

# CS-380 Final Project Report

- Gary Miller

## Abstract

The goal of this research was to determine the effect that the average height of an NBA team's lineup will have on the success of the team in that season. There is no question to the fact that being tall has a significant advantage when playing basketball, but how significant does this advantage effect the success of a team at the sport's highest level of competition? This research analyzes a sample of 5 NBA season from 2006-2007 season through the 2010-2011 season in order to identify some interesting trends focused around the starting lineup's average height. Some key points that are focused on is how much height can effect a team's ability to win regular season games, make the playoffs, and win championships. This research also focuses on identifying some interesting trends regarding the height of starters at specific positions on the roster. Though it is clear that there are more significant factors that go into how successful a team will be, such as how skilled a team's bench players are, the number of all star players on the team, or the level of competition within the team's division and conference; it is still a very interesting concept to consider.

## Data

In order to be able to identify any interesting information, the relevant data to the question and where to obtain this data had to be decided. Due to the fact that the variables in questions were height and team performance, obvious values to be located were the name of each team, its wins and losses for each season within the sample, the team's starting lineup for each season and the each individual player's height, and finally if the team made the playoffs. With data on sports being entirely public information, a very large number of websites offering the desired information were easily found. Unfortunately, no public database containing data on NBA player's height and team's season records was located and therefore the data had to be manually retrieved from several different locations. First, data related to each team's record (wins and losses) was retrieved from the NBA's official website. Though plenty of other relevant data that was needed could have be retrieved from this site, other databases that structured this data and a way that was easier to obtain were found and utilized. Two of these sites included basketball-reference.com, which gave information about each player's height, and realgm.com, which listed the starters at each of the 5 positions for every team and for every year. It is important to note that a starting lineup is something that is subject to high variability throughout a season for a large number of reasons, injuries and rest for example, and therefore it was determined this site was the best to use because the starting lineup was

based off of the number of starts throughout the entire season. The final site used to retrieve data was ESPN.com; which was used to determine what teams made the playoffs, won either conference, and finally who won the championship. ESPN was decided to be used for this data because it displayed the playoff results for each season in a visually appealing and easy to read bracket. After all of the necessary data was retrieved it was entered into a spreadsheet and saved as a csv file to easily be inserted into the database.

## Schema

Due to the relatively straightforward nature of the information only team tables were needed to be constructed to efficiently analyze trends in the data. The first table that was created was named team, where each row contained information about all of the team's wins, loses, each starting player's name, their respective height, and if the team made or missed the playoffs. The second table that was created was named NBA, and contained information pertaining to what 16 teams qualified for the playoffs in each of the 5 sample year, which team won either conference, and finally who won the champions series. By utilizing the two tables containing this information, queries could be entered in an attempt to find interesting information about the season results. While there was only two tables used to analyze the table, each table had a more than 15 columns. The schema for both of these tables can be viewed in Figure I.

## Database Software & Interface

Due to the fact that there were only two tables used in this project, it was determined that a simple software with little implementation to initially setup the database would be the optimal choice for this research and therefore DB Browser was selected. The steps taken to construct the database were as follows, first the data was obtained and manually entered into two spreadsheets, each containing the relevant data for one of the aforementioned tables (described more in data and schema sections). The next step taken was to create a text file which contained the code entered to create each of the two tables and is depicted in Figure I. It can be seen that immediately before creating either table, that respective table is dropped or deleted. This is for the event in which changes are made within the spreadsheets and the data should reflect these changes. If the table is not first dropped, an error will occur because SQLite3 already recognizes that table as existing and no changes will be made. Following the creation of the two tables and the initialization for each attribute with its associated integrity constraints, two commands that simply check to make sure the tables were created properly and also that the schema is correct are given. The command ".tables" simply lists the names of all created tables and the next command ".schema" outputs the schema that was assigned to each existing

table. Next for the purpose of importing the data from the two spreadsheets, a separator was selected. This occurs on the line:

```
.separator ",";
```

This command simply signifies to the SQLite3 software that if sequential data is entered into the database, every comma parsed signifies a the value of the next attribute and every new row in the spreadsheet represents a new entity. Finally two commands are given on the lines that read:

```
.import data/NBAHeight5Year.csv Team
```

```
.import data/nbaSeasonData.csv Nba
```

These commands simply tell SQLite3 to import all the data from the file, who's file path is given in the first argument into the table that is given as the second argument.

Finally in order to actually run the commands explained above a command was entered into the terminal within the directory that the set up text file was located and where the SQLite3 file will be created. The command entered was:

```
cat nbaSetUp.txt | nba.sqlite3
```

This command is used because the terminal command “cat” tells the operating system to sequentially read the file given as the first argument into the file given as the second argument. The pipe symbol , “|”, is used to signify that argument 1 should be read into argument 2. After this commands successfully executes the database file can be opened and utilized through the DB Browser software that is installed on the Alden Hall lab computers. Since a simple software was used for this database, there was no source code from a programming language such as python required.

## Queries & Data Analysis

After the database was constructed and the software used for querying was decided, a series of queries were entered to find interesting trends within the data. Due to the fact that the main focus of the research was determine the effects of the height of players within a team's starting lineup, the first information extracted from the data was the average height of players at each position and then the average height of team's entire starting lineup throughout each of the 5 years included within the sample. This was found by entering the commands:

```
SELECT AVG(pgH), AVG(sgH), AVG(sfH), AVG(pfH), AVG(cH) FROM TEAM;  
SELECT AVG(avgh) FROM TEAM;
```

The first command outputs the average height of each player and is shown in Figure II, it is important to note that size is evenly and normally distributed among each position. The second command selects the average for the value

of the attribute avgh, which represents average height of an individual lineup, and the value was found to be 79.331. This means that the average height of an NBA starting lineup in the years 2006-2010 was equal to 79.331 inches or equivalent to 6' 7.331". Next, given the average lineup height as a benchmark to start with, the number of teams that are above or below this value were found by entering the two queries:

```
SELECT COUNT(ID) FROM TEAM WHERE avgh > 79.331;
SELECT COUNT(ID) FROM TEAM WHERE avgh < 79.331;
```

The results of these queries showed that 81 lineups among the 150 sample lineups were above this average and 69 lineups were below this average. This suggests that the distribution of height of skewed right and the shortest players within the sample could be potential outliers that heavily weigh down the average height of a team's lineup. With this in consideration the next information that was obtained from the database was what teams had the shortest average height and using that information, what proportion of those teams made the playoffs. In order to find this out the 15 shortest teams (bottom 10%) were identified using the query:

```
SELECT name, year, playoff FROM TEAM ORDER BY avgh LIMIT 15;
```

The output of this query is depicted in Table I and shows that only 5 teams (only 4 distinct organizations) within the bottom 10% of height were able to clinch a playoff birth, of those 5 teams none of them won a conference championship game. In order to determine the significance of this result, the top 10% tallest teams were identified using the query:

```
SELECT name, year, playoff FROM team Order BY avgh DESC LIMIT 15;
```

The output of this query is depicted in Table II and surprisingly shows that 5 teams as well (also 4 distinct teams) within the top 10% of height clinched playoff births, and of those 5 teams 2 teams, the 2006-2007 Cavaliers and the 2010-2011 Mavericks managed to win the conference playoff game (Mavericks also won championship title). These results suggest that perhaps height isn't as significant in regular season than it is in playoff play. Though due to a small sample size no conclusions can be made. By referring to Table I and II, the minimum and maximum values for average height were found to be 77.4, 1.931 inches less than average, and 81.2, over 1.869 inches more than average (the tallest team, the 2006 Cavaliers also won the Eastern conference). This shows that the average values only have a range of 3.8 inches and either value only differs by .062 inches with respect to distance from the mean. This suggests that the distribution is not a standard normal distribution, but is relatively close. The next series of queries that were entered determines the number of teams who fall into the respective categories of clinching playoffs and being above or below the average height. the series of queries entered were as follows:

```
SELECT COUNT(ID) FROM team WHERE avgh > 79.331 AND playoff = "yes";
SELECT COUNT(ID) FROM team WHERE avgh < 79.331 AND playoff = "yes";
```

```
SELECT COUNT(ID) FROM team WHERE avgh > 79.331 AND playoff = "no";
SELECT COUNT(ID) FROM team WHERE avgh < 79.331 AND playoff = "no";
```

When considering the sample size of 150 teams and the fact that 16 teams make the playoff, and average value should be around 38 teams for each category. Interestingly enough the results show this to be true for situation where teams were not above average and made the playoffs, 42 teams, or below average and missed the playoffs, 31 teams. The fact that these values both spread 6.5 away from 37.5 (one fourth of the sample size, with respect to four possible outcomes). 38 teams that were above average height and made the playoffs and 39 teams were above average and missed the playoffs; it is interesting to see how close the values are when only the two conditions don't agree as the hypothesis of this research assumes. Finally, the average height of the team that won the NBA championship was analyzed by the following queries:

```
SELECT, eastChamp, westChamp, nbaChamp FROM NBA;
```

The results are shown in Figure III. By referring to Figure III, a series of more specific queries were entered to analyze the average height of each championship team.

```
SELECT name, avgh, FROM TEAM WHERE ID = "spurs2006";
SELECT name, avgh, FROM TEAM WHERE ID = "celtics2007";
SELECT name, avgh, FROM TEAM WHERE ID = "lakers2008";
SELECT name, avgh, FROM TEAM WHERE ID = "lakers2009";
SELECT name, avgh, FROM TEAM WHERE ID = "heat2010";
```

The results of these queries showed that the 2006 champion Spurs had an average height of 79.4 inches, almost exactly the overall average by only being .069 inches above average. The 2007 champion Celtics had an average height of 79.2 inches, only being .131 inches below average. The 2008 champion Lakers had an average height of 79.8 inches, exceeding the overall average by .469 inches. The 2009 champion Lakers had an average height of 79.4 inches, almost exactly the overall average by only being .069 inches above average. Finally the 2010 champion Heat had an average height of 78.6 inches, a value that was .731 inches below the overall average. By looking at these results, no true pattern emerges which suggests either average height is insignificant when it gets to the championship playoff series or a larger sample size is necessary.

## Conclusions

This research was able to conclude that the average height of an NBA team's starting lineup does not have significant impacts on how well that team will perform. It is likely that other factors such skill of bench/role players, number of all-star players, coaching, and strength of other teams within a division and conference most likely play a more significant role on the team's overall performance. Though, the impact of average height is not nearly as significant

as originally hypothesized, it certainly does present some interesting trends. If this project was to be extended in the future, the most likely changes would be to increase the number of seasons in the sample, and possibly include more statistics than height such as points per game, offensive/defensive efficiency, or include the entire team's roster.

```

CREATE TABLE Team (
    ID                PRIMARY KEY NOT NULL,
    name              VARCHAR(25),
    year              INTEGER NOT NULL,
    wins              INTEGER NOT NULL,
    losses            INTEGER NOT NULL,
    avgH              REAL NOT NULL,
    pgH               INTEGER NOT NULL,
    pgN               VARCHAR(75) NOT NULL,
    sgH               INTEGER NOT NULL,
    sgN               VARCHAR(75) NOT NULL,
    sfH               INTEGER NOT NULL,
    sfN               VARCHAR(75) NOT NULL,
    pfH               INTEGER NOT NULL,
    pfN               VARCHAR(75) NOT NULL,
    cH                INTEGER NOT NULL,
    cN                VARCHAR(75) NOT NULL,
    playoff            VARCHAR(4) NOT NULL
);

DROP TABLE Nba;

CREATE TABLE Nba (
    year              VARCHAR(10) PRIMARY KEY,
    west1             VARCHAR(25) NOT NULL,
    west2             VARCHAR(25) NOT NULL,
    west3             VARCHAR(25) NOT NULL,
    west4             VARCHAR(25) NOT NULL,
    west5             VARCHAR(25) NOT NULL,
    west6             VARCHAR(25) NOT NULL,
    west7             VARCHAR(25) NOT NULL,
    west8             VARCHAR(25) NOT NULL,
    east1             VARCHAR(25) NOT NULL,
    east2             VARCHAR(25) NOT NULL,
    east3             VARCHAR(25) NOT NULL,
    east4             VARCHAR(25) NOT NULL,
    east5             VARCHAR(25) NOT NULL,
    east6             VARCHAR(25) NOT NULL,
    east7             VARCHAR(25) NOT NULL,
    east8             VARCHAR(25) NOT NULL,
    westChamp         VARCHAR(25) NOT NULL,
    eastChamp         VARCHAR(25) NOT NULL,
    nbaChamp          VARCHAR(25) NOT NULL
);

```

Figure 1:

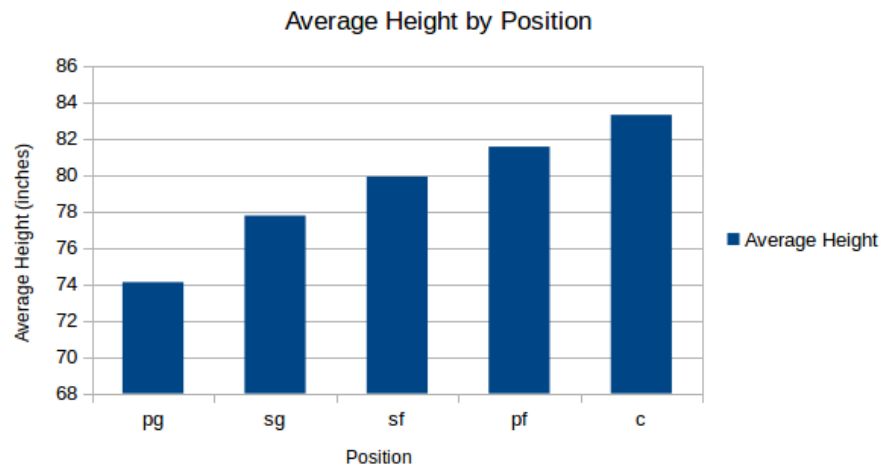


Figure 2:

	year	eastChamp	westChamp	nbaChamp
1	2006-2007	cavaliers	spurs	spurs
2	2007-2008	celtics	lakers	celtics
3	2008-2009	magic	lakers	lakers
4	2009-2010	celtics	lakers	lakers
5	2010-2011	heat	mavericks	mavericks

Figure 3:



1	Team	Year	Playoff	Average Height
2	bucks	2006	no	77.4
3	hornets	2008	no	77.4
4	warriors	2009	no	77.6
5	timberwolves	2008	no	77.8
6	hawks	2007	yes	78
7	hornets	2007	no	78
8	rockets	2010	no	78
9	bucks	2008	no	78.2
10	bucks	2009	yes	78.2
11	heat	2007	no	78.2
12	rockets	2009	no	78.2
13	timberwolves	2007	no	78.2
14	warriors	2006	yes	78.2
15	76ers	2007	yes	78.4
16	76ers	2008	yes	78.4

Figure 4:

1	Team	Year	Playoff	Average Height
2	cavaliers	2006	yes	81.2
3	nets	2007	no	81
4	kings	2009	no	80.8
5	wizards	2008	no	80.8
6	cavaliers	2007	yes	80.6
7	pacers	2007	no	80.6
8	thunder	2008	no	80.6
9	<u>celtics</u>	2006	no	80.4
10	kings	2010	no	80.4
11	<u>knicks</u>	2006	no	80.4
12	mavericks	2010	yes	80.4
13	<u>raptors</u>	2009	no	80.4
14	rockets	2006	yes	80.4
15	wizards	2010	no	80.4
16	blazers	2008	yes	80.2

Figure 5: