CS170A Course Project

Bolun Hu

UID: 104056403

Professor D. S. Parker

**NBA Game Schedule Analysis and Assessment**

**Abstraction**

The project is focused on the problems existing in the design of NBA season schedule, and discusses its potential effect (positive/negative) on each team and its season performance.

By comparing and analyzing the data of past NBA seasons’ game schedule, I first summarized the three main influential factors that could be affecting a team’s performance: time interval between each two games, transitions between home games and away games, power ranking of opponents. For compiling and processing statistical data, I used PASCAL and Excel as my main assistants. And I used different methods when solving values for the quantified factors mentioned earlier.

When considering the effects of the time interval between games, I used the method for computing variance in statistics, and evaluate the team’s schedule based on the fluctuation. When considering the effects of transition between home games and away games, and the power ranking of opponents, I quantified different factors. And after comparing the effects of different factors, I choose the data for transition between arenas as the standard, and normalized other two factors through the use of parameter constant , obtained the three index values , (j representing team number). After that, based on the degree of effect of each factor, I calculated out the effect ratio for each factor. Finally, I multiplied the each factor’s normalized value with its corresponding effect ratio, and added them up, which gives me the final “rating” for the team’s schedule. The results showed that the team with most advantageous schedule is Boston Celtics; the most “unlucky” team is the Milwaukee Bucks.

After the above analysis, I further did a more detailed analysis on the schedule of my favorite team Houston Rockets, and got the conclusion that the schedule for the Rockets is fair, not easy nor too tough, overall.

Finally, I focused on the question of deciding which teams would play against each other 3 times instead of 4 times during a season (for detail of the question please part 3 of **Topics / Domain of Analysis**). I used the idea of greedy algorithm and gave my own choosing method.

**Keywords**: Data Processing, Variance, Analytic Hierarchy Process, Statistical Analysis, Greedy Algorithms, PASCAL.

**Motivation for this project**

NBA is one of the most famous sport associations in the world, with hundreds of millions of fans around the world, and I am one of the fanatics. This fact became especially true since the joining of Yao Ming into the Houston Rockets in 2002 (and as well Yi Jianlian joining the Milwaukee Bucks), and it was also the time I became obsessed with NBA computer games like NBA Live and NBA 2K series. At first I was only interested in the real game playing; and things like graphics cards, CPU speeds, face patch, shoes patch were the most fascinating topics to me. However, when the dynasty mode was introduced in NBA Live 2004, which enables me to really manage a club, the issues relevant to data came into my mind: how to pick a good rookie among hundreds of candidates around the nation? How to design an efficient rotation plan so that not only my players can win the game but also have enough amount of recovering time? Is the season schedule really fair to all the teams? How much money should I pay for a player like Kevin Love or DeAndre Jordan in the free agent market? What is team chemistry? Can it be represented by data? All of these kinds of questions bothered me every time I played the game. And today I am really happy to be given such an opportunity to analyze an interesting topic that confuses but as well excites me for years.

**Topics / Domain of Analysis**

For organization like NBA, coming up with a complete, fair-to-all season schedule is really a huge and complex project, and the schedule would greatly affect each team’s overall performance and records. My focus for this paper is to use the method of mathematical modeling to analyze and assess the fairness of the schedule quantitatively. The season schedule I chose is 2008 – 2009, which is a little bit far from 2014, for the reason that it was the last season that Yao Ming, my idol, participated in. (After that season he was stricken by a serious ankle injury and unfortunately retired from NBA in 2011) Sorry if my obsession for Yao Ming and Houston Rockets bothers you.

My analysis would be focusing on the following 3 main aspects:

1. Determine the factors and parameters that would be decisive and influential when evaluating a season schedule. And based on those factors, I would try to convert the schedule into a numerical format that could be analyzed mathematically, and finally give out the mathematical expression which is valid for rating a team’s schedule.
2. Based on the results in part 1’s computation, I would try to do a more detailed evaluation of the schedule of Houston Rockets for which Yao played. Also I would try to find out the team with most advantageous game schedule, and the team with the least advantageous schedule, based on the results of my computation.
3. There is an interesting fact that confused me for a long time: every team in the NBA would play against other teams within a different Conference for 2 times (e.g. LA Lakers would play against Boston Celtics for 2 times every season) and would play 4 times against teams within same Conference and as well the same Division (e.g. Golden States Warriors would play against LA Clippers 4 times every season because they both are in Pacific Division). **But for teams that are in the same Conference but different Divisions, some of them would play 4 games against each other throughout the season, and some of them would play only 3 games against each other (either 2 away, 1 at home or 2 at home, 1 away).** Let’s call this situation as **some-3-some-4 policy**. As a result, every team in the end would play 41 at home and 41 away, and teams within same Conference but different Divisions would stay in a balanced state. (i.e. for teams in Pacific Division, say, Lakers, they would play the same number of games at home and away in total against teams in Northwest Division and Southwest Division. The same situation holds for all other teams in the league.) **What is the mathematical idea/method behind the some-3-some-4 choosing?** With this question in mind and the idea of greedy algorithm, I would try to find out a reasonable mathematical model for choosing teams that would play against each other for 3 times instead of 4. And I would try to evaluate the method to explore the possibility of improvement / optimization.

**Source of Data**

All data associated with this project can be downloaded or accessed from the following websites:

<http://www.hoopchina.com/teams/>

[www.basketball-reference.com](http://www.basketball-reference.com)

[www.sports.sina.com.cn/nba](http://www.sports.sina.com.cn/nba)

<http://espn.go.com/nba/statistics/player/_/stat/>

Also, the commands and scripts needed for this project, including extracting those data would be appended after this project. I will upload the initial raw data for your reference as well.

**Analysis of the Topics and Questions**

In order to design a game schedule, there are multiple factors needed to consider: the time interval between different games, number of back to back games, miles traveled throughout the season, strong / weakness of opponents, etc. Under such limits, the following 2 main tasks are what I plan to focus on:

1. Design programs to create a satisfying schedule according to the constraints
2. Given the schedule data, analyze its fairness to all teams, and make a detailed analyze on its effects on a particular team

Firstly I would focus on the first two topics. For the first two topics, I would try to deal with it using hierarchical approach: first find out the factors that would be influential to all teams; and then for every factor, analyze its effect on each team: is it a positive effect or negative effect? After that I would analyze on a higher level: evaluate the integrated effects of different factors on a team, and try to compute out an appropriate ratio for each factor. Finally, based on the computation I can assess the “rating” of the schedule for each team, and find out the team that is most advantageous and team that is most “unlucky”.

For the detailed discussion of Houston Rockets, I plan to take step further based on my previous conclusions, by adding more specialized factors that may be discarded in the previous analysis, such as back to back games, the support rate of fans, etc. And I would analyze as detailed as possible to figure out the goodness or badness of the schedule from the perspective of Rockets, and draw a reasonable solution if possible.

For the 3rd topic, I think it would be a good attempt to use greedy algorithm. First we can figure out the possible factors that may influence the choosing of 3 teams. After that I can digitalize those factors and build a mathematical model if possible. Finally I would compare the results with the given one, and formalize the solution for choosing which teams would play against each other for 3 times instead of 4.

**Assumptions before Analysis**

1. All teams would be on-time for all games, and there is no mid-season adjustment for any team.
2. When analyzing the effect of a single factor, I assume all other factors are mutually independent and thus would not affect each other.
3. Assume the strong / weakness of each team stays constant and could be represented by its winning percentage.

**Notes on Symbols and Definitions**

The following definitions are prepared for the later modeling. It might seem to be cumbersome. Skip this part if you want, because they would make more sense as you read the later parts.

*Aj* --------------normalized index for the time interval of team j,j=1,2,3,…30;

*k*1 --------------parameter representing the degree of effect of time interval;

*Bj* -------------- normalized index index for the transition of arena for team j, j=1,2,3,…30;

*C j* -------------normalized index for the power rating of opponents for team j, j=1,2,3,…30;

k2 --------------parameter representing the ratio of effect of opponents’ power;

*Wj* --------------index for the rate of advantageous for team j, j=1,2,3,…30;

*Q*1 --------------ratio representing the effect of time interval (0<Q<1) ;

*Q*2 -------------- ratio representing the effect of transition of arena (0<Q<1) ;

*Q*3 --------------ratio for the effect of opponents’ power (0<Q<1) ;

*Ei* --------------the difference of winning percentage between team i and team j;

*Pj* --------------representing team j’s winning percentage;

*Gi* --------------representing team i’s winning percentage against team j that is in the same Conference but different Division from team i;

*Li* --------------representing team j’s winning percentage against team i;

**Build and Analysis of the Model**

***(a) Analysis of Model***

Aiming at the negative and positive effects that a certain schedule has on different teams, I considered the following factors:

First factor to be considered is the time interval between two games. If the time interval is too short, players would feel exhausted and that would definitely downgrade their performance. On the other hand, if the time interval is too long, that would as well hinder player from playing high quality games, and thus affect the team’s overall performance. After some internet search and analysis, I think the effect of time interval should be major, and decide to choose it to be the main factor, with the ratio Q1.

Second factor is the transition between arenas (i.e. transition of home to away game), it is easy to realize that if a team play multiple away games in a roll, it would cause players to travel continuously, and therefore affect their performance in a negative way. We set the ratio of this factor to be .

Finally is the effect of opponents’ power ranking. For instance, if Sacramento Kings play against LA Clippers today, Houston Rockets tomorrow, and Dallas Mavericks next Monday, such painful scheduling would greatly affect players’ mood, since a losing stream against opponents that are far stronger than yourself would hurt your confidence and give you more pressure, which will restrict your future performance for almost certain. I set this ratio of this factor to be .

Other factors including weather, foods, and transportation method would also objectively affected a team’s performance. But because the degree of those effect seems to be much smaller than what I have mentioned above, I will not include these into my model.

Therefore, I surfed on different NBA data websites and gathered information about the factors I have mentioned, and computed out the ratio of effect for each of those factors. The following diagram shows the results of my research:

Other 5%

Opponents’ power

20% 45% Time interval between games

30%

Change of arenas

***(b) Building the Model***

**(1) Analyze each factor’s influence on a team’s performance**

* Model for Time Interval

Given the schedule, I analyzed the data with the help of PASCAL programming and Excel which helps me to coordinate and neaten the data. The resulting diagram for time interval is shown as follows:

(how to interpret: for example, on the first row: Magic 15, 52, … means there are 15 games with no days of rest (2nd game of back-to-back), 52 games with 1 day of rest, 14 games with 2 days of rest prior to the gamme)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Index | Team Name | Time interval before a game (days) | | | | | | |
| 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| 1 | Magic | 15 | 52 | 14 | 0 | 1 | 1 | 0 |
| 2 | Wizards | 19 | 48 | 10 | 5 | 0 | 1 | 0 |
| 3 | Hawks | 23 | 41 | 11 | 7 | 0 | 1 | 0 |
| 4 | Bobcats | 22 | 40 | 17 | 3 | 0 | 1 | 0 |
| 5 | Heat | 19 | 45 | 15 | 3 | 0 | 1 | 0 |
| 6 | Celtics | 19 | 48 | 13 | 1 | 1 | 0 | 1 |
| 7 | Raptors | 17 | 50 | 11 | 4 | 0 | 0 | 1 |
| 8 | 76ers | 22 | 46 | 8 | 3 | 3 | 1 | 0 |
| 9 | Nets | 22 | 41 | 16 | 3 | 0 | 0 | 1 |
| 10 | Knicks | 19 | 47 | 13 | 2 | 1 | 1 | 0 |
| 11 | Pistons | 18 | 47 | 16 | 1 | 0 | 1 | 0 |
| 12 | Cavaliers | 21 | 44 | 13 | 4 | 0 | 0 | 1 |
| 13 | Pacers | 23 | 38 | 19 | 2 | 0 | 1 | 0 |
| 14 | Bulls | 24 | 35 | 21 | 2 | 0 | 1 | 0 |
| 15 | Bucks | 24 | 38 | 15 | 5 | 0 | 1 | 0 |
| 16 | Hornets | 20 | 47 | 10 | 4 | 1 | 1 | 0 |
| 17 | Spurs | 19 | 47 | 13 | 2 | 1 | 1 | 0 |
| 18 | Rockets | 21 | 44 | 14 | 3 | 0 | 1 | 0 |
| 19 | Mavericks | 17 | 49 | 14 | 1 | 1 | 1 | 0 |
| 20 | Grizzilies | 23 | 40 | 14 | 5 | 0 | 1 | 0 |
| 21 | Jazz | 21 | 41 | 18 | 2 | 0 | 1 | 0 |
| 22 | Nuggets | 22 | 45 | 9 | 5 | 1 | 0 | 1 |
| 23 | Trail Blazers | 19 | 47 | 13 | 3 | 0 | 1 | 0 |
| 24 | Timberwolves | 23 | 37 | 21 | 1 | 0 | 0 | 1 |
| 25 | Super Sonics | 19 | 45 | 16 | 2 | 0 | 1 | 0 |
| 26 | Lakers | 20 | 46 | 12 | 3 | 1 | 1 | 0 |
| 27 | Suns | 20 | 46 | 13 | 1 | 2 | 1 | 0 |
| 28 | Warriors | 16 | 52 | 11 | 3 | 0 | 1 | 0 |
| 29 | Kings | 24 | 43 | 9 | 4 | 2 | 0 | 1 |
| 30 | Clippers | 22 | 44 | 11 | 4 | 0 | 2 | 0 |

Chart 1

Since the number of games for regular season is a constant, the total interval time (rest time) added up should also be constant. My idea is that if the rest time is distributed evenly between every two games, then the team with that schedule should be the most advantageous team. In other words, the smaller the fluctuation of the time interval values is, the better the schedule is. Therefore, I use the concept of variance, and computed out the fluctuation of the time interval.

Based on the above data chart, I calculated out the variance for the time interval with help of Lingo (For the detailed program script, please check out appendix 2):

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Index | Team | Variance | Multiplied by K | Index | Team | Variance | Multiplied by K |
| 1 | Magic | 0.648 | 517.6 | 16 | Hornets | 0.786 | 628.0 |
| 2 | Wizards | 0.944 | 754.4 | 17 | Spurs | 0.798 | 639.3 |
| 3 | Hawks | 0.745 | 596.8 | 18 | Rockets | 0.835 | 667.2 |
| 4 | Bobcats | 0.82 | 664 | 19 | Mavericks | 1.033 | 825.6 |
| 5 | Heat | 0.912 | 726.4 | 20 | Grizzlies | 0.784 | 628.0 |
| 6 | Celtics | 0.653 | 520.8 | 21 | Jazz | 0.895 | 715.2 |
| 7 | Raptors | 0.798 | 639.2 | 22 | Nuggets | 0.83 | 656 |
| 8 | 76ers | 0.893 | 715.2 | 23 | Trail Blazers | 0.860 | 687.1 |
| 9 | Nets | 0.786 | 628 | 24 | Timberwolves | 0.762 | 610.4 |
| 10 | Knicks | 0.710 | 568.8 | 25 | Super Sonics | 0.884 | 706.3 |
| 11 | Pistons | 0.773 | 619.2 | 26 | Lakers | 1.031 | 825.5 |
| 12 | Cavaliers | 0.712 | 568.8 | 27 | Suns | 0.884 | 706.5 |
| 13 | Pacers | 0.712 | 568.8 | 28 | Warriors | 0.835 | 667.2 |
| 14 | Bulls | 0.860 | 687.2 | 29 | Kings | 0.662 | 528.8 |
| 15 | Bucks | 1.13 | 896 | 30 | Clippers | 0.981 | 785.5 |

Chart 2

To unify the order of magnitude of each factor so that I can make comprehensive analysis later, I decide to multiply the variance values by a constant value =800, and got *Aj*.

* Model for Transition between arenas (Home/Away Transition)

According to different locations of different games, I roughly divide the situation into 4 main categories: (later I would build an automata based on this 4 categories)

**1. At Home**

**2. Away but in the same Division**

**3. Away in the same Conference but different Division**

**4. Away in different Conference and different Division.**

When transitioning between these 4 states, I give each of these transitions different values according to its geographical distance: the farer the locations is from home arena, the worse this transition would be, and the larger this transition value of would be. Namely:

1. Transition from Home to Away, within the same Division: Transition value = 4
2. Transition from Away to Home, within the same Division: Transition value = 2
3. Away to Away, within the same Conference, Transition value = 14
4. Home to Away, within the same Conference, different Division, Transition value = 6
5. Away to Home, within same Conference, different Division, Transition value = 5
6. Transition from away where is in the same Conference, but different Division, to another away where is in a different Conference, Transition value = 18
7. Transition from away where is in different Conference, to another away where is in the same Conference but different Division, Transition value = 18
8. Home to Away, different Conference, Transition value = 8
9. Away to Home, different Conference, Transition value = 7
10. Transition from away where away is in different Conference, to another away, where away is in the same Division, Transition value = 16
11. Transition from away where away is in the same Division, to another away where away is in a different Division, Transition value = 16

I then built an automata representing the transition as follows:

Different Conference ------------18----------🡪 Same Confer. Diff. Division

20 16

16

8 7 5 14 14

6

Home 2 Same Division

0 4 10

According to this transition diagram, I computed out the transition values *Bj* for all teams, data chart is shown as follows: (source code please refer to appendix 3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| index | Team | Transition Value | index | Team | Transition Value |
| 1 | Magic | 562 | 16 | Hornets | 557 |
| 2 | Wizards | 552 | 17 | Spurs | 567 |
| 3 | Hawks | 572 | 18 | Rockets | 569 |
| 4 | Bobcats | 561 | 19 | Mavericks | 572 |
| 5 | Heat | 564 | 20 | Grizzilies | 560 |
| 6 | Celtics | 569 | 21 | Jazz | 565 |
| 7 | Raptors | 560 | 22 | Nuggets | 566 |
| 8 | 76ers | 557 | 23 | Trail Blazers | 590 |
| 9 | Nets | 540 | 24 | Timberwolves | 564 |
| 10 | Knicks | 572 | 25 | Super Sonics | 563 |
| 11 | Pistons | 550 | 26 | Lakers | 587 |
| 12 | Cavaliers | 549 | 27 | Suns | 572 |
| 13 | Pacers | 545 | 28 | Warriors | 577 |
| 14 | Bulls | 573 | 29 | Kings | 569 |
| 15 | Bucks | 560 | 30 | Clippers | 569 |

Chart 3

* Effect of Opponents’ Power Ranking

For any team, it would play 3 games with some opponents, and 4 games with other ones (i.e. some-3-some-4 policy). So in such circumstance, there always exist some teams that would play more games with you than some other teams do. If luckily, those teams that play more times are underdogs (low power ranking/low winning percentage), then the schedule is a good one for your team right?

Therefore, using Lingo I calculated out the degree of effect that opponent’s power ranking has on each team. The resulting chart shows the values I got: (source code are appended in Appendix 3)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Team | result | Multiplied by const. K | Team | result | Multiplied by const. K |
| Magic | 2.438 | 487.45 | Hornets | 2.527 | 505.46 |
| Wizards | 2.798 | 559.41 | Spurs | 2.216 | 443.08 |
| Hawks | 3.102 | 620.19 | Rockets | 2.525 | 505.20 |
| Bobcats | 3.161 | 632.43 | Mavericks | 2.566 | 513.35 |
| Heat | 4.471 | 894.10 | Grizzlies | 3.908 | 781.62 |
| Celtics | 2.214 | 443.06 | Jazz | 2.667 | 533.31 |
| Raptors | 2.772 | 554.54 | Nuggets | 2.836 | 567.13 |
| 76ers | 2.690 | 537.86 | Trail Blazers | 3.017 | 603.46 |
| Nets | 2.944 | 588.87 | Timberwolves | 4.173 | 834.50 |
| Knicks | 3.842 | 768.13 | Super Sonics | 3.97 | 794.08 |
| Pistons | 2.148 | 429.76 | Lakers | 2.429 | 485.85 |
| Cavaliers | 2.593 | 518.44 | Suns | 2.46 | 492.00 |
| Pacers | 3.275 | 654.82 | Warriors | 2.75 | 549.97 |
| Bulls | 3.100 | 620.25 | Kings | 3.058 | 611.66 |
| Bucks | 3.452 | 690.69 | Clippers | 3.764 | 752.84 |

Chart 4

In order to unify the order of magnitude, each resulting value is multiplied by a constant = 200. And I got *C j.*

**(2) Comprehensive Analysis of integrated effects on teams’ performance**

After analyzing each factor individually in part 1, the integrated equation for the overall effect on a team’s performance based on all 3 factors can be formalized as the following:

*Wj* = *Aj* \* K1 \* *Q*1 + *Bj* \* *Q*2 + *C j* \* K2 \* *Q*3

Using the Microsoft Excel, I computed out the final integrated index for each team’s schedule, and sorted them in the ascending order. Also it is worth noting that the smaller the W value is, the better the corresponding schedule is. Think why it is so if you feel not convinced. Following is the resulting chart:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rank | Team | Total index | ） Rank | Team | Total index |
| 1 | Celtics | 493.673 | 16 | Suns | 587.883 |
| 2 | Magic | 499.009 | 17 | Bobcats | 593.585 |
| 3 | Cavaliers | 524.36 | 18 | 76ers | 596.514 |
| 4 | Pistons | 529.589 | 19 | Jazz | 598.003 |
| 5 | Kings | 530.993 | 20 | Bulls | 605.20 |
| 6 | Spurs | 546.354 | 21 | Grizzilies | 606.926 |
| 7 | Pacers | 550.423 | 22 | Trail Blazers | 606.92 |
| 8 | Hornets | 550.793 | 23 | Timberwolves | 610.783 |
| 9 | Nets | 562.371 | 24 | Wizards | 616.97 |
| 10 | Hawks | 564.195 | 25 | Lakers | 644.794 |
| 11 | Raptors | 566.549 | 26 | Super Sonics | 645.595 |
| 12 | Rockets | 571.981 | 27 | Mavericks | 645.793 |
| 13 | Nuggets | 578.423 | 28 | Clippers | 674.785 |
| 14 | Knicks | 581.187 | 29 | Heat | 674.902 |
| 15 | Warriors | 583.333 | 30 | Bucks | 709.341 |

Chart 5

From the above char, I can draw the claim that the most “lucky” team (with theoretically the most advantageous schedule) for the 2008-2009 season is the Boston Celtics, and the most “unlucky” team is the Milwaukee Bucks. Better luck next season Bucks.

Here comes my favorite part!

**(c) Detailed Analysis of pros and cons of Houston Rockets’ schedule**

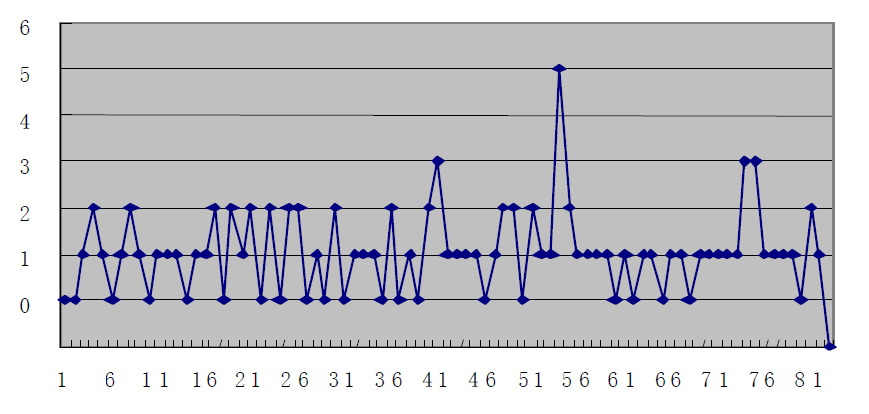
(1) Analyzing each factor individually

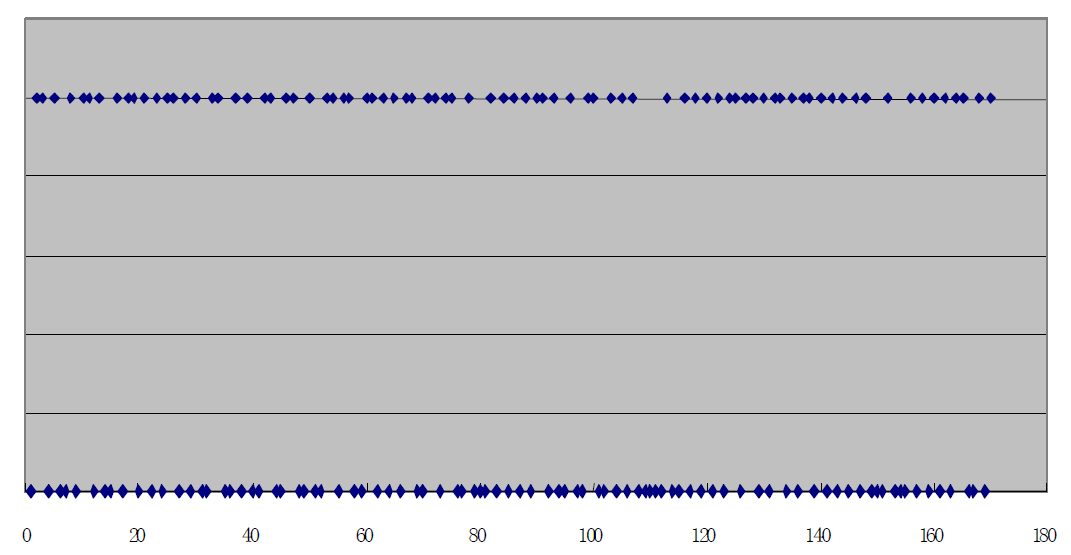
(a) Effect of time interval on Houston Rockets

By checking the variance chart above, I got the variance of time interval is 0.834.

After data compilation, the broken line graph describing effect of time interval on Houston Rockets is drawn as follows:

(Horizontal axis representing number of game; vertical axis representing the number of days of rest prior to that game)

 As we can see, the broken line fluctuates frequently, which indicates that the time interval is not distributed very uniformly. However the good news is, for the most part, the time interval fluctuates within a relatively small region, which indicates that the factor of time interval does not play a huge role in the overall effect. I also plotted out the scatter diagram for the games, as follows:



Note: the lower spotted line indicate the days without games. Upper spotted line indicates the opposite.

From the above scatter diagram we can see that the Rockets had a quite tough opening, which included a 6-game-in-9-days agenda. But the middle and later period is relatively easy compared to the beginning, and such schedule is beneficial for player’s recovery.

(b) The effect of transition between home game and away game on Rockets

From the schedule I observed that in November after the game versus Boston Celtics, the Rockets went through a 5-game road trip, and the road trip was connected by a back-to-back home game versus the New Orleans Hornets (now Pelicans). This period of time was the toughest moment of the season. After that period, situation became a bit better in the mid-late season, since the Rockets had 23 home games out of 41 total. And this fact is especially true in February, during which Rockets had 9 home games out of 11 games in total, including a 6 home-game stream. Such home game stream is very beneficial for team’s recovery and facilitating team chemistry. After checking chart 3, I got the transition value for Rockets is 569, which is in the middle of all 30 teams.

(c) Effect of opponents’ power ranking on Rockets

In the first 18 games, there were 12 away games. More importantly, in those 12 away games, only 3 opponents were not among last season’s (07-08) playoffs. During the 5-game road trip, Rockets’ opponents are respectively: Portland Trail Blazers, LA Clippers, LA Lakers, Phoenix Suns, and San Antonio Spurs. After checking chart 4, I got the Rockets’ value C18 being 505.21, which is also around the middle of all teams.

(2) Comprehensive Analysis of Rockets’ schedule

After checking chart 5, I got the overall ranking of Rockets: 12, not a very top-notch rank. But considering its tough opening, it is a satisfying one in my opinion.

**Building Model for Topic 3 (some-3-some-4 policy)**

(a) Build of the Model

First let me remind you about what the topic is in case you already forget right now: Based on the schedule we can find that for teams that are in the same conference but different divisions, some teams will play against each other only 3 times, instead of 4 times. My concern is to find out the method/model that could decide which teams will be playing with each other for only 3 times in a season. I begin with the following 3 main aspects of analysis:

1. From the audience’s point of view

NBA is a league with strong commercial background. And from the stand point of being as much eye-catching as possible, the people scheduling the season definitely would take the audience perspective as a big factor into consideration. In other words, when people make schedule, they would take teams’ overall ability and power ranking into consideration. If two teams’ squads have similar overall ability and power ranking, then their difference would be minor, which means game between such teams would be much more competitive than those with larger power differences. Therefore, the task for finding teams that play 3 games against each other could be converted into the task for finding teams that have the largest power difference (pairing strong team and lesser team).

Therefore, to measure this type of effect, I used greedy algorithm. First I computed out all 30 teams’ winning percentage for last (07-08) season as the base for computing the power difference between each two teams. Then for each team, I used the following formula and computed out the power difference between itself and each of the 10 teams that are within the same Conference but different Division:

*Ei* = *Pj* − *Gi*

After I gathered all the results for each team, I sorted the difference values in descending order. (For detail, please refer to appendix 4) Finally, for each team, I picked the first 4 corresponding opponents which have the largest power difference values. And those 4 would be the teams that will play against us for only 3 games.

1. From each team’s point of view

NBA is competitive, and every team is aiming for the playoffs. Therefore, from each team’s stand point, if it plays more games with teams that are weaker than itself, then it is a good schedule. Thus for each team, finding out which teams would play 3 games against itself is the same as finding which teams will be the most unlikely ones for us to beat. Such thinking makes sense because if we play only 3 games against those teams that we are afraid of the most, then the rest of the schedule would be somewhat easier and would give us more opportunities of winning.

I used the same greedy algorithm thinking. First I used given data to compute out each team’s winning percentage against those within same Conference but different Division. The formula I used is:

After I gathered the results, I sorted them in ascending order. (For specific values please refer to appendix 5) Then for each team, the first 4 corresponding teams in the chart after the sort are those will play against itself for 3 times this season.

1. Effects of other factors

Other than the above two main factors, there are many other important factors that can affect the choosing, such as the historical records between two teams, trading of players during offseason, etc. But because those historical data is somewhat incomplete hard to compile, and the effects of transfer of players is hard for me to quantify in this example, I did not include those factors in this project.

1. Assessment

As you can imagine, my results is still not a perfect match for the real one, with about 50% of “hit rate”. This is not a surprise because the real scheduling would consider far more factors than I did for this project. But overall, this result is satisfying enough to me.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Eastern Conference | | | | |
| Team | Opponents that will play 3 games with itself | | | |
| Magic | 76ers | Nets | Pistons | Cavaliers |
| Wizards | Bucks | Pacers | Raptors | Celtics |
| Hawks | 76ers | Bulls | Knicks | Pistons |
| Bobcats | Cavaliers | Bulls | Nets | Celtics |
| Heat | Raptors | Bucks | Pacers | Knicks |
| Team | Opponents that will play 3 games with itself | | | |
| Celtics | Bucks | Bulls | Pacers | Wizards |
| Raptors | Pistons | Cavaliers | Wizards | Heat |
| 76ers | Bucks | Pistons | Hawks | Magic |
| Nets | Magic | Bobcats | Cavaliers | Pacers |
| Knicks | Bulls | Pacers | Heat | Hawks |
| Team | Opponents that will play 3 games with itself | | | |
| Pistons | Magic | Hawks | Raptors | 76ers |
| Cavaliers | Magic | Bobcats | Raptors | Nets |
| Pacers | Knicks | Nets | Heat | Wizards |
| Bulls | Hawks | Bobcats | Celtics | Knicks |
| Bucks | 76ers | Celtics | Heat | Wizards |
| Western Conference | | | | |
| Team | Opponents that will play 3 games with itself | | | |
| Hornets | Jazz | Timberwolves | Suns | Clippers |
| Spurs | Kings | Lakers | Jazz | Nuggets |
| Rockets | Trail Blazers | Super Sonics | Suns | Kings |
| Mavericks | Warriors | Lakers | Super Sonics | Trail Blazers |
| Grizzlies | Nuggets | Timberwolves | Warriors | Clippers |
| Team | Opponents that will play 3 games with itself | | | |
| Jazz | Clippers | Lakers | Spurs | Hornets |
| Nuggets | Spurs | Grizzilies | Suns | Warriors |
| Trail Blazers | Kings | Warriors | Mavericks | Rockets |
| Timberwolves | Hornets | Grizzilies | Suns | Clippers |
| Super Sonics | Kings | Lakers | Mavericks | Rockets |
| TeamTeam | Opponents that will play 3 games with itself | | | |
| Lakers | Spurs | Mavericks | Jazz | Super Sonics |
| Suns | Timberwolves | Nuggets | Rockets | Hornets |
| Warriors | Mavericks | Grizzilies | 快拖着 | Nuggets |
| Kings | Spurs | Rockets | Trail Blazers | Super Sonics |
| Clippers | Hornets | Grizzilies | Jazz | Timberwolves |

After comparing my results with the real one. I realize that the game schedule is not designed simply based on one factors or several factors, but mathematical models’ results that integrate huge sets of factors including objective and subjective reasons, and quantified deliberation that considers each team’s own situation. Such results minimize the potential unfairness and as well minimize the schedule’s effects on the game, at the same time increasing the competitiveness and excitement of the game.

**Conclusion and Lessons Learned**

After all of the above analysis, I am amazed by the level of complexity that is hiding inside a seemingly ordinary NBA game schedule, and the experience this project brings me would definitely encourage me to discover more interesting and technical aspects of NBA in the future.

The biggest challenge in this project for me was to gathered enough data and organize them in a nice way. Fortunately, with the help of PASCAL programming and high performance software like Excel and Lingo, my workloads in processing the data was largely reduced. Since I did not purchase MATLAB or use remote desktop during most of my time working on this project, MATLAB was not one of my options. But I believe that with the help of MATLAB, this project would become even less cumbersome and more doable.