[[1]](#footnote-1)

The Los Angeles Lakers: A Return to Dominance

Patrick Nichols, Author, Adam Soto, Author, and Michelle Arriaga, Author

***Abstract*—****Once the most dominant teams in the NBA, the Los Angeles Lakers, are now in a stagnant position at the bottom of the western conference. One objective of this research was to evaluate NBA game averages from the last 4 years, to determine what game variables contributed to wins and of those variables which were most influential. The second objective, create a visual tool that provided Laker coaches with the ability to predict the teams chances at making it to the Western Conference playoffs.**

# INTRODUCTION

W

ins are the single most important factor that allow a team to participate in the NBA playoffs. By evaluating team statistics in the NBA for the previous four seasons, we set out to identify which statistics contribute to wins. Secondly, of those strategic statistics, which were the most influential? Thirdly, what type of utility or tool could be provided to coaches to assist with managing team progress; or more importantly, equip coaches with information to evaluate incoming free agents in the upcoming season?

The major deliverable for this project is our online prediction tool that takes into account the most influential factors leading to the Los Angeles Lakers’ wins total and a flexible prediction model that the management can use to boost the Lakers chances at making it to the Western Conference playoffs.

Section 2 describes the methodology used to obtain the NBA data for the regression analysis and the database infrastructure setup. Section 3 describes the regression analysis methodology and findings. Section 4 describes the integration and effort to visualize the regression findings on the web platform and Section 5 discusses future work.

# Data Gathering and Database Creation

In order to perform analysis and identify the significant variables that contribute to team success, a comprehensive data set was necessary. Sportradarus1 provided over thirteen (13) different type of data extracts of NBA statistics, however only three (3) extracts met the needs of our research.

## SportradarUS

Sportradarus externalizes a RESTful API that provides access to data extracts in a JSON format. Utilizing Python and Pandas, we were able to use three (3) data extracts consisting of team information, season statistics, and standings information. However, this was not without it’s challenges. Each of the extracts contained far more statistical information than was necessary to perform our analysis. Also, each of the extracts has its own structure in the JSON file which made parsing through this data challenging. . A significant amount of work went into determining which data elements were necessary and how to retrieve them, so that only the necessary elements were consumed. For example, some of the data we needed for our analysis was located in nested structures with in the JSON file. After studying and understanding the format we were receiving our data, we were able to write a python function to retrieve the data we needed.

## Data Framework Creation

Once evaluations of the three independent data extracts were performed, we moved onto setting up the MySQL environment and identifying the relative mapping of the SQL tables. A secondary challenge was determining how to parse multiple, large JSON files. Each JSON file required an iteration of evaluating schema, python indexing and relational attributes mapping. Once accomplished, each JSON file could be consumed and translated into the respective SQL table. An overview of the data flow is provided in Fig 1. Corresponding to the Sportradarus and NBA DB swimlanes.

## Python and NBA Database Creation

In order to orchestrate the integration of the API calls and the creation of the MySQL database and respective tables, Python was used.

By utilizing the Pandas Library for Python, we were able to: retrieve the NBA Statistics in JSON from the SportRadar API, extract and format the data from the JSON structure, build multiple data frames and their respective tables, and create tables in MySQL. While using Pandas allowed us to extract and analyze the data easier than if we used the standard Python Library, there was a learning curve to Pandas that we had to overcome.

Each JSON schema had to be investigated thoroughly. Further, only selective data was required for analysis, which required extensive indexing and JSON file manipulation once each JSON file was loaded to a Pandas dataframe. Standard procedural programming methods were used once a dataframe was established for each JSON file. Loops gave us the ability to select the data necessary, along with element indexing on JSON values and keys.

The final deliverable for this portion of the project was the NBA database containing three (3) tables: teams, teamstats and standings. The teams table consists of conference information, location, stadium name and other descriptive features. The teamstats table contains traditional NBA statistics and standings allowed us to capture wins and losses information by team, by year. Once these tables were established, a master view called statsmaster was created. Tables were joined on unique IDs for each team given by SportsRadar’s API and the year (season) where applicable. Figure 2 illustrates the final entity relationship diagram for the NBA database. The previously mentioned tables and view were stored locally in a MySQL RDBMS.

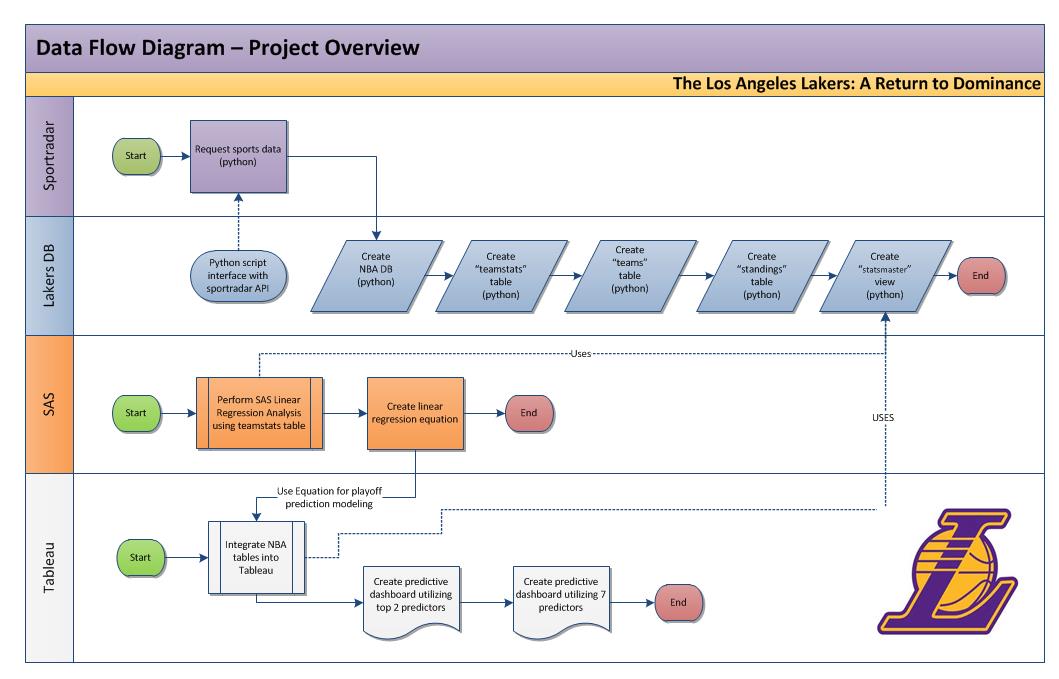


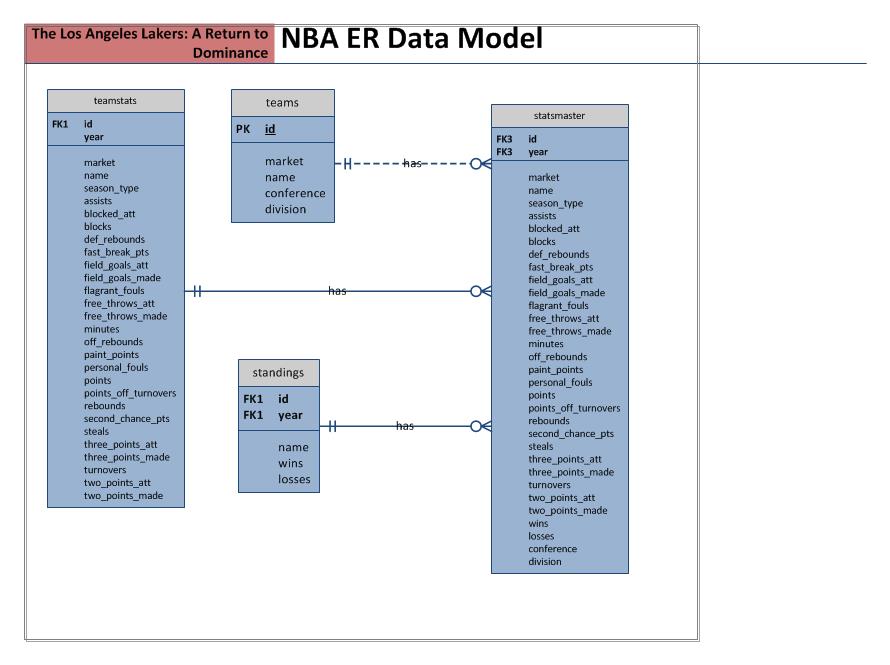
Fig. 1. Represents the data flow as it moves through each of the respective layers from the external API site (Sportradarus) to the   
internal NBA mySQL database, the SAS platform and finally through Tableu.

Fig. 2. The ER Data Diagram illustrates the tables (teamstats, teams and standings) and custom view (statsmaster) contained in the NBA MySQL database.

# Statistical Analysis

Once data was cleaned and integrated, data analysis commenced. The primary goal of the statistical analysis was to determine what traditional statistics were the most linearly correlated with wins. Specifically, which statistics played the most significant role in a team’s ability to win basketball games.

## Statistical Methodology

In order to determine which statistics showed substantial significance in relation to a teams wins, we evaluated four (4) years worth of NBA team statistics utilizing the statsmaster view as seen in Figure 1. Overall team statistics for all 30 teams in the Western, Northwest, Southwest, Pacific, Eastern, Southeast, Atlantic and Central divisions were evaluated. A list of the entire set of variables evaluated in the multiple linear regression model is listed in Table 1.

|  |  |
| --- | --- |
| TABLE 1 | |
| VARIABLES EVALUATED IN MULTIPLE  LINEAR REGRESSION MODEL | |
| \*Assists | Paint Points |
| Blocked Attempts | Personal Fouls |
| \*Blocks | Points |
| \*Defensive Rebounds | Points Offensive Turnovers |
| Fast Break Points | Rebounds |
| Field Goals Attempted | Second Chance Points |
| Field Goals Made | \*Steals |
| Flagrant Fouls | Three Point Attempts |
| Free Throws Attempted | \*Three Points Made |
| \*Free Throws Made | \*Turnovers |
| Minutes | Two Point Attempts |
| Offensive Rebounds | Two Points Made |

List of 24 individual variables evaluated in a multiple linear regression model to determine which variables were statistically correlated to wins. \*Indicates the variables that were highly correlated to wins.

Using an alpha level of 0.05, each of the 24 variables were investigated for linearity utilizing a correlation matrix and scatterplot matrices in both Python and SAS. Of the total 24 variables tested, only seven (7) were moderately to strongly correlated with wins as seen in Table 2. Multiple iterations of models were investigated, eliminating variables along the way with extra sum of squares F tests and individual variable T-tests for significance. To ensure linearity was true and outliers were not impacting our analysis, scatterplot matrices were used extensively.

|  |  |  |
| --- | --- | --- |
| TABLE 2 | | |
| MULTIPLE LINEAR REGRESSION VARIABLE SIGNIFICANCE | | |
| VARIABLE | PARAMETER ESTIMATE | P-VALUE |
| Three Points Made | 2.23 | <.0001 |
| Defensive Rebounds | 2.65 | 0.0001 |
| Steals | 4.15 | <.0001 |
| Turnovers | -3.65 | <.0001 |
| Free Throws Made | 1.4 | 0.0024 |
| Blocks | 2.99 | 0.0043 |
| Assists | 1.34 | 0.0164 |

Results of the multiple linear regression analysis demonstrating the statistical relevance in the P-value as well as the standard error used in the prediction equation.

***Prediction Equation for Wins***

WINS = (2.23)Three Points Made + (2.65)Defensive Rebounds + (4.15)Steals + (-3.65)Turnovers + (1.4)Free Throw Made + (2.99)Blocks + (1.34)Assists

The final regression model produced a significant overall F (p <.0001) with an r-squared value of 0.59. We consider this an acceptable model for predicting wins.

As a result of our model, the Los Angeles Lakers average win / loss record was predicted at 37 wins and 45 losses for the past four years. As a frame of reference, the Los Angeles Lakers’ actual average win / loss record was 35 wins and 47 losses. Based on this test, we considered our multiple linear regression model acceptable.

# Data Visualization

To easily disseminate the final regression model to Lakers management, a visual tool was developed that provides users the ability to predict the team’s chances at making it to the Western Conference playoffs. Tableau was selected as our online visualization tool in part because of the variety of data visualization displays available and because of the flexibility with connecting to our local database. Tableau is the market leader in visual analytics and business intelligence and provides an easy to use and easy to publish interface.

## Tableau Connection

Tableau natively connects to MySQL via an ODBC driver. Our tableau dashboard is directly connected into a local NBA database in MySQL. Data from the statsmaster view drives the visualization interfaces for the project. As seen in Fig. 1. Tableau utilizes the wins prediction equation as a parameter that the bar graph uses to readjust the bar graph. Data input fields and slider tools give users the ability to adjust any of the influential variables which triggers recalculation of the predication equation to dynamically show the effect on the number of wins.

## Tableau Visualization Interfaces

The first visualization incorporates the original seven (7) variables. Each variable is pre-populated based on last seasons standings and adjustments to any of the seven (7) variables dynamically changes the wins prediction graph to show how close or far away the team is from making it to the Western Conference playoffs. Upper and lower boundaries were constructed for each variable so that a user would not make invalid or unrealistic adjustments to a variable. However, minimal extrapolation was allowed for unique prediction situations where applicable.

The second visualization is a much more specific and targeted prediction model that utilizes the two (2) most influential variables for wins: three points made and defensive rebounds. Similar boundaries were placed on the variables to prevent invalid selections. A slider mechanism is provided to change the number of the variables.

## Added Business Value

A statistics widget from NBA.com was integrated into both visualization interfaces that provides coaches access to real-time data: Team Roster, Team Leaders, Team Stats and Player Stats. The granular statistical information, along with the predictive tools can assist coaching staff with predicting how small changes in a few or many areas can have a big effect on the Lakers chances at wins and making the Western Conference playoffs.

# Future Work

In the future, it would be reasonable to improve on the work already performed by utilizing automation to import real-time data, so that our prediction model is kept up to date. Additionally, it would be beneficial to add in the ability to predict the effect of a player, whom a team is currently scouting or looking to trade, on a teams past season.

A list of potential players could be developed based on the previously mentioned analysis, allowing Lakers management easy access to players with the biggest potential impact on wins.

# Conclusion

Our goal was to conduct research on game statistics for all the teams in the NBA in effort to identify the most influential variables that contribute to wins. We found that there were seven (7) out of twenty-four (24) variables that were statistically significant and correlated with wins. Additionally, of the seven (7) variables, we drilled down even more to find that two (2) variables: three points made and defensive rebounds were the most influential variables. With this statistical information, a visualization model was constructed utilizing our local database of NBA statistics and multiple linear regression equation that provides a prediction tool for coaches to use to assist at predicting wins.

On a technical level, our project traversed 4 layers. We initiated our data research by retrieving data through SportradarUS, a free, publically available API. Once we retrieved the data, we transformed the data and inserted it into a local NBA database and into multiple statistical tables using Python and Pandas. A custom view was created to demonstrate MySQL proficiency in multiple joins. The newly created custom view would then be used in the multiple linear regression statistical analysis utilizing SAS in order to develop a prediction equation, which would finally get utilized in Tableau to render a prediction visualization that predicts wins and chances at making the playoffs.

Appendix

Appendixes, if needed, appear before the acknowledgment.

References

*Basic format for books:*

[1] <http://developer.sportradar.us/docs/NBA_API>

[2] <http://public.tableau.com/s/>

**http://thesportjournal.org/article/an-examination-of-the-moneyball-theory-a-baseball-statistical-analysis/**

1. Date Submitted: 08/20/2015.

   This work was supported in part by Dr. Daniel Engels, Associate Professor – Computer Science and Engineering. Southern Methodist University, Dallas, TX 75205.

   Patrick Nichols is with Southern Methodist University, Dallas, TX 75205 USA. (e-mail: pcnichols@SMU.edu).

   Adam Soto is with Southern Methodist University, Dallas, TX 75205 USA. (e-mail: ajsoto@SMU.edu).

   Michelle Arriaga is with Southern Methodist University, Dallas, TX 75205 USA. (e-mail: marriaga@SMU.edu). [↑](#footnote-ref-1)