

NBA Statistics – Cost per Victory, Players’ Investment on Return, and Correlation to Victories

Final Project
CSE 140
Daniel W. Chen

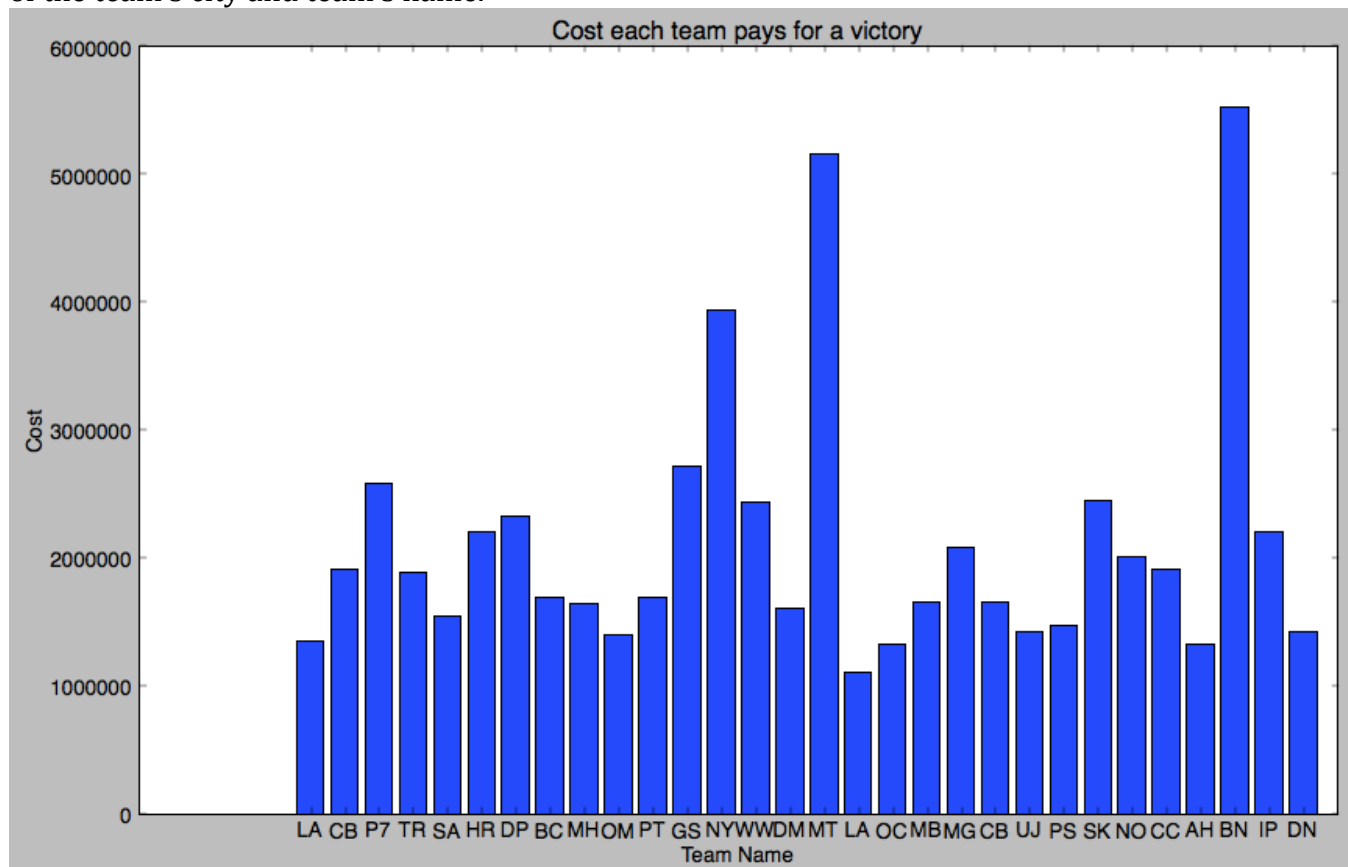
Summary of Research Questions

Question:

How much does a team have to pay to get a win? The answer to this question will calculate the total investment over the number of wins for each team and calculate the average for each of the win. We want to see if we can monetize the value of a victory and what does the value entail?

Answer:

The amount of money each team pays for a victory is displayed below. Each team name is an abbreviation of the team’s city and team’s name.



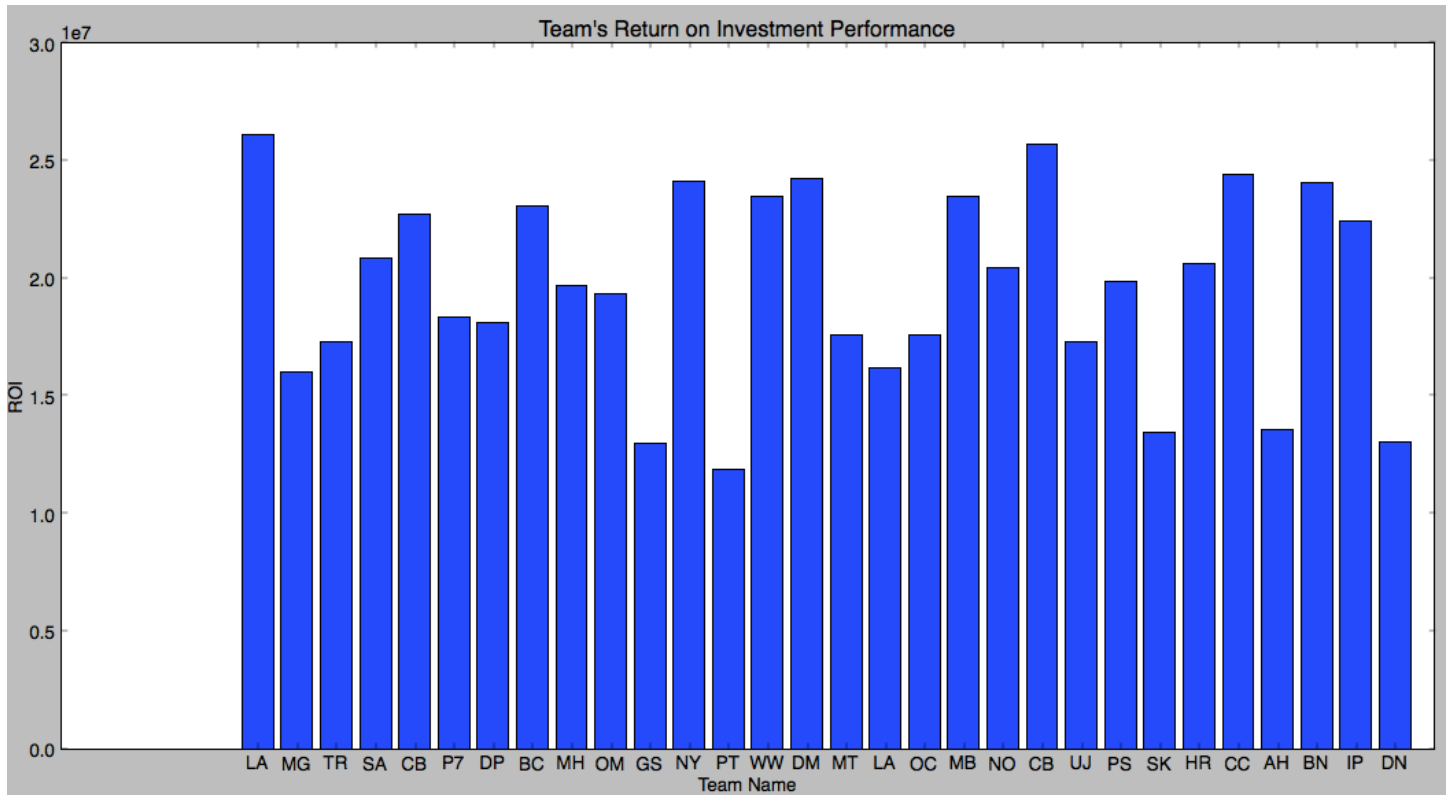
Question:

Which team made the best investment decisions for each of the player they paid to hire on the team? The answer to this question will calculate the return on investment with the salary information and their regular and playoff season record for each of the team. We want to see if there’s a better way to improve making investment in players and what leads to good and bad decisions.



Answer:

Every team's return on investment is presented below. Return on investment of each player is calculated by using player's salary divided by players' contribution score. The low the ROI is, the more valuable the player is.

**Question:**

What type of players provides the highest return? The answer to this question will analyze the top 25 players with the highest "Return on Investment". We want to identify the best type of player to invest in.

Answer:

It's usually non-star player that provides the highest value to the team. From the top 25 player with lowest ROI. We do not see any star player.

Top 25 players with the lowest ROI:

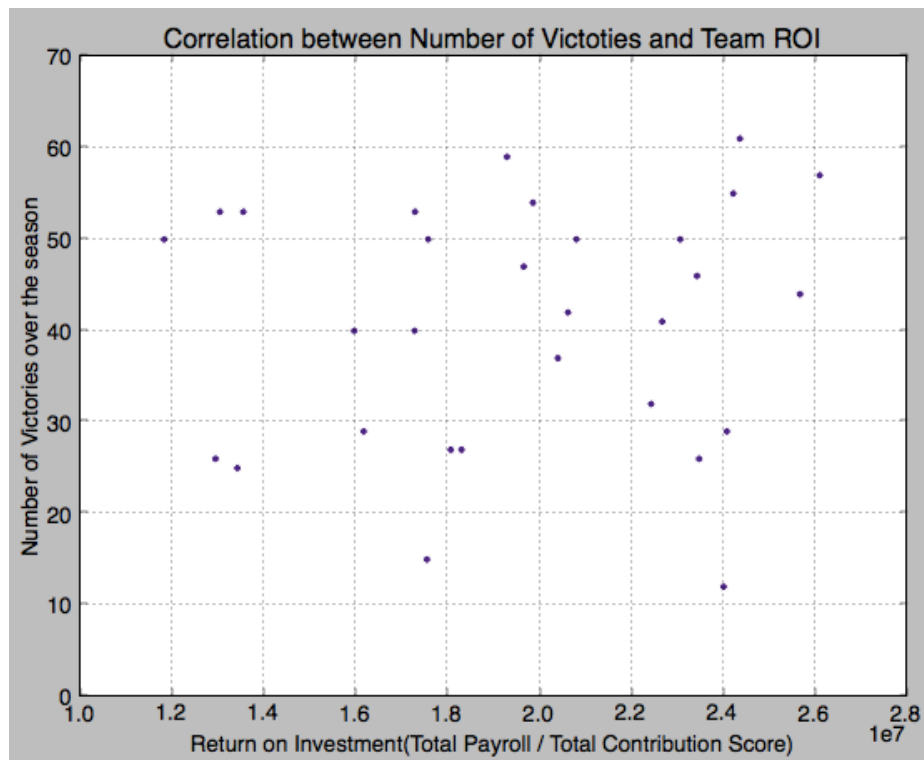
[('D. Blair', 238764.04), ('M. Chalmers', 235514.02), ('C. Budinger', 230061.35), ('B. Walker', 224518.29), ('W. Chandler', 218717.77), ('S. Gaines', 214063.78), ('C. Douglas-Roberts', 204561.11), ('W. Bynum', 197960.91), ('C. Hunter', 196672.64), ('A. Price', 168230.88), ('A. Brooks', 163765.74), ('A. Morrow', 159744.03), ('W. Matthews', 129262.15), ('R. Bowen', 122985.0), ('M. Wilks', 121569.66), ('J. Jerebko', 113264.36), ('S. Livingston', 105672.38), ('M. Thornton', 95930.4), ('P. Davis', 77850.0), ('T. Gilder', 61363.64), ('D. Mason', 53402.88), ('A. Tolliver', 13208.56), ('M. Harris', 9345.79), ('J. Crawford', 8561.64), ('M. Williams', 2345.89)]

Question:

Would ROI correlate with number of victories?



Answer:



Motivation and background

Inspired by the book *Moneyball: The Art of Winning an Unfair Game* in which a team with disadvantaged revenue situation was able to assemble a competitive baseball team with an analytical, evidence-based, sabermetric approach. The similar approach could be applied to basketball by measuring players' performance metrics, team's season record, and how they compare to the salary each of the players are paid by the team. The result of this analysis might potentially provide useful information from a business management perspective in deciding what type of player is worth investing from a pure statistical angle and other potential applications in making business decisions.

Datasets

Below is the list of available datasets during 09-10 NBA Season that are used in the analysis:

They can be downloaded http://basketballreference.com/stats_download.htm

The salary information can be accessed:

<http://www.draftexpress.com/nba-player-salaries/>

1. player_allstar.csv
2. player_career.csv
3. player_playoffs_career.csv
4. player_playoffs_csv
5. player_regular_season.csv
6. players_csv
7. teams.csv



From these datasets, I have cleaned up the data both in python and excel to recreate and rename the datasets into the below datasets. They are included in the zip file submitted. Below is the list of datasets renamed in the zip file.

1. 09-10PlayerPlayoffStats.csv
2. 09-10PlayerSalaryInfo.csv
3. 09-10PlayerStats.csv
4. 09-10TeamRoster.csv
5. 09-10TeamSalaryInfo.csv
6. 09-10TeamStats.csv

Methodology (Algorithm or Analysis)

(I)

The first part of analysis includes computing the cost per victory for each of the team in the NBA 2009 league. We will do this by taking the total amount of salary each team pays to its players and divide that number by the number of victory the team has over the season.

(II)

The second part of the analysis includes calculating the returns on investment for each of the player on each of the team. Each player's return on investment will be calculated by using the contribution each player has during a team's victory. A player's contribution to a team's victory can be broken down into the points, rebounds, assists, steals, and field goal percentage. A weight is given to each of the factors and each player will receive a contribution score. Above is a table with the weight given to each of the factor. After the contribution score is being assigned to each of the player, we will divide the annual salary of each of the player by the contribution score to calculate each player's return on investment. The lower the ROI is, more valuable a player is and the better investment decision a team has made. Because a team's getting more contribution from every dollar they invest, so they would want a lower score. We will calculate the contribution scores for both regular season and playoffs season to see if there's any major differences.

Points	25%
Rebounds	22.5%
Assists	22.5%
Steal	15%
Block	15%

(III)

The third part of analysis includes aggregating the results from part two and identifying the top 25 players with the highest return on investment.

(IV)

For the final part of our analysis, we will look the relationship between number of victories versus the each team's ROI to determine if a team's investment decision is correlated to the outcome of the games.

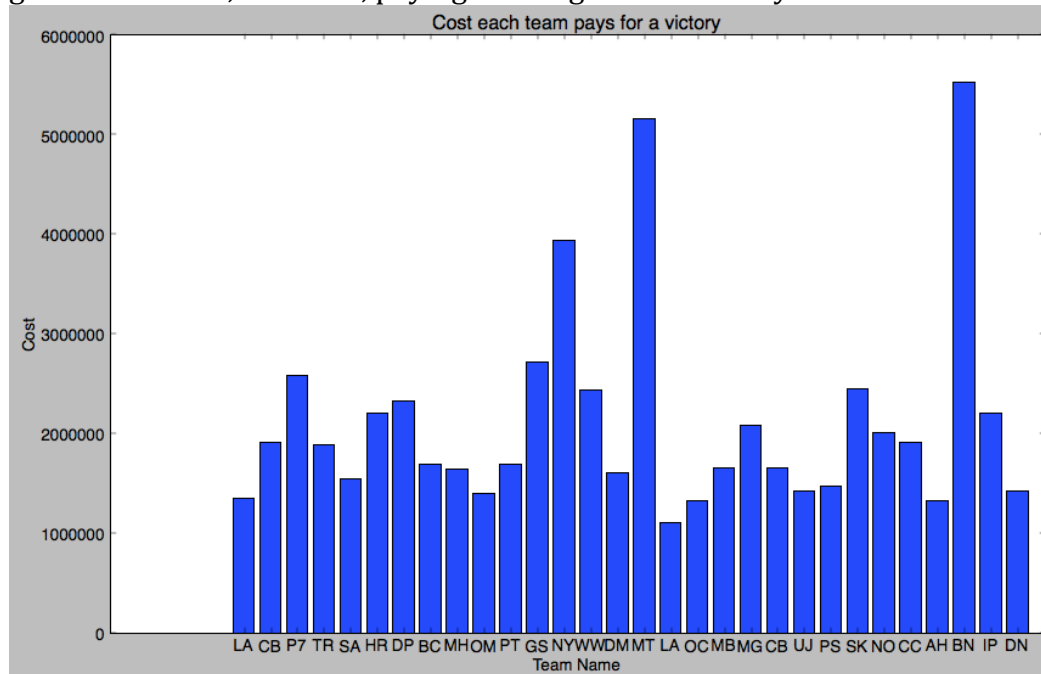
Results

The results and findings of this project can be divided into several sections:

Cost of Victory



As presented in the below graph, the amount of money a team pays for a victory varies on the teams' season record. Obviously, if a team wins more game then a team would cost less for every team. L.A. Lakers has the lowest cost for each of the victory. However, the second place Boston Celtics does not have one of the lowest cost for each of the victories. There are a lot of teams who pay less for each of the victory they gain through out the season. This data tells us that paying more in the team's payroll do not guarantee a win, however, paying less might have make your team weaker.



Is Contribution score a fair estimate of the players' performance and value?

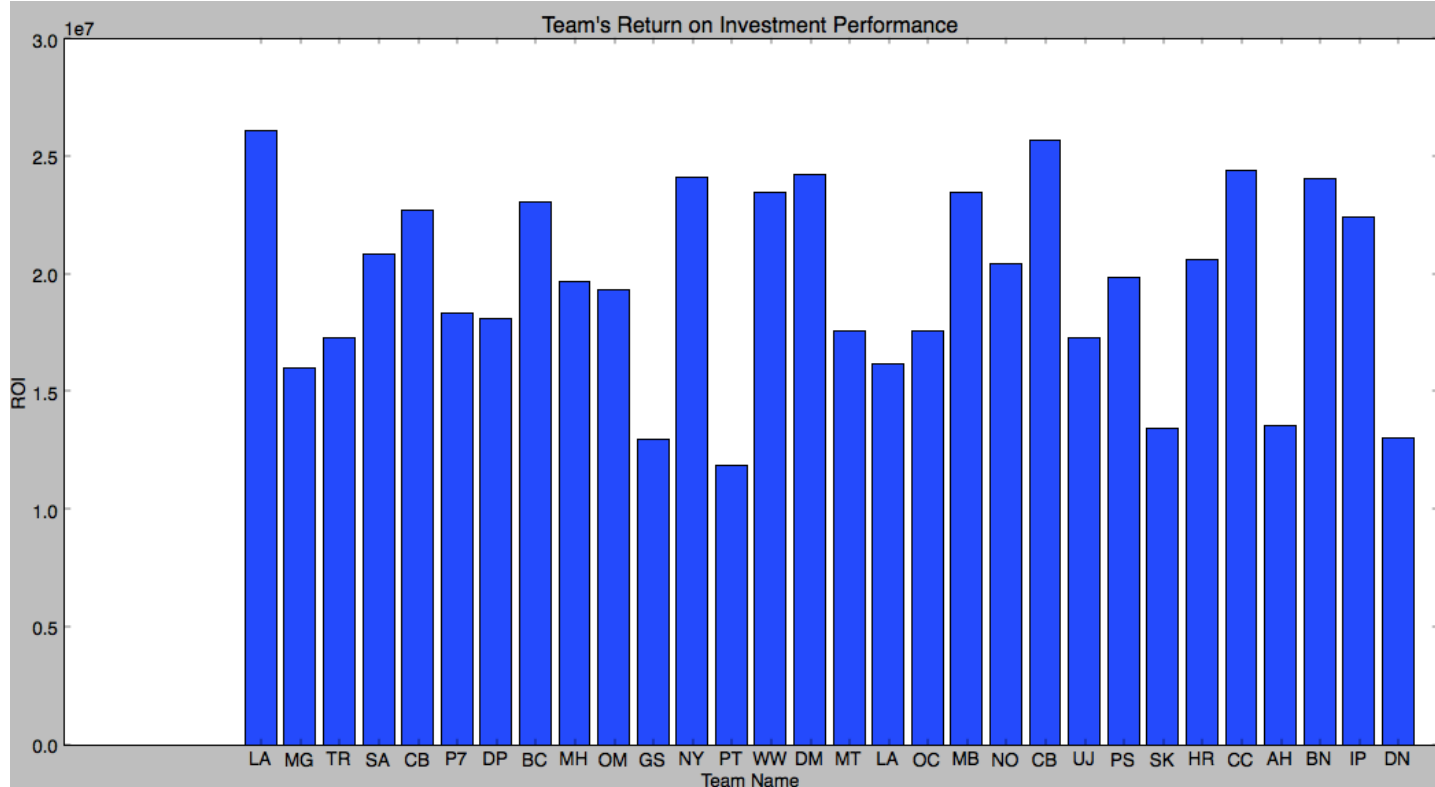
By looking at the top 25 players with the highest contribution score, we can conclude that it's a fair metric to score the player's performance.

[('L. James', 11.38), ('K. Durant', 10.14), ('D. Wade', 9.75), ('C. Anthony', 9.58), ('K. Bryant', 9.55), ('C. Bosh', 9.22), ('D. Nowitzki', 8.9), ('M. Ellis', 8.87), ('D. Lee', 8.72), ('G. Arenas', 8.46), ('Z. Randolph', 8.45), ('D. Howard', 8.43), ('C. Boozer', 8.38), ('C. Paul', 8.37), ('P. Gasol', 8.28), ('D. Williams', 8.25), ('D. Granger', 8.24), ('A. Stoudemire', 8.16), ('T. Duncan', 7.78), ('T. Evans', 7.77), ('G. Wallace', 7.66), ('D. Rose', 7.63), ('B. Lopez', 7.51), ('S. Nash', 7.44), ('B. Roy', 7.42), ('S. Jackson', 7.38), ('A. Iguodala', 7.37), ('C. Kaman', 7.33), ('R. Westbrook', 7.23), ('A. Bogut', 7.14), ('R. Rondo', 7.13), ('A. Jefferson', 7.07), ('C. Billups', 7.05), ('S. Curry', 7.01), ('R. Gay', 7.0), ('L. Aldridge', 6.96), ('A. Jamison', 6.9), ('A. Brooks', 6.83), ('C. Maggette', 6.83), ('B. Davis', 6.78), ('L. Deng', 6.76), ('M. Gasol', 6.67), ('K. Martin', 6.67), ('K. Love', 6.66)]

We can see most of the players on the list are names that can be easily identified since they are all 'Star Player'. However, having the highest performing players does not guarantee a great season.



Which team made the best investment decision based on the contribution score?



According to the graph, Portland Trail Blazers made the best decision in investing by having the lowest ROI (highest value). However, their season record is only top 10 in the league. L.A. Lakers made the worst investment decision but had won the championship. So ROI is not the best metric to predict the teams' chance on championship or season record.

Is there a correlation between number of wins and a team's ROI

Correlation between number of victories and teams' ROI is as below:

Moderate Correlation: 0.090949184727789484

P-Value 0.63266817602931669

With the p-value being so high, showing that this is not the statistically significant result. The graph shows there's only a very moderate correlation between investment decisions versus teams' season results.

The program also produces results such as (They are not being displayed due to space limit):

1. Sorted list of team ranked by the amount of money a team pays for a victory from high to low:
2. Sorted list of player ranked by contribution score from highest to lowest:
3. Sorted list of player ranked by contribution score in playoff with highest to lowest:
4. Sorted list of player ranked by return on investment with highest to lowest:
5. Sorted list of player ranked by return on investment in playoff with highest to lowest:
6. Top 25 players with the lowest ROI (Highest Value):
7. Top 25 players with the lowest ROI(Highest Value) in playoff:
8. A sorted list of teams ranked by teams' total ROI ranked from highest to lowest:



Reproducing your results

Since the data is downloaded from the Internet through the links provided and modified after downloaded. The steps are as follow:

1. Place the NBA_Project.py file with all the csv files(listed in the dataset section) in the same folder
2. Open up Python IDEL
3. Open NBA_Project.py and make sure this file is in the same folder as all the other csv files listed earlier
4. Run the Module
5. All the results would be printed as well as three charts would be created. Please expand the charts to see clearly

Work Plan Evaluation

Part 1: Import, Clean, and Store

An estimate of 3-4 hours is bad. The entire data cleaning and storing process took more than 6 hours because the complexity and the variety of datasets and data formats

Part 2: Cost per victory

30 mins – 40 mins was a good estimate.

Part3: Return on Investment

It took 2 hours to calculate ROI, only 30 mins more than the predicted.

Part4: Highest Return Type & Correlation between metrics and victories

2 – 4 hours was a good estimate.

Part 5: Visualization and Graphing

2 – 3 hours was a good estimate.

Collaboration

Did not collaborate with anyone

