



ETL PROJECT PROPOSAL

11.02.2019

HYEJIN JEON
JAE KICK
RUTA HAILE

Overview

Based on a historical dataset on the modern Olympic Games, we will find out which country won most medals through Extract, Transform, Load process. The dataset consists of 3 csv files: Summer, Winter, and Dictionary; It shows game information including year, city, sport and information of medal winners such as country, gender, name, medal color between 1896 and 2014.

Datasets

Olympic Sports and Medals, 1896-2014: The dataset contains Olympic information including year, city, sport, and athlete information.

<https://www.kaggle.com/the-guardian/olympic-games>

Milestones

- I. DataFrame CSV FILE (Pandas)
- II. Create Database Connection (SQL Alechemy)
- III. Create Schema (PostgreSQL)
- IV. Query to get the specific data meeting condition (Query)

Questions

- I. Which country won most Gold/Silver/Bronze medal for summer olympic?
- II. Which country won most Gold/Silver/Bronze medal for winter olympic?
- III. Will most medal winning country have most GDP?
- IV. Will most medal winning country have most population?
- V. Who joined olympic both summer and winter for same year?