Database Project - NBA Stats

mh6069@nyu.edu Ming-Han Huang
wz2055@nyu.edu Wenyuan Zhang

Description

This project will collect and integrate NBA players' personal information, stats, and all their gossip, news and crime record into a well visualized website/app. Also, it will show the recent NBA matches. Users can search and find the whole information about their favorite NBA players, or the players they are interested in, on this website/app.

For the UI, there will be a list of players with their name and photo in the main page. Users can choose filters, e.g. teams, positions, to find the players. Or they can enter the player's name to search for a specific player.

Users can then click the player's name to connect to the player's page. In the player's page, there will be a player's photo followed by their personal information and stats, e.g. name, height and weight. Next, the page shows the teams the player served in his career, including the periods, the player number and his salary. After that, there will be a list of awards and honors the player has received. Here shows the award's name and the time he received the award. Then, a section shows a list of endorsements, including the company name and the products, e.g. shoes. At the end of the page, users can find the gossip, news and crime records of this player.

ER diagram & business rules
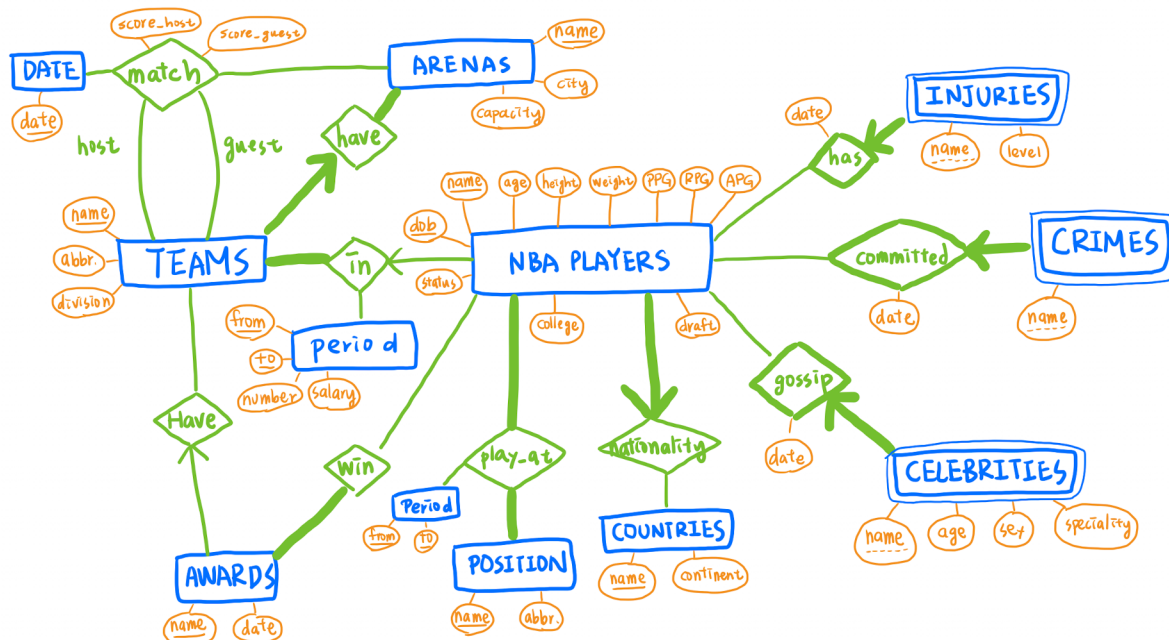
E1: *NBA Players* is uniquely identified by a combination of name and dob, and has attributes age, height, weight, PPG, RPG, APG, status, college, draft

E2: *Teams* are uniquely identified by name and have attributes abbr. and division.

E3: *Period* is uniquely identified by a combination of from and to, and has attributes number and salary.

E4: *Arenas* is uniquely identified by name and has attributes city and capacity.

E5: *Date* is uniquely identified by date.

E6: *Awards* is uniquely identified by name and has attribute date

E7: *Position* is uniquely identified by name and has attribute abbr

E8: *Countries* is uniquely identified by name and has attribute continent

E9: *Celebrities* is uniquely identified by name and has attributes age, sex, speciality

E10: *Crimes* is uniquely identified by name

E11: *Injuries* is uniquely identified by name and has attribute level

R1: *Players* have exactly one nationality and *Country* can have multiple players.

R2: All *players* play at some position at a specific period and we record every period for each player, each *position* might have one or more players.

R3: *Players* can have *awards*, one *award* belongs to at least one *player*.

R4: Great *teams* can have many *awards* but not all *Teams* have an *award*. An *award* could belong to at most one *team*.

R5: *Players* play for at most one *team* at a period. *Teams* have at least one *player*.

R6: *Teams* can compete with each other in some *arenas* at a specific date, as we call it a match.

R7: *Teams* have exactly one *arena* as their home. All *arenas* have a *team*.

R8: A *celebrity* will not occur independently in a gossip relation with some *players*. If a *player* is no longer recorded in our database then we do not track the *celebrity* either.

R9: *Players* can commit *crimes* at some date. A *crime* will not be committed independently in our records. If a *player* is no longer recorded in our application then we do not track his *crime* status.

R10: *Injury* is dependent on the *player* themselves and it cannot exist without a *player*.

Data acquisition

NBA team & player data: we will fetch the data from https://www.nba.com/. The NBA website has complete data of each player's stats and personal information, including their age, height, weight, point per game (PPG), rebound per game (RPG), assist per game (APG) and the teams.

Awards: wikipedia has a full list of NBA awards

Sponsorships: there is a list of sponsored players in each company's wikipedia page. Also, each player's wikipedia page has full data of their endorsements.

Gossip and crime records, https://nbacrimelibrary.com/ has a completed record of all NBA crimes. Also, each player's wikipedia page has all the data we need.

Match records, https://www.espn.com/nba/scoreboard has a complete record of each day's contest results, including top performance player, highlighted video link, and score in each quarter of the match etc.

Injury records, https://www.espn.com/nba/injuries has a complete record of NBA players. Including specific date and injury comments.

User interaction

default show:
There will be a default view on the main page of the website/app based on either teams, cities, or hot players or even today's top players information. We could provide as many default features as we can implement.

search show:
There will be a search box on top of the website/app, users can input relevant players name or teams, and we can then fetch the related date to show to the users.

recommend show:
We can show popular players' information for users who are looking through the website/app, based on https://trends.google.com/trends/?geo=US or other available metrics.

random show:
Users can just click a random button on the website/app, we will randomly retrieve some players name, photo and their main page link.

Relational schema

drop table if exists Play_at;
drop table if exists Celebrities;
drop table if exists Crimes;
drop table if exists Injuries;
drop table if exists Awards;
drop table if exists Position;
drop table if exists Period;
drop table if exists NBA_Players;
drop table if exists match;
drop table if exists Date;
drop table if exists Teams;
drop table if exists Arenas;
drop table if exists Countries;

create table Countries (
  name varchar(32) primary key,
  continent varchar(32)
);

```sql
create table Arenas (
  name varchar(32) primary key,
  city varchar(16) not null,
  capacity integer
);

create table Teams (
  name varchar(64) primary key,
  abbr varchar(16),
  division varchar(16) not null,
  arena varchar(32) not null,
  foreign key (arena) references Arenas(name)
);

create table Date (
  date date primary key
);

create table match (
  host varchar(64),
  guest varchar(64),
  m_date date,
  primary key(host, guest, m_date),
  score_host integer,
  score_guest integer,
  arena varchar(32) not null,
  foreign key (host) references Teams(name),
  foreign key (guest) references Teams(name),
  foreign key (m_date) references Date(date),
  foreign key (arena) references Arenas(name)
);

create table NBA_Players (
  name varchar(32),
  dob date,
  status boolean not null,
  college varchar(64),
  draft date,
  APG real,
  RPG real,
  PPG real,
  weight real,
  height real,
  age integer,
```

```sql
    nationality varchar(32) not null,
    p_name varchar(64) not null,
    primary key (name, dob),
    foreign key (nationality) references Countries(name),
    foreign key (p_name) references Teams(name)
);

create table Position (
    name varchar(16) primary key,
    abbr varchar(32)
);

create table Period (
    from_date date not null,
    to_date date not null,
    number integer,
    salary integer,
    primary key (from_date, to_date)
);

create table Play_at (
    player_name varchar(32) not null,
    player_dob date,
    position_name varchar(16) not null,
    from_date date,
    to_date date,
    primary key(player_name, player_dob, position_name, from_date, to_date),
    foreign key (from_date, to_date) references Period(from_date, to_date),
    foreign key (player_name, player_dob) references NBA_Players(name, dob),
    foreign key (position_name) references Position(name)
);


create table Awards (
    name varchar(32) primary key,
    date date,
    t_name varchar(64),
    p_name varchar(32) not null,
    p_dob date not null,
    foreign key (t_name) references Teams(name),
    foreign key (p_name, p_dob) references NBA_Players(name, dob)
);
```

```sql
create table Celebrities (
  name varchar(32),
  age integer,
  sex char(1),
  speciality varchar(32),
  date date,
  player_name varchar(32),
  player_dob date,
  primary key (name, date, player_name, player_dob),
  foreign key (player_name, player_dob) references NBA_Players (name, dob) on delete
cascade
);

create table Crimes (
  name varchar(32),
  date date,
  player_name varchar(32),
  player_dob date,
  primary key (name, date, player_name, player_dob),
  foreign key (player_name, player_dob) references NBA_Players (name, dob) on delete
cascade
);

create table Injuries (
  name varchar(32),
  level varchar(32),
  date date,
  player_name varchar(32),
  player_dob date,
  primary key (name, date, player_name, player_dob),
  foreign key (player_name, player_dob) references NBA_Players (name, dob) on delete
cascade
);

insert into Countries (name, continent) values ('USA', 'North America');
insert into Countries (name, continent) values ('Canada', 'North America');
insert into Countries (name, continent) values ('Australia', 'Australia');

insert into Arenas (name, city, capacity) values ('Staples Center', 'Los Angeles', 18997);
insert into Arenas (name, city, capacity) values ('Barclays Center', 'Brooklyn', 17732);

insert into Teams (name, abbr, division, arena) values ('Brooklyn Nets', 'BKN', 'Atlantic',
'Barclays Center');
```

insert into Teams (name, abbr, division, arena) values ('Los Angeles Lakers', 'LAL', 'Pacific', 'Staples Center');

insert into Date (date) values ('2021-10-03');
insert into Date (date) values ('2022-01-25');

insert into match (host, guest, m_date, score_host, score_guest, arena) values ('Los Angeles Lakers', 'Brooklyn Nets', '2021-10-03', 97, 123, 'Staples Center');
insert into match (host, guest, m_date, arena) values ('Brooklyn Nets', 'Los Angeles Lakers', '2022-01-25', 'Barclays Center');

insert into NBA_Players (name, dob, status, college, draft, APG, RPG, PPG, weight, height, age, nationality, p_name) values ('Kevin Durant', '1988-09-29', TRUE, 'The University of Texas at Austin', '2007-06-28', 5.0, 10.0, 29.8, 240, 2.08, 33, 'USA', 'Brooklyn Nets');
insert into NBA_Players (name, dob, status, college, draft, APG, RPG, PPG, weight, height, age, nationality, p_name) values ('LeBron James', '1984-12-30', TRUE, 'St. Vincent-St. Mary High School', '2003-06-26', 5.3, 6.3, 26.0, 250, 2.06, 36, 'USA', 'Los Angeles Lakers');

insert into Position (name, abbr) values ('Small forward', 'SF');
insert into Position (name, abbr) values ('Power forward', 'PF');
insert into Position (name, abbr) values ('Point guard', 'PG');
insert into Position (name, abbr) values ('Shooting guard', 'SG');
insert into Position (name, abbr) values ('Center', 'C');

insert into Period (from_date, to_date, number, salary) values ('2003-9-02', '2021-10-29', 23, 5000);
insert into Period (from_date, to_date, number, salary) values ('2007-9-02', '2021-10-29', 7, 4000);

insert into Play_at (player_name, player_dob, position_name, from_date, to_date) values ('LeBron James', '1984-12-30', 'Small forward', '2003-9-02', '2021-10-29');
insert into Play_at (player_name, player_dob, position_name, from_date, to_date) values ('Kevin Durant', '1988-09-29', 'Power forward', '2007-9-02', '2021-10-29');

insert into Awards (name, date, t_name, p_name, p_dob) values ('2019-2020 NBA Championship', '2020-10-11', 'Los Angeles Lakers', 'LeBron James', '1984-12-30');
insert into Awards (name, date, t_name, p_name, p_dob) values ('2019-2020 NBA Final MVP', '2020-10-11', null, 'LeBron James', '1984-12-30');

insert into Celebrities (name, age, sex, speciality, date, player_name, player_dob) values ('Scarlett Johansson', 36, 'F', 'Actress', '2011-01-16', 'Kevin Durant', '1988-09-29');

insert into Injuries (name, date, player_name, player_dob) values ('Right Ankle', '2021-10-25', 'LeBron James', '1984-12-30');

insert into Injuries (name, date, player_name, player_dob) values ('Ankle', '2021-05-25', 'LeBron James', '1984-12-30');
insert into Injuries (name, date, player_name, player_dob) values ('Ankle', '2021-05-16', 'LeBron James', '1984-12-30');
insert into Injuries (name, date, player_name, player_dob) values ('Thigh', '2021-04-18', 'Kevin Durant', '1988-09-29');
insert into Injuries (name, date, player_name, player_dob) values ('Rest', '2021-04-14', 'Kevin Durant', '1988-09-29');