What the Puck: Building a Successful NHL Team Using Data

Dustin Casey
Bellevue University
Bellevue, NE 68005 USA
dcasey@my365.bellevue.edu

Abstract

This project will explore how data science and analytics can be used to try to make an existing National Hockey League (NHL) franchise successful by optimizing their use of funds on player salaries. This will be done by exploring how a franchise can build the best roster while staying under budget and create an exciting team to watch.

Author Keywords

NHL; Data analytics; Roster Optimization; Sabermetrics; Statistics;

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; See http://acm.org/about/class/1998 for the full list of ACM classifiers. This section is required.

Introduction

Creating a well-functioning, winning, and profitable sports franchise can be a challenge even in the best markets like Los Angeles or New York, but even more challenging in a small market or when a team does not have a large budget to work with. Optimizing and utilizing every dollar can make an average team good and a good team great. There is also a level of balance that needs to be achieved so that your team doesn't

win every game and bore people. As Miller says in *Sports Analytics and Data Science: Winning the Game with Methods and Models*:

> "Gaining market share and becoming the dominant player is a goal of firms in many industries. Not so in the business of Professional sports. If one team were assured of victory in almost all of its contests, interest in those contests would wane. A team benefits by winning move often than losing, but winning all the time may be less beneficial than winning most of the time." [1]

This project will explore how data science and analytics can be used to try to make an existing National Hockey League (NHL) franchise successful by optimizing their use of funds on player salaries. This will be done by exploring how a franchise can build the best roster while staying under budget and create an exciting team to watch.

Sports Statistics Use Today

The way we view player statistics and evaluate their abilities is always evolving, especially as we have more and more data to work with. One example involving statistics in basketball from Brian Skinner states,

"Thus far, publicly-available studies using player tracking data have largely focused on augmenting or refining the use of conventional statistics. For example, recent studies have examined the effect of a defender's proximity on shooting percentage, broken down shooting percentage based on how many dribbles are taken before the shot, characterized the effect of defender proximity on shooting percentage, and examined the dependence of rebound rate on spatial location. These studies are certainly illuminating, and they suggest significant improvements that can be made to the conventional statistics by which players are evaluated." [2]

The use of these new statistical viewpoints can help us make more well-rounded decisions when evaluating players.

Brian McDonald also discusses the use of new statistics and why they are prevalent in hockey. McDonald comments on how the randomness and scarcity of goals in hockey really limits our ability to judge players' performance. This also makes it a challenge to use goals to predict future outcomes for teams. [3] This is why we must use other performance metrics and stats to better evaluate how well teams and players are doing. These new statistics are always important to consider as we move forward into evaluating players and their worth.

Another revolution in sports modeling and predicting outcomes comes from using new technology and equipment we have access to now. An example of this can be seen in fitness tracking technology offered by a variety brands such as Apple and Fitbit. Being able to track the physical health and performance of athletes as they train and compete can help coaches and trainers view their progress over time.

Along with equipment made to track performance and training, many researchers and companies are trying to implement Internet of Things (IoT) into sports. By connecting multiple devices that players use along with new equipment to track performance and training,

teams hope to be able to get a better idea of what influences higher performance for players and why others don't see as great of results. [11]

Another instance of tech used in sports can be seen with new optical tracking being used in the NBA to predict points and decision making in real time. In a paper on the subject, Dan Cervone, Alex D'Amour, Luke Bornn, and Kirk Goldsberry discuss the usefulness and benefits of this technology. Because basketball is a game of split second decision making, it would be useful to be able to track decisions made and if they resulted in positive outcomes. [9]

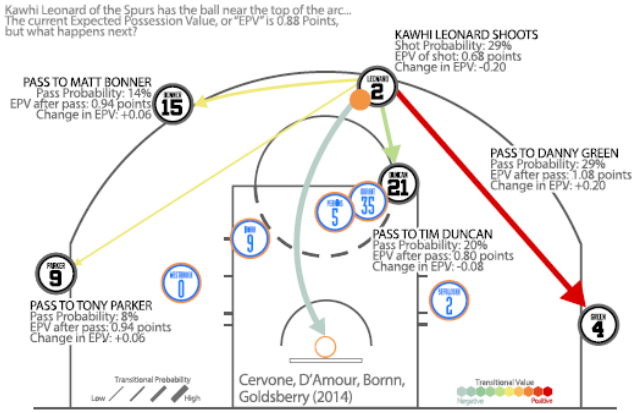

Figure 1. Diagram of EPV as a weighted average of the values of the ballcarrier's (Leonard's) decisions and the probability of making each decision. We also consider the possibility of Leonard dribbling to a different area or driving toward the basket, as well as turning the ball over, but these are omitted from the above diagram for conceptual clarity.

The authors of the paper describe how they came up with a new metric they created called expected possession value (EVP) by scoring/assigning point values to each moment of an offensive position and predicting the likelihood that the possession will end in a made basket. [9]

This technology and way to break down the game will lead to immense advances in how we see the game and evaluate performance. Using this tech along with new metrics and equipment, teams and organizations will be able to better predict the value of players and make better decisions when considering who to play, who to bench, who to sign and who to cut.

Statistics in the National Hockey League

Statistics in the NHL recorded across many different categories from offensive to defensive, from individual statistics/performance to team performance. This can make it more difficult to track a player's performance, especially as hockey is such a fast-moving sport and there isn't always an easy way to quantify every move the player is making. Some great examples of this can be seen in positioning and how the player performs as a member of his line or unit on the ice. Or how often he screens the goalie or protects the puck from leaving the zone over the blue line.

This makes hockey a much tougher sport to track and quantify the impact individual performance has on the team as a whole when there are many different factors, both individual and group oriented, that can affect the success of the team. This differs from baseball which is much more individualistic (batter versus pitcher) with a team aspect to it.

Mason and Foster do a great job of summarizing these differences as they talk about how baseball players act autonomously or semi-autonomously throughout a game. Each player can either make a play or prevent a play that would score runs. Knowing this, players' performance can be isolated and measured individually. [5]

Figure 1: Zach Parise of the Minnesota Wild

Basic NHL Stats Used Currently: [8]

- Corsi – Plus/minus for shot attempts instead of goals.
- Fenwick – same as Corsi, but only shots on goal and missed shots count
- (any statistic) for Percentage – (Corsi, Fenwick, Goals) as a ratio.
- (any statistic) relative to Teammates – how did a player perform in a given statistic compared to his teammates.
- Score-Adjusted – newer statistic. Adjusts a player's stats depending on the score of the game and how they would play given the situation.
- "Per 60" Metrics – takes ice time into consideration when evaluating stats.

Mason and Foster also discuss what they call complex sports/organizations such as basketball and hockey. Every outcome in these sports cannot be attributed to one player but the effort of the whole team working as a unit to produce and prevent goals from being scored. Because of the interdependence of the players, it is much hard to track and measure each individual player's performance and affect on the outcome of the game. [5]

Another challenge facing analysts is the availability of statistics due to feuds between the NHL and the National Hockey League Players' Association (NHLPA). GMs and Teams were concerned players would use statistics to arbitrate for better contracts. This led to the league discontinuing keeping expanded statistics at the beginning of the 2002-03 NHL season. Players were trying to use newly made and expanded statistics to win arbitrations over higher salaries and better contracts. The NHLPA ended up filing a grievance and the statistics were reinstituted. But for the 2003-04 season, the statistics were not made available to the media. They could only be seen by teams and players. [4]

This made it much more difficult to get the statistics needed to analyze how individuals can affect team outcomes. This set the NHL behind other major sports in North America.

Methodology for Building the Most Successful Hockey Team

Using NHL player statistics, salary data, and scholarly sources on the subject, a model will be built to evaluate and rate players on their ability compared to their salary. This method is trying to find the players with the best value in each position.

Multiple machine learning algorithms will be used to evaluate players and the one with the best results will be used in the final output.

One idea we want to keep in mind as we build out these models comes from Andrew Brunette, a winger for the Colorado Avalanche and baseball nut. Andrew welcomes increased statistical analysis but believes that there are more nuances and variables to the game of hockey that cannot be captured by statistics. These intangible attributes are very hard to measure and quantify. [6]

Andrew believes that we need to move past looking at metrics like goals and assists and focus on asking questions like *Moneyball* does – "what is undervalued?" If we can find what areas of the game are undervalued, we can utilize and focus on those areas. Andrew sees hockey moving in that direction. [6]

Changing our thinking and not trying to predict what players' stats improve wins and outcomes but thinking more about what areas of the game are undervalued, we may find more answers. Brunette describes this perfectly and displays his knowledge for the game. If the penalty kill is the area our team struggles most in or is the most undervalued area of the game, we want to exploit that to our advantage and find the guys who excel in that area at the lower price. This will improve our team and allow us to invest money in some of the other areas our team needs more capital.

Results

The results of this project will provide a full 23-man NHL roster that is under the salary cap restrictions of

the National Hockey League. The goal is to have the best players who will already either be on the team, picked up via free agency, or drafted in the 2019 NHL season. Without being able to actually build the team and have them compete, testing the "best" roster possible will be tough and somewhat objective. But, the evaluation will be based on prior stats and college stats for players who are drafted.

The main idea behind this would to find what areas of the game are undervalued around the NHL and also to see how much those areas affect the success of the team. If we find that a strong powerplay or penalty kill is undervalued and provides positive affects to game outcomes, we want to target those areas in the players we choose to sign. We also want to find a good value to keep the team even and not lacking in any one area.

As we evaluate many different areas and statistics, we can see why this has not been easy for any analyst to solve and is constantly a struggle for owners and general managers to use to evaluate players before signing/drafting them. With this in mind, there may not be one easy solution or answer, but we can use this analysis to make better decisions overall and support our signings.

Craig Button of the Toronto Maple Leafs talked about making the transition to using data analysis to make decisions and how it will always be a challenge no matter what business you are in. Craig believes that change and moving towards more data analysis will clash with traditional methods but will provide better results and proof to stakeholders on why decisions are being made. The idea is to challenge the status quo and try to open the eyes of some of those individuals who may be more close-minded and reluctant to change. [7]

It may be tough for teams to make the transition to making personnel decisions using data analytics and statistics, but challenging the current system is important in growing the game and the way these tools are used to make better decisions.

Conclusions

It can be challenging to make decisions for sports franchise as an owner or general manager no matter what. But, with the wealth of data and the ability we have today to analyze and make decisions from it, their jobs can be a bit easier.

It is also good to keep in mind that data isn't the end-all be-all and that many decisions teams make about player personnel transcend data as there are many intangible attributes that players poses that won't show up in the stats book.

Knowing this, it is always a good idea to use the data we have but also trust people in our organizations who have been around the game for many years and are experts in their field.

Acknowledgements

I want to thank my mom and my dad for supporting me with this project and teaching me to love the great game of hockey from a young age. Some of my favorite and most cherish childhood memories were at Minnesota Wild games and they are the reason/inspiration for this project.

References

1. Thomas W. Miller. 2016. *Sports Analytics and Data Science: Winning the Game with Methods and Models (1st Edition).* Pearson Education, Inc., Old Tappan, NJ.
2. Skinner, B., & Guy, S. J. 2015. *A method for using player tracking data in basketball to learn player skills and predict team performance.* PLoS One, 10(9) doi:http://dx.doi.org.ezproxy.bellevue.edu/10.1371/journal.pone.0136393
3. Brian Macdonald. 2012. *An Expected Goals Model for Evaluating NHL Teams and Players.* MIT Sloan Sports Analytics Conference, Boston, MA.
4. Ballard, C. (2005, October 24). *Measure of success. Sports Illustrated, 103(16), pp. 78-82.* Morton L. Heilig. 1962. Sensorama Simulator, U.S. Patent 3,050,870, Filed January 10, 1961, issued August 28, 1962.
5. Daniel S. Mason and William Foster. 2007. *Putting Moneyball on Ice?*, International Journal of Sport Finance (November 2007).
6. Dater, A. (2006, October 15*). A new ice age: NHL plays numbers game Hockey taking broader approach in analyzing statistics to determine a player's net worth in team game*. Denver Post, p. BB.01
7. David, S. (2006, October 5). *Baseball-style statistical analysis starts creeping into hockey world.* The Canadian Press.
8. Charlie O'Connor (2017, October 9). *An advanced stat primer: Understanding basic hockey metrics.* The Athletic.
9. Dan Cervone, Alexander D'Amour, Luke Bornn, and Kirk Goldsberry. 2014. *POINTWISE: Predicting Points and Valuing Decisions in Real Time with NBA Optical Tracking Data*. MIT Sloan Sports Analytics Conference.
10. Coates, D., & Humphreys, B. R. 1999. *The growth effects of sport franchises, stadia, and arenas.* Journal of Policy Analysis and Management, 18(4), 601-624.
11. Mahanth Gowda, Ashutosh Dhekne, Sheng Shen, and Romit Roy Choudhury. 2017. *Bringing IoT to Sports Analytics.* USENIX Symposium on Networked Systems Design and Implementation, Boston, MA.