Design Document

Comprehensive Data from 1896 to 2016 on Olympic Athletes and their Performances

Data Science

120 Years of Olympic History: Athletes and Results

Author:
Shubh Desai

Submitted to:
Dr. Junaid Qaiz

Shubh Desai

Introduction

Purpose

The purpose of Olympics is to design and development of a data visualization project that explores 120 years of Olympic history, athletes, and results. Aim is to provide an interactive and informative dashboard that allows users to explore and analyse Olympic data.

Scope

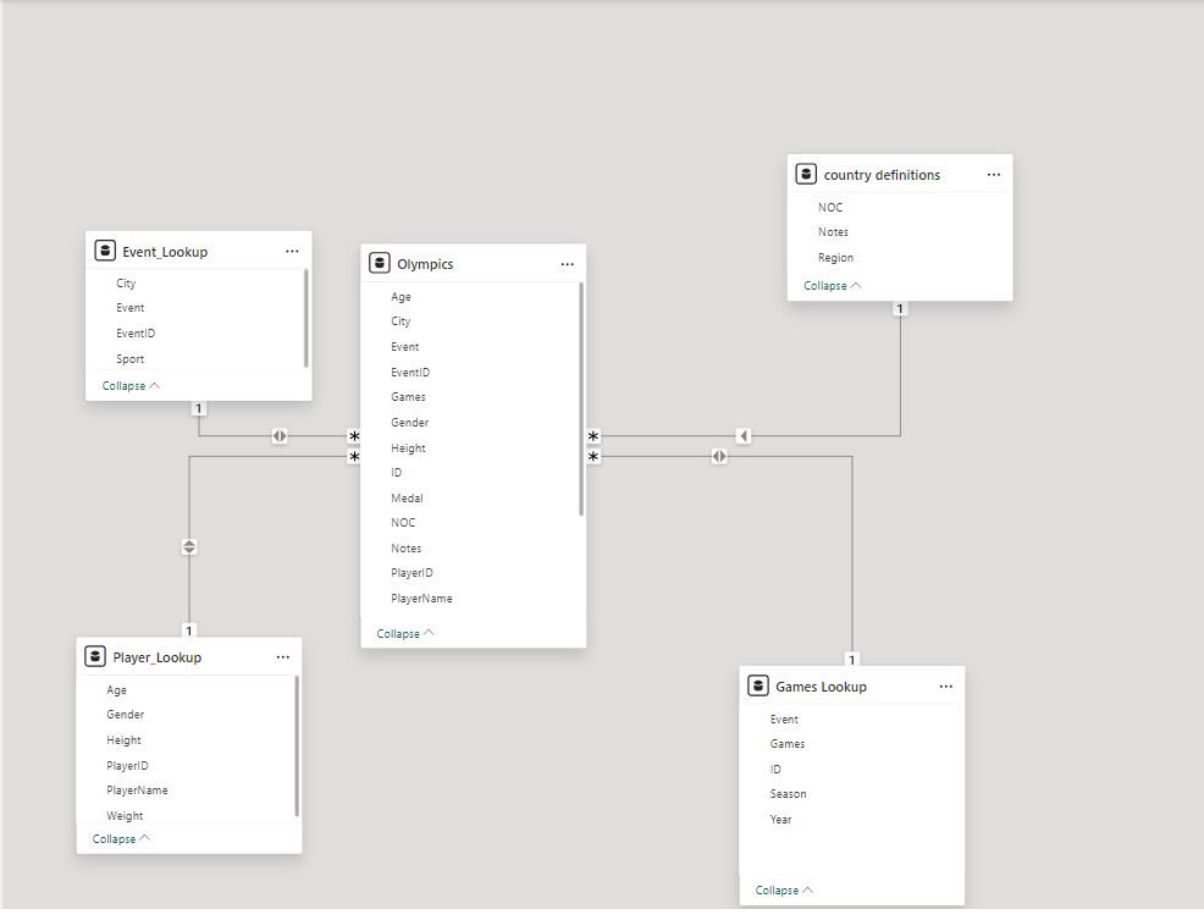
Collection, processing, and visualization of Olympic data from 1896 to 2016. Focus on athlete and result data, including all three medals, events, and countries. The main dashboard will provide an interactive interface for users to understand and analyse the dataset.

System Overview

System Architecture

The system architecture will consist of the following components:

- Data Sourcing: Kaggle dataset "120 Years of Olympic History: Athletes and Results"
- Data Modelling: Entity-Relationship Diagram (ERD) and Star schema design
- Visualization: Power BI for interactive dashboard development



System Components

The major components of the system are:

- User Interface: Interactive dashboard for data analysis

Data Management

Data Sourcing

The data source for this project is the Kaggle dataset "120 Years of Olympic History: Athletes and Results".

ETL Process

The ETL process will involve the following steps:

- Extract: Download the dataset from Kaggle and load it into a Power BI
- Transform: Transforming the data, including handling missing values and remove duplicates
- Load: Load the transformed data into a Power BI

Data Modelling

Entity-Relationship Diagram

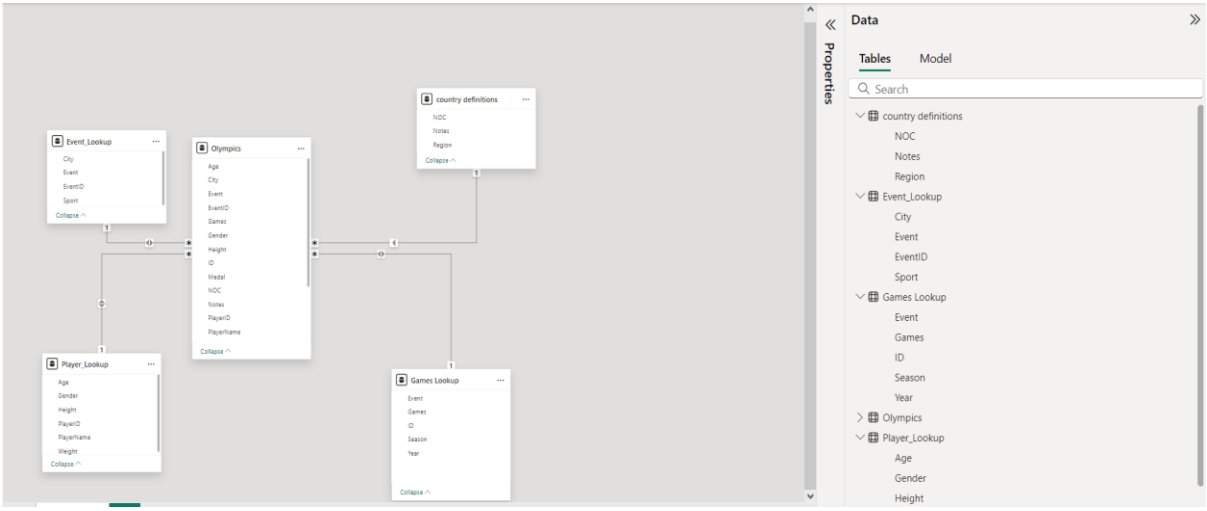
The ERD will consist of the following entities:

- Olympics
- Event_lookup
- Countries_definitions
- Games_lookup
- Player_lookup

Schema Design

The database schema will consist of the following tables:

- Olympics: games_ID, event_ID, NOC, medal, player_ID, team
- Event_lookup: city, event, event_ID, sport
- Countries_definitions: NOC, notes, region
- Games_lookup: event, games, games_ID, season, year
- Player_lookup: player_ID, player_name, age, gender, height, weight

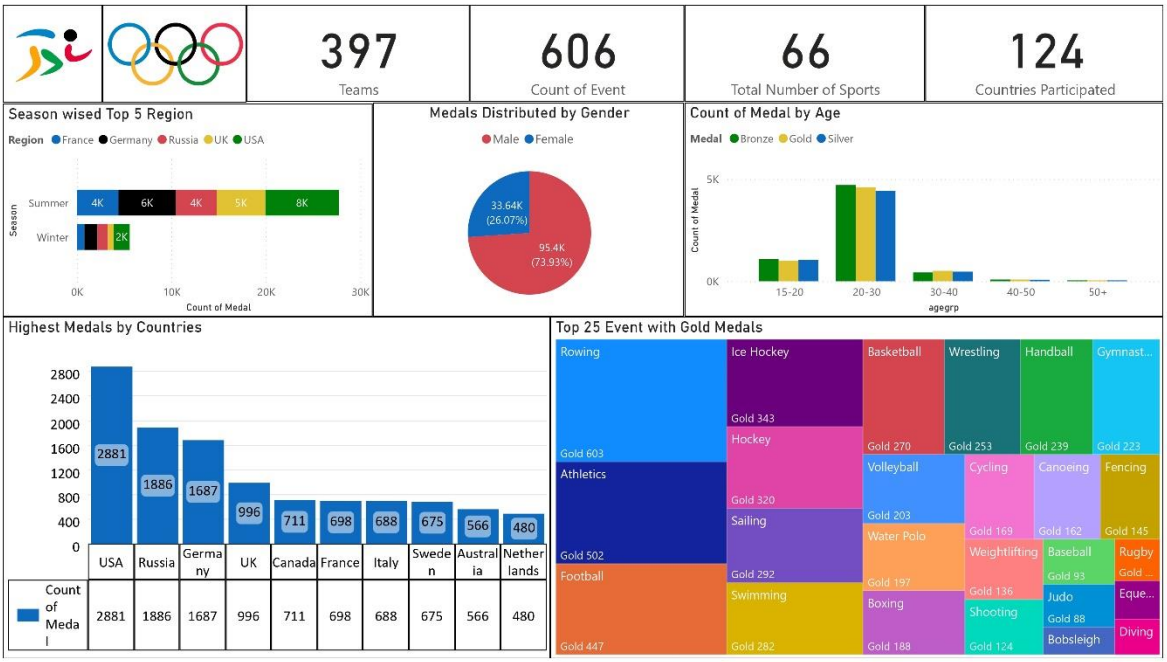


Visualization & User Interface

Layout

The dashboard will consist of the following sections:

- Introduction: Overview of the project and dataset
- Athlete Analysis: Informative representation of athlete data, including medals, events, age, and gender
- Event Analysis: Interactive visualization of event data, including results and gold medals
- Country Analysis: Interactive representation of country data, including top medal holders and participation



Features & Functionalities

The dashboard will offer the following features and functionalities:

- Interactive Elements: Filtering, sorting, and drill-down capabilities
- Filter & Search Option: Search for specific athletes, events, or countries
- Navigation: Easy navigation between sections and visualizations

Conclusion

Outlines the scope, architecture, and technical requirements for the Olympic 120 year's records. Provide an interactive and informative dashboard for users to explore and analyse Olympic data.