

GLOSSARY

For this capstone, I am using hockey statistics that have been downloaded from Evolving Hockey and financial information that has been scraped from CapFriendly. After wrangling and cleaning the data, there are **147** columns of data for each data point, and the data points represent unique Player/Years.

With so many statistics, it is difficult for a layperson to easily decipher the meanings of all of them. Here, I will categorize the statistics and give a brief description of each one in order to provide a better understanding of the results of the models and interpretations from this project.

Player Information

These features are descriptive to the player's history, financial and biographical information.

Player: The player's full name.

Year: The season that the player played during. Because hockey seasons run before and after January 1st, seasons are typically listed as a two year range (ie. 2014-15). For this data, a single year is listed which is the year that the season *ended*. A season that is listed as 2015 represents the 2014-15 season.

Team: The team the player played for in the season. Multiple players will have more than one team listed in this category as a result of being traded to another team mid-season.

Birthday: The player's birthday, formatted as yyyy-mm-dd.

Age: The age of the player at the *start* of the season they played during.

Draft Yr: The year the player was drafted.

Draft Rd: The round the player was drafted. In 1993 and 1994 there were 11 rounds, from 1995 to 2004 there were 9 rounds and since 2005 there are 7 rounds in the draft.

Draft Ov: The pick overall the player was drafted. The total number of picks per draft varies as a result of league rules, but since 2005 there are typically 212-224 picks in the draft.

Salary: This is the total amount of money that a player received for the season as part of their contract.

Cap Hit: This is the total amount of money that a player's contract counts against the salary cap of their team and is calculated by the *average* salary of a player for the duration of their contract. This is the primary target for the regression models.

From a team-building perspective, this is a more important factor than salary as it is the cap hit of a player that will determine how a team can collect talent. For the 2021-22 season, that amount is \$81.5 million spread across a minimum of 20 active players and a maximum of 23.

Take the following two examples:

Player A has a 4-year contract with the following salary breakdown:

Year 1 = \$8 million, Year 2 = \$6 million, Year 3 = \$4 million, Year 4 = \$2 million
This averages to \$5 million per year, so the cap hit is equal to \$5 million

Player B also has a 4-year contract, but with a different salary breakdown:
Year 1 = \$5 million, Year 2 = \$5 million, Year 3 = \$5 million, Year 4 = \$5 million
This averages to \$5 million per year as well, so the cap hit is also equal to \$5 million

For the purposes of team-building, these two players are the same in regards to their impact on the salary cap despite the differing salary structures. This is why I have chosen Cap Hit as the target for my models as it more directly impacts that construction of an NHL roster.

Left-Handed: This is the handedness of the player. In this dataset this is a binary feature, 1 for left-handed and 0 for right-handed.

Defenceman: This represents the position a player plays. In hockey generally, there are Goalies, Defencemen and Forwards. Forwards are further categorized as Left Wing, Center and Right Wing.

For the purposes of this project, we have chosen not to include Goalies as it is a small sample size with completely different statistics. I have also decided to just recognize players as just Forwards, rather than their specific forward position. This is a binary feature, so 1 is for a player that is a Defenceman and 0 represents a player that is a forward.

Basic Statistics

These statistics are the common statistics that are tracked in regards to player performance, typically found in an old-fashioned 'box score'.

GP: Games Played. This is the number of games a player was active for through the course of a season.

G: Goals. Specifically, this is credited to the offensive player who last touched the puck prior to the puck crossing the goal line.

A1: First Assist. This is credited to the last offensive player who touched the puck before the goal scorer. In the event that a goal is scored and the last person to touch the puck before the goal scorer is a defensive player, then no assist is rewarded.

A2: Second Assist. This is credited to the second-to-last offensive player who touched the puck before the goal scorer. In the event that a goal is scored and the second-to-last person to touch the puck before the goal scorer is a defensive player, then no assist is rewarded.

Points: This is the sum of a player's goals, first assists and second assists.

In this dataset, there are several statistics that are expressed as a rating, specifically as $\frac{\text{statistic}}{60}$. This expresses the rating at which a player accumulated the stat over the course of 60 minutes of on-ice time.

For example, if a player scores 1 goal in 20 minutes of ice time, then their $G/60 = 3$.

G, A1, A2 and Points all have /60 statistics in this dataset.

Time On Ice Statistics

These statistics represent the amount of time a player is on the ice in specific situations.

TOI: Time On Ice. The total number of minutes a player was on the ice over the course of an entire season.

TOI_EV: Time On Ice, Even-Strength. This is the total number of minutes a player was on the ice over the course of a season when there were the same number of players for both teams (5-on-5, 4-on-4, 3-on-3).

TOI_PP: Time On Ice, Power Play. This is the total number of minutes a player was on the ice over the course of a season when their team had a man advantage as a result of the opposing team taking a penalty.

TOI_SH: Time On Ice, Shorthanded. This is the total number of minutes a player was on the ice over the course of a season when their team had a man *d*isadvantage as a result of their own team taking a penalty.

EV (Even-Strength), PP (Power Play) and SH (Short-handed) are game situations that will be revisited in other more advanced statistics. When listed, the statistics refer specifically to when the stat is being accumulated in that situation.

Possession Statistics

These statistics are an attempt to evaluate and measure the amount of possession a player and team has over the course of the game. Calculating this was really the first 'wave' of advanced statistics in the NHL in hopes of measuring other factors that led to (or prevented) goals being scored.

In order to do this, evaluators tried to use the number of shots a player/team took to judge how many more scoring chances a team would get over the course of a game and a season. How this was measured differs and becomes important for later advanced statistics. In order to understand these stats, some definitions are needed first.

Shots on Goal: Any shot attempt that would have resulted in a goal had a goalie not saved it.

Blocked Shots: A shot attempt that is stopped by a defensive player other than the goalie.

Missed Shots: A shot attempt that misses the net and is not blocked by either a goalie or a defensive player.

These definitions help define these possession statistics.

GF: Goals For. This is the total number of goals scored by a team while a player was on the ice. The player just has to be on the ice for the goal, not necessarily earning a point.

GA: Goals Against. Total number of goals scored against a team while a player was on the ice. Again, ALL players on the ice accrue a Goal Against when a goal is scored against their team.

SF: Shots For. This is the total number of Shots on Goal and Goals accrued while a player was on the ice.

SA: Shots Against. The total number of Shots on Goal and Goals given up while a player was on the ice.

CF: Corsi For. This is the total number of Shots on Goal, Goals, Blocked Shots and Missed Shots (essentially ALL shot attempts) that a team accrued while a player was on the ice.

CA: Corsi Against: The total number of Shots on Goal, Goals, Blocked Shots and Missed Shots that a team gave up while a player was on the ice.

FF: Fenwick For: The total number of Shots on Goal, Goals and Missed Shots (all shot attempts except for Blocked Shots) that a team accrued while a player was on the ice.

FA: Fenwick Against: The total number of Shots on Goal, Goals and Missed Shots that a team gave up while a player was on the ice.

In this dataset, as well as these gross totals, these measures are also expressed as a differential:

G+/- = Goals For - Goals Against

S+/- = Shots For - Shots Against

C+/- = Corsi For - Corsi Against

F+/- = Fenwick For - Fenwick Against

As well, these stats are also expressed as a percentage, where 50% would represent an equal number of For and Against in each of these stats:

GF% = Goals For/(Goals For + Goals Against)

SF% = Shots For/(Shots For + Shots Against)

CF% = Corsi For/(Corsi For + Corsi Against)

FF% = Fenwick For/(Fenwick For + Fenwick Against)

All of the 'For' stats are also calculated at the individual level, seeing how many of each statistic an individual player is responsible for:

iSF: Individual Shots For: Number of shots on goal and goals accrued by an individual player.

iCF: Individual Corsi For: Number of ALL shot attempts accrued by an individual player.

iFF: Individual Fenwick For: Number of all shot attempts, minus blocked shots, accrued by an individual player.

There is no iGF stat, that would be the same as Goals for a player.

These individual statistics are also translated into a percentage to measure the relative success of their shots.

Sh%: Shooting Percentage: Goals/Individual Shots For

FSh%: Fenwick Shooting Percentage: Goal/Individual Fenwick For

Lastly, the total stats (ie. SF/SA), the differential stats (ie. S+/-) and the individual stats (ie. iSF) are all also expressed as a /60 minutes rating (ie. SF/60, S+/-/60, iSF/60) to normalize the data for players that have different levels of ice time.

In addition to these shot-based possession statistics, there are other stats that are believed to influence the amount of puck possession a team has leading to potentially better outcomes.

FOW: Face-off Wins. The number of times a player wins a face-off, giving their team possession after a stoppage in play.

FOL: Face-off Losses. The number of times a player loses a face-off, giving the opposing team possession after a stoppage in play.

FOW and FOL both have rating stats (FOW/60, FOL/60) and a differential stat (FO+/-).

IBLK: Individual Blocked Shots. The number of times a player blocks a shot attempt and prevents it from becoming a shot on goal.

GIVE: Giveaways. The number of times a player, in the flow of play, loses possession of the puck to the other team.

TAKE: Takeaways. The number of times a player, in the flow of play, takes possession of the puck from the other team.

iBLK, GIVE and TAKE all have rating stats as well (iBLK/60, GIVE/60, TAKE/60)

Penalty Statistics

Through the course of a game, players will be penalized for infractions. These come as either minor penalties (2:00 long) or major penalties (5:00 long). Evolving Hockey tracks not just players who committed the penalties (such as tripping a player) but also the player who *draws* the penalty (the player who got tripped).

iPENT2: Individual Minor Penalty Taken. The number of minor penalties a player committed.

iPEND2: Individual Minor Penalty Drawn. The number of minor penalties a player draws.

iPENT5: Individual Major Penalty Taken. The number of major penalties a player committed.

iPEND5: Individual Major Penalty Drawn. The number of major penalties a player draws.

iPEN+/-. Penalty Differential. The number of total penalties drawn minus the number of total penalties taken.

Advanced Statistics

This is where things get much more complex.

To put it simply, the advanced hockey analytics community, similar to other sports, has attempted to calculate the contributions of a player down to a single number. For Evolving Hockey (and many others in the community) this statistic is:

GAR: Goals Above Replacement. This attempts to measure how many more goals a player would contribute over a 'replacement level' player.

At Evolving Hockey, this is the primary advanced stat, that is then further broken down to understand a player's contribution to a team in all phases of the game. I will not be going into the math of how GAR is calculated, but if you would like to understand the process of how it was derived, the full analysis can be found [HERE](#).

To use a specific example, the highest recorded GAR in this dataset is Connor McDavid in 2017, with 33.6. This suggests that McDavid would contribute 33.6 more goals than a replacement level player. At the other end of the spectrum, Derick Brassard in 2019 had a GAR of -14.6, suggesting that he contributed to 14.6 fewer goals than a replacement player would have.

Another important factor that the hockey analytics community recognizes is that the outcomes produced in hockey are more random than in other sports. Scoring goals, especially within a smaller sample size of games, may be more of a result of luck than skill. Therefore, utilizing the possession statistics described (and an increasing amount of player tracking on the ice), models are developed to estimate how much a player would be *EXPECTED* to score given the other stats he has accrued. This has led to the development of these primary stats:

xGF: Expected Goals For. The number of goals a team would expect to score based on the individual player's possession and player tracking statistics.

xGA: Expected Goals Against. The number of goals a team would expect to give up based on the individual's player possession and player tracking statistics.

Having an understanding of these three primary advanced statistics (GAR, xGF, xGA) are the main building blocks for understanding the contributions of players in all facets of the game. I will now list the remaining advanced statistics that are based on these few.

WAR: Wins Above Replacement. This is a conversion of GAR that calculates the number of Wins a player's team earns, over a replacement level player, based on that player's contribution.

SPAR: Standing Points Above Replacement. Similar to WAR, this calculates the number of points their *team* would earn in the standings based on the player's contributions.

xG: Expected Goals. The number of goals an individual player would get based on their other performance statistics.

GAR can be broken down into specific phases of play. Adding these sub-statistics up will get to the total contribution reflected in GAR.

Off_GAR: Offensive GAR. Total offensive contributions to GAR.

Def_GAR: Defensive GAR. Total defensive contributions to GAR.

EVO_GAR: Even Strength Offensive GAR. Contributions to GAR made at even-strength (ie. 5-on-5) on offence.

EVD_GAR: Even Strength Defensive GAR. Contributions to GAR made at even-strength on defence.

PPO_GAR: Power Play Offensive GAR. Contribution to GAR made while playing on the power play with a man advantage.

SHD_GAR: Shorthanded Defensive GAR. Contribution to GAR made while playing on the penalty kill while at a man disadvantage.

Take_GAR: Taking Penalties GAR. The contribution to GAR as a result of committing penalties. A positive number here would indicate situations where other players would have committed a penalty where this player did not.

Draw_GAR: Drawing Penalties GAR. The contribution to GAR as a result of making the opposing team commit penalties.

Pens_GAR: Penalties GAR. The total contribution to GAR as a result of penalties. This is the sum of Take_GAR and Draw_GAR.

GAR is an accumulative statistic that is ultimately based on the number of goals scored by the player and by the team when the player is on the ice. As mentioned, expected goals for (xGF) has become an important measure for the analytics community to measure the potential offensive output. GAR models use Goals For (GF) as the foundation for the model, so xGF can also be used in order to calculate an *expected* GAR (xGAR) statistic. I will list these stats, which have the same definitions as above with the only difference being the use of xGF in the calculations.

xWAR

xSPAR

xOff_GAR

xDef_GAR

xEVO_GAR

xEVD_GAR

xPPO_GAR

xSHD_GAR

xTake_GAR

xDraw_GAR

xPens_GAR

Lastly, as all of these stats are accumulated totals through the course of the year, they are also expressed as ratings (ie. GAR/60). Every GAR and xGAR statistic listed above has a correlated /60 statistic as well.

Other Statistics

The last few statistics do not fit neatly into the other categories of statistics, so they will be listed here.

oiSh%: On-Ice Shooting Percentage. This is the *team's* shooting percentage while the player is on the ice.

Sv%: On-Ice Save Percentage. This is the save percentage of the goalie for the player's team.

oiSh% and Sv% are considered 'puck-luck' statistics. Over the course of the year, these two numbers should add up to 100%. Players higher than 100% could be considered 'lucky' whereas those below 100% could be considered 'unlucky'.

iHF: Individual Hits For. The number of hits (bodychecks) a player delivers to a player of the opposing team.

iHA: Individual Hits Against. The number of hits a player takes from an opposing player.

iHF and iHA both have /60 rating statistics as well.

Additional Resources:

[Evolving Hockey Statistics Glossary](#)

[Hockey Analytics Background](#)