NHL Player Valuation System: Using Performance Metrics to Understand and Predict Player Salary

Michael Kuby, Sangmun Kim, Yuyang Chen, and Haichen Sun

Table of Contents

Motivation

The NHL is commonly referred to as a results-based business. Players get paid to perform at the highest levels, and some players can transcend barriers and do the unimaginable on the ice, year after year. Those who find a way into the league and succeed are rewarded handsomely come contract time. However, the methods and models for explaining a player's value could be more clear. For those directly involved in the contract negotiations, including general managers, player agents, and players themselves, one or more tools must exist to determine a player's worth based on their performance. However, what those tools or models are or look like is not common knowledge. This motivates the project herein, which attempts to both peel back the curtain for those on the outside looking in and offer a product for those at the negotiation table.

Existing Solutions: Player Cards from The Athletic

Predicting a player's value is not a new phenomenon. It is an integral element of all professional sports and is done regularly by general managers, player agents, fans and players alike.
One tool that attempts to value players similarly to the project herein is The Athletic Player Cards (https://theathletic.com/5015509/2024/03/30/nhl-player-cards-pacific-division/). As stated on their website, the "value is according to Net Rating, our all-in-one player value stat that's based on each player's Offensive and Defensive Rating. The cards showcase all the states that determine how strong each player's ratings are on and off the puck, a weighted combination of their production (goals, assists, expected goals, blocks, penalty differential) and their play-driving (on-ice expected and actual goal stats)."

While interesting from a fan perspective, it is rather complicated to determine the utility of the Player Cards due to their seeming lack of reasonableness. For example, The Athletic Player Card for Connor McDavid values him at $19.0 M, representing 22.75% of the salary cap, which is hard to reconcile with the fact that no player in the salary cap era has ever been paid above 16% of the cap, nor could they be since the most a player can be paid under the collective bargaining agreement is 20%. And this valuation extends to numerous other players in the game that are considerably less prominent in the sport. Hence, our model tries to reconcile the valuation of a player based on their performance in a practical manner that can be used to make decisions in the real world.

Problem Statement

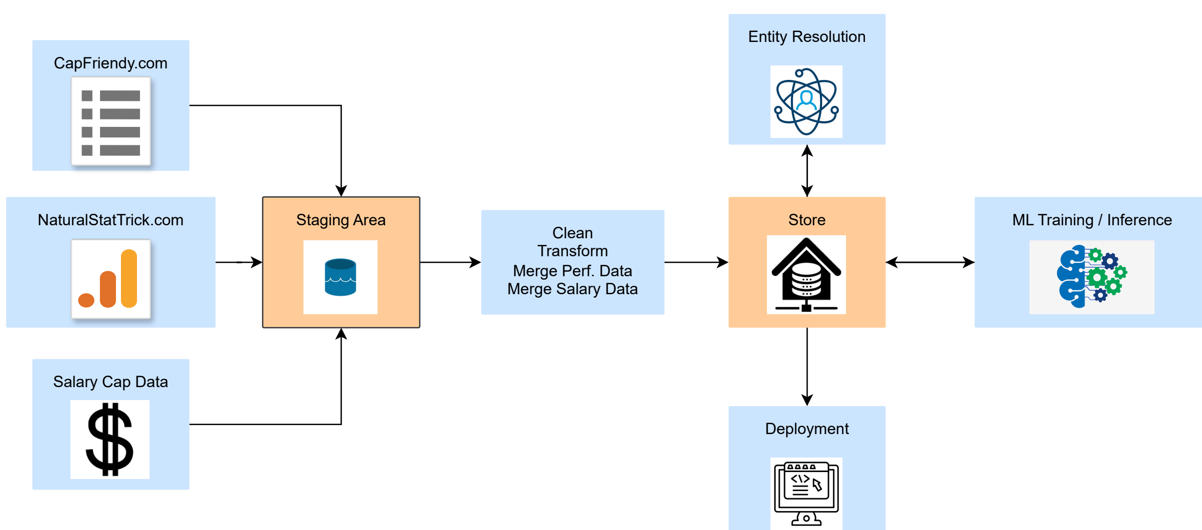
Given our motivation for the project, we try to answer the following questions:

- What is player A worth? Why?
- Why does player A get paid X and player B get paid Y?
- What do general managers find valuable in a player?
- Who is overperforming relative to their cap hit? Who is overperforming?
- What explains the valuation of a Forward (Center, Left-wing, Right-wing)?
- What explains the valuation of a Defenceman?

Limitations of Quantitative Analysis

Given the problem statements we wish to answer, it is vital to recognize the limitations of quantitative analysis. Many elements lead to success in team sports that numbers cannot capture. Leadership, mentorship, spirit, camaraderie, sportsmanship, humour, supportiveness and kinship are all necessary, desirable, valuable, and sought-after characteristics that cannot be captured by performance metrics alone. As such, any quantitative model that uses statistics, machine learning, or otherwise inherently fails to capture aspects of a player's value. As such, we recognize the limitations of our model in the sense that it strictly uses quantitative analysis to drive value and should be used as a starting point, not an endpoint.

Data Science Pipeline



Extract

Our data comes from two sources: CapFriendly.com, which tracks player salary data, and NaturalStatTrick.com, which houses player performance metrics—both standard and advanced metrics—and biographical information. Data regarding the Salary Cap is found here. We use BeautifulSoup to procure the data which is moved to a staging area for further processing.

Transform

Salary Data

After procuring the data, the first task was to formulate our target values. Deciding on how to formulate our **y** values was not trivial. Each year, each player has many data points regarding their salary, including their cap hit, AAV, salary, base salary, signing bonus, and performance bonus. A complete discussion of each of these figures is beyond the scope of this paper. Still, we will discuss what is necessary to explain our reasoning for formulating our target labels.

Cap hit is calculated by dividing the total salary plus signing bonuses by the contract's length. Performance bonuses are included in a player's cap hit.

AAV is calculated by dividing the total salary plus signing bonuses by the contract's length. Performance bonuses are *not* included in a player's AAV.

Salary is the actual amount of money the player earns each year, not including any signing bonuses. It can vary from year to year in the contract.

Base salary is the guaranteed amount a player earns each year and does not include signing bonuses, performance bonuses, or any other forms of additional compensation.

Signing bonuses are guaranteed money lump sum payments paid right when a contract is signed or on a specific date outlined in the contract. They are prorated over the contract term, meaning the cap hit from a signing bonus is spread evenly across each year of the contract's duration.

Performance bonuses in the NHL are additional compensation players can earn by achieving specific milestones outlined in their contracts. These bonuses are over and above the base salary and can be tied to various individual and team performance metrics.

After some deliberation, we used cap hit as our base value. Cap hit and AAV are equivalent for most players in the league; however, a subset of players, primarily young players just entering the league, are required to sign entry-level minimums at roughly 1% of the salary cap maximum. For young but highly skilled players primed to be difference-makers on the ice, performance bonuses offset their entry-level salaries, allowing even first-year players to make around 4% of the salary cap. It is also important to note that both restricted and unrestricted free agents cannot earn performance bonuses, that is, after their age 35 season in the latter case.

Since AAV would be likely to deflate the predicted value of a subset of players who we can reasonably assume to be performing at a high level (as symbolized by the fact that their signing team offered them performance bonus incentives, which tend only to be offered to certain players), we decided that cap hit was the most appropriate metric for tracking value year over year.

Inflation is another hurdle that needs to be addressed regarding our target labels. Since inflation causes a meaningful change in the interpretation of raw totals every year, we require some form of normalization. This transformation was done by dividing the players' cap hits by the respective salary cap for that year, producing a float value between 0 and 1, representing the percentage of the salary cap it costs a team to roster that player in the given year. When you see values of 0.03 or 0.06 in the graphs later in this document, they represent these normalized percentages.

Performance Data

Each of the three performance metric datasets undergoes a cleaning, refinement, and transformation process. The standard and advanced metrics were extracted as rates in a 'metric x per 60 minutes played' format, i.e., x/60. While useful, per-game metrics were also of interest, and so all features from the standard and advanced metrics were converted first to totals, then to per-game played, i.e., x/GP.

There was no actual relational structure between the three performance datasets, and indeed, NaturalStatTrick.com split them for logical display reasons. At this point, the decision was made to combine all three datasets before performing entity resolution with our salary data.

Entity Resolution

The 'Player' and 'Year' columns were identified as the primary keys for merging, with 'Date of Birth' included as an additional key to ensure uniqueness, given the potential for players with the same name to be in the league simultaneously.

The total number of performance data records was 15219, while the total number of salary data records was 20719. Taking the minimum of these two, the total number of records we could hope to retain after resolution was just over 15000. An inner join on the two datasets retained 12152 records, representing a 79.8% retention rate.

Cleaning after entity resolution

Unfortunately, after entity resolution, it was found that duplicate columns existed between the two datasets with wide discrepancies in values. At this point, a second heavy data cleaning was done to ensure that values were as desired. Duplicate columns were removed, relevant information was extracted, new features were created, all the features were converted to ensure they were the appropriate type and outlier values were removed.

Methodology

The general approach taken was:

1. Use of domain knowledge to select relevant features.
2. Correlational analysis to select and identify important features.
3. Model training.
4. Model analysis.
5. Inference.

Model training and analysis, however, underwent many iterations.

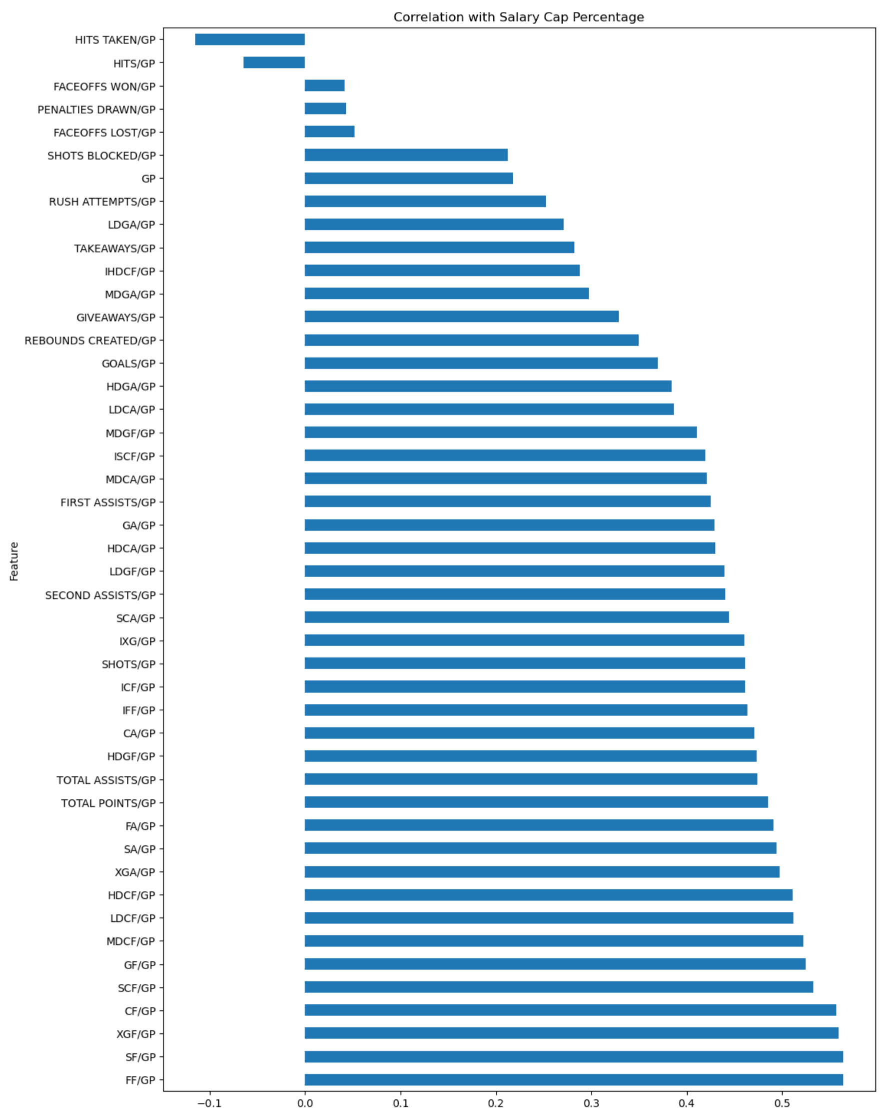
Note: Many features analyzed in the following section will not be known to the casual reader. An overview of each of these features is beyond the scope of this paper, but I encourage you to visit [this web page] to get a detailed breakdown of each feature discussed.

Domain Knowledge

Obvious features to be selected for training include Total Points/GP, Total Assists/GP, Goals/GP, First Assists/GP, Second Assists/GP, Shots/GP, Goals For/GP, XGF/GP, and IXGF/GP. To win a hockey game, you must score goals and record points, so we expect each of these columns to be highly positively correlated with our target labels.

Correlational Analysis (Defence)

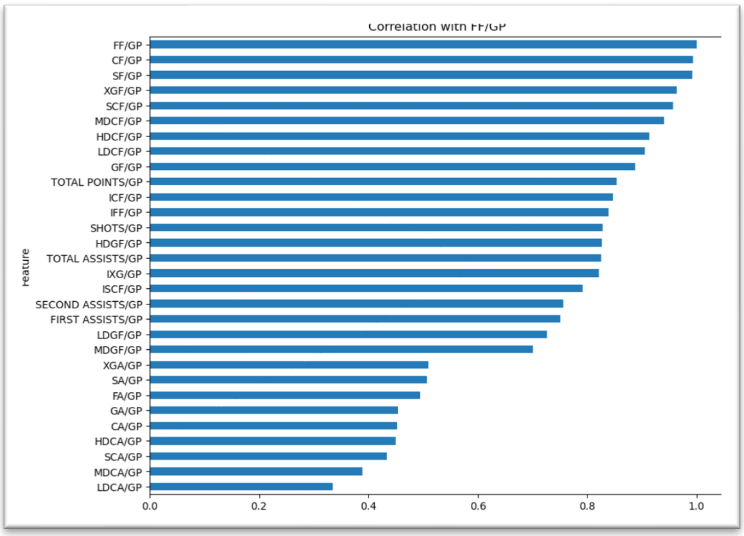
Before performing correlational analysis for defence players, we analyzed only those who had played more than 41 games in a given season to ensure that a proper sample size of performance data corresponds to our labels. The next step was to create a horizontal bar plot identifying the correlations between each feature.



Our findings were that a huge number of features have a reasonably high positive degree of correlation with our target labels. Features hovering around positive 0.5 of particular interest include FF/GP, SF/GP, XGF/GP, CF/GP, GF/GP, MDCF/GP, LDCF/GP and SCF/GP. Notably, most of these features are advanced metrics that do not track standard aspects of the game, such as goals and assists. We performed extensive covariance analysis to identify which of these features to focus on.

FF/GP

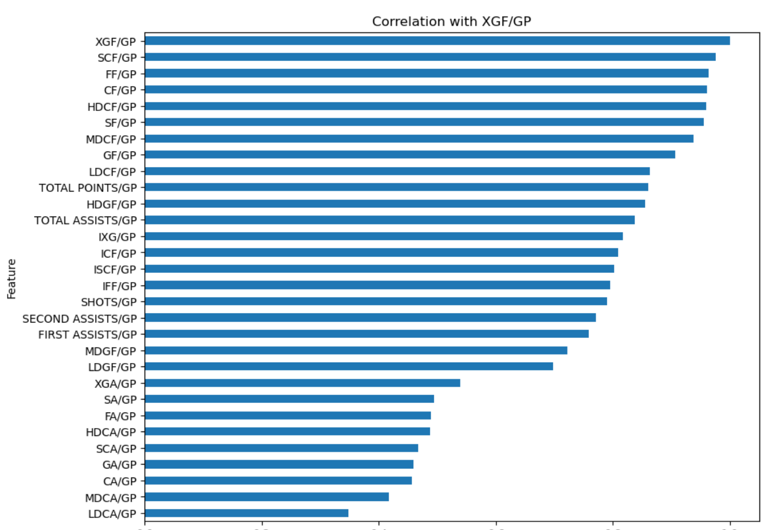
A bar graph of FF/GP with each of the other 20 most correlated features (sorted by absolute value) produced the following figure:



From this, we see that FF/GP is nearly perfectly correlated with CF/GP and SF/GP and has correlational coefficients above 0.8 for over half of the features we have to work with. We may remove several of these features and only use FF/GP during training.

XGF/GP

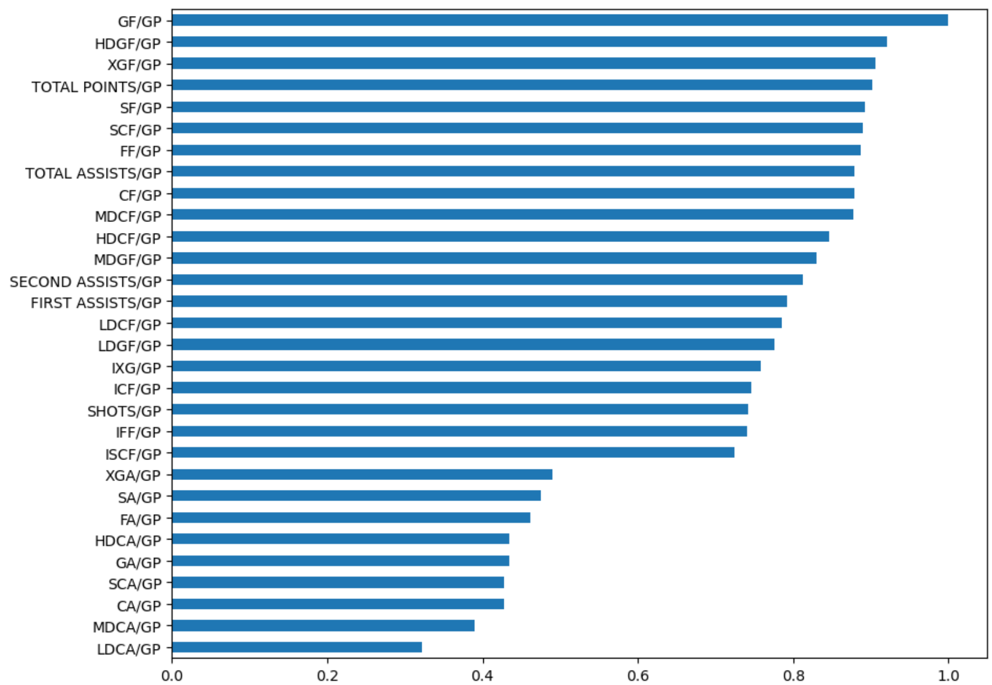
Next, we analyzed XGF/GP for covariance among the most correlated features.



With XGF/GP, we start to see a similar trend: its correlational coefficients are above 0.6 for all features except for metrics ending in ‘A’, which stands for ‘Against’.

GF/GP

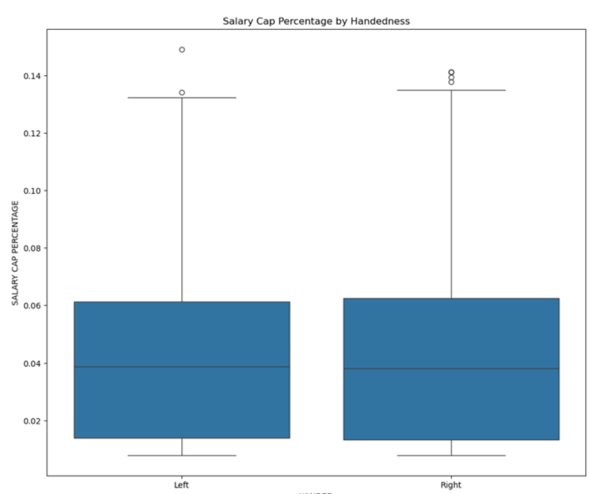
The final feature we analyzed for correlational analysis for defensive players was GF/GP. Rather interestingly, compared to the previous features analyzed, it had a somewhat reduced level of

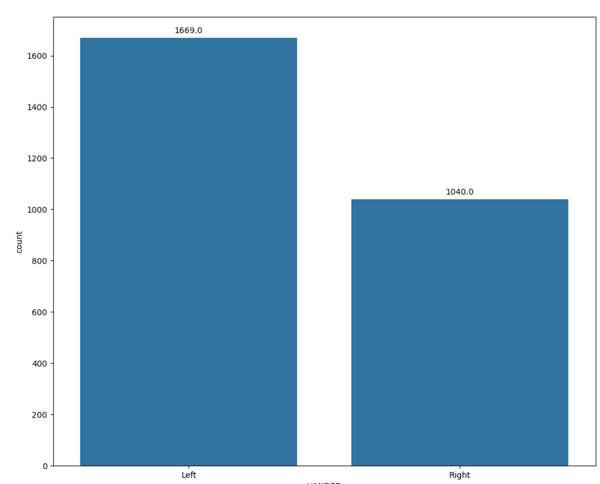


correlation with a number of the other features. Given its importance in the results of actual gameplay, this is a rather promising feature to include during training.

Non-Numerical Feature Analysis

The only non-numerical feature identified for analysis was handedness, that is, the way a player holds his stick. Our results show that their cap hit distributions remain similar despite the large

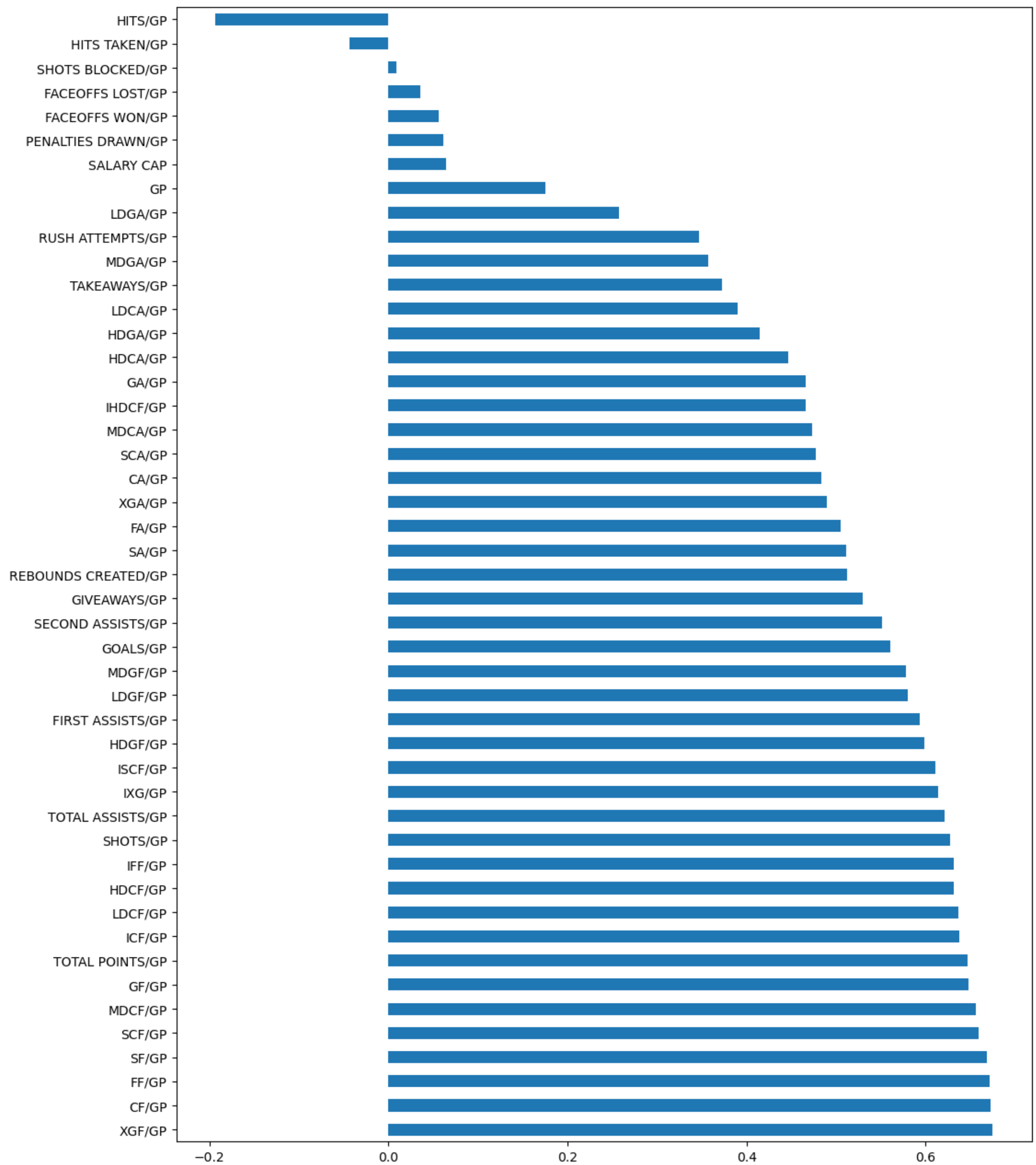




discrepancy in the number of left-handed defensemen relative to right-handed defensemen. Given these findings, it is unclear whether their inclusion as a feature during training would be helpful.

Correlational Analysis (Wingers)

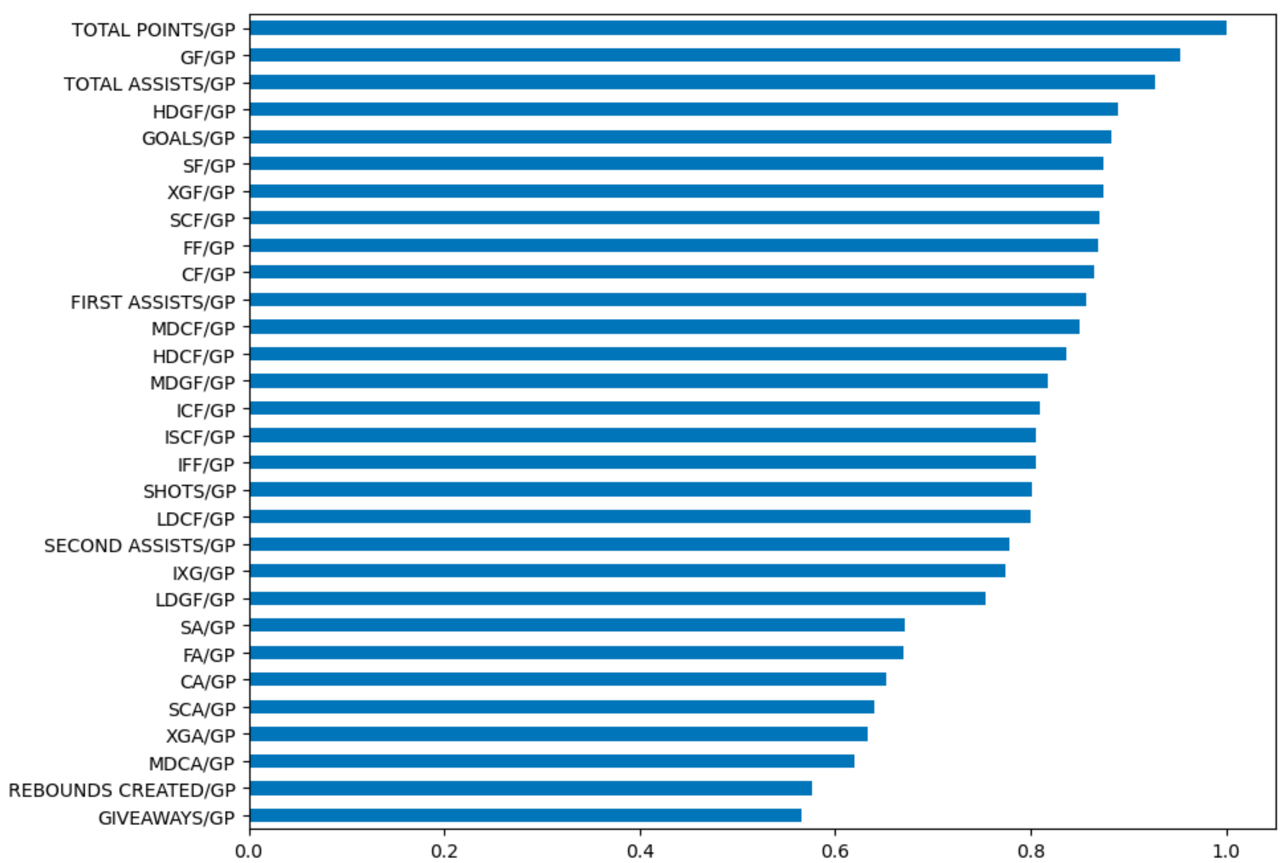
Again, we decided to perform a correlational analysis on only those players who had played more than 41 games in a given season. The following shows each feature's correlational coefficient with respect to our target labels.



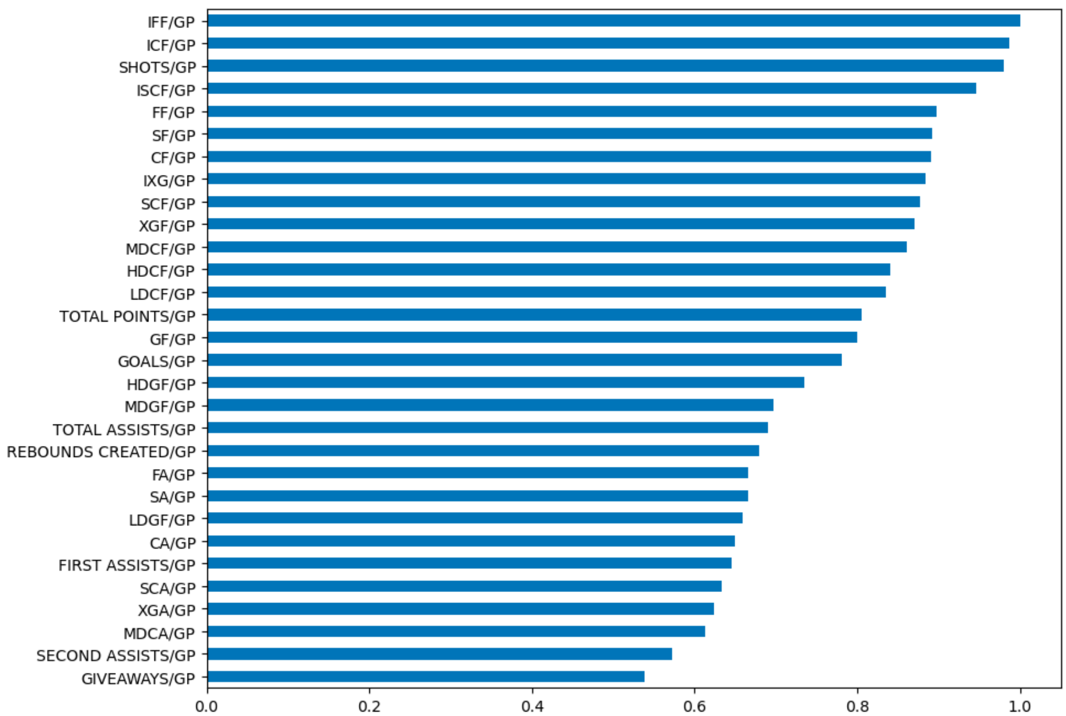
The immediate thing to note here is that the immediate results of our correlational analysis for wingers look nearly identical to that of the defence. The first seven features with the highest correlation with our labels are all the same, permuted. The other thing to note is how many of these features are strongly positively correlated with our targets. Since we already understand the collinearity between the top seven features, we dug into a few other features we have yet to explore.

Given the trend that we witnessed with such a high degree of correlation between all of the features, at this point, we abandoned correlational analysis. We decided to continue with a new approach for feature identification: recursive feature elimination while training. Out of interest to the viewer, we have included some additional correlational charts but will refrain from commenting on them.

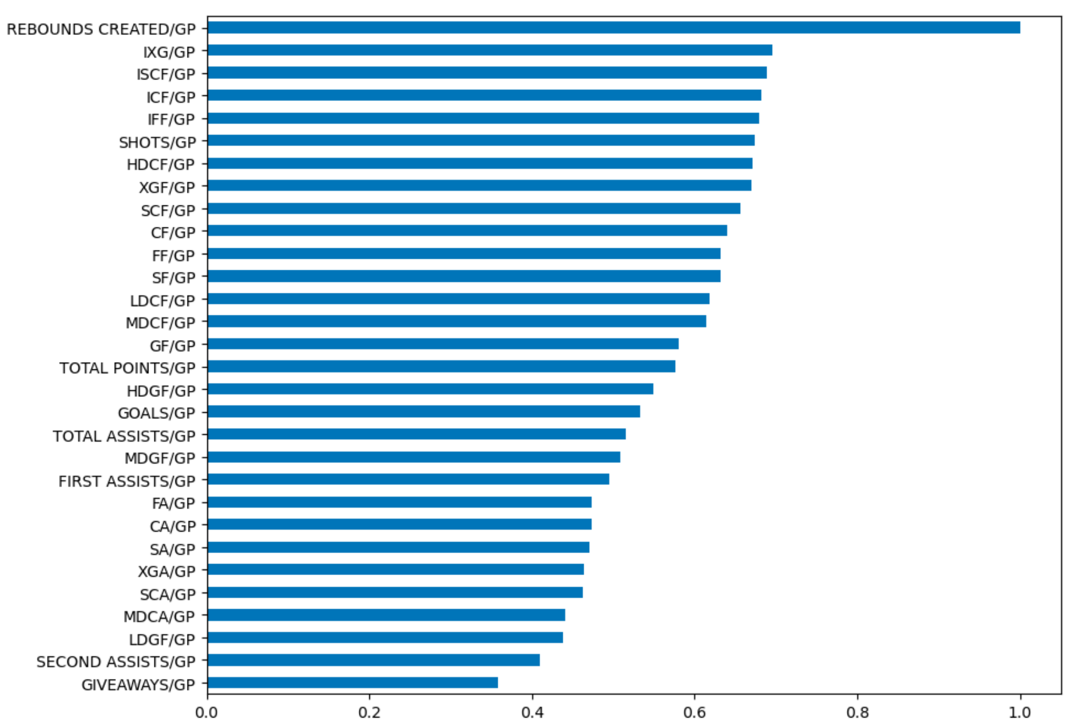
Total Points/GP



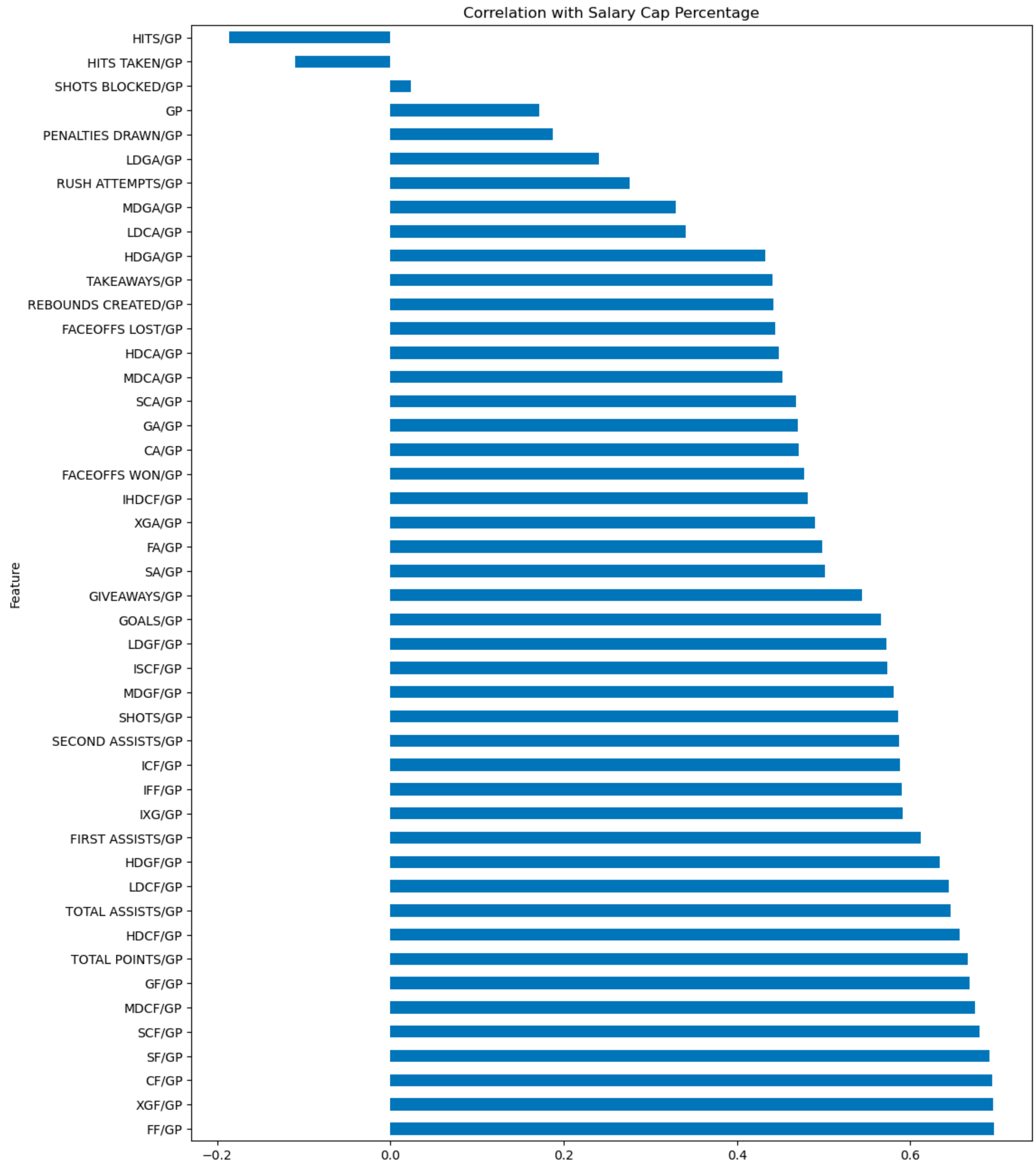
IFF/GP



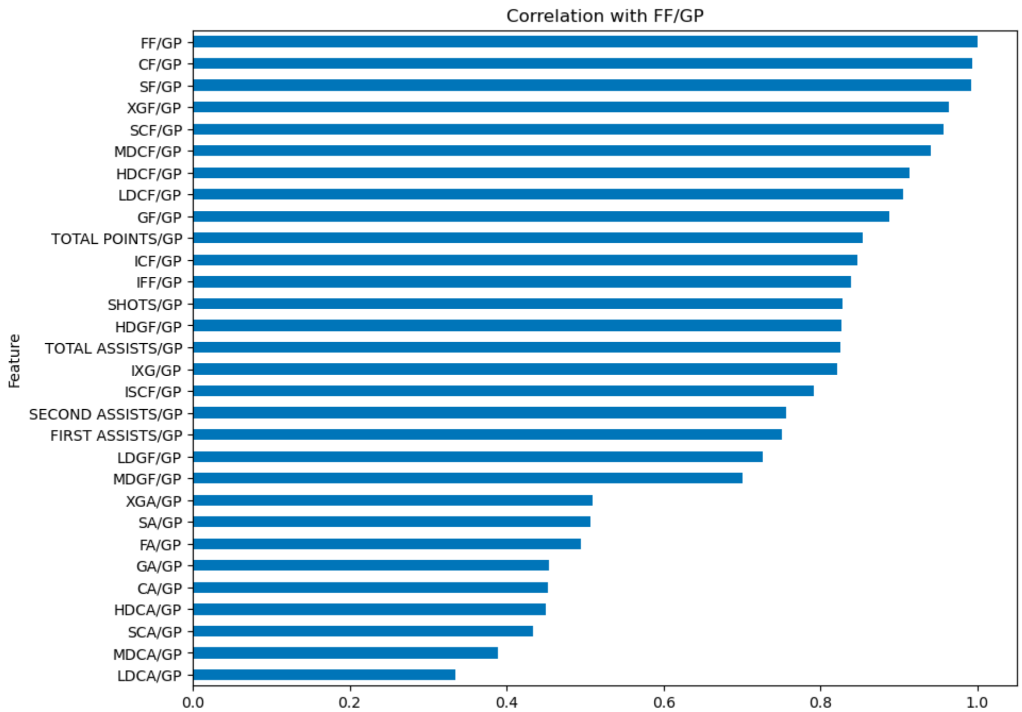
Rebounds Created/GP



Correlational Analysis (Centers)

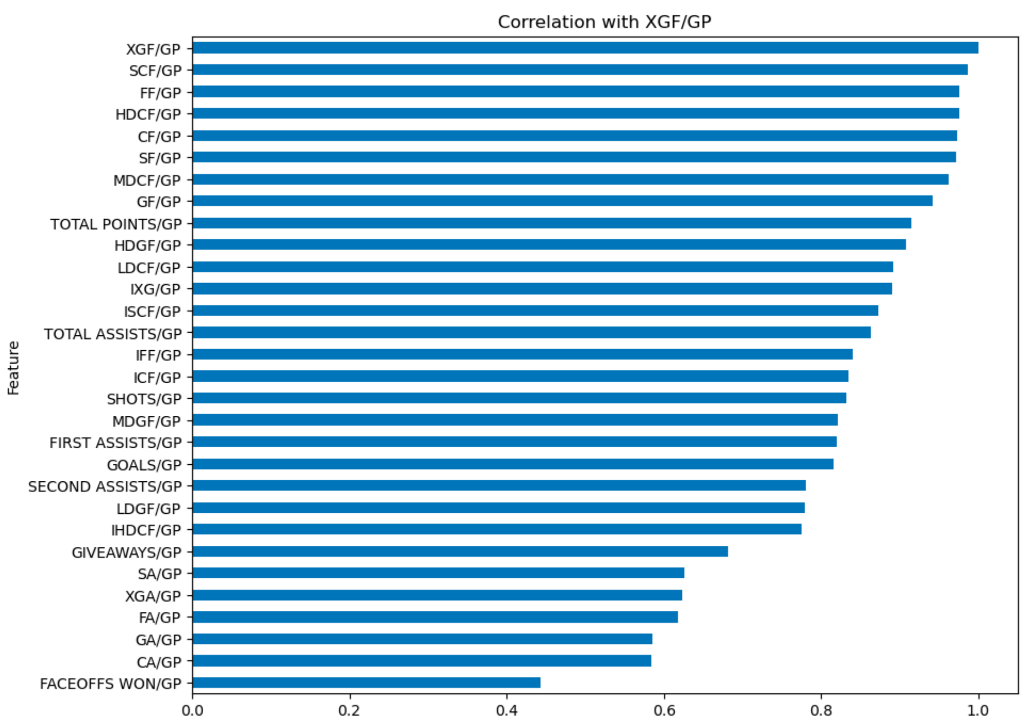


FF/GP



Correlation with FF/GP
Feature
FF/GP
CF/GP
SF/GP
XGF/GP
SCF/GP
MDCF/GP
HDCF/GP
LDCF/GP
GF/GP
TOTAL POINTS/GP
ICF/GP
IFF/GP
SHOTS/GP
HDGF/GP
TOTAL ASSISTS/GP
IXG/GP
ISCF/GP
SECOND ASSISTS/GP
FIRST ASSISTS/GP
LDGF/GP
MDGF/GP
XGA/GP
SA/GP
FA/GP
GA/GP
CA/GP
HDCA/GP
SCA/GP
MDCA/GP
LDCA/GP
0.0
0.2
0.4
0.6
0.8
1.0

XGF/GP



Correlation with XGF/GP
Feature
XGF/GP
SCF/GP
FF/GP
HDCF/GP
CF/GP
SF/GP
MDCF/GP
GF/GP
TOTAL POINTS/GP
HDGF/GP
LDCF/GP
IXG/GP
ISCF/GP
TOTAL ASSISTS/GP
IFF/GP
ICF/GP
SHOTS/GP
MDGF/GP
FIRST ASSISTS/GP
GOALS/GP
SECOND ASSISTS/GP
LDGF/GP
IHDCF/GP
GIVEAWAYS/GP
SA/GP
XGA/GP
FA/GP
GA/GP
CA/GP
FACEOFFS WON/GP
0.0
0.2
0.4
0.6
0.8
1.0

Model Training and Feature Selection

Given our features' high dimensionality and multicollinearity, we decided to leverage tree-based models, which are known to handle this scenario well. Specifically, we chose to use XGBoost (Extreme Gradient Boosting) Regression, a powerful and efficient implementation of gradient-boosted decision trees.
The following sections apply to each of the three models developed.

Feature Selection

Feature selection was identified through recursive feature elimination with cross-validation. Cross-validation was set at 5.

Parameter Selection

Parameter selection was done using grid-search with the following parameters:

- Number of estimators: [500, 1000, 1500]
- Max depth: [5, 10, 15]
- Lambda Regularizer: [.1, 1, 10]
- Learning Rate (eta): [0.01, 0.1, 0.3]

Due to the run time of grid-search, it was only executed on the model for defensive players. The optimal parameters found were as follows:

- Number of estimators: 500
- Max depth: 5
- Lambda Regularizer: 1
- Learning Rate (eta): 0.01

It is important to mention that these were not the final parameters used; however, they offered an excellent starting point for model assessment.

Model Performance Metrics and Parameters

Centers

Training performance score: -0.020667422528004353
Testing performance score: - 0.0969631894712717

Mean Squared Error: 0.00048072938974367055

Parameters:

- Number of estimators: 1000
- Max depth: 5
- Lambda Regularizer: 2
- Learning Rate (eta): 0.01

Wingers

Training performance score: -0.6332180983027547
Testing performance score: -0.5635567300526567

Mean Square Error: 0.0004629816415939871

Parameters:

- Number of estimators: 1000
- Max depth: 5
- Lambda Regularizer: 2
- Learning Rate (eta): 0.01

Defence

Training performance score: -1.6686577385274926
Testing performance score: -1.693932693882489

Mean Squared Error: 0.0005237915819316516

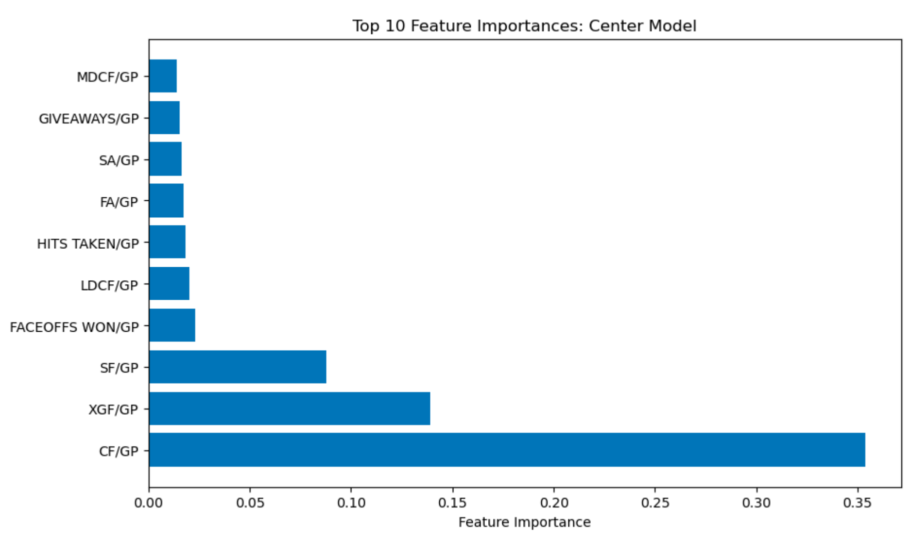
Parameters:

- Number of estimators: 500
- Max depth: 5
- Lambda Regularizer: 0
- Learning Rate (eta): 0.01

Model Analysis and Evaluation

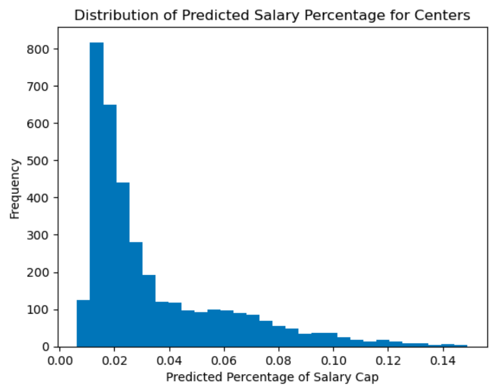
Centers

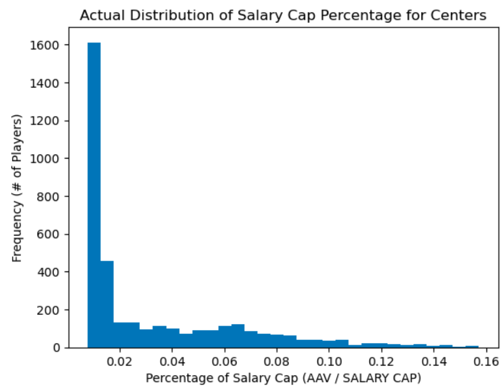
Let us start by looking at the most important features the model found.



Understanding the model's preference for these features is not hard to grasp. Given our correlational analysis done earlier, we know that a player's corsi for per game played (CF/GP) was one of the most highly correlated features with our target labels. Similarly, we should expect to see expected goals for per Game Played (XGF/GP) to be an impactful variable, given that goals are highly desirable, so to expect a lot of them is a good thing. An interesting metric on this list is faceoffs won/GP, given that one of the primary responsibilities of a center is to take faceoffs. Mind you, it is somewhat surprising to see it so high on the list.

Another way of evaluating the performance of our model is to compare the actual distribution of center cap hits relative to our predictions.

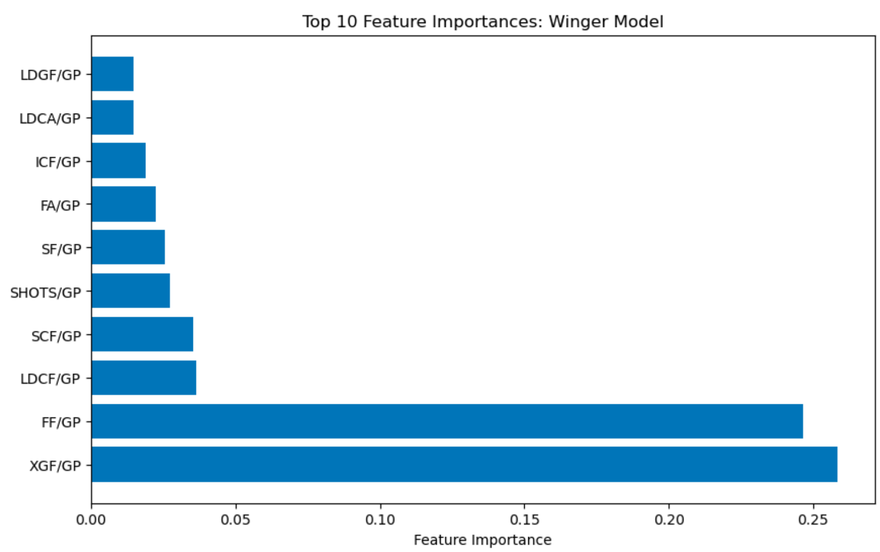




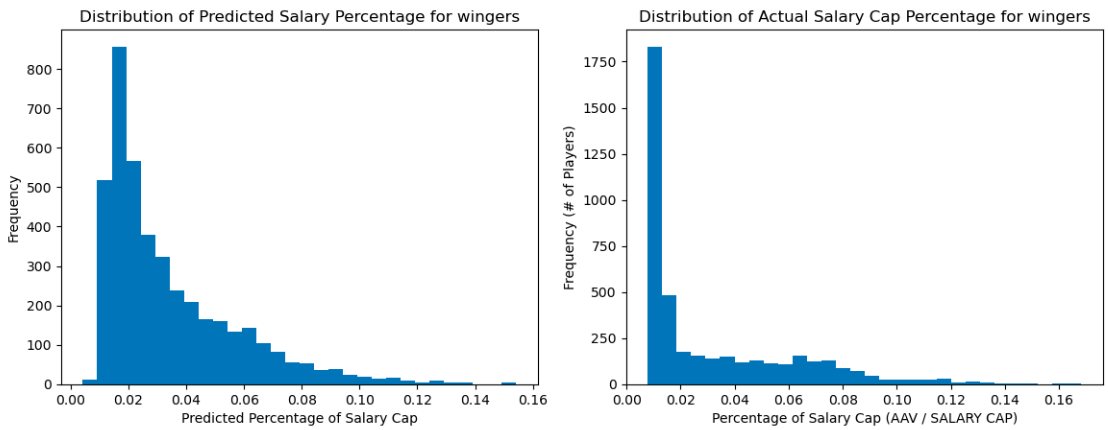
Our predictions' shape, results and range emulate the actual distribution with significant accuracy. A sharp eye will notice that the actual distribution experiences a slight peak right around 0.06, which the distribution of the model mirrors.

Wingers

We can follow a similar approach for analyzing the winger's model. The features chosen by the model as being of primary importance are unsurprising.

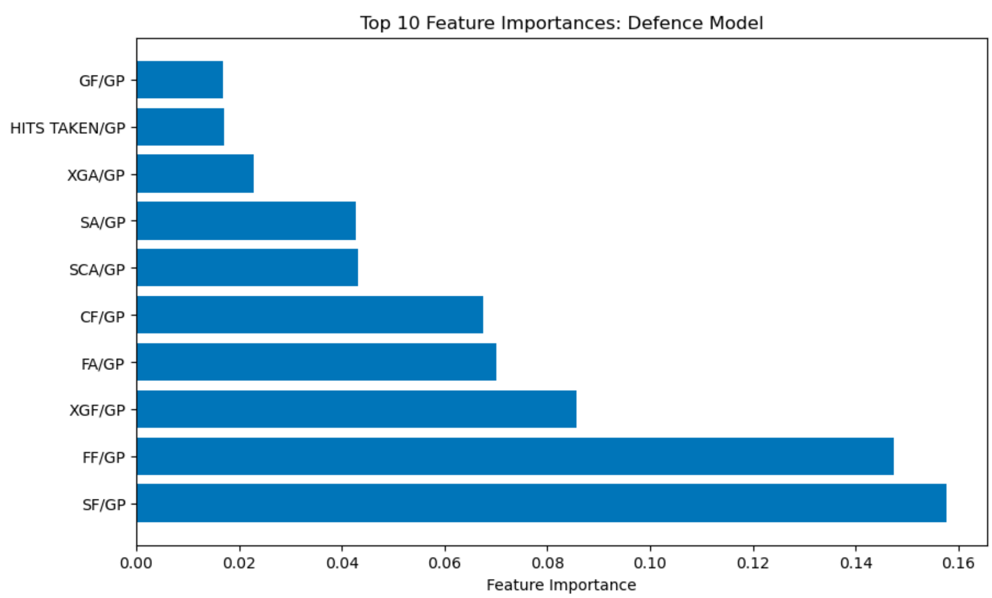


The two features it finds most important, Fenwick for per game and expected goals per game, are two features our correlational analysis would have put right near the top of the list. Perhaps of more interest then is the feature that comes right after: LDCF/GP, or low danger chances for per game. As it sounds, this metric tracks the number of times the player's team can produce a scoring chance, given that the play is considered low-danger for the opposing team. In other words, it tracks a player's ability to produce a scoring chance, given that the probabilities show they shouldn't get one. This feature stands out as one that intuitively makes a lot of sense despite not being a data scientist's first choice, even with a healthy amount of domain knowledge.

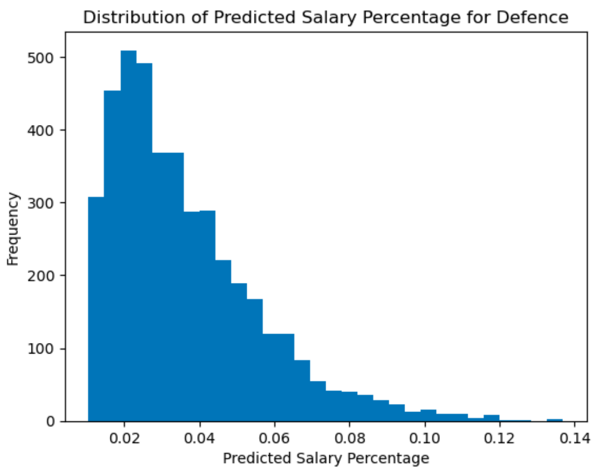


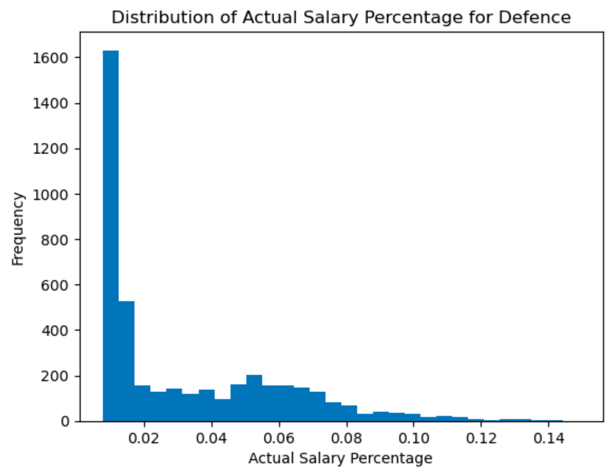
Regarding the comparison between distributions, we see a slightly larger deviation relative to the center's model. The peak occurring just after .06 is present in the predictions' distribution but is considerably more muted. Regarding the overall shape and range, however, the model seems to be acting reasonably.

Defence



The defensive model's feature importance's suggests that, given the features available for training, predicting the value of a defensive player is considerably more nuanced than predicting the value of wingers and defensemen. Nevertheless, we see some familiar friends on this list, including FF/GP, XGF/GP, and CF/GP. On the contrary, however, we have some new friends too, including the introduction of some primarily defence-related statistics, including expected goals against per game played (XGA/GP), scoring chances against per game (SCA/GP), shots against per game (SA/GP), and Fenwick against per game (FA/GP). The defensive model recognizes the value in defensive players both scoring and *not getting scored against.* This is an excellent thing to be seeing.





Concerning the distributions, the shape of our defensive model's predictions reflects the poorer performance attained during training (given by a higher MSE relative to the two other models). However, the range of predictions makes sense, and if you squint, the shape isn't too bad.

Absolute Mean Errors

Given that the mean squared errors are all around 0.0005, we can square this value and consider the absolute mean error instead, which is approximately 0.0224. This value represents the mean absolute error across all predictions and calculates to about 2.24%. In other words, the model over or underpredicts a player's cap hit on average by about 2.24% of the proportion of the salary cap that it costs to roster them.

Entry-level Contracts and Deflated Salaries

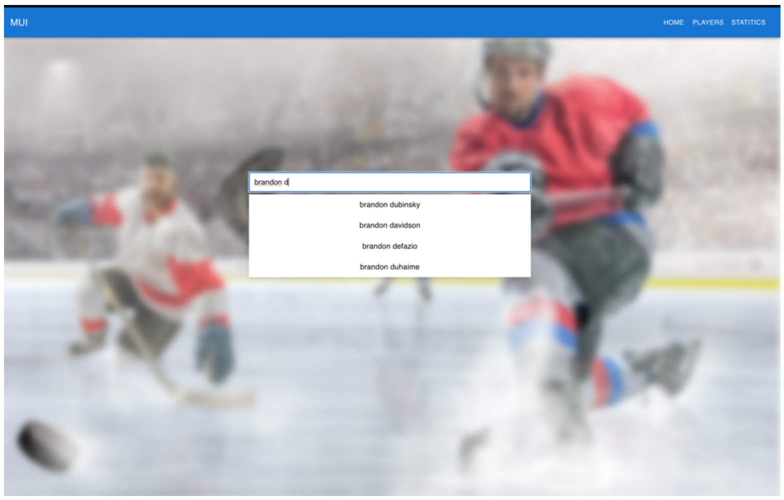
One final point about the distributions: a keen eye will notice that all three actual distributions have a mode at the leftmost bin. In other words, most of the players in the league are making the league minimum. This is a by-product of entry-level contracts and young players just entering the league having very little power to negotiate. The result, then, is that quite a few players in the league perform at a level above the league minimum despite that being what they are paid.

When we consider what this means concerning our predictions, it makes sense that we see some thickness in the middle of the predictive distributions that are not as present in the actual predictions.

Data Product

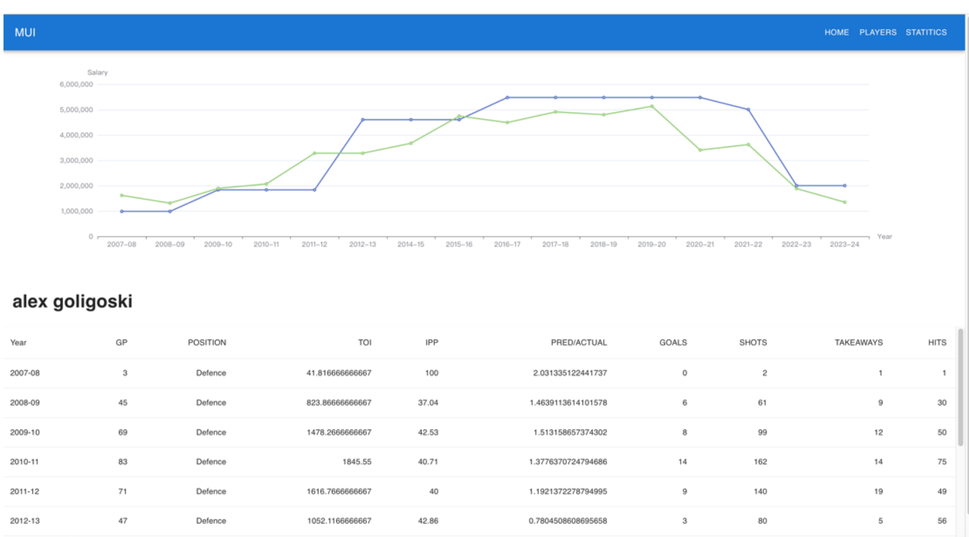
General

To make the fruits of the labour of this endeavour easily accessible, we decided to create a web-based interactive environment. The environment is searchable by either team or player and is designed to lead the user to specific player pages. Each page has both performance and cap hit statistics for each year, as well as a line graph displaying the player's actual cap hit, as well as their predicted cap hit, over the course of their career.



Player Pages

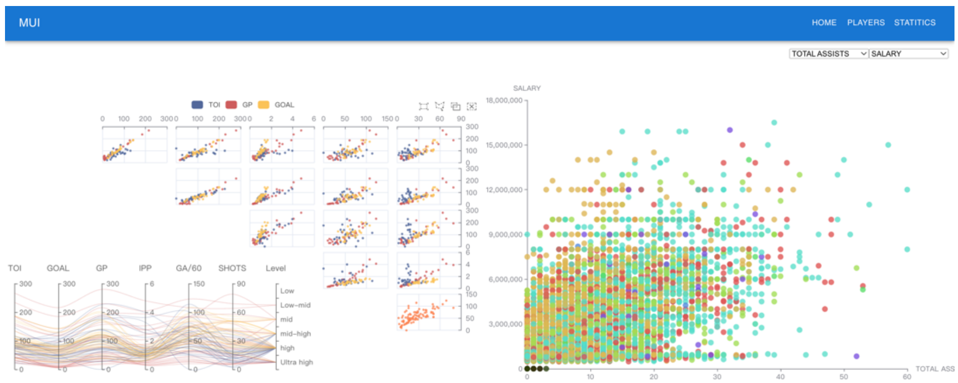
Below, we can see an example of the player page for Alex Goligoski, who happens to be included in every year of our data. The green line in the plot shows his predicted cap hit, whereas the blue line represents his actual cap hit.



Year	GP	POSITION	TOI	IPP	PRED/ACTUAL	GOALS	SHOTS	TAKEAWAYS	HITS
2007-08	3	Defence	41.816666666667	100	2.031335122441737	0	2	1	1
2008-09	45	Defence	823.86666666667	37.04	1.4639113614101578	6	61	9	30
2009-10	69	Defence	1478.2666666667	42.53	1.513158657374302	8	99	12	50
2010-11	83	Defence	1845.55	40.71	1.3776370724794686	14	162	14	75
2011-12	71	Defence	1616.7666666667	40	1.1921372278794995	9	140	19	49
2012-13	47	Defence	1052.1166666667	42.86	0.7804508608695658	3	80	5	56

League-wide Statistics

The statistics page of our interactive web environment allows the user to graph any two variables in our database against one another. In the image below, we see the graph of Total Assists in relation to Salary.



Lessons Learned

The Challenge of Multicollinearity

When embarking on this project, our team did not anticipate the challenge imposed during correlational analysis. While the analysis did prove to be insightful to some degree, it was considerably less helpful than was originally anticipated and mostly reinforced our understanding of the models after they had been trained, as opposed to helping us figure out what features specifically to use during training itself.

Cleaning, and Cleaning, and Cleaning, and …

Those just embarking on their data science journey are often told—and warned—of the amount of time spent on the data cleaning process. While this warning is one thing, it is altogether something else to experience it. Despite the fact that we felt our data had been procured in a rather clean manner, a tremendous amount of time went into the cleaning process, much of which was unanticipated.

Limitations of Domain Knowledge

Given that one of our group members—Michael Kuby—came into this project with a deep knowledge and love for the NHL, model training proved considerably more challenging than was anticipated. Training models solely on the features one might anticipate to predict player cap hit proved much more fruitless than had been anticipated, which was a rather large surprise.

Non-linearity

When embarking on this project, it was idealized that a linear regression model could serve our purpose. However, the training process seemed to unveil the fact that linear assumptions were a limiting factor. As a general observation, we found that a linear model could do a reasonable job of predicting the average cap hit of a given player but performed quite poorly for a variety of edge cases. Players that command extremely high salaries, for example, were generally seen to be well over-paid by a linear model, which was to some degree accounted for by using tree models that allow for non-linearity.

Summary

This report has endeavoured to construct a robust system for the valuation of NHL players, aiming to demystify the factors influencing and predicting player cap hits. We have drawn from extensive datasets obtained from CapFriendly.com and NaturalStatTrick.com, harnessing performance metrics and salary data to forge a predictive model that stands to serve general managers, player agents, and players themselves in the nuanced arena of contract negotiation.

We initiated our project by evaluating existing models, such as The Athletic Player Cards. While innovative, these models fall short of practical applicability due to their complex and sometimes unrealistic valuation methods. Our model seeks to bridge this gap, offering a tool grounded in the realities of the NHL's salary cap system.

The core of our methodology is a data science pipeline that meticulously extracts, transforms, and analyzes data through a multi-phased process. Recognizing the limitations of purely quantitative analysis—namely, its inability to encapsulate intangible qualities such as leadership and team spirit—we have nevertheless striven to quantify player worth through objective metrics.

Employing domain knowledge alongside comprehensive correlational analysis, we have delineated performance indicators across positions—defence, wingers, and centers—culminating in the strategic employment of XGBoost Regression for model training. This approach not only accommodates the high dimensionality and multicollinearity inherent in our data but also enables us to glean nuanced insights into the constituents of player value.

Our models have been rigorously tested and refined, yielding predictions with a mean absolute error of approximately 2.24% of the salary cap—a testament to the model's precision. The utility of these findings is encapsulated in our interactive web-based environment, which offers accessible, user-oriented insights into player valuations and projections.

In conclusion, this report documents our journey and the lessons therein. From the pivotal role of data cleaning to the pitfalls of relying solely on domain knowledge and the imperative for non-linear modelling approaches, we have charted a course through complex analytical terrain. The culmination of our work is a model that not only enhances transparency in player valuation but also reflects the intricate interplay between performance metrics and market realities.