ETL | Technical Report

Joey Ashcroft, David Mathews + Brooke Rodriguez

We combined historical Olympic games data, UN health data and World Bank economic data to build a diverse demographic picture of the countries that participate in the Olympics.

## EXTRACT

**Data Sources**

|  |  |  |  |
| --- | --- | --- | --- |
| **Data Set** | **Meta Data** | **Data Type** | **URL** |
| Olympic History | Olympic athletes, physical traits, country, medals | CSV | <https://www.kaggle.com/dsfgfgdsuioghg/120-years-of-olympic-history-basic-graphics/data> |
| UN Health | Healthcare professionals | CSV | <http://data.un.org/> |
| World Bank | economic indicators | JSON | <https://databank.worldbank.org/data/home.aspx> |

## TRANSFORM

**Methods**

Python notebooks to clean data, performing the following:

* Created data frames
* Filtered/Cleaned fields
* Pivot

MySQL to join data, performing the following:

* Queried from multiple tables
* SQL operators
* Aggregate functions

## LOAD

**Production Database**

Relational - MySQL