NHL Players Time on Ice

Data analysis of NHL players time on the ice

This Capstone Project is submitted in partial fulfillment of the requirements for the course Data-Driven Decision-Making (MDA 620) during the Fall Semester of 2022.

While writing this Capstone Project, I have not witnessed any wrongdoing, nor have I personally violated any conditions of the LIU Honor Code.

Adam Goodsir
December 15, 2022

Table of Contents

Figures

Tables

Background

The NHL is one of the four major professional sports leagues in North America; the others are the NBA, NFL and MLB. The NHL is the second oldest professional sports league and came into existence in 1917, just fourteen years after the MLB was created in 1903. The NHL has grown tremendously over the past 10 years in every aspect and especially when it comes to data analytics. Every NHL team has an analytic team to help break down the game film, stats, see trends in their play and their opponents play (Larkin, 2021). Stats and numbers have become an integral part of the game and allowed for teams to make business decisions, tactical decisions and personal decisions based on the numbers and what they are telling the teams. There are countless stats and metrics that NHL teams collect everyday in practice and in games but one of the most important ones is time on ice, or also known as "TOI". This stat is extremely important since in hockey, the players are always rotating on and off the ice and understanding who is on the ice allows teams to have better insight to who is helping or hurting the teams success.

Problem Scenario/Business Issue

As much as hockey is a sport and a viewing pleasure for millions of fans, NHL teams are a business and treated as such when the annual operating cost of an NHL team is around 70 million USD (Baird, 2005). The amount of money that is at stake for the various teams and owners is very significant and that is why as much as it's a sport, it is a business. Just like any other business, owners want all the possible information to help their business succeed and make sure they have the right personal operating and running the business. This analysis will be conducted from a business standpoint to see how players' production is related to TOI.

Figure 1. Highest Single-Game Toi

Objective/Goals of the Project

The purpose of this project is to examine their productivity and efficiency of players based on their time on ice (TOI). We want to see if there is correlation between TOI to Games Played (GP), Assists and PTS(Points) to have a better understanding of the stats that are collected which can represent which stat shows a player efficiency.

Data Exploration

The NHL comprises 32 teams which span from east coast to west coast and include teams in Canada. The 32 teams are divided into four divisions and. “The four division names are Atlantic, Metropolitan, Central and Pacific. The Atlantic and Metropolitan divisions make up the Eastern Conference and each division has eight teams. The Central and Pacific divisions make

up the Western Conference and have seven teams in each division." (NHL, 2013). Similar to other professional sports teams and leagues, besides winning games, the main focus is making money and having a successful business on and off the ice, field or court. The benefit to making more money on your professional team is that money can be used to acquire the best players, improve facilities and provide the fans with the best possible experience so they will come back for more games and support the team. Much of this success depends on the quality of the team and the success they have on the ice.

As technology has increased, NHL teams and the analytic teams they have hired are able to track any stat you could think of. Technology allows players shifts to be tracked with GPS ability, on-ice performance as well as their physical performance with information such as heart rate, calories burned, blood lactate and many others. This sport is much more than a game and hobby, this is a job for the players and a multi-million dollar business that needs to be run to the best of its abilities in order to succeed. The data is comprised of every single player who played in a NHL game in 2018, the stats collected were Player, Age, TM(Team), Pos(Position), GP(Games Played), GPG(Goals Per Game), A(Assists), PTS(Points), +/-, TOI and ATOI(Average Time on Ice).

Data Visualization

Figure 2 looks at the relationship between "GP' (Games Played) and "TOI" (Time on Ice). As we can see there is a linear relationship between the two variables and we are able to conclude that the more games played in a NHL season, the greater time on ice an individual player will have. This is to be expected since when a player has more game experience, a coach's trust in them and their ability tends to increase thus allowing the player to be on the ice more and better suited to fill the necessary role and position in order to help the team win.

Figure 2. GP relationship to TOI

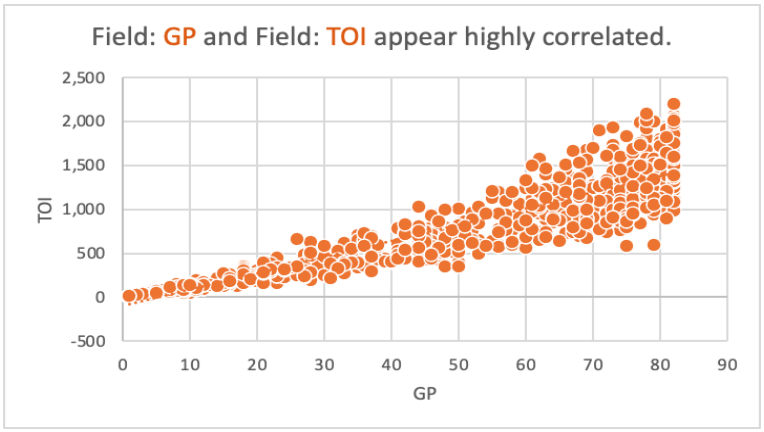


Figure 3 looks at the relationship between Assists and TOI. Based on the data and the trend, we can see that as TOI increases, the number of assists increases as well. This is important for NHL coaches and teams to look at since if TOI is correlated to the number of assists, coaches will want to try to increase the TOI for players that are known for setting people up to score goals which will result in more goals and a greater chance of winning games.

Figure 3. Assists compared to TOI

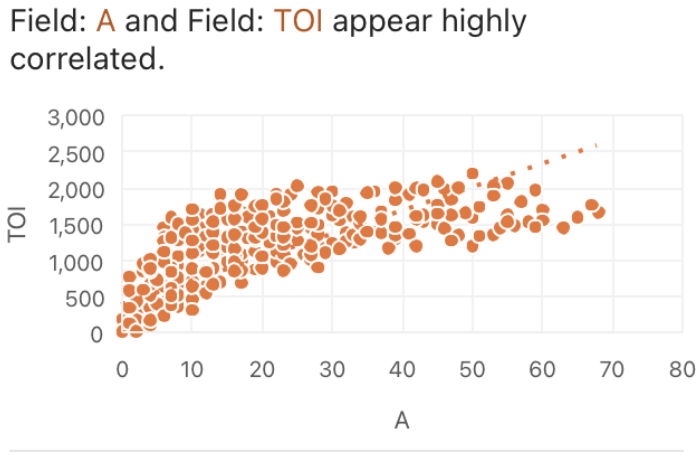


Figure 4 depicts the relationship between PTS and TOI. PTS represents points a player has which is composed of both goals and assists. Based on the data, we can see there is a strong relationship between PTS and TOI. There is a point that more TOI is not correlated with PTS and

this could be related to athletes fatigue and at some point, their physical abilities decline but the general trend is more TOI is correlated with more points.

Figure 4. PTS compared to TOI

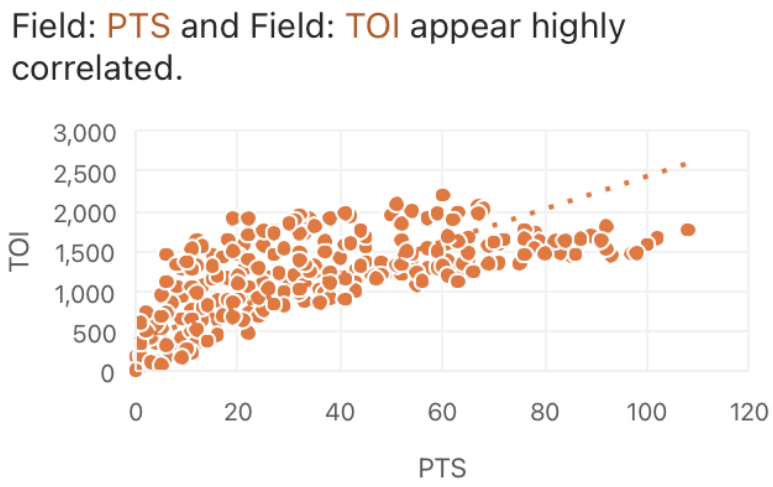


Figure 5 looks at the various positions in hockey and their respective TOI. It is interesting to see that 'D' , which are the defenseman, have a noticeably higher TOI than any other position but when looking at overall PTS, forwards 'C','LW','RW' tend to lead this category. This could be the nature of the position and being more defensive minded while the other positions tend to be more offensively focused.

Figure 5. 'Pos D' TOI

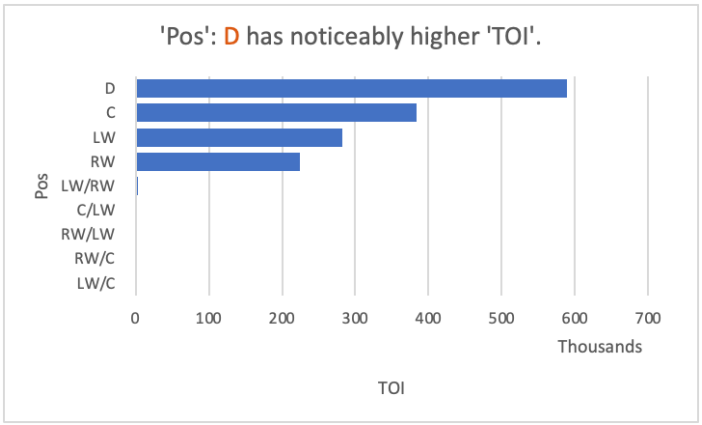
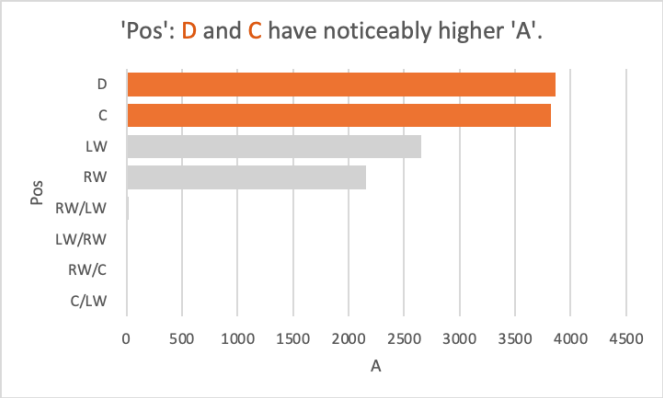


Figure 6 looks at the number of assists based on the position of the players. Based on the data, we see that Centers and Defensemen have an increased number of assists compared to Left and Right Wingers. In my opinion, this is caused by the nature of these positions and the roles they provide on a team as well as the increased TOI.

Figure 6. Assists sorted by position



Data Manipulation

Before I was able to analyze any of the data or run various models, I had to understand that data that was available, determine which data that I wanted to use and that was relevant and finally format it so when it was time to use it, the data was ready. I began with filling in the rows that contained " -" which represented no value or zero and I changed them to zero in order for R Studio to better read the data. Next, I had to delete data that was in columns and rows that was not relevant to the topic of looking at players productively related to TOI. After deleting these columns, I had to determine the time frame that I will be analyzing this data. The data set that I originally downloaded was extremely large and contained data from 1940. This dataset was too large and much of the data that I wanted to look at was unavailable due to the advances in technology. The dataset contained every single player that had played a single game in the NHL from years 1940-2018. With this extremely large amount of data, I decided to look only at the

most recent year, 2018. This data contained almost 900 names so I determined that this was enough information to make conclusions about players TOI and their production.

Methodology/Model Building/Analysis

Linear Regression Model:

For the first model in this analysis, I used a Linear Regression Model and looked at the various independent variables regressed on the dependent variable which was TOI (Time on Ice). The independent variables included GP, PTS and Assists. I used the linear regression model equation which is listed below.

$$TOI = \beta_0 + \beta_1(GP)$$
$$TOI = \beta_0 + \beta_1(PTS)$$
$$TOI = \beta_0 + \beta_1(Assists)$$

Table 1. GP Linear Regression Results

	TOI
GP	18.37 (0.2529)
Cons	-96.23 (14.88)
Observations	888
R-Squared	0.856

Table 2. PTS Linear Regression Results

	TOI
PTS	20.29 (0.5406)
Cons	392.2498 (16.9970)
Observations	888
R-Squared	0.6133

Table 3. Assists Linear Regression Results

	TOI
Assists	32.28 (0.8096)
Cons	392.83 (16.1664)
Observations	888
R-Squared	0.6133

Tables 1,2 and 3 confirm what the data visualization showed us. There is a very strong correlation between GP, PTS, Assists and TOI. We see the beta coefficient being significant in all three models. For GP, when a player's GP increases by one game, their TOI increases by 18.37 minutes. When a player's number of assists increases by one, their TOI increases by 32.28 minutes. Finally, when a player's PTS increases by one, their TOI increases by 20.29 minutes.

Multiple Regression Model:

For the second model, I used a Multiple Regression Model. For this model, TOI will be the dependent variable and my independent variables will be GP, PTS and Assists. The equation used for the Multiple Regression Model is listed below.

$TOI = \beta_0 + \beta_1(GP) + \beta_2(PTS) + \beta_3(Assists)$

Table 6. Multiple Regression Model

	TOI
GP	14.81 (0.3033)
PTS	-7.52 (1.0196)
Assists	22.18

	(1.5422)
Cons	-57.0937 (12.4484)
Observations	886
R-Squared	0.9035

After performing a Multiple Regression Model, we see that when a player increases their GP by one, their TOI will increase by 12.81 minutes. Similarly, as a player's number of assists increases by one, their TOI increases by 22.18 minutes. The interesting thing about this Multiple Regression model is that, when a player's PTS increases by one, their TOI decreases by 7.52 minutes. This is very interesting since in the linear regression model, as PTS increase, TOI increases as well but opposite in this model. This could be explained by the fact that PTS is composed of both goals and assists and when looking at GP, PTS and Assists all together compared to TOI, the goals portion in PTS did not have a strong correlation to TOI.

Conclusions/Recommendations

After performing both models, I would recommend that NHL teams would use a Multiple Regression Model. Based on the data and the results, the multiple regression model takes into account more factors than just looking at one compared to TOI. It is extremely important for NHL teams to have all the information and be able to see the full picture of a player when making decisions about them and their involvement with the team and organization. GP has a correlation, Assists has a correlation but PTS does not in the multiple regression model and this could be due to the fact that PTS is comprised of goals and assists and based on my expert opinion, assists are harder to create for players at a high level due to the speed of the game and decisions that must be made in order to result in a goal.

Limitations

Even though the year 2018 contained almost 900 players that logged games in the NHL, this analysis could be even further studied by looking at 5-10 years of data and looking at the relationship between players TOI and their performance/productivity. One could say that with increased TOI, a player will have more chances to produce and this could be true but a players individual skill is in play and we can quantify some aspects of their performance but in sports, there are intangibles that play a factor in performance and production. With this being said, our results can only be generalized and looked at with correlation but not causation in terms of TOI and productivity of players.

Works Cited

Baird , M. (2005). *Marquette University Law School*. Sports Facility Reports | Marquette University Law School. Retrieved December 13, 2022, from https://law.marquette.edu/national-sports-law-institute/sports-facility-reports

Larkin, M. (2021, November 25). *Secret formula: Analytics in the NHL*. The Hockey News. Retrieved December 13, 2022, from https://thehockeynews.com/news/secret-formula-analytics-in-the-nhl

NHL.com, & Rosen, D. (2013, July 19). *NHL introduces new division names with Schedule*. NHL.com. Retrieved December 13, 2022, from https://www.nhl.com/news/nhl-introduces-new-division-names-with-schedule/c-678456

Varner, J. (2019, October 3). *NHL skater stats*. Kaggle. Retrieved December 13, 2022, from https://www.kaggle.com/datasets/ace184/nhl-skater-stats