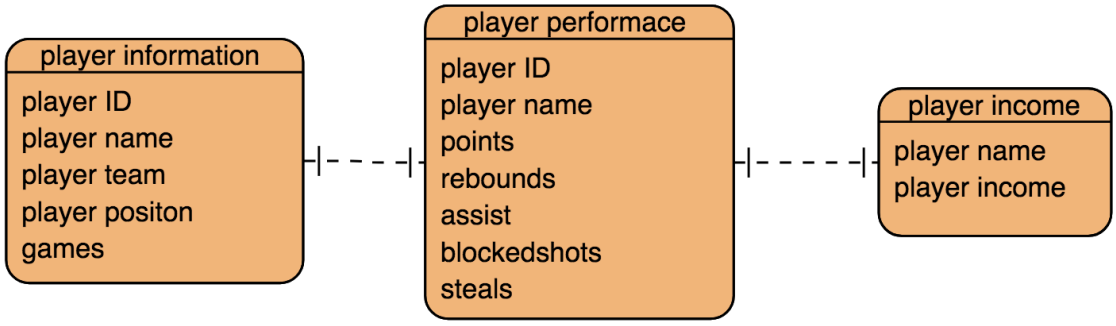
INFO 6210 Data Management and Database Design
The report for homework 1
Quanhan Sun / Guoyan Li

Abstract:
The fantasy data is a community built NBA players' database. Every piece of data has been added and we just want to extract data about the season 2016-207. Here, our aim is to extract NBA player data and reformat it to fit in a Conceptual database schema.

Object: In this assignment, we choose the topic about NBA players. By the joint of all data that we are gathering, we want to find the relationship between the player's performance and player's income, and finally, make a list of the top 200 income players.

Data resource: We collected data from three different sources:
1.NBA data feed API
2.The raw data downloaded from the website of the fantasyData
3.The website of ESPN: http://www.espn.com/nba/salaries/_/year/2017/page/

Object model diagram:



Conceptual schema explanation:
Firstly, we can easily find that this is a one to one model. In the first section, we define the entity of players information, which including four attributes: player ID, player name, player team, and player position. In the second section, we define the entity of player performance, every player ID and name can identify a player's performance. We include seven attributes in the second section, which can evaluate the player's performance such as the total points in 2016-2017 season, the total number rebounds, the total number of assist, the number of blockedshots and steals. The data collected in the first two section are form the same website, so they enjoy the same player ID. According to the player performance, we set up the third section, where the entity is the player income, to find the connection between the player income and the player performance.

Conclusion:
We can make a conclusion in the final csv file named top200NBAplayer_income that the player's income has strong relationship with their performance. But at the meantime, we cannot just use the total point use measure the player's performance. Players' income is based on the overall performance in their team positions.

Percentage contribution in assignment:
QUANHAN SUN and GUOYAN LI did the 50% of the assignment, respectively.

** The details about the code are explained in the Jupyter Notebook file.