Database Term Project Report

Title: Which Colleges Produce the Best NBA Players?
Team: 2019NationalChamps
Team Members: Alexander Kimbrell, Obediah Blair, Pravallika Nallamotu

Introduction:
The NBA, alongside most major sports, continues to dive deeper into the realm of data analytics. Professional basketball teams are constantly trying to discover the next great basketball player. Who will be the next Michael Jordan? In an effort to solve this question, our project aims to answer just where these players come from. Specifically, what collegiate basketball programs are able to produce the top NBA talents that we hear about on ESPN on a nightly basis? My team and I have collected data which contains NBA players' statistics from 1978 to 2017 (40 seasons). This dataset includes statistics such as points scored, assists, rebounds, steals, blocks, turnovers, field goal percentage, and games played along with many others.

Original Contributions:
Our team aims to rank players based on their career statistics by creating a metric known as the 'KNB Score'. The KNB Score is:

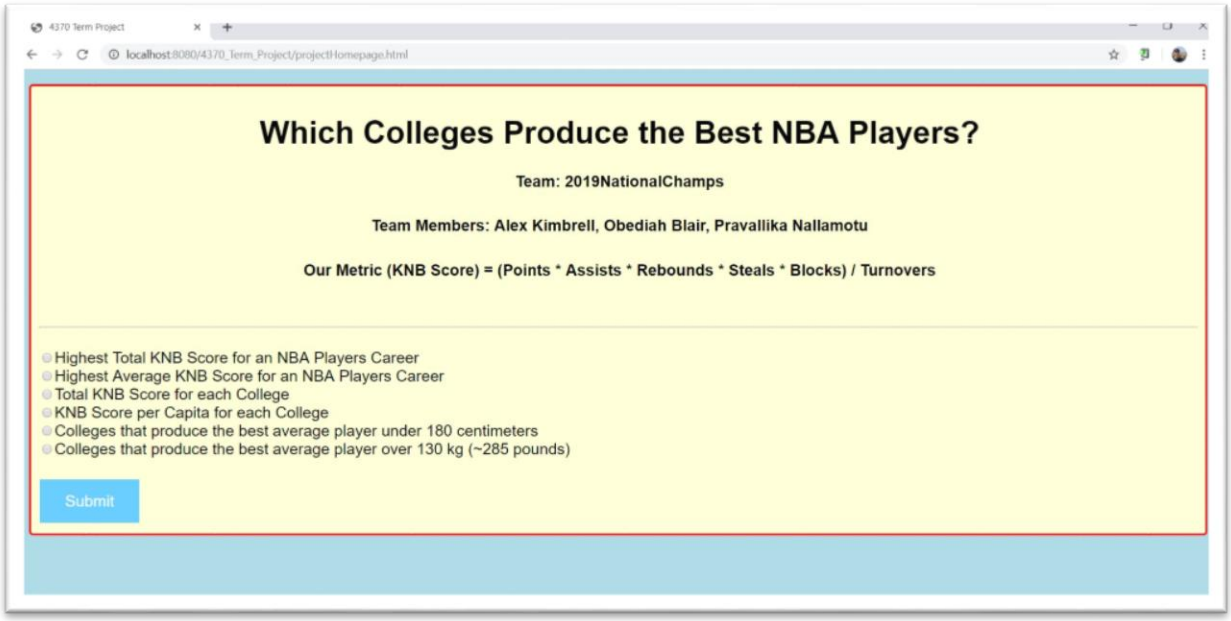
$$\text{KNB Score} = \frac{\text{Points}*\text{Assists}*\text{Rebounds}*\text{Steals}*\text{Blocks}}{\text{Turnovers}}$$

Every NBA player in our dataset has a calculated KNB Score for their career. In order to find out which colleges produce the best players, we combined the KNB Score for every player that has attended The University of X. For example, Michael Jordan has a career KNB Score of 450 billion which contributes to his Alma Mater North Carolina's Total KNB Score of 975 billion. We also created several ways to rank each college. For instance, North Carolina has one of the highest Total KNB Scores (2nd only to UCLA). We also thought that colleges could be normalized by the number of NBA players they've had. North Carolina has put over 500 players in the NBA. Therefore, it makes sense that UNC would have a higher Total KNB Score than The University of Georgia which has 104 NBA players. With this in mind we divided the Total KNB Score by the number of NBA players from each university to create colleges' Average KNB Score. The rankings for colleges with the highest Total KNB Score is listed below:

College	Total Score
University of Houston	1.0376856972276669E12
University of North Carolina	9.761897284109253E11
University of California, Los Angeles	8.0540351865094E11
United States Naval Academy	5.8959888757E11
Georgetown University	4.655163887266918E11
Duke University	4.284268561361171E11
Clemson University	4.150226178729256E11
University of Kentucky	3.695155219509291E11
Louisiana Tech University	3.634272625456859E11
Wake Forest University	3.622904824531206E11

Broader Impact:
With a better idea of what college basketball programs produce the best NBA players, we feel that our results suggest which basketball programs will be more successful in the future based on the trend of NBA talent they're producing. This could also lead to predicting the outcomes of individual college basketball games by comparing which team is more talented. Our findings may also have an impact on collegiate basketball recruiting efforts. Many of the top high school players are looking to attend a college that best prepares them for an NBA career. Our results give recruits an idea of which schools they should consider and which schools they shouldn't.

System Architecture:
Our project utilized many software components. MySQL Workbench was used to develop our basketball related queries. We used basic HTML and CSS for the Web-Frontend. Java was used for the backend along with JDBC in order to connect to our MySQL tables and data. We also used Java Server Pages (JSP) along with Servlets in an effort to coordinate our HMTL form element and java backend by supporting simple GET requests to fetch the query results. Our project homepage looks like this:



Users can select any of the six available queries that include the NBA Players with the highest career KNB Score and Colleges with the highest Total KNB Score. We used our metric to find out the colleges that produce the best players with a height under 5'10'' and which colleges produce the best players with a weight over 285 pounds.

Technical Details:
Our dataset initially contained over 50 attributes for each player. We removed several attributes from the dataset since we knew which statistics we wanted to focus on. Once we preprocessed our dataset, we began the normalization process. We applied BCNF normalization to divide our data into two tables using Player Name as the primary key for both. The Season Stats table contains season-by-season statistics for each NBA player while the Players table contains the college the player attended as well as the player's height and weight measurements.

Related Work:
Sports Analytics is a growing industry, and basketball is no exception. There are many metrics that have been created and several are used by NBA teams to measure how well players are playing. One example is Player Efficiency Rating (PER) which measures a player's statistical performance on a minute-by-minute basis. However, no metric is perfect. PER is biased toward offensive statistics rather than balanced between offensive and defensive statistics. After analyzing our metric, the KNB Score, we've noticed that one weakness of our metric is that it grows exponentially. For instance the difference between Hakeem Olajuwon (ranked first) and David Robinson (ranked second) is much greater than the difference between Grant Hill (ranked fifty-first) and Terry Cummings (ranked fifty-second). This has a strong effect on our Per Capita College KNB Score. Colleges that have produced on a few NBA players, but one of whom was an all-time great player will skew that colleges Per Capita KNB Score. For example, Indiana State is ranked second in Per Capita KNB Score. Indiana State has only produced 18 NBA players, but one of those players is Larry Bird. Larry Bird has a very high individual KNB Score which skews the average for Indiana State.

Team Member Responsibilities:

- Pravallika Nallamotu:
    - Analyzing data
    - Formula for ranking players
    - Website Design (HTML, CSS)

- Alexander Kimbrell:
    - Analyzing data
    - Formula for ranking players
    - SQL Query for ranking colleges
    - Website Backend (Java, JSP)
    - Connecting Database to Website (Java/JDBC)

- Obediah Blair:
    - Analyzing data
    - Formula for ranking players
    - Optimizing SQL Queries
    - SQL Query for ranking players

Conclusion:
Our results have shown that The University of Houston, The University of North Carolina, and The University of California, Los Angeles are the best collegiate basketball programs based on the NBA talent they've produced. The United States Naval Academy and Indiana State had the highest College Per Capita KNB Score, and Hakeem Olajuwon and David Robinson were the highest rated individual players. Overall, we felt like this project was a great opportunity to get to work with a large-scale real world dataset. We were also able to put together several of the skills that we learned throughout the semester such as forming queries and optimizing them, performing normalization using techniques such as BCNF to reduce redundancy in our database schema, and connecting our database to java using the JDBC.

References:

1. https://www.kaggle.com/drgilermo/nba-players-stats
   This is the link to the dataset and also provides some general information about the data we are using for both tables in this project.

2. https://www.basketball-reference.com
   This is an additional website that contains a lot of both simple and advanced data metrics for NBA players and teams. Much of the data that is in our dataset was scraped from this site. However, this website contains also contains several other features that will be very useful for analyzing our results and determining the validity of our own ranking metric.