NBA Statistical Analysis

Logan T. Donlan

Western Governors University

Table of Contents

A. Proposal Overview 4

A.1 Research Question or Organizational Need 4

[A.2 Context and Background 4](#_Toc1384268585)

[A.3 Summary of Published Works and Their Relation to the Project 4](#_Toc2079458267)

[Review of Work 1 4](#_Toc694593955)

[Review of Work 2 4](#_Toc475195288)

[Review of Work 3 4](#_Toc1156669356)

[A.4 Summary of Data Analytics Solution 5](#_Toc374341712)

[A.5 Benefits and Support of Decision-Making Process 5](#_Toc1116633891)

[B. Data Analytics Project Plan 6](#_Toc1561262715)

[B.1 Goals, Objectives, and Deliverables 6](#_Toc1217623361)

[B.2 Scope of Project 7](#_Toc427690838)

[B.2.A Included in Project Scope 7](#_Toc1243851001)

[B.2.B Not included in Project Scope 7](#_Toc2036717847)

[B.3 Standard Methodology 7](#_Toc204762054)

[B.4 Timeline and Milestones 8](#_Toc1049840195)

[B.5 Resources and Costs 8](#_Toc313275103)

[B.6 Criteria for Success 8](#_Toc1337049307)

[C. Design of Data Analytics Solution 9](#_Toc764513894)

[C.1 Hypothesis 9](#_Toc1169566000)

[C.2 and C.2.A Analytical Method 9](#_Toc1622461191)

[C.3 Tools and Environments 12](#_Toc53928856)

[C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance 10](#_Toc1386867513)

[C.5 Practical Significance 14](#_Toc157862960)

[C.6 Visual Communication 14](#_Toc1263386613)

[D. Description of Dataset 14](#_Toc34323025)

[D.1 Source of Data 14](#_Toc1867631323)

[D.2 Appropriateness of Dataset 14](#_Toc418252330)

[D.3 Data Collection Methods 14](#_Toc1665164903)

[D.4 Observations on Quality and Completeness of Data 15](#_Toc57902398)

[D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances 15](#_Toc1649406101)

References………………………………………………………………………………………………... 16

# A. Proposal Overview

## A.1 Research Question or Organizational Need

The organizational need here is to find out which team statistics correlate to winning, be it positive or negative. It is also important to break this down into eras, as the game has evolved over time.

## A.2 Context and Background

Teams are always looking for an edge in the National Basketball Association. Sports analytics has exploded in the past decade and has become an integral part of decision making within these organizations. This information also leads to insights for companies that host sports gambling and the gamblers themselves. Sports journalism relies heavily on these types of analyses. The insights learned from this type of project are invaluable to a plethora of businesses and people. It will drive revenue in multiple facets.

## A.3 Summary of Published Works and Their Relation to the Project

Multiple public works were used to explore and inspire insights into this project.

1. “9 Stats That Every Serious Basketball Coach Should Track” (Published by Jeff Haefner in 2016): This article outlines key statistics essential for basketball coaches to monitor to enhance team performance and increase winning chances. The stats include Field Goal Attempts (FGA), Effective Field Goal Percentage (EFG%), Free Throw Attempts (FTA), Free Throw Percentage (FT%), Turnovers, Defensive Rebounding Percentage (DReb%), Offensive Rebounding Percentage (OReb%), Fouls, and Attempts in the Paint. Tracking these stats can help identify positive and negative correlations with winning, crucial for addressing the organizational need to understand how team performance metrics contribute to winning (Haefner 2016).
2. “NBA’s Most Valuable Statistic Discovered: How To Predict Team Wins With 95% Accuracy” (Published by Ryan Sullivan in 2020): This work explores the correlation between various team statistics and win percentage in the NBA. Sullivan compiled data from the past 20 NBA seasons, identifying that the Team Performance Index (PIE) and Dean Oliver's Four Factors strongly correlate with winning. He developed his own advanced metric, the Sully Four Factor Rating, which proved more accurate in predicting wins, considering the evolution of the game over time (Sullivan 2020).
3. “How Different Metrics Correlate with Winning in the NBA over 30 Years” (Published by David Peterson in 2020): Here, Peterson shows how various team performance metrics have correlated with winning in the NBA over the past three decades. Peterson highlights several key metrics, including two-point field goal percentage, three-point field goal percentage, assists-to-turnover ratio, and offensive rebounding rate. The study shows that two-point field goal percentage has the highest correlation with winning in the current season, while metrics like offensive rebounding rate have shown a negative correlation over time (Peterson, 2020).

## A.4 Summary of Data Analytics Solution For this project I will use a dataset of NBA team statistics from the regular seasons spanning from 2001 to 2024. The dataset consists of 29 columns and 717 rows that contain fields like win percentage, field goal percentage, rebounds, blocks, and personal fouls drawn. This dataset is downloadable as a zip file. I will use a Jupyter notebook and python to import, clean, and analyze the dataset. The machine learning method of Lasso regression will be used to analyze the data. The model will then be bolstered using a bootstrap method, along with calculating the coefficients within 95% confidence intervals. After I have gained the regression coefficients for the target variable (win percentage) I will derive secondary regression coefficients for each of the primary features. From here I will be able to say which features are the most important to winning in the regular season across the entire dataset. I will use a correlation matrix, horizontal bar charts, and axes horizontal line charts to help with the visualizations for the analysis. Next, I will split the dataset into 2 eras (2001-2012 & 2013-2024) and repeat the process to see if there are any statistically significant differences between the eras. The libraries used to perform these actions include pandas, matplotlib, numpy, seaborn, sklearn, and math. When the analysis is complete, I will be able to identify any changes to feature importance between the 2 eras. To help display these findings I will have an interactive dashboard that shows the regression coefficients for each feature by era. I also will create line and shaded graphs to help represent the changes in scoring over the years.

## A.5 Benefits and Support of Decision-Making Process

This analysis will provide multiple benefits across a variety of organizations. First, it will provide insight to teams looking to gain an edge against their opponents. This can help them decide which players to bring in, which coaches to choose, and how they should modify their playstyle. Second, businesses like ESPN BET and FanDuel will be able to predict lines more accurately on things like individual game lines and season plus/minus on win totals. This analysis will also be able to allow the informed gambler about the same type of variables when placing their bets. The methods used to perform this analysis fit with the main task of finding features that influence win percentage. Calculating results within 95% confidence intervals can help assure the users that this information is accurate and predictable.

# B. Data Analytics Project Plan

## B.1 Goals, Objectives, and Deliverables

**Goal:** Explain how features influence win percentage over the entire dataset and identify any variance in the different eras.

* **Objective A:** Calculate advanced statistics and choose which features to use in the model by examining the data through various methods, including a correlation matrix.
  + **Deliverable A1:** This objective will provide us with a dataset that avoids using features with high multicollinearity so that the model can be more accurate.
* **Objective B:** Use Lasso regression and bootstrapping on the selected features to identify regression coefficients with the target of win percentage.
  + **Deliverable B1:** After the model is fitted and refined with the bootstrapping method, a dataframe of primary regression coefficients will be returned.
* **Objective C:** Generate secondary regression coefficients on the primary features to identify which features lead to higher performance of the primary features using the Lasso regression model and bootstrapping.
  + **Deliverable C1:** This objective will deliver a dictionary containing individual dataframes for each primary feature. These dataframes will show the regression coefficients of each primary feature to the other features.
* **Objective D:** Split the dataset into 2 eras and repeat the process to discover variance in the coefficients.
  + **Deliverable D1:** The model will be fitted to the era 1 dataset and return a dataframe of primary regression coefficients.
  + **Deliverable D2:** Using the methods from objective c, secondary feature coefficients will be stored in a dictionary holding dataframes for each primary feature.
  + **Deliverable D3:** The era 2 dataset will be fit with the Lasso model and the bootstrapping method will be applied. This will give me the primary feature regression coefficients for the second era.
  + **Deliverable D4:** Repeating the methods used in objective c, I will gain insight into the secondary coefficients through a dictionary containing dataframes for each individual feature. These dataframes will explain the coefficients of the primary features to the other features.
* **Objective E:** Display these findings using simple, accurate visualizations.
  + **Deliverable E1:** Using Tableau, create multiple visualizations that accurately depict the outcome of the complete analysis.
  + **Deliverable E2:** Create a slideshow to present the findings and visualizations.

## B.2 Scope of Project

### B.2.A Included in Project Scope

The scope of this project includes calculating advanced statistics, selecting primary features, identifying how they influence with the datasets, discovering how primary features interact with each other, determine if there is any variance between eras, and display these findings using various visualizations. The analysis will be done in a Jupyter notebook using python as the language and the pandas, numpy, seaborn, matplotlib, sklearn, and math libraries. The supporting visualizations will be created using Tableau. Finally, a slideshow will be used to present the results.

### B.2.B Not included in Project Scope

It is important to note that this is regular season data only. To truly identify the features that lead to winning a championship, a similar analysis will need to be done on post season data. This will come in a future project.

## B.3 Standard Methodology

I am using an agile approach to this project. It’s iterative and flexible nature will allow me to go piece by piece, going back and refining parts as needed. This methodology can be broken down into 6 parts:

1. Requirements: The requirements of this project are to discover the features that influence winning in the regular season and display those findings for multiple audiences.
2. Design: I will come up with an initial design based on my knowledge of the source material and the desired outcomes. This step will be revisited multiple times throughout the process and revised using the sources provided as inspiration.
3. Development: This is where I will write the code, including creating the model and custom functions, that will be used to generate the data for the analysis. This process will be repeated as many times as needed based on multiple phases.
4. Testing: Once the development is complete, it will be tested. Based on the outcomes of the test phase, other phases will be reworked, and this phase will be done again.
5. Deployment: Once the testing is completed successfully and returns the desired objectives, the analysis will be done, and the entire project will be delivered.
6. Review: Once the project is complete, I will review all aspects of its delivery to gain insights on future project development.

## 

## B.4 Timeline and Milestones

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone | Projected Time | Start Date | End Date |
| Collect/Import/Study Data | **4 Hours** | **5/27/2024** | **5/28/2024** |
| Clean/Process Data | **8 Hours** | **5/29/2024** | **5/31/2024** |
| Model Training/Testing | **6 Hours** | **6/1/2024** | **6/3/2024** |
| Analyze/Tune Model | **8 Hours** | **6/4/2024** | **6/7/2024** |
| Develop Visualizations | **4 Hours** | **6/6/2024** | **6/8/2024** |
| Create Slideshow | **2 Hours** | **6/9/2024** | **6/9/2024** |
| Present Findings | **1 Hour** | **6/30/2024** | **6/30/2024** |

## B.5 Resources and Costs

|  |  |
| --- | --- |
| Item | Cost |
| Custom Built Desktop Computer | **$3,000** |
| Jupyter Notebook Software | **Free** |
| Python Libraries | **Free** |
| Dataset | **Free** |
| Work | **$1,155 (33 Hours @ $35/Hour)** |

## B.6 Criteria for Success

Multiple outcomes will result in the success of this project. To start, performing accurate and informative calculations of advanced statistics will help to further my knowledge of NBA analytics and provide deeper insights into the data. Next, the efficient refining of a Lasso regression model will bolster my abilities to analyze all types of data and ensure that the results are accurate and predictable. To measure the effectiveness of the model I will use Root Mean Squared Error and R2 scores. Root Mean Squared Error will measure how accurately the model predictions are to actual outcomes. This value needs to stay within twenty percent of the feature mean. R2 measures how well the independent variables explain the variability of the target feature. I do not have a set value for this measure to determine the success of the model. I have taken out strong determining features in some of the secondary feature analysis to determine less obvious variable influences. The visualizations will all be considered successful if they plainly depict accurate and informative data. The interactive dashboard should be easily navigable and understood. Finally, the presentation will be considered successful if findings are appropriately shown and insights, or a lack there of, can be determined.

# C. Design of Data Analytics Solution

## C.1 Hypothesis

This is a 2-part hypothesis:

1. **Full Dataset Hypotheses:**
   1. **Null Hypothesis-** No features will significantly influence the target feature.
   2. **Alternative Hypothesis-** Certain features will influence the target feature.
2. **Split Era Datasets Hypotheses:**
   1. **Null Hypothesis-** There will be no variance in the regression coefficients between the 2 different eras.
   2. **Alternative Hypothesis-** There will be some variance in the regression coefficients between the 2 different eras.

**C.2 and C.2.A Analytical Method**

The method used to perform this analysis will be a Lasso regression model. This model will be developed and refined through the agile project development methodology. I will test multiple alpha values to ensure the most accurate model possible. I will also bootstrap the data and enter this into the same model for more robust results. Then, confidence intervals of 95% will be calculated to provide precise insights. The metrics of Root Mean Squared Error (MSE) and R-Squared (R²) score will be used to evaluate the reliability of the model.

The Lasso regression model is great for NBA team statistical analysis for multiple reasons:

1. Feature Selection - Lasso regression performs automatic feature selection by shrinking some coefficients to exactly zero. This means it can effectively identify the most important statistics (features) that influence the target variable (e.g., team performance, wins, player efficiency).
2. Handling Multicollinearity: In datasets with highly correlated variables, such as NBA team statistics where many performance metrics are interrelated, traditional regression methods can struggle. Lasso regression helps mitigate multicollinearity by penalizing the absolute size of the regression coefficients.
3. Interpretability: By reducing the number of non-zero coefficients, Lasso regression simplifies the model, making it easier to interpret. This can be particularly useful for coaches, analysts, and decision-makers who need clear insights into which factors most impact team performance.
4. Regularization: The regularization aspect of Lasso (the L1 penalty) helps prevent overfitting, ensuring that the model generalizes well to new, unseen data. This is crucial for predictive modeling in sports analytics, where past performance may not always perfectly predict future outcomes.
5. Sparsity: In datasets with many variables but limited observations (common in sports analytics), Lasso regression’s tendency to produce sparse models (models with few non-zero coefficients) is advantageous. It reduces the risk of overfitting and enhances model robustness.
6. Optimization: Lasso regression provides a balance between bias and variance, optimizing the predictive performance of the model. This balance is essential for making accurate predictions in the dynamic and sometimes unpredictable environment of sports.

Bootstrapping the data with 95% confidence intervals has multiple benefits:

1. Estimating the Stability of Coefficients: Bootstrapping allows us to assess the variability and stability of the coefficients estimated by the Lasso regression. By repeatedly resampling the data and fitting the model, we can observe how much the coefficients change, providing insights into which statistics are consistently important.
2. Confidence Intervals: Bootstrapping enables the estimation of confidence intervals for the regression coefficients. This is particularly useful in sports analytics, where understanding the range within which the true effect of a statistic lies can help in making more informed decisions.
3. Bootstrapping helps in validating the robustness of the Lasso regression model. By generating multiple samples and fitting the model on each, we can evaluate how well the model performs across different subsets of the data, ensuring that the model's performance is not overly dependent on a particular sample.
4. Bias Reduction: Bootstrapping can reduce the bias of the Lasso regression model's estimates. By averaging the results over many bootstrap samples, we can obtain a more accurate estimate of the model parameters.
5. Improving Predictive Performance: By using bootstrapping, we can improve the predictive performance of the model. The aggregated results from multiple bootstrap samples can lead to better generalization and more reliable predictions.
6. Resilience to Overfitting: Bootstrapping helps in assessing and mitigating overfitting. If a model performs well on multiple bootstrap samples, it is less likely to be overfitting the original data and more likely to generalize well to new data.
7. Uncertainty Quantification: In the context of sports analytics, quantifying uncertainty is crucial. Bootstrapping provides a straightforward way to quantify the uncertainty in the model's predictions and the importance of different statistics.

Using Root Mean Squared Error (RMSE) to evaluate the model is relevant for the following reasons:

1. Sensitivity to Large Errors: RMSE, like Mean Squared Error (MSE), squares the errors before averaging them, meaning that larger errors have a disproportionately higher impact on the metric. This is beneficial in contexts like NBA team statistics, where large deviations from actual performance (e.g., predictions of wins, player efficiency) are particularly undesirable and should be heavily penalized.
2. Interpretability: RMSE provides a clear and interpretable measure of the average squared difference between predicted and actual values, then takes the square root. This makes it straightforward to understand how well the model is performing in the same units as the target variable, enhancing interpretability.
3. Continuous Output Evaluation: RMSE is well-suited for regression tasks where the output is continuous, such as predicting the number of wins or points scored by a team. It directly measures the accuracy of the model's predictions by comparing the predicted values to the actual values.
4. Optimization: RMSE is commonly used as the evaluation metric for the loss function (MSE) minimized during the training of regression models, including Lasso regression. This means that the model is optimized to perform well according to this metric, making it a natural choice for evaluation.
5. Differentiability: The mathematical properties of the underlying MSE (being differentiable) facilitate gradient-based optimization methods used in training regression models, including Lasso regression. This helps in efficiently finding the optimal model parameters.
6. Penalty for Bias and Variance: By considering the squared errors and then taking the square root, RMSE implicitly balances bias and variance. A model with high bias will have high RMSE due to consistently poor predictions, while a model with high variance will also have high RMSE due to large errors on certain predictions. Thus, RMSE helps in selecting models that generalize well.
7. Alignment with Business Objectives: In sports analytics, minimizing prediction errors is often directly aligned with the objectives of teams and analysts, such as making accurate performance forecasts or strategic decisions. RMSE, by penalizing large errors, ensures that predictions are as close to actual values as possible, supporting better decision-making.

The evaluation of the R2 score will benefit my analysis in many ways:

1. Proportion of Variance Explained: R² represents the proportion of the variance in the dependent variable that is predictable from the independent variables. In the context of NBA team statistics, it provides a measure of how well the model captures the overall variability in team performance metrics (e.g., wins, points scored).
2. Interpretability: R² is easy to interpret, as it ranges from 0 to 1. An R² value closer to 1 indicates that a large proportion of the variance in the dependent variable is explained by the model, whereas a value closer to 0 indicates that the model explains very little of the variance.
3. Goodness of Fit: R² is a direct measure of the goodness of fit for the model. It helps in assessing how well the model’s predictions match the actual data. This is crucial in sports analytics, where understanding the fit of the model can help in making reliable predictions and strategic decisions.
4. Model Improvement: By tracking changes in R², analysts can assess the impact of adding or removing variables from the model. An increase in R² suggests that the model is improving in its ability to explain the variance in the dependent variable, guiding the model refinement process.
5. Business Relevance: In sports analytics, explaining the variance in team performance is often a key objective. R² provides a clear metric to quantify how well the model achieves this goal, supporting decision-making processes for coaches, analysts, and management.
6. Model Simplicity: While R² doesn’t directly penalize model complexity like some other metrics, it offers a straightforward assessment of how well the independent variables collectively explain the dependent variable. This can guide the development of simpler models that still provide a good fit to the data.

## C.3 Tools and Environments

The coding for this project will be done using the python language within a Jupyter notebook environment. The libraries used will be pandas, seaborn, matplotlib, math, statsmodels, and sklearn. The visualizations used for the presentation will be created using Tableau Public.

Note: No third-party code will be used for this project.

**C.4 and C.4.A Methods and Metrics to Evaluate Statistical Significance**

* **Model Type:** Asupervised learning Lasso regression model will be used.
* **Algorithms:** 
  + A Lasso regression algorithm will be applied to develop the model.
  + A bootstrapping algorithm will be used to solidify the predictability of the model.
* **Evaluation Metrics:**
  + Root Mean Squared Error
  + R2 Score
  + Regression Coefficients
  + Bootstrapping with Confidence Intervals
* **Criteria for Successful Model Deployment:**
  + The RMSE value will need to be less than twenty percent of the mean of the target feature to indicate high predictability. A reliable RMSE will allow me to accept or reject both null hypotheses.
  + The R2 score from the bootstrapped model will need to be greater than 0.5 to indicate that the model accurately explains a significant portion of the variance in the results. Models with lower R2 scores will still be looked at to gain general insights into possible secondary feature influences. A sufficient R² score will recognize the ability of the model to provide accurate results that will enable me to accept or reject both null hypotheses.
  + Based on the regression coefficients derived from the analysis, I can conclude if:
    - The null hypotheses of no features will significantly influence the target feature, and there will be no variance of regression coefficients between eras.
    - The alternative hypotheses of certain features will influence the target feature, and there will be a variance of regression coefficients between eras.
  + Bootstrapping with Confidence Intervals of 95% will allow me to assuredly deduce whether the regression coefficients can be used to reject or accept both null hypotheses.
* **Justification of Metrics and Methods:**
  + Lasso regression is well-suited for NBA team statistics analysis due to its ability to handle large, complex datasets, emphasize the most relevant features, and produce interpretable and generalizable models.
  + Bootstrapping with confidence intervals enhances the reliability and interpretability of Lasso regression models in the analysis of NBA team statistics by providing robust estimates, validating model stability, and improving predictive accuracy.
  + RMSE is a suitable and effective metric for evaluating Lasso regression models in NBA team statistics analysis due to its sensitivity to large errors, interpretability, optimization compatibility, and alignment with the goals of accurate prediction and model comparison.
  + R² is a valuable metric for evaluating Lasso regression models in NBA team statistics analysis due to its ability to measure goodness of fit, interpretability, ease of comparison, and alignment with the objectives of explaining variance and making accurate predictions.

## C.5 Practical Significance

The practical significance of this solution will be measured in multiple ways. First, the RMSE and R² scores must show model accuracy. I have stated that the RMSE should be less than twenty percent of the target mean. The R2 scores should be above 0.5, but this also won’t disqualify the integrity of the model completely. With these metrics displaying a functional model, we can have confidence in the model and the results it produces. Second, the regression coefficients and their confidence intervals will provide reliable data to base our model significance on. Lastly, practical significance can be gained by reviewing the results. Whether they show some features influencing the target stronger than others, no obvious influences, variance between eras, or no variance between eras, we will be able to confidently steer organizations in the right direction when it comes to decision making and meeting organizational needs.

**C.6 Visual Communication**

I will provide multiple types of visualizations to both analyze and display my findings for the project. I will use matplotlib and seaborn to generate graphics within the file to help analyze the data. I will present my results with visualizations generated using Tableau Public.

* Analytical Visualizations:
  + Correlation Matrix with Heatmap - I will create multiple correlation matrices with heatmaps using the built-in corr() function in the python language library.
  + Horizontal Bar Chart - A horizontal bar chart will be used in the analysis to display the correlation coefficients. It will be created using the matplotlib.pyplot library.
  + Axes Horizontal Line Chart - An axhline chart will be generated to visualize the correlation coefficients within their 95% confidence intervals.
  + Tableau Dashboard - I will create a dashboard in Tableau that shows bar charts representing correlation coefficients for all 3 datasets (full, era 1, & era 2). There will also be a story containing scoring data across the 3 datasets. This will include line and shaded graphs to represent possible trends.

# D. Description of Dataset

## D.1 Source of Data

The NBA Team Stats dataset will be downloaded from Kaggle.com. The dataset was authored by user Michael H. A link to this dataset will be available in the references section.

## D.2 Appropriateness of Dataset

This dataset contains all the necessary fields to calculate advanced statistics and run the analysis. Features include win percentage, field goal percentage, rebounds (offensive and defensive), steal, blocks, free throw attempts, and more.

**D.3 Data Collection Methods**

The dataset will be downloaded as a CSV file from the fore mentioned source. It will then be imported into the Jupyter notebook environment using the pandas read\_csv method.

## D.4 Observations on Quality and Completeness of Data

## This dataset was exceptionally clean and complete. There was no missing data or extreme outliers. I did use the original dataset to do some custom calculations of advanced metrics that would lead to a more complete analysis.

## 

## D.5 and D.5.A Data Governance, Privacy, Security, Ethical, Legal, and Regulatory Compliances

In reference to data governance, this dataset was free and available to the public from Kaggle.com. There were no privacy or security concerns using this dataset. I had to register an account with the site and accept the terms and conditions. The only real term that applies here is that I am of legal age to sign a binding contract, which I am. In terms of ethical and regulatory compliance standards, there is nothing within this dataset or analysis that could compromise myself or the organization this project is for.

**References**

H, M. (2024, April). *NBA Team Stats.* Retrieved from Kaggle: https://www.kaggle.com/datasets/mharvnek/nba-team-stats-00-to-18

Haefner, J. (2016). *9 Stats That Every Serious Basketball Coach Should Track.* Retrieved from

Basketball Breakthrough: https://www.breakthroughbasketball.com/stats/9\_stats\_basketball\_coach\_should\_track.html

Sullivan, R. (2020, April). *NBA’s Most Valuable Statistic Discovered: How To Predict Team*

*Wins With 95% Accuracy.* Retrieved from SPGN: https://www.sportsgamblingpodcast.com/2020/04/20/nba-most-valuable-statistic/

Peterson, D. (2020, May). *How Different Metrics Correlate with Winning in the NBA over 30*

*Years.* Retrieved from Medium: https://towardsdatascience.com/how-different-metrics-correlate-with-winning-in-the-nba-over-30-years-57219d3d1c8

# 