Rishabh Wason, Eugene Lee, Omar Buenostro, Jin Kweon, Jiyoon (Clover) Jeong

INFO 257

Spring 2018

Database Description

Database Intro:

For the database we will be using a Kaggle dataset on FIFA 18 players. FIFA is an association that is the international governing body of soccer (football). Kaggle is a website on which there are a multitude of datasets of different types. This dataset that we are using represents all the players in the FIFA 18 game that represents international teams, leagues for each country and many other smaller, independent teams. The main entity that we will be working with will be the players entity that represents a player and their statistics. Some of these statistics include attacking statistics, defense statistics, position statistics, salary statistics, and other special characteristics dependent on the specific player at hand. The Kaggle dataset that we are working with is a singular relation that contains players in FIFA 18 but we would like to separate the dataset into multiple entities so that it can make more sense logically. Since we will separate the data into multiple relations that each represent an entity, we decided to create the following relations: player bios, overall statistics, attacking statistics, defense statistics, goalkeeper statistics, money statistics, teams, leagues, nationality, positions, traits, and specialties. These relations will be discussed further in detail in the data dictionary and the entity relationship diagram that shows how these entities relate to each other.

Database Uses/Purpose:

The database that will contain the separated entities and relations will allow a user of the database to see players, their statistics and their relations to other teams and leagues. The main use for the database will be to see player statistics and what teams/leagues they are part of. A user of this database can also query the relations to find players that match a certain criteria such as players with pass ratings greater than 90 or even more complicated queries.

Database Issues:

There are two main issues with the initial dataset that we are working with to produce a database composed of multiple relations. The first issue is that some of the records in the dataset are not in English. This might cause an issue when a user is accessing the database as they may not be able to understand the language being used in the dataset but to resolve this issue we might use a translate API to convert the text to English if

possible. The second issue at hand is that the dataset we are working with is composed of a single table and not separate relations. We will have to process the data into separate relations so that we can make logical relationships between the entities as described in our data dictionary and entity relationship diagram.

Data Dictionary

<Entities>

Logical connections between them:

Player's have **overall stats**, **attacking stats, defense stats**, **goal keeper stats, position stats, money stats, traits,** and **specialties.**

Players play on **teams,** play in **leagues**, and play for **countries**.

Players_bio

text **full_name**: A player's full name
numeric **ID: Primary Key**: A player's FIFA ID
text **club**: **Foreign key**: The club team a player plays for.
text **nationality**: **Foreign key**: The country a player plays for.
DateTime **birth_date**: Player's birthdate.
numeric **age**: Player's age.
numeric **height_cm**: Player's height in cm.
numeric **weight_kg**: Player's weight in kg.
hyperlink **photo**: Link to a photo of a player.

Overall_stat

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

text **preferred_foot**: Player's preferred foot.
numeric **overall**: A player's overall rating.
numeric **potential**: A player's rating for potential.
numeric **pac**: Rating assigned to pace/speed.

text **body_type:** Player's body type: normal, lean, stocky, etc
numeric **weak_foot**: Rating of how good the player's weak foot is from 1 to 5 (the best)
text **international_reputation**: Reputation of the players from 1 to 5 (the most famous)
numeric **stamina**: Rating of a player's stamina (1-100).
numeric **strength**: Rating of a player's strength (1-100).
numeric **balance**: Rating of a player's balance (1-100).
numeric **reactions**: Rating of a player's reactions (1-100).
numeric **heading_accuracy**: Rating of a player's header accuracy (1-100).
numeric **interceptions**: Rating of how good a player's interceptions are (1-100).
numeric **positioning**: Rating of how good a player's positioning is (1-100).
numeric **vision**: Rating of how good a player's vision is (1-100).
numeric **penalties**: Rating of how good a player is at penalties (1-100).
numeric **composure**: Rating of how good a player's composure is (1-100).

Attacking_stat

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

numeric **sho**: Rating assigned to shooting (1-100).
numeric **pas**: Rating assigned to passing (1-100).
numeric **dri**: Rating assigned to dribbling (1-100).
numeric **crossing**: Rating assigned to crossing (1-100).
numeric **finishing**: Rating assigned to finishing (1-100).
numeric **short_passing**: Rating assigned to short passing (1-100).
numeric **volleys**: Rating assigned to volleys (1-100).
numeric **dribbling**: Rating assigned to volleys (1-100).
numeric **curve**: Rating assigned to curve (1-100).
numeric **free_kick_accuracy**: Rating assigned to free kicking accuracy (1-100).
numeric **long_passing**: Rating assigned to long passing (1-100).
numeric **ball_control**: Rating assigned to ball control (1-100).
numeric **acceleration**: Rating assigned to acceleration (1-100).
numeric **sprint_speed**: Rating assigned to sprint speed (1-100).
numeric **agility**: Rating assigned to agility (1-100).
numeric **shot_power**: Rating assigned to shot power (1-100).
numeric **jumping**: Rating assigned to jumping (1-100).
numeric **shot_power**: Rating assigned to acceleration (1-100).
numeric **long_shots**: Rating assigned to long shots (1-100).
numeric **aggression**: Rating assigned to aggresion (1-100).
text **work_rate_att:** How active a player is on offense while not in possession of the ball (Low, Medium, High).

Defense_stat

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

numeric **def**: Rating assigned to defense (1-100).
numeric **phy**: Rating assigned to physical (1-100).
text **work_rate_def:** How active a player is on defense while not in possession of the ball (Low, Medium, High).
numeric **marking**: Rating assigned to marking (1-100).
numeric **standing_tackle**: Rating assigned to standing tackle (1-100).
numeric **sliding_tackle**: Rating assigned to sliding tackle (1-100).

Goalkeeper_stat

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

numeric **gk_diving**: Rating of how good a goal keeper is at diving (1-100).
numeric **gk_handling**: Rating of how good a goal keeper is at handling the ball (1-100).
numeric **gk_kicking**: Rating of how good a goal keeper is at handling the ball (1-100).
numeric **gk_positioning**: Rating of how good a goal keeper is at positioning (1-100).
numeric **gk_reflexes**: Rating of how good a goal keeper's reflexes are (1-100).

Money

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

numeric **eur_wage**: Player's wage in euros.
numeric **eur_value**: Player's value in euros.
numeric **eur_release_clause**: Player's release buyback cost in euros.

Teams

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

text **club**: A club team's name.

text **league:** Name of league.
hyperlink **clublogo**: Link to a photo of a club's team logo.

Nationality

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

text **country_name**: Name of a country.
hyperlink **flag**: Link to a photo of a country's flag.

Positions

numeric **playerID**: **Foreign key:** References ID in player_bio entity.
numeric **rs**: Rating as a right sweeper.
numeric **rw**: Rating as a right winger.
numeric **rf**: Rating as a right forward.
numeric **ram**: Rating as a right attacking midfielder.
numeric **rdm**: Rating as a right defensive midefielder.
numeric **rcb**: Rating as a right center-back defender.
numeric **rm**: Rating as a right midfielder
numeric **rb**: Rating as a right full-back defender.
numeric **rwb**: Rating as a right wide-back defender.
numeric **cf**: Rating as a center forward.
numeric **cam**: Rating as a center attacking midfielder.
numeric **cdm**: Rating as a center defensive midfielder.
numeric **cm**: Rating as a center midfielder.
numeric **cb**: Rating as a center-back defender.
numeric **ls**: Rating as a left sweeper.
numeric **lw**: Rating as a left winger.
numeric **lf**: Rating as a left forward.
numeric **lam**: Rating as a left attacking midfielder.
numeric **ldm**: Rating as a left defensive midefielder.
numeric **lcb**: Rating as a left center-back defender.
numeric **lm**: Rating as a left midfielder
numeric **lb**: Rating as a left full-back defender.
numeric **lwb**: Rating as a left wide-back defender.
numeric **gk**: Rating as a goal keeper.

Trait

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

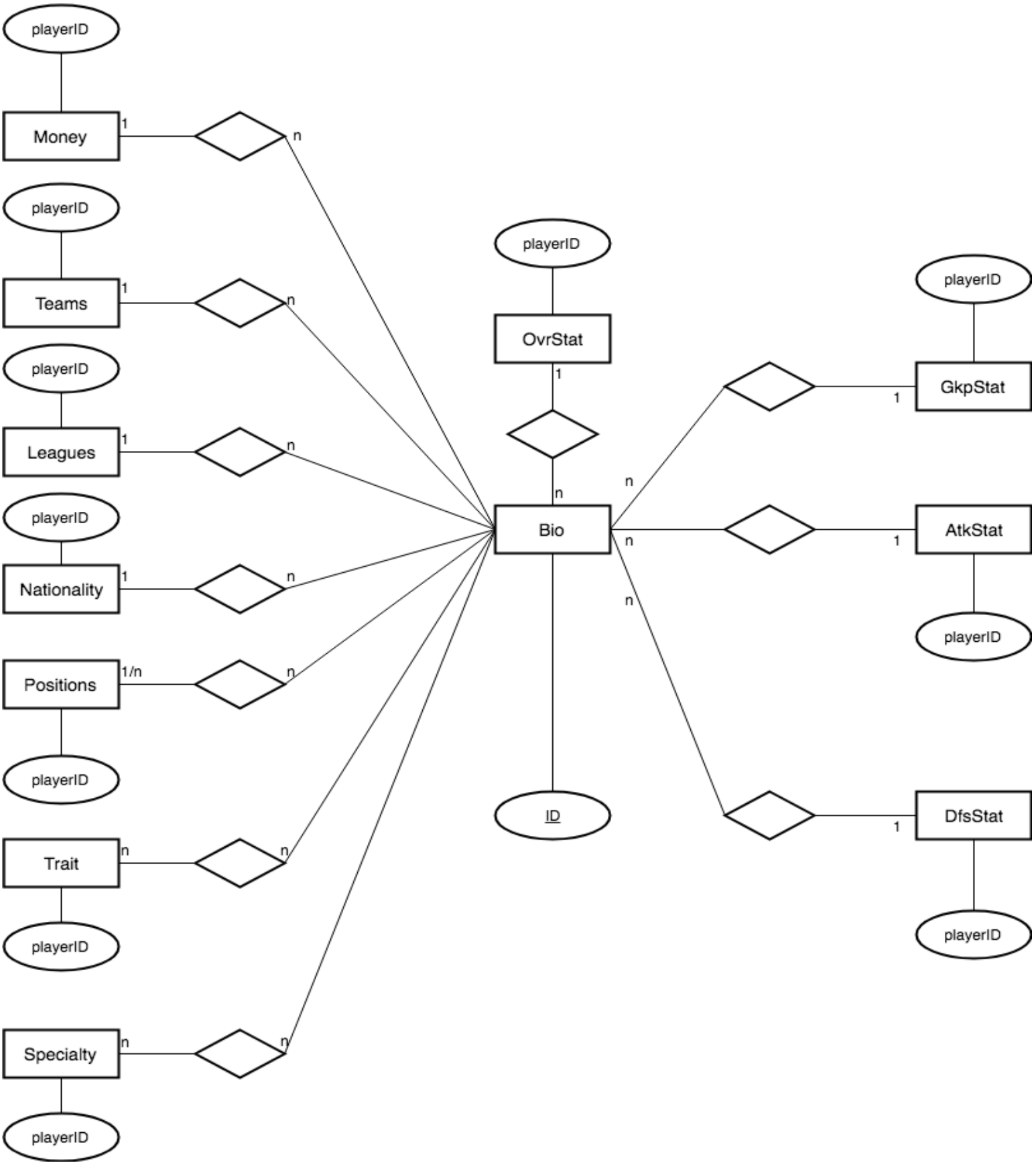
boolean **chip_shot_trait**: Does the player have the chip shot trait?
boolean **corner_specialist_trait**: Does the player have the corner shot trait?
boolean **diver_trait**: Does the player have the diver trait?
boolean **finesse_shot_trait**: Does the player have the chip shot trait?
boolean **gk_long_throw_trait**: Does the player long throw as a goalkeeper?
boolean **gk_up_for_corners_trait**: Does the player up for corners as a goalkeeper?
boolean **injury_free_trait**: Is the player injury free?
boolean **injury_prone_trait**: Is the player prone to injury?
boolean **leadership_trait**: Has the player been shown to be a leader?
boolean **long_passer_trait**: Does the player long pass?
boolean **long_shot_taker_trait**: Does the player take long shots?
boolean **one_club_player_trait**: Does the player only play for 1 club?
boolean **playmaker_trait**: Does the player make plays?

boolean **power_free_kick_trait**: Does the player power free kick?
boolean **power_header_trait**: Does the player power header?

Specialty

numeric **playerID**: **Foreign key:** References ID in player_bio entity.

boolean **speedster_speciality**: Is strength this player's specialty?
boolean **dribbler_speciality**: Is dribbling this player's specialty?
boolean **engine_speciality**: Is engine this player's specialty?
boolean **distance_shooter_speciality**: Is distance shooting this player's specialty?
boolean **free_kick_specialist_speciality**: Is the player a free kick specialist?
boolean **tackling_speciality**: Is tackling this player's specialty?
boolean **strength_speciality**: Is strength this player's specialty?


playerID
Money
1
n
playerID
Teams
1
n
playerID
Leagues
1
n
playerID
Nationality
1
n
Positions
1/n
n
playerID
Trait
n
n
playerID
Specialty
n
n
playerID
playerID
OvrStat
1
n
Bio
ID
n
n
n
playerID
GkpStat
1
playerID
AtkStat
1
playerID
DfsStat
1
playerID