

**Project - Module 02**

**Knowledge Graphs and Linked Data**

SUBMITTED BY:

**Areesha Maqsood - i180600  
Shujaa Marwat - i180432  
Abdul Mannan - i180557**

Section:  
**CS-A**

Submitted To:  
**Ma’am Amna Basharat**

DATE:

**24/11/2021**

Table of Contents

[**Domain Description**](#_yje5rqyo0ytm) **2**

[**Competency Questions**](#_wtzz88pbv646) **2**

[**Conceptual Model**](#_tpn8s0yxxv10) **3**

[**Dataset Description**](#_8zhc0ijq5z36) **3**

[**External Linked Data sets identified for Linking**](#_kf9hji82s6mo) **4**

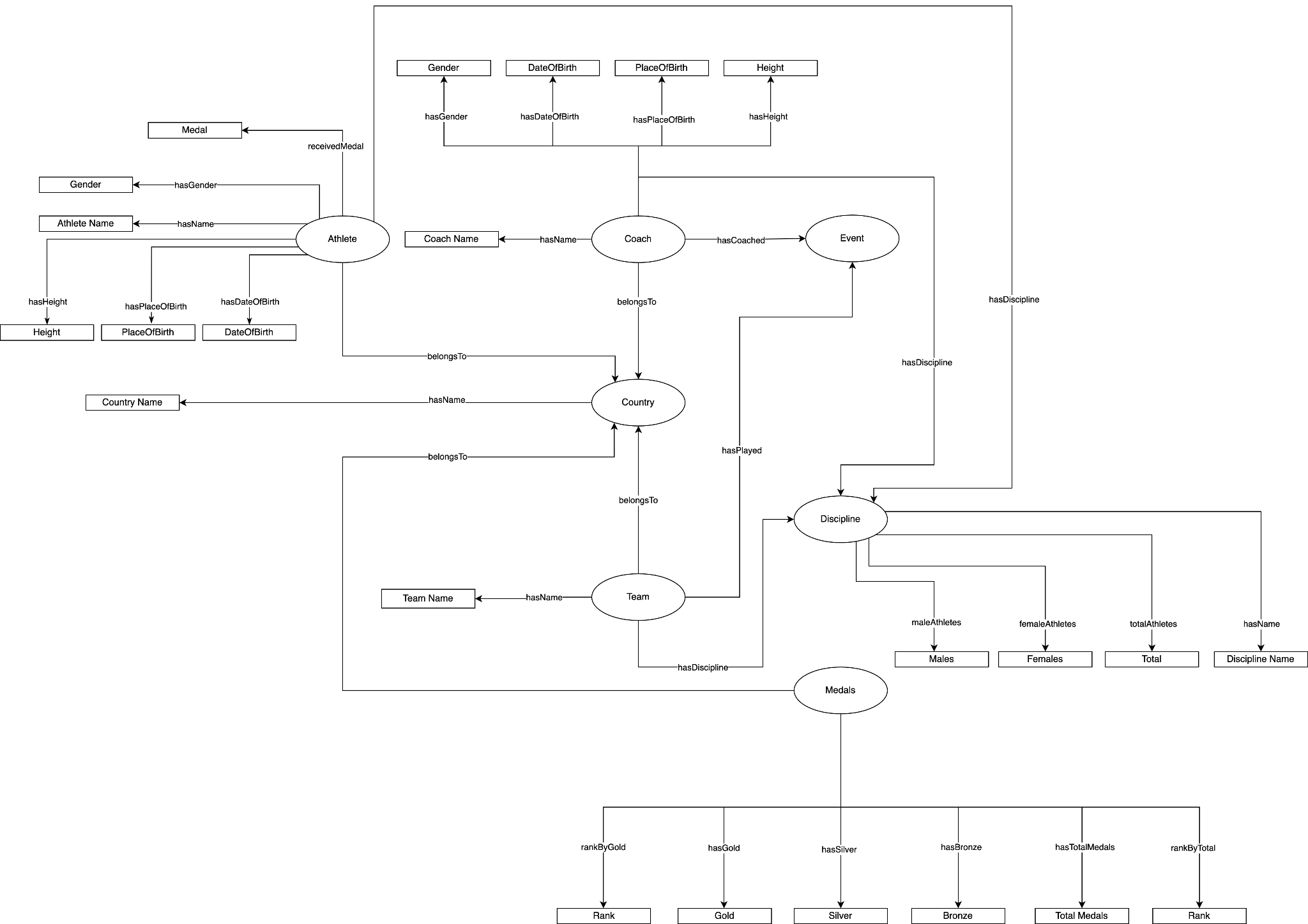
# Domain Description

Our project domain revolves around making a knowledge graph of an existing sports dataset. This will involve making the design for the knowledge graph (conceptual model), implementing it through code and then querying the knowledge graph. We are working on the Olympics 2021 dataset taken from [Kaggle](https://www.kaggle.com/arjunprasadsarkhel/2021-olympics-in-tokyo).

# Competency Questions

* How many (M or Total) medals did Country X win?
* Which country won the most (M) medals?
* What rank did Country X get in general?
* What rank did Country X get by total?
* Which team(s) had only (Men, Mixed, Women, Duet, Team, Relay, Softball) in X sport?
* Which Disciplines did X team participate in?
* Which team(s) participated in X Discipline?
* Which Countries had more than one Team participating in a sport?
* Which country(ies) participated the least in a sport?
* Which Discipline had the most X gender participants?
* Order the Disciplines by number of (X gender) participants?
* Which team does Coach X belong to?
* Which Coaches are coaching more than one Discipline?
* Who is Athlete X's Coach?
* Which Athletes are being coached by Coach X?
* What Disciplines were played in 2021's Olympics?
* Which Countries participated?
* Who were the Coaches?
* Which Teams were there?
* Which Athletes participated from X Country?
* Which Athletes participated in X Discipline?

# Conceptual Model



# 

# Dataset Description

We chose the Olympics 2021 dataset from the sports domain. We picked up the data available on [Kaggle](https://www.kaggle.com/arjunprasadsarkhel/2021-olympics-in-tokyo) which contains details of over 11,000 athletes, with 47 disciplines, along with 743 Teams that were a part of the 2021 Tokyo Olympics.

This dataset contains the details of the Athletes, Coaches, Teams participating, the Medals won and Entries of total count of genders. It contains the athletes’ names, countries represented, discipline, gender of competitors, name of the coaches.

# External Linked Data sets identified for Linking

We are going to be linking our knowledge graph with DBPedia and WikiData to add to their as well as our relevant information regarding Athletes, Coaches, etc.

Along with this, the data available on kaggle was insufficient for athlete specific medal information, for which we used a publicly available set of tables from github.