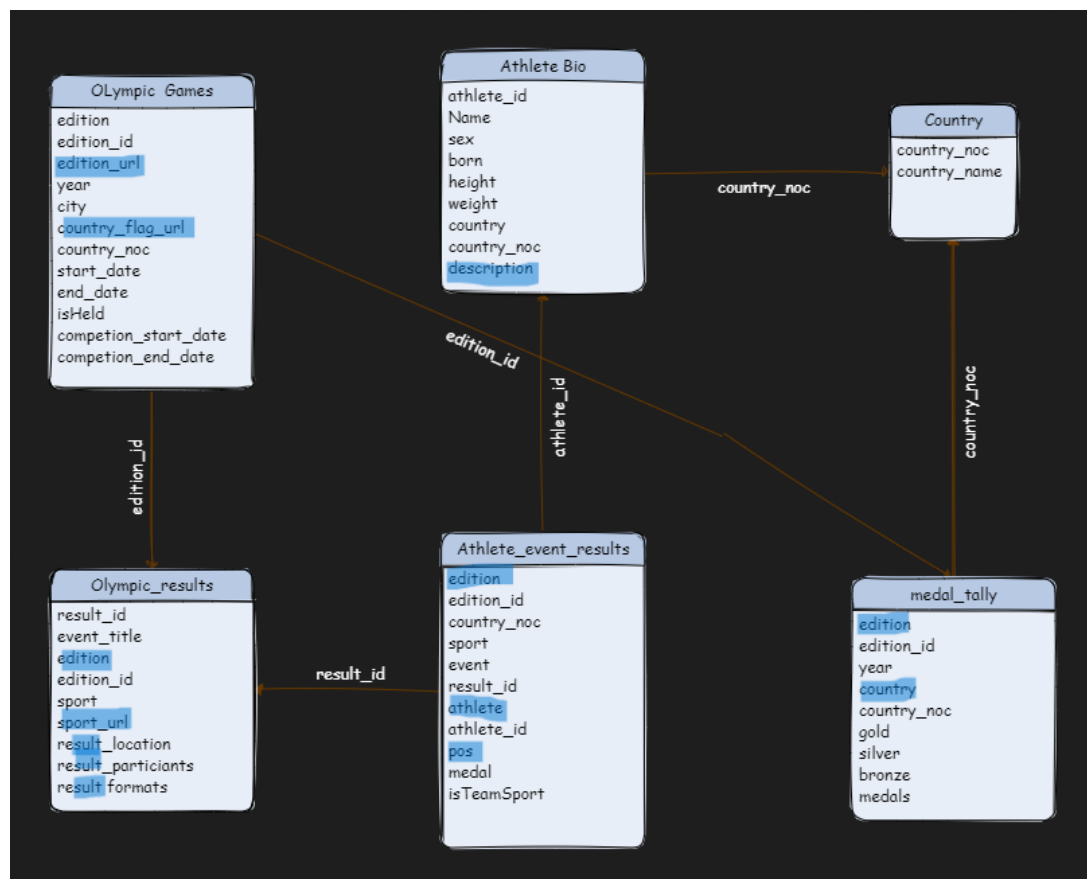
Objective

By gathering, cleansing, and analyzing data that we have obtained from Kaggle (please click on link to access our data source), we aim to deduce key data insights that will show us

- Trends seen in the summer editions of the Olympics since the start of the millennia.
- Deduce/Predict how the top nations stack against each other in their respective dominant sport(s)
- Recognize the role that factors such as sex, age etc play in some of the sport contests

Data Description

Our data set contains 6 tables which are related to one another via multiple connections. This data source contains information from Olympic editions (both Winter & Summer) as far back as the 1900s but for the sake of our study (for relevance), we have chosen to focus on data from the Summer Olympic editions starting from the 1980s.

Please find below a graphical summary of all tables in our data set and how they are related.

Data-Cleansing Steps

1. Remove all irrelevant data (unwanted Olympic editions, participating spoertsmen/women who do not make the podium etc)
2. Remove all incomplete records
3. Identify outliers (if present)
4. Create a composite table for normalization

Some of the questions we seek to answer via of analysis include:

1. What is the medal quota for the different sport types?
2. Which are the best performing nations?
3. Which demographic (age, sex) is responsible for these performances (in general, per nation)?
4. Baring any major changes/constraints which nations are still likely to perform best in the coming edition?