Troy Ang Ko – Rupert Ed Bana – Manuel Jose Samaniego

Project Report Paper

A. Big Data Problem

- **Question**
  - Given the statistical records of NBA players, how do we determine the average number of points, rebounds and assists per game of a player's age?
- **Short Background**
  - National Basketball Association, also known as NBA, the world's leading professional men's basketball association located in North America. The problem will look into the number of the main 3 stat line of a basketball player: points, rebounds, and assists. The goal is to group the statistics of the players depending on their age, and determine their average points, rebound, and assists. Through this, the interval of each stat as the player ages is calculated to see the trend line of the growth of the NBA players. Several conclusions can come up with the data such as the peak age of an NBA player, the productivity length, and the whether the stat line production shifts from points, rebounds, and assists depending on age. The statistics will be compared as the average stat per age of player per game. The data that will be used for this research will be composed of all the game production statistical line of the NBA players from 1950-2017.
- **Data Resources**
  - The group shall make use (and make necessary edits) of the data given in the statistics taken from the Kaggle website.

B. Description of source dataset

The dataset is obtained via Kaggle[1] It contains all statistics from every player in the National Basketball Association (NBA) history. This includes the number of years and seasons they played, their position, teams they played for, number of 3, 2 or 1-point shots, number of assists and rebounds, etc. It is worth noting that a player may appear more than once in a year to account for him playing for different teams in that same year.

[1] http://wagesofwins.com/nba-players-age-like-milk/

C. The Output

The output contains all ages from 18-44 (denoted by 'Age'), the number of games played (denoted by 'G'), and the number of points (denoted by 'PTS'). The team decided to divide the ages into the following age range:

- 18-21
- 22-25
- 26-29
- 30-33
- 34-37
- 38-41
- 42-44

Total points were subsequently divided with the total number of games to find the average points per game on each of said age range.

Based from the code done, along with research gathered, the following are the average points per game on each age range (using non-decimal numbers):

- 18-21 – 775 Points
- 22-25 – 801 Points
- 26-29 – 682 Points
- 30-33 – 694 Points
- 34-37 – 343 Points
- 38-41 – 177 Points
- 42-44 – 125 Points

The points above show that the age range of 30-33 has the highest average points per game, which is likely due to the 'prime' age principle, where a player's peak performance is dependent on his prime age[2].

[2] http://wagesofwins.com/nba-players-age-like-milk/

D. Output Visualization

Below is a histogram depicting the average points per game on each age range.