McMaster University

PROJECT RAPTORS REBOUND

DAT 205 - CAPSTONE

**WINTER 2021 – GROUP 1**

Bhavika Patil

Cindy Guo

Dennis Hung

Table of Contents

[Executive Summary 3](#_Toc67521069)

[Business Goal 4](#_Toc67521070)

[Objectives 4](#_Toc67521071)

[Data 5](#_Toc67521072)

[Gathering data – nba api 5](#_Toc67521073)

[Understanding the Data – NBA player salary data 7](#_Toc67521074)

[Project Planning 8](#_Toc67521075)

[Implementations 9](#_Toc67521076)

[References 11](#_Toc67521077)

Figures

[Figure 1 Example PlayerGameLogs data subset 4](#_Toc66743728)

[Figure 2 Available features for data between 2005 to 2020 (NBA Media Ventures, LLC., 2021) 6](#_Toc66743729)

[Figure 3 Example NBA Salary data subset 6](#_Toc66743730)

[Figure 4 Available features for salary data between 2015 to 2025 (ESPN and HoopHype) 6](#_Toc66743731)

[Figure 5 High level project schedule and milestones 7](#_Toc66743732)

[Figure 6 Gantt Chart of the overall project plan 7](#_Toc66743733)

# Executive Summary

# Business Goal

The Toronto Raptors are an NBA championship team with a $2.5 billion market capitalization. Key players added in 2019 contributed to critical wins during the season and championship playoff run. The team has been in decline since being crowned 2019 NBA champions.

The team’s General Manager (GM) and owners are seeking recommendations on improving the team’s chances at another championship run while preserving the team’s long-term future. The recommendations will be in the form of marginal/minor improvements to the player roster. The stakeholders intend to either maintain or improve on the number of wins during the regular season and more importantly during the playoffs.

The organization wants to retain a public interest in the team while improving the team’s performance without rebuilding the entire team. A spot in the NBA finals while maintaining the number of regular-season wins would be ideal.

# Objectives

To attain the requirements outlined by the management, we are looking to determine what features affect team wins/losses and predict how this can be improved/reduced through changes in the team’s roster.

The primary objective of this analysis is to determine which features are important to the Toronto Raptors team to produce more wins in the regular season and the playoffs.

The secondary objective is to review team players’ performance-to-cost effectiveness.

The scope of this analysis will be to focus on the historical player data from 2005 to 2020 as this closely represent the players of the current era. The performance baseline of the team and players will be established using the data for The Toronto Raptors 2019-2020 season. These statistics will be used to indicate performance features for the team and the individual players.

To simplify the analysis the following assumptions were made:

* Changes are only applied to the Toronto Raptors; the statistics from other teams are not affected or changed.
* Obtaining any players is up to the Raptors’ GM to determine if they will pursue the player via free agency, trading of players, etc.

# Data

### Gathering data – nba api

The data for this analysis is hosted by [nba.com](http://www.nba.com); It is made available via webservices through [stats.nba.com](http://stats.nba.com). nba\_api is an open access API client library for Python developed by Swar Patel. There are more than 250 endpoints (the method to request information through the API) available.

#### Understanding the Data - PlayerGameLogs

PlayerGameLogs endpoint gathered the statistical data of each player in each game by season (figure 1). This consists of 34 possible features from the year 1946 to the present day for roughly 1.3 million records. This is a live dataset which is updated periodically.



Figure Example PlayerGameLogs data subset

Feature WL in the above figure indicates win or loss of the game. As an exercise of this analysis, the win or loss is to be predicted based on players’ performances, which makes WL the target feature. Classification models such as Logistic Regression, Decision Tree, or Random Forest will be applied for the analysis.

|  |  |  |  |
| --- | --- | --- | --- |
| **Attributes** | **Data Type** | **Example value** | **Definitions** |
| **SEASON\_YEAR** | STRING | 2015-16 | NBA season |
| **PLAYER\_ID** | INT | 202710 | Player Unique ID |
| **PLAYER\_NAME** | STRING | Jimmy Butler | Player name |
| **TEAM\_ID** | BIG INT | 1610612741 | Team Unique ID |
| **TEAM\_ABBREVIATION** | STRING | CHI | Team 3-character abbreviation |
| **TEAM\_NAME** | STRING | Chicago Bulls | Team name |
| **GAME\_ID** | STRING | 0011500103 | Unique ID for the game |
| **GAME\_DATE** | DATETIME | 2015-10-23T00:00:00 | Date/Time game is played |
| **MATCHUP** | STRING | CHI vs. DAL | Teams featured in the game |
| **WL** | STRING | W | The player who was on the team who won/lost |
| **MIN** | FLOAT | 24.25 | Minutes Played The number of minutes played by a player |
| **FGM** | INT | 2 | Field Goals Made The number of field goals that a player has made. This includes both 2 pointers and 3 pointers |
| **FGA** | INT | 7 | Field Goals Attempted The number of field goals that a player has attempted. This includes both 2 pointers and 3 pointers |
| **FG\_PCT** | FLOAT | 0.286 | Field Goal Percentage The percentage of field goal attempts that a player makes Formula (FGM)/(FGA) |
| **FG3M** | INT | 0 | 3 Point Field Goals Made The number of 3 point field goals that a player has made |
| **FG3A** | INT | 0 | 3 Point Field Goals Attempted The number of 3 point field goals that a player has attempted |
| **FG3\_PCT** | FLOAT | 0 | 3 Point Field Goal Percentage The percentage of 3 point field goal attempts that a player makes Formula (3PM)/(3PA) |
| **FTM** | INT | 0 | Free Throws Made The number of free throws that a player has made |
| **FTA** | INT | 0 | Free Throws Attempted The number of free throws that a player has attempted |
| **FT\_PCT** | FLOAT | 0 | Free Throw Percentage The percentage of free throw attempts that a player has made Formula (FTM)/(FTA) |
| **OREB** | INT | 0 | Offensive Rebounds The number of rebounds a player has collected while they were on offense |
| **DREB** | INT | 3 | Defensive Rebounds The number of rebounds a player has collected while they were on defense |
| **REB** | INT | 3 | Rebounds A rebound occurs when a player recovers the ball after a missed shot. This statistic is the number of total rebounds a player has collected on either offense or defense |
| **AST** | INT | 6 | Assists The number of assists -- passes that lead directly to a made basket -- by a player |
| **TOV** | INT | 1 | Turnovers A turnover occurs when the player on offense loses the ball to the defense |
| **STL** | INT | 1 | Steals Number of times a defensive player takes the ball from a player on offense, causing a turnover |
| **BLK** | INT | 1 | Blocks A block occurs when an offensive player attempts a shot, and the defense player tips the ball, blocking their chance to score |
| **BLKA** | INT | 0 | Blocks Against The number of shots attempted by a player that are blocked by a defender |
| **PF** | INT | 3 | Personal Fouls The number of personal fouls a player committed |
| **PFD** | INT | 0 | Personal Fouls Drawn The number of personal fouls that are drawn by a player |
| **PTS** | INT | 4 | Points  The number of points scored |
| **PLUS\_MINUS** | INT | 11 | Plus-Minus The point differential when a player is on the floor |
| **DD2** | INT | 0 | Double Doubles The number of double-doubles (double-digit number total in two of the five categories in a game) a player achieves |
| **TD3** | INT | 0 | Triple Doubles The number of triple-doubles (double-digit number total in three of the five categories in a game) a player achieves |

Figure Available features for data between 2005 to 2020 (NBA Media Ventures, LLC., 2021)

### Understanding the Data – NBA player salary data

For the secondary objective, we have retrieved historical player salary data from [espn.com/nba/salaries](http://www.espn.com/nba/salaries). We have merged the future salary data from [hoopshype.com/salaries/](https://hoopshype.com/salaries/) to the salary dataset.

The Salary data consists of 16 possible features from the year 2015 to the 2025(If contracted) for approximately 500 players.

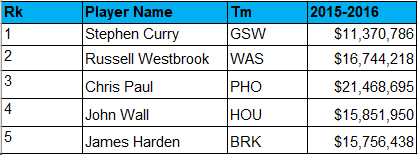


Figure Example NBA Salary data subset

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Data Type** | **Example value** | **Comments / Definitions** |
| Rk | INT | 1 | Current NBA rank of the player |
| PLAYER NAME | STRING | Stephen Curry | Player name |
| Tm | STRING | GSW | Team’s 3-character abbreviation |
| YYYY-YYYY | FLOAT | 11,370,786 | Salary (in US dollars) for the given NBA season year  Column YYYY-YYYY represents the NBA season year |

Figure Available features for salary data between 2015 to 2025 (ESPN and HoopHype)

# Project Planning

The overall implementation will be in 5 phases with a deliverable provided at the end of each phase. Completion of the final report is targeted for Apr 15th along with a presentation to summarizing the recommendations outlined in the report.

High Level Milestones: Summary of important deliverable dates

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ID** | **Deliverable / Action Item** | **Milestone (Due Dates)** | **Start Date** | **End Date** |
| **S1** | **Team Case Proposals** | **2021-03-04** | **2021-02-25** | **2021-03-04** |
| **S2** | **Final Plan** | **2021-03-18** | **2021-03-04** | **2021-03-13** |
| **S3** | **Interim Presentation** | **2021-03-25** | **2021-03-13** | **2021-03-25** |
| **S4** | **Final Report** | **2021-04-08** | **2021-03-13** | **2021-04-03** |
| **S5** | **Final “Boardroom” Presentation** | **2021-04-15** | **2021-04-03** | **2021-04-15** |
|  | **Implementation** | **2021-04-15** | **2021-03-01** | **2021-04-12** |
|  |  |  |  |  |

Figure High level project schedule and milestones

Detailed Project Plan: Detailed project schedule



Figure Gantt Chart of the overall project plan

# Implementations

The model from this analysis will be used by Business or Data analysts working for a given basketball team to improve the performance looking at any set of years. The analyst can choose which performance factors they would like to focus on.

This solution will need tweaking based on team / management to run for the real time. The API provides almost real-time data for the performances which can be configured with parameters.

The salary data, however, may need some manual effort for each new season.

The tools used:

* Visual studio code as an Integrated Development Environment
* Python as the programming language
* Microsoft Access, Microsoft Excel and csv as file formats for scraping data.
* Git as a collaboration repository: <https://github.com/bhavikapatil/Capstone> (currently, a private repository)
* GitHub desktop for a Git client
* Microsoft Office as reporting tools
* Tableau desktop for data visualization

# Model configuration

# Review of the results

Conditions Applied to the model

Correlation Matrix



Figure Heat map for Toronto Raptor (Regular Seasons 2004 to 2020)

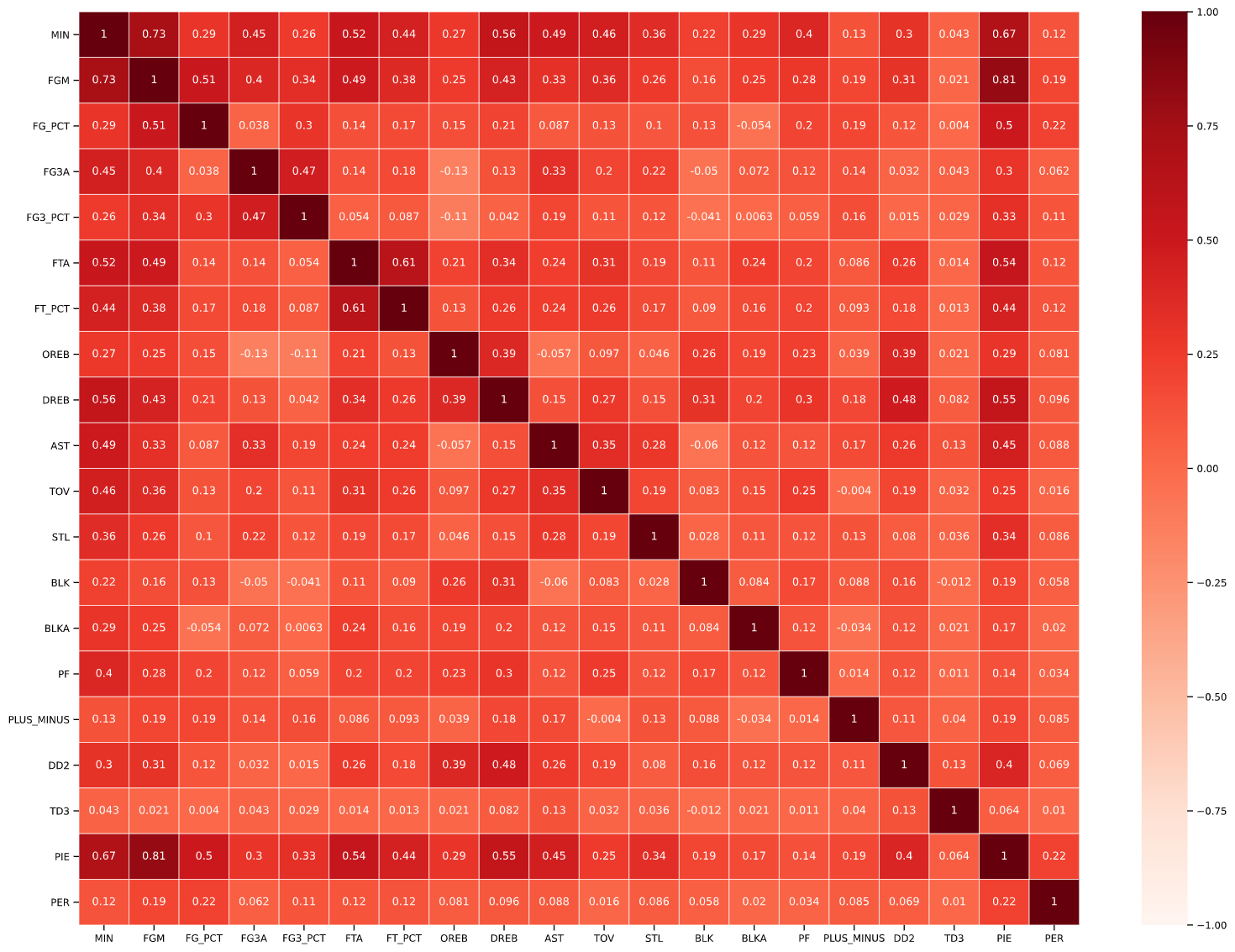


Figure Heat map for Toronto Raptor (Regular Seasons 2004 to 2020) after removal of several features with values > 0.84

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Logistic Regression** | | **Decision Tree** | | **Random Forest** | |
| **Feature** | **Feature Importance** | **Feature** | **Feature Importance** | **Feature** | **Feature Importance** |
| FG3\_PCT | 13.8% | PLUS\_MINUS | 28.6% | PLUS\_MINUS | 39% |
| TD3 | 12.2% | MIN | 14.0% | PER | 18% |
| FG\_PCT | 11.9% | PIE | 13.5% | PIE | 13% |
| FGM | 6.7% | PER | 11.2% | FG\_PCT | 6% |
| FT\_PCT | 6.5% | DREB | 3.9% | FG3\_PCT | 6% |
| DREB | 6.0% | FG\_PCT | 3.5% | MIN | 4% |
| STL | 5.8% | PF | 3.4% | PF | 3% |
| BLK | 5.7% | AST | 2.8% | DREB | 2% |
| AST | 5.3% | FG3A | 2.7% | FG3A | 2% |
| PLUS\_MINUS | 4.7% | FGM | 2.2% | BLKA | 1% |
| FTA | 4.4% | FTA | 2.2% | AST | 1% |
| OREB | 3.6% | OREB | 1.9% | FGM | 1% |
| FG3A | 3.1% | TOV | 1.9% | BLK | 1% |
| PER | 3.1% | FT\_PCT | 1.8% | TOV | 1% |
| MIN | 2.5% | FG3\_PCT | 1.8% | FTA | 1% |
| DD2 | 2.2% | STL | 1.6% | OREB | 0% |
| PIE | 1.0% | BLKA | 1.5% | FT\_PCT | 0% |
| TOV | 1.0% | BLK | 1.5% | STL | 0% |
| BLKA | 0.5% | DD2 | 0.2% | DD2 | 0% |
| PF | 0.0% | TD3 | 0.0% | TD3 | 0% |

Figure Model Feature Importance

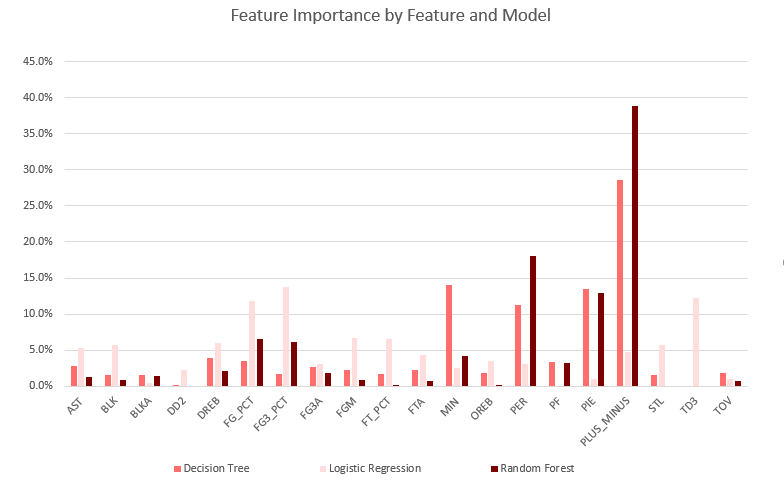


Figure Plot of Feature Importance by Model

# Recommendations

# References

Bleacher Report, Inc. (2021). *Advanced NBA Stats for Dummies: How to Understand the New Hoops Math*. Retrieved from bleacherreport.com: https://bleacherreport.com/articles/1813902-advanced-nba-stats-for-dummies-how-to-understand-the-new-hoops-math

ESPN. (2020-2021). *NBA Player Salaries 2020-2021*. Retrieved from ESPN.com: http://www.espn.com/nba/salaries

HoopsHype. (2020-2021). *NBA Player Salaries*. Retrieved from hoopshype.com: https://hoopshype.com/salaries/players/

Jee, K. (2019, December 22). *How to Get NBA Data Using the nba\_api Python Module (Beginner).* Retrieved from Playing Numbers: https://www.playingnumbers.com/2019/12/how-to-get-nba-data-using-the-nba\_api-python-module-beginner/

NBA Media Ventures, LLC. (2021). *NBA Advanced Stats*. Retrieved from NBA.com: https://www.nba.com/stats/

NBA Media Ventures, LLC. (2021). *NBA Advanced Stats - Glossary*. Retrieved from nba.com: https://www.nba.com/stats/help/glossary/

Patel, S. (2020, August 19). *swar / nba\_api.* Retrieved from GitHub: https://github.com/swar/nba\_api/blob/master/docs/table\_of\_contents.md

SportRac. (2021). *Toronto Raptors 2020 Salary Cap*. Retrieved from spotrac.com: https://www.spotrac.com/nba/toronto-raptors/cap/

Sports Reference. (2020-2021). *2020-21 NBA Player Contracts*. Retrieved from Basketball Reference: https://www.basketball-reference.com/contracts/players.html

Sports Reference. (2020-2021). *Basketball Reference*. Retrieved from Basketball Reference: https://www.basketball-reference.com/