

1. Which client/dataset did you select and why?

I had selected Olympics Dataset - 120 years of data for analysing key insights in the data. My client was SportsStats which is a sports analysis firm. Insights could be patterns/trends highlighting certain groups/events/countries, etc. for the purpose of developing nations to perform well in upcoming Olympics events and the dataset is of compact size with the necessary information for drawing important information.

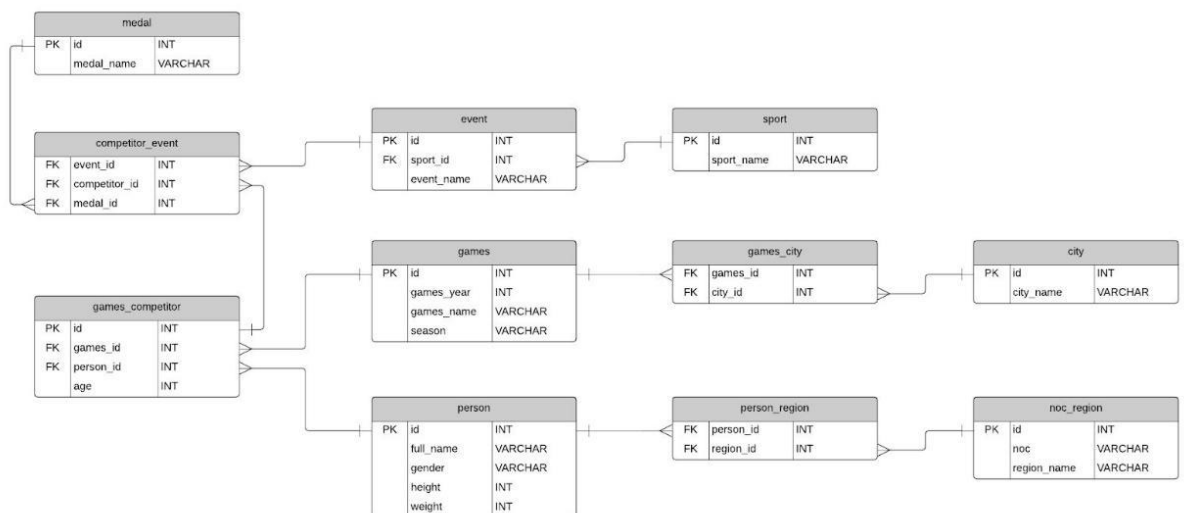
2. Describe the steps you took to import and clean the data.

I have used pandas library to import the Olympics dataset files into jupyter notebook environment, then checked for the basic info about null values in the dataset and used mean and most frequent methods to fill the null values.

3. Perform initial exploration of data and provide some screenshots or display some stats of the data you are looking at.

I have used pandasql library to derive some initial insights of the data like number of medals of each category are won till now, the top 10 players who won more medals, the number of participants in each year of olympics....

4. Create an ERD or proposed ERD to show the relationships of the data you are exploring.



Description

Write a 5-6 sentence paragraph describing your project; include who might be interested to learn about your findings. Who might be your audience?

The Olympic games are international sports events with more than 200 nations participating in various competitions. The Sports persons from various countries participate in competitions and make their countries proud of their excellence in sports. Despite the massive population, many most populous countries fail to grab many medals at the Olympic games. The primary objective is to analyse the Olympic dataset using python to compare the overall performance of countries and to evaluate the contribution of each country in the Olympics. I would like to work on men and women contribute to the Olympics, the sports that most of the people willing to participate and how to win more medals for the countries by analysing the winners. I would like to analyze the contribution of India in the Olympics to the country that won the most number of medals. Research analysts, sports analysts and newspapers are the most interested people who could go through my findings.

Questions

Create 2-3 questions that you want to answer with the data:

1. What is the contribution of men and women in the Olympics games?
2. Which country has won number of medals?
3. Which sport has more willingness to participate in the Olympics games?
4. What was the contribution of India in the Olympics compared to the country that won most medals?
5. What was the best-performed sport for each country?
6. Is there any differences between the sports happening in summer to winter olympics?

Hypothesis

What are your initial hypotheses about the data?

I think men are the most participants in the Olympics compared to women as women were least encouraged during that time. But the ratio of winners to gender would be more for women to men as the women who participated in the Olympics during that difficult times could be more passionate and want to prove themselves. The country that won most medals could be USA or Russia those are the countries with serious competition between them. I think the number of participants are more likely to increase over the years as it became more and more outreach to the people by time.

Approach

I would be looking at the name (ID) to see the persons with more winning contributions, then the ROC column to compare the country with more winnings. Most important column to find the contributions of men and women to the Olympics is Gender. To check the relation between the summer and winter Olympics. I would like to evaluate the measure on simple comparisons and test my hypothesis using AB metric.

