Big Data Application Final Project Guide

Visualization and Analysis with NBA Player's Statistics

Liang Su

33551247

1. Target Users and Task

The target users of the project are basketball enthusiasts, who would be interested in knowing a basketball player's performance from statistics without understanding excessive professional statistical attributes. In this project, a user should be able to search for the player interested in and select proper time ranges. Probably I will provide comparison tools also for data exploration. The tasks of project can be summarized as the following:

Analysis Task:

- Display players' basic information including image, name, weight and height, etc.
- Display players' basic performance data, such as points, assists, rebounds, field goal attempted, and turnovers.
- Display data details based on users' focus, for instance, displaying number of points made for each year or game within the given period.
- Display players' ranking of an attribute.
- Display data on both time base and spatial base, for example, scoring data in terms of both time and position on court.

Comparison Task:

- Compare two different players' performances
- Compare two different teams' performances

2. Questions

In my project, I am trying to analyze basketball players' performance. The first question I want to answer is related with simple date derived. For example, "What is A player's points in B time period?" or " How many is A player's assists in B time period?". Moreover, I want to analysis the balance of a player's skills. What is more, answers for the detail of a specific aspect, such as points, assists, rebounded, should be included in our design. For example, I need to derive data detail of rebound in order to let users to analyze or compare.

Based on previous questions I desire to answer in our design, I also come up with a new idea to show data as both time based and spatial based. Therefore,

with analysis the data on time line, users could also analysis data by position of basketball yard.

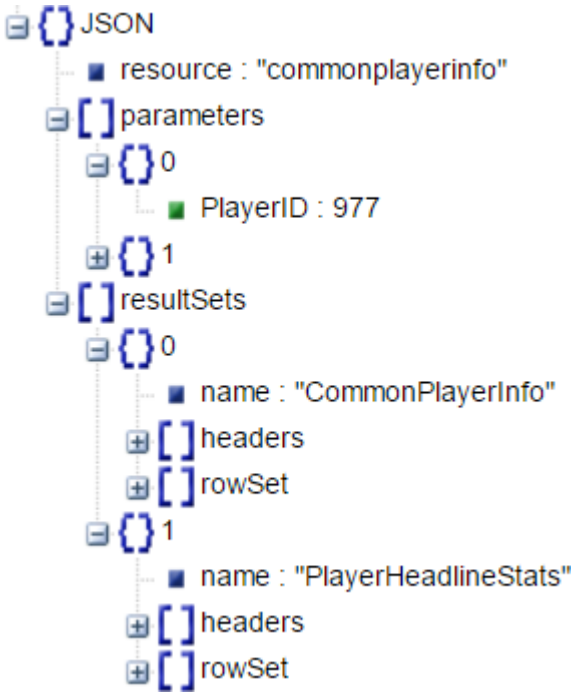
3. Must Have Features

- Filter different players, year.
- Display players' basic information.
- Display player's team transfer.
- Derive data to show general players' points, defense rebounds, assists, offense rebounds and show the balance of players' performance.
- Derive data to show the ranking of a specific player based on the users' selecting aspect.
- Derive data to show the detail data of each aspect of a player.
- Compare two players' performance.

4. Data and data process

Data Source:

I will collect data from the API provided by www.nba.stats.com. The API generates JSON files based on query parameters I provided. I plan to include data in terms of both each player and each team. Player data consists of mainly four parts: player's general information, player's career summaries, player's season summaries and player's play-by-play records. All data has similar structures like the following figure:



Scraping Method

I wrote python scrapers for grabbing data from the website. The program starts by downloading the player list, and then it will grab data for each player based on the list. All scripts that I used have similar structures.