DS 501: CASE STUDY 2

Business Intelligence in the NBA: valuation of player performance

Group 6:
Elizabeth Karpinski
Jingnan Xu
Qian Wang
Xiaoshuai Li

1. DATA COLLECTION

We collected a variety of game-oriented data on NBA players that we believe is impactful on the players combined ability to win or lose games as a team. To obtain useful data, we accessed the API that the NBA uses for stats.nba.com. We started by collecting data about individual NBA players via shots API, (see chart in part 3) which gave us a relatively random sample (though vulnerable to alphabetical selection bias) of NBA players. The table we created gave us information on their team, status, and years active. We also used playerProfile API and playerDashPtShotLog API to explore the data by running additional experiments.

2. OUR INTEREST IN THE TOPIC/ MOTIVATIONS

After data collection, we went more in depth in our analysis and focused on a specific NBA player. We chose Chris Paul for this, as one of our group members who has previously followed NBA remembered him from several years ago. He is a particularly interesting player because he is older than average: this makes him more experienced, but at a handicap of being less youthful. He

is one of the highest paid athletes in the world, and top companies from around the world have endorsement deals with him. He has also won an Olympic gold medal, which is one of the highest honors an athlete can achieve. He is 6' tall and approximately 175 lbs in weight. We believe this combination of factors makes him an excellent NBA specimen.

To get a better idea of what makes him such a valuable player (he is especially renowned for his assists and steals) we analyzed data on where he took his shots for the goal.

Another important and complex aspect of basketball is the relative movements of the players. At any given moment, there is a unique arrangement of players on either team with respect to the ball. Nonetheless, certain patterns appear, which make it possible for this to be analyzed and for conclusions to be drawn. The movements of basketball players are of particular interest for data scientists because they can be evaluated in so many possible ways, including regression and cluster analysis, which gives us many different possible ways of understanding the data. In addition, the NBA now is able to provide an unprecedented amount of data on the movements of the players in each game, allowing us to know the exact location and speed of each player and of the ball at each second in the game. This is a truly tremendous amount of data when looked at in the aggregate. In order to deal with it, our team broke it down into smaller segments, and focused on a particular moment in a particular game: Chris Paul taking a serious risk by taking a shot himself, while being guarded by two defenders, despite three open teammates. We used this data, and combined it with our earlier analysis of Chris Paul's shooting abilities, to evaluate his decision. This is also of particular interest because it allows us to take the data under an additional consideration: decision theory. This is an

important tool for data scientists, especially from a business intelligence standpoint, because a data scientist must not only be able to present results, but to interpret them in context and give advice accordingly, especially to management and people of non data science backgrounds. It is for this reason that we picked a particularly controversial moment in the game, rather than one in which a player made an obvious or uncontested decision.

3. Data Analysis & Results

We looked at the collected data through a variety of lenses. For the first part of the project, where we looked at NBA player data (a sample of 20 players can be seen below), we conducted simple analysis. From the full list of 493 players, we started by analyzing the years in which they were active. The first way we looked at this was through a simple chart graph showing of the difference between their start and end years (fig 1).

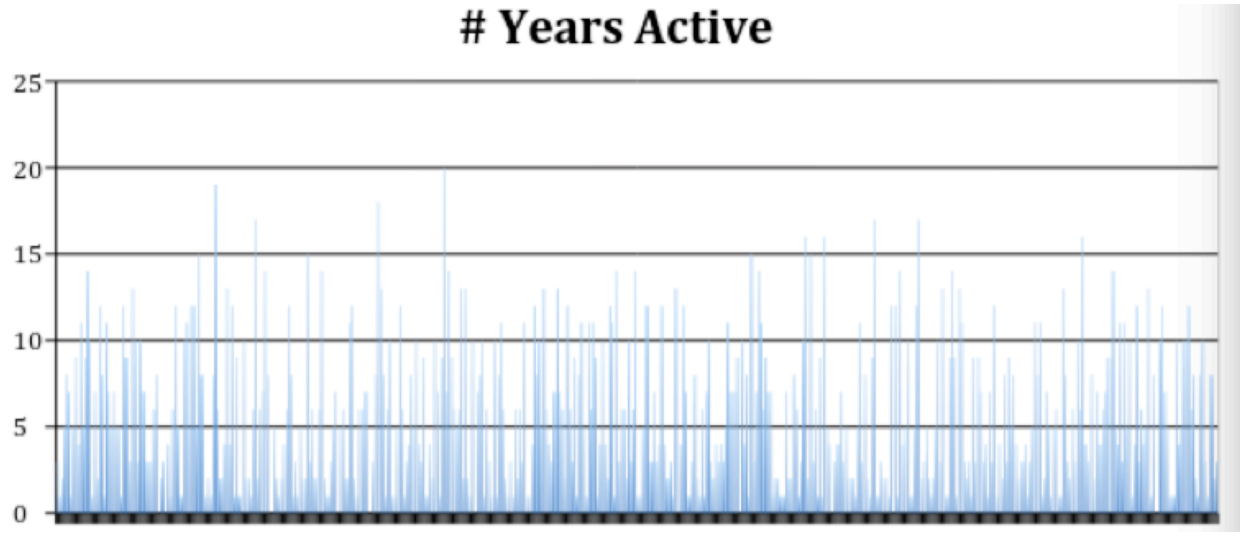

Fig 1: Years Active of NBA Players

From this we can see the approximate distribution of experiences through these NBA players. Another object of curiosity were start years. We knew from the first graph that there were nineteen possible years that these players could have started, from 1995-2014. We counted the number of players who had started in each year and created the following graph to further investigate and visualize the distribution.

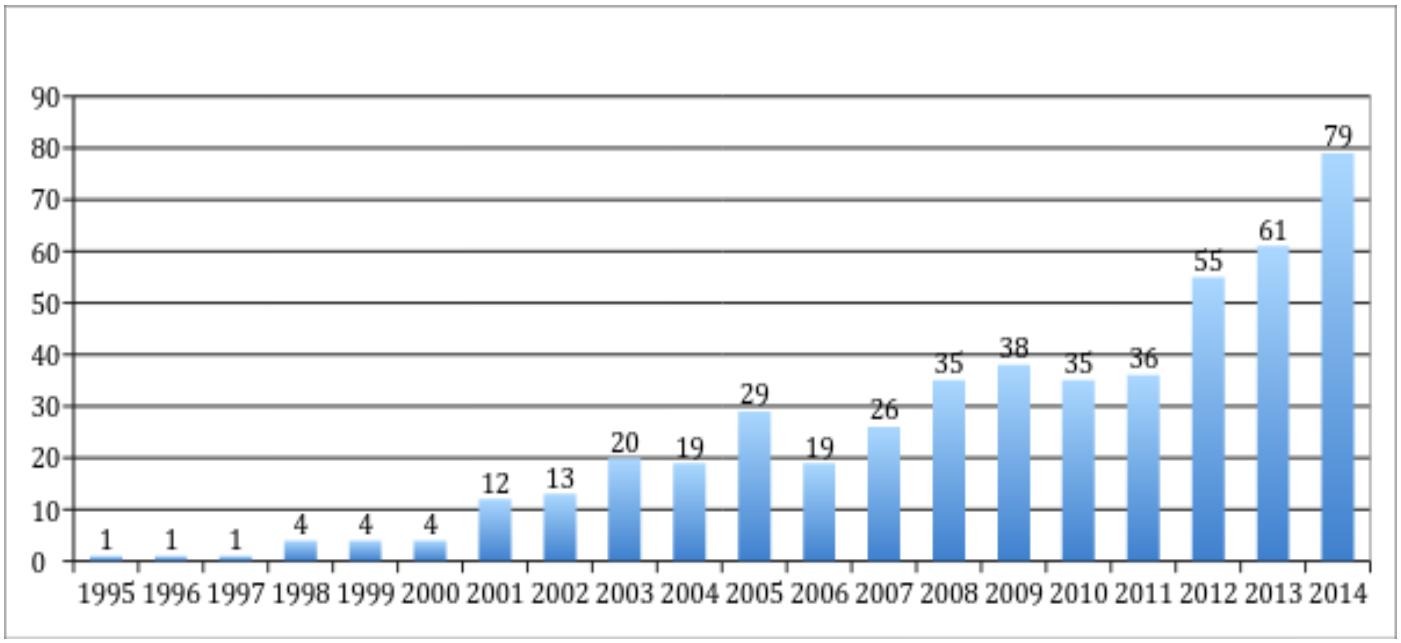

Fig 2: NBA Players Start Year Distribution

This is a better way to visualize the data if one is trying to understand the general trend of NBA players' careers. Despite the prominence of outliers on the previous graph, on fig2 it is much easier to see how comparatively few players started before the year 2000. However, while this graph helps to show the relative proportions of NBA player start years and give us a better sense of the relative proportion of player experiences. For perspective, Chris Paul has been a professional NBA player since 2005.

Other data we collected and analyzed include Chris Paul's record of shots based on position. A sample of the raw data we collected in order to create this plot can be seen in fig 3. Based on fig 3, coordinate values for each shot attempted by Chris Paul were plotted onto a set of axes in gif 4. In order to make the data more meaningful, we added the lines and circles that

	GRID_TYPE	GAME_ID	GAME_EVENT_ID	PLAYER_ID	PLAYER_NAME	TEAM_ID	TEAM_NAME	PERIOD
0	Shot Chart Detail	0021400020	42	101108	Chris Paul	1610612746	Los Angeles Clippers	1
1	Shot Chart Detail	0021400020	48	101108	Chris Paul	1610612746	Los Angeles Clippers	1
2	Shot Chart Detail	0021400020	97	101108	Chris Paul	1610612746	Los Angeles Clippers	1
3	Shot Chart Detail	0021400020	229	101108	Chris Paul	1610612746	Los Angeles Clippers	2
4	Shot Chart Detail	0021400020	268	101108	Chris Paul	1610612746	Los Angeles Clippers	2
5	Shot Chart Detail	0021400020	299	101108	Chris Paul	1610612746	Los Angeles Clippers	3
6	Shot Chart Detail	0021400020	302	101108	Chris Paul	1610612746	Los Angeles Clippers	3
7	Shot Chart Detail	0021400020	303	101108	Chris Paul	1610612746	Los Angeles Clippers	3
8	Shot Chart Detail	0021400020	320	101108	Chris Paul	1610612746	Los Angeles Clippers	3
9	Shot Chart Detail	0021400020	332	101108	Chris Paul	1610612746	Los Angeles Clippers	3
10	Shot Chart Detail	0021400020	350	101108	Chris Paul	1610612746	Los Angeles Clippers	3
11	Shot Chart Detail	0021400020	371	101108	Chris Paul	1610612746	Los Angeles Clippers	3
12	Shot Chart Detail	0021400020	456	101108	Chris Paul	1610612746	Los Angeles Clippers	4
13	Shot Chart Detail	0021400020	467	101108	Chris Paul	1610612746	Los Angeles Clippers	4
14	Shot Chart Detail	0021400020	469	101108	Chris Paul	1610612746	Los Angeles Clippers	4
15	Shot Chart Detail	0021400020	490	101108	Chris Paul	1610612746	Los Angeles Clippers	4
16	Shot Chart Detail	0021400020	493	101108	Chris Paul	1610612746	Los Angeles Clippers	4
17	Shot Chart Detail	0021400020	515	101108	Chris Paul	1610612746	Los Angeles Clippers	4
18	Shot Chart Detail	0021400026	10	101108	Chris Paul	1610612746	Los Angeles Clippers	1
19	Shot Chart Detail	0021400026	39	101108	Chris Paul	1610612746	Los Angeles Clippers	1

Fig 3: Chris Paul Shot Chart

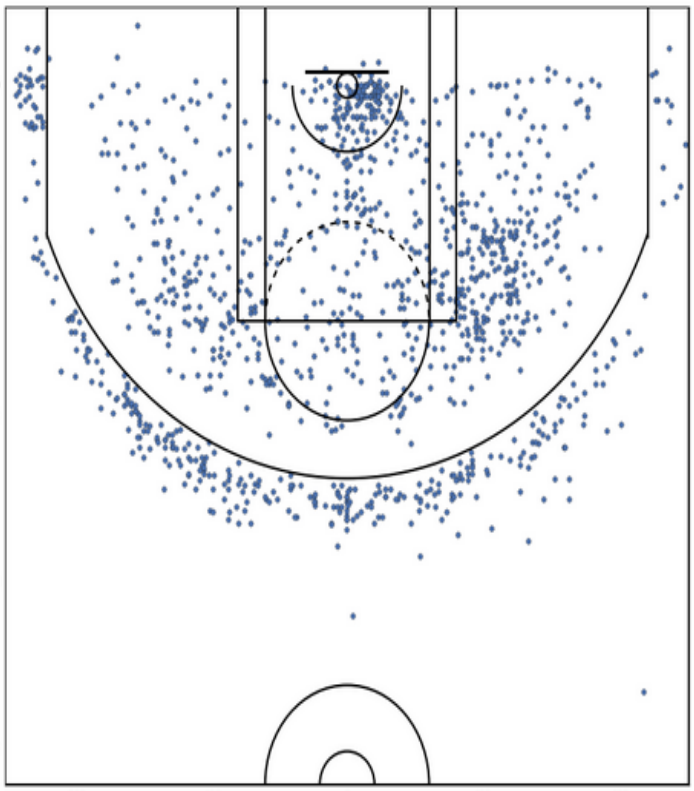

Fig 4: Chris Paul Shot Chart in court

detonate areas on the basketball court, so as to give a better representation of the locations with respect to real world items.

Another aspect of the game we looked at in depth was the specific moments and movements of it. In the moment of the game we collected, Chris Paul makes a risky shot instead of passing to open teammates. Without having seen the game, we can see this in the data, which tells us precisely what is happening. A plot of Chris Paul's movements during this

moment is shown in fig 5. More importantly, we also created a plot of the distance between players. As you can see in fig 6, Chris Paul was being heavily guarded by two opposing players, while his teammate (seen in the orange) was wide open, at a considerable distance from the nearest defender.

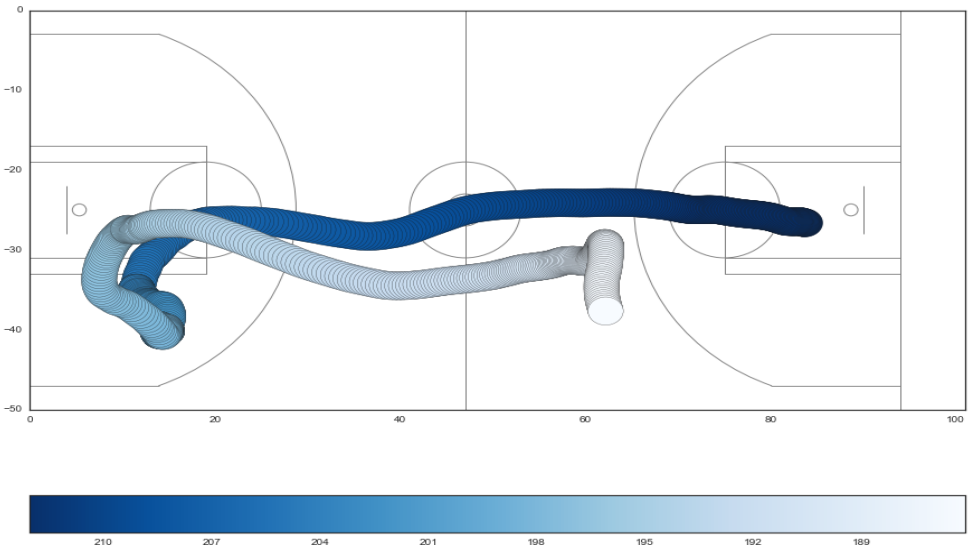

Fig 5: Chris Paul Movements during the moment

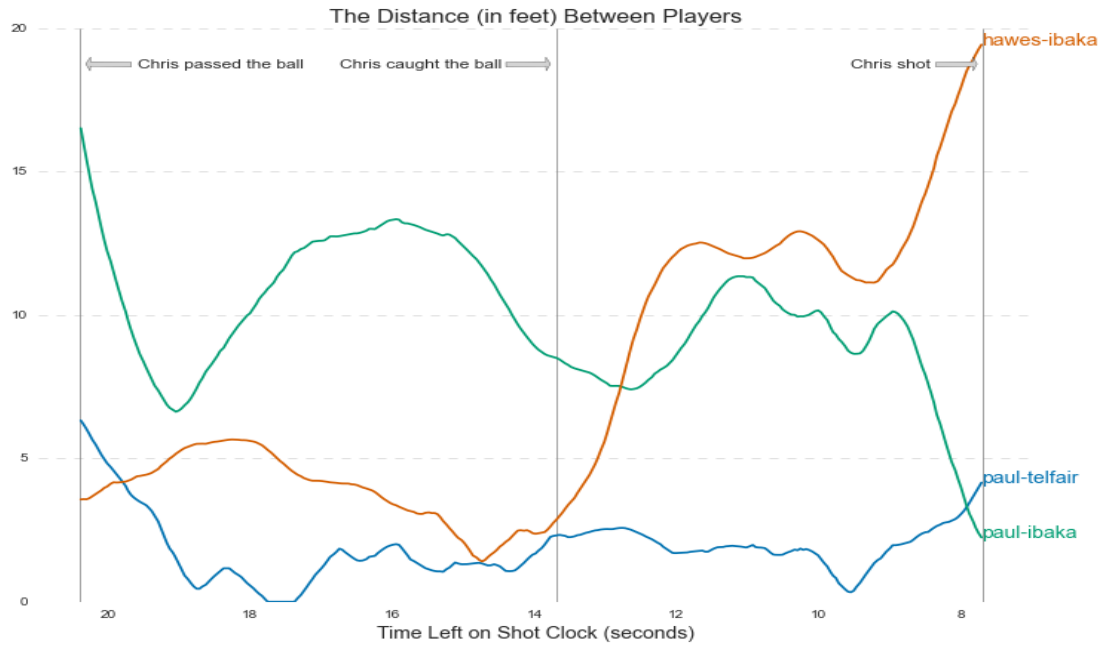

Fig 6: Distance between players during the moment

Another way to think about the data is to compute statistics on the players involved. Some of the statistics were calculated where the Euclidean distance the players travelled, which we then used to find the speed of the players (see fig 7 and fig 8). The players and their order are the same in both lists because we looked at a single ~19 second segment, which made the denominator constant.

player_name	
Andre Roberson	195.583820
Chris Paul	172.508199
Ekpe Udoh	155.586541
Jamal Crawford	189.809648
Matt Barnes	152.804275
Perry Jones	149.819474
Sebastian Telfair	150.657166
Serge Ibaka	153.816636
Spencer Hawes	160.409444
Steven Adams	142.483852
ball	260.935171

dtype: float64

Fig 7: Distance traveled by each player

player_name	
Andre Roberson	5.146761
Chris Paul	4.539529
Ekpe Udoh	4.094238
Jamal Crawford	4.994814
Matt Barnes	4.021023
Perry Jones	3.942478
Sebastian Telfair	3.964522
Serge Ibaka	4.047663
Spencer Hawes	4.221152
Steven Adams	3.749442
ball	6.866472

dtype: float64

Fig 8: Average speed for each player

Given this information, we now turn to a more pressing question. Why did he make the shot instead of attempting to pass to his teammates? We believe the answer lies in the previous graph of his shot attempts. Referencing that, we can clearly see that at the moment in question, he was in a position where he habitually made shots. It seems likely that instinct and practice took over, and urged him to make the shot.

Based on the previous observation, we concluded that whether a NBA player makes a shot or passes the ball to others is related to his experience and habits. So we did another analysis to find out some interesting facts about NBA players habits. We took all of the 2014 NBA players and from this we built a list of their average statistics in terms of movements, using

this kind of data from the NBA. From this we found their average distance from defenders, average dribbles, average shot distance, and average touch time throughout all of their games. These represent four crucial metrics of NBA players. From this we found Jerret Grant, has the longest defender distance, at 7.5', and Andrei Kirilenko had the shortest defender distance at 1.7'.

	name	avg_defender_distance	avg_dribbles	avg_shot_distance	avg_touch_time
0	Adams, Jordan	4.545349	1.174419	13.524419	1.784884
1	Acy, Quincy	4.924159	0.318043	14.162080	1.544343
2	Adams, Jordan	4.545349	1.174419	13.524419	1.784884
3	Adams, Steven	2.731122	0.375000	4.996684	1.508163
4	Adrien, Jeff	2.704651	0.511628	6.527907	1.793023

Fig 9: Average statistics for each player

Touch time is another useful measure: it measures how many seconds the player has control of the ball before they take the shot. In this, a long touch time represents a long time before taking the shot. According to the blog *Basketball Analytics,* a shorter touch time is better than a longer one: It is correlated with more successful scoring: shots taken after more than two seconds are significantly less likely to score. This correlation helps to make touch time a valuable way of assessing the offensive strength of a player. From these metrics we can search for a theoretical ideal offensive player: somebody with short touch times, high defender distances and short average shot distances.

To explore the NBA data in another way, we collected NBA player performance data and predict the MVP for a specific season. We used a formula called Performance Index rating to

analyze all NBA players for the MVP in the 2014-2015 season. Some factors in the formula could not be found in the data we collected and were discounted. The modified formula is: (Points + Rebounds + Assists + Steals + Blocks) - (Missed Field Goals + Missed Free Throws + Turnovers). Fig 10 shows the top 10 candidates for MVP in the 2014-2015 season.

	name	PTS	REB	AST	STL	BLK	PFD	FTA	FTM	FGA	FGM	TOV	Score
103	Cousins, DeMarcus	24.1	12.7	3.6	1.5	1.7	8.8	9.2	7.2	18.1	8.4	4.3	36.4
116	Davis, Anthony	24.4	10.2	2.2	1.5	2.9	5.5	6.8	5.5	17.6	9.4	1.4	35.8
469	Westbrook, Russell	28.1	7.3	8.6	2.1	0.2	6.6	9.8	8.1	22.0	9.4	4.4	34.2
196	Harden, James	27.4	5.7	7.0	1.9	0.7	6.7	10.2	8.8	18.1	8.0	4.0	33.9
234	James, LeBron	25.3	6.0	7.4	1.6	0.7	6.0	7.7	5.4	18.5	9.0	3.9	31.3
138	Durant, Kevin	25.4	6.6	4.1	0.9	0.9	4.9	6.3	5.4	17.3	8.8	2.7	30.7
111	Curry, Stephen	23.8	4.3	7.7	2.0	0.2	3.7	4.2	3.9	16.8	8.2	3.1	29.7
360	Paul, Chris	19.1	4.6	10.2	1.9	0.2	3.6	3.9	3.5	14.3	6.9	2.3	29.5
189	Griffin, Blake	21.9	7.6	5.3	0.9	0.5	5.9	6.4	4.6	17.1	8.6	2.3	29.5
166	Gasol, Pau	18.5	11.8	2.7	0.3	1.9	4.2	4.7	3.8	14.8	7.3	2.0	29.0

Fig 10: Candidates for MVP in 2014-2015 season

Stephen Curry, who is the winner of MVP for the 2014-2015 season, ranks 7th in our result. As NBA MVP reward is cast by fans throng online voting, we should take more factors into consideration to achieve a more accurate prediction of NBA MVP.

References

How to Track NBA Player Movements in Python. Retrieved from:

http://savvastjortjoglou.com/nba-play-by-play-movements.html

How to Create NBA Shot Charts in Python. Retrieved from:

http://savvastjortjoglou.com/nba-shot-sharts.html - Plotting-the-Shot-Chart-Data

The NBA's stats API for the browser or Node. Retrieved from:

https://www.npmjs.com/package/nba

NBA Most Valuable Player Award. Retrieved from:

https://en.wikipedia.org/wiki/NBA_Most_Valuable_Player_Award