Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
May 2, 2024
A Regional Conference of the Society for Industrial and Systems Engineering

Identifying Undervalued "Playmakers" in the NHL Utilizing Passing Metrics and AAV

Jacob Felker and James Enos

Department of Systems Engineering,
United States Military Academy,
West Point, New York 10996

Corresponding author's Email: jacob.h.felker.mil@army.mil

Author Note: Cadet Jacob Felker is a member of the Army hockey team as well as a senior (Firstie) in the Department of Systems Engineering at the United States Military Academy (USMA). CDT Felker is originally from Omaha, NE. This paper is written as a component for one of the requirements of graduating with honors in Systems Engineering. Thank you to COL Enos, Mr. Kloo, and LTC Geisler for the assistance in developing R Code to clean the initial data and analyze aspects of the data. Thank you to COL Enos for advising me throughout the entirety of this independent research study at USMA.

Abstract: This research aims to identify undervalued "playmakers" in the NHL by comparing their average annual value (AAV) with passing metrics derived from Corey Sznajder's "All Three Zones" Project. These undervalued "playmakers" offer high value to teams but may be undervalued due to their total number of points throughout the season. The data utilized in this research is from the 2022-23 NHL season, however, the dataset only accounted for 966 NHL games. Data analysis offers hockey scouts, coaches, and general managers a unique lens to identify valuable players. K-means clustering was utilized to focus specifically on the "playmaker" cluster to compare "playmakers" based off total points, AAV, scoring chance ratio, and shot on goal ratio. The research found numerous players who have similar, and at times, better scoring chance and shot on goal ratios, but have a significantly lower AAV.

Keywords: Competitive Advantage, Statistical Metrics, Data Analysis, Scoring Chance Ratio, Shot on Goal Ratio, AAV

1. Introduction

As the NHL continues to develop and become more competitive, the analytics behind the game are also becoming more complex. Analytics seeks to perform large-scale analysis by quantifying various levels of data to provide players, teams, and organizations with different perspectives that traditional hockey minds cannot see (Shea et al., 2017). Analytics in the game of hockey is rapidly evolving and becoming increasingly popular in professional organizations around the world (Alamar, 2023). Sports analysts take current and historical data to develop predictive analytics through advanced statistics, data management, data visualization, and several other statistical strategies (Alamar, 2023). Statistics seek to differentiate and compare teams' performance, and players skills and abilities.

For decades, hockey teams have relied on the basic statistics such as goals, assists, hits, and plus-minus to compare players. Up until the early 2000s, the plus-minus statistic was a metric used to analyze a player's contribution towards their team (Nandakumar, 2019). If a player had a positive plus-minus rating, they were on the ice for more goals scored by their team than goals scored against their team. On the other hand, players with a negative plus-minus rating were on the ice for more goals against their team then goals for their team. While this measure has its significance, one issue that arose from the plus-minus metric was differentiating an individual players performance from their teammate's performance (Nandakumar, 2019). For instance, if a player is playing on a bad team, their plus-minus has a greater chance of being negative versus a player who is on a very good team. Additionally, plus-minus does not account for the entirety of a hockey player's game. Since the plus-minus metric is not a complete metric of a player's ability on the ice, statisticians and hockey analysts are transitioning into shot based statistical metrics (Nandakumar, 2019). These shot based metrics are more beneficial in predicting the value that a hockey player brings to their team. Data analysts and researchers continue to seek and develop predictive analytics to gain a competitive advantage.

This research builds on previous hockey analytics by utilizing passing metrics and AAV to gain a unique insight into a player's value (OpenAI, 2023). To provide a thorough overview for the rest of the paper, the literature review will discuss a range of models and statistical measures currently used in analytics (OpenAI, 2023). The methodology will break down the data collection, and how the data was filtered and cleaned to effectively utilize k-means clustering. The methodology will be followed by the results of the research, recommendations, and conclusion.

2. Literature Review

Statistics in the hockey world are utilized to gain a competitive advantage. As technology continues to modernize and develop, hockey analysts and researchers are transitioning from basic statistics to modern, more complex models. These complex models are seeking to find aspects of the game that are hidden from a scout or coach's eye. Analysts are developing complex models and metrics that can accurately predict players and teams' success. These newly developed metrics and models are beginning to change the way hockey teams view the game. NHL organizations are beginning to hire analysts to develop new metrics and models that help gain a competitive advantage across the league. The benefit of having hockey analysts attached to NHL teams is their ability to break down a team's performance and gain insight on players and teams' tendencies and abilities.

2.1 Fenwick and Corsi Rating

The Fenwick Rating is a statistical measure commonly used in the hockey world. The Fenwick equation is displayed below in equation 1.

$$Fenwick = (the\ number\ of\ shots\ on\ goal\ for\ your\ team + the\ number\ of\ missed\ shots\ on\ goal\ for\ your\ team) - (the\ number\ of\ shots\ on\ goal\ against\ your\ team + the\ number\ of\ missed\ shots\ on\ goal\ against\ your\ team) \quad (1)$$

The Corsi Rating is another statistical measure that helps provide insight into a player's performance. The Corsi Rating builds on the Fenwick Rating by accounting for blocked shots. The Corsi Rating is displayed in equation 2.

$$Corsi = (the\ number\ of\ shots\ on\ goal\ for\ your\ team + the\ number\ of\ missed\ shots\ on\ goal\ for\ your\ team + the\ number\ of\ blocked\ shots\ for\ your\ team) - (the\ number\ of\ shots\ on\ goal\ against\ your\ team + the\ number\ of\ missed\ shots\ on\ goal\ against\ your\ team + the\ number\ of\ missed\ blocked\ shots\ against\ your\ team) \quad (2)$$

Hockey analysts use both the Fenwick Rating and the Corsi Rating to overcome the difficulty of analyzing individual player performances due to low scoring rates in hockey (Krueger, 2021). The benefit of using the Fenwick or Corsi Rating is that there are more shots during a 60-minute game compared to the number of goals during a 60-minute game (Krueger, 2021). These statistics are great metrics for predicting goal differential (Krueger, 2021). While these rating have shown to be a great metric, they still do not account for everything that contributes to player's performance. Thus, many analysts utilize a combination of both traditional goals statistics and modern metrics to understand an individual players overall performance.

2.2 Weighted Shots Model

The weighted shots model by Mason Krueger utilized "shot-zones" with varying weights for each zone (Krueger, 2021). By using weighted values for different areas of the offensive zone, Krueger was able to add a shot value for different shots. In this model, rebound shots were weighted heavier because rebounds have a greater likelihood of leading to a goal compared to other shot attempts (Krueger, 2021). Rebounds are shots on net after the goalie makes the original save and tend to be within a few feet from the net.

$$Fenwick = \sum [Zone1A_M(W_{1A}) + Zone1A_{MR}(W_{1A} + W_R)] + [Zone1A_{SoG}(W_{1A}) + Zone1A_{SoGR}(W_{1A} + W_R)] + \cdots \quad (3)$$

In the weighted Fenwick equation above, Krueger calculated the weighted summation for the product of shots on goal for the team and missed shots for the team, subtracted by the product of shots on goal against the team and missed shots on goal against the team (Krueger, 2021). In the equation above, $\mathrm{Zone1A_M}$ represents the total number of missed shots in Zone 1A (Krueger, 2021). $\mathrm{W_{1A}}$ stands for the weight that Zone 1A is classified as (Krueger, 2021). $\mathrm{Zone1A_{MR}}$ represents the Zone 1A missed rebound shots (Krueger, 2021). $\mathrm{W_R}$ stands for the additional weighted value of rebound shots (Krueger, 2021). $\mathrm{Zone1A_{SoG}}$ represents the total number of Zone 1A shots on goal (Krueger, 2021). Lastly, $\mathrm{Zone1A_{SoGR}}$ represents the shots on goal that were rebounds in Zone 1A (Krueger, 2021).

$$Corsi = \sum [Zone1A_M(W_{1A}) + Zone1A_{MR}(W_{1A} + W_R)] + [Zone1A_B(W_{1A}) + Zone1A_{BR}(W_{1A} + W_R)] + [Zone1A_{SoG}(W_{1A}) + Zone1A_{SoGR}(W_{1A} + W_R)] + \cdots \quad (4)$$

ISBN: 97819384962-4-0

In the weighted Corsi equation above, Krueger utilized an equation very similar to the weighted Fenwick equation, however, Krueger also added in blocked shots for the team and blocked shots against the team (Krueger, 2021). Krueger iterated these equations for every weighted zone in the rink (Krueger, 2021). This allowed Krueger to gain an understanding of how many scoring chances Army West Point generated in high priority scoring areas compared to weaker scoring areas (Krueger, 2021). The weighted shots model by Krueger is very flexible and reliable. The same process can be applied to every Army hockey game and every hockey organization (Krueger, 2021). Additionally, the equations used in the model are flexible due to the weight factors applied. The weight factors can easily be changed to various weights to gain key insight to different areas of the ice (Krueger, 2021).

2.3 Goals versus Threshold (GVT) Model and Goals versus Salary (GVS) Model

The Goals versus Threshold (GVT) model measures an individual players contribution regarding goals relative to how the team would do with a replacement player (Vollman, 2016). The GVT model developed a metric to provide teams and organizations with a measure that could compare current players contributions in terms of goals to possible trade opportunities (Vollman, 2016). Vollman then created a more complex model that took the GVT model and compared values with the players' salaries (Vollman, 2016). The GVS model builds off the GVT model by accounting for a player's cap hit. The cap hit is acquired by dividing a player's total salary and signing bonuses by their contract's length (Sportsnet, 2024). As seen in equation 5, GVS represents the number of goals an individual player scored or prevented relative to a player with a similar cap hit (Vollman, 2016).

$$GVS\ (Goals\ versus\ Salary) = GVT - (cap\ hit - league\ minimum)x\ 3 \tag{5}$$

To calculate the GVS, multiply the difference between the individual players cap hit and the league minimum cap hit. Then multiply that value by three goals per million dollars (Vollman, 2016). Then subtract the GVT value from the value calculated after multiplying the difference by three. Previous models used a 3-1-1 rule which represented 3 goals, on average, got a team at least one point in the standings, but costed around one million dollars (Vollman, 2016). The GVS model is a more flexible and dynamic model. Even though the 3-1-1 model utilizes fixed numbers, the GVS model uses numbers that continuously get updated and changed (Vollman, 2016). In terms of contracts, Vollman found that younger players are more likely to meet or exceed an organizations expectations compared to older players (Vollman, 2016). It is important to note, however, Vollman saw a correlation between a players age and the value of their contract. Statistical trend lines displayed that as an older players resigned and obtained a deal, the value of their contract rapidly decreased (Vollman, 2016).

2.4 Passing and Pressure Metrics

Due to the development of puck and player tracking (PPT) systems, new ways to quantify players' skills and abilities have been developed (Radke et al.). The Pass Overtaking Model utilizes shaded regions that surround the passer and receiver (Radke et al.). The shaded region, or lane from the passer to receiver, is scaled to the nearest opposing team's player. From there, the values are combined into different metrics such as overtake total (OVT), beaten total (BTT), and passing plus-minus (PPM). In hockey, OVT refers to the total number of times you take the puck from an opposing player, whereas BTT refers to the numbers of times the opposing team take the puck from you.

$$PPM = OVT - BTT \tag{6}$$

PPM helps derive insight into whether a player overtakes other players more than getting overtaken (Radke et al.). A specific limitation to this model is that players who rarely overtake opponents and rarely get overtaken have the same value as players who routinely overtake opponents but get routinely overtaken (Radke et al.). Additionally, this model does not consider unique factors such as coaching style, manpower, or play location (Radke et al.). In the PPM model, Radke concluded that forwards tend to complete shorter passes and overtake more opponents than defenseman (Radke et al.).

3. Methodology

By utilizing passing data collected from a subset of NHL games in the 2022-23 season, statistical analysis and clustering were performed to discern various player types (OpenAI, 2023). Following statistical analysis and clustering, each players AAV was compared within cluster 1, characterized as "playmaker" (OpenAI,2023).

3.1 Data Collection and Breakdown

The data used in this research was obtained from Corey Sznajder's "All Three Zones" Project and Cap Friendly. The "All Three Zones" project provided detailed microstats from NHL games. When cleaning the datasets, the raw data seemed to be manually inputted by Sznajder and his team for 966 NHL games during the 2022-23 NHL season. Thus, an assumption made throughout this research was that the trends observed within the dataset mirror those from the unaccounted-for games. Assuming Sznajder manually inputted every item into the raw data, the first step was to clean and verify the data before conducting statistical analysis. While the data in this research was verified for four games within the dataset, it was not verified for every game. The first game verified was the Tampa Bay Lightning vs New York Rangers game on October 11, 2022. When comparing the box scores statistics with Sznajder's dataset, the final score of the game was correct, along with the correct jersey numbers of the players who tallied a goal or an assist during the game. Moreover, the total entries (scoring opportunities) for each team in Sznajder's dataset was more than the total shots for each team. Thus, it was reasonable to assume that every shot on goal was properly tracked and accounted for in the dataset. The Colombus Blue Jackets vs Carolina Hurricanes game on October 12, 2022, was verified by watching the first period to confirm the data matched the dataset by Sznadjer. This process was repeated for the 2nd period of the San Jose Sharks vs New Jersey Devils game on October 22, 2022, and the 3rd period of the San Jose Sharks vs Pittsburgh Penguins game on January 28, 2023.

Overall, while subjective, the data revealed to be approximately 90% accurate. On average, each period reviewed had 35-40 inputted scoring chances. However, on average, there were 3-4 missed opportunities per period. These missed opportunities are due to the subjectivity of whether a scoring chance was generated or not. For example, a pass that creates a scoring chance is subjective to the individual collecting the data. The "house" as displayed in Figure 1, is not a definite spot from a viewer's eye. Thus, a pass into the house creating a scoring chance could be depicted differently for different viewers. In hockey, a scoring chance occurs when the team with puck possession scores, the shot missed the net, or the shot was saved by the goaltender (OpenAI, 2023). Additionally, the dataset did not account for goals called back. In one of the games verified, one of the players created a prime scoring chance leading to a goal, but the goal got called back. In this situation, the dataset did not account for the scoring chance because the goal got called back. Here, one could make the argument that the scoring chance should still be accounted for even if the goal was called off. Overall, by verifying and making assumptions for the data collected, one can assume with 90% accuracy that the data is correct. The microstats that Sznajder tracked were:

a. Period: Period of the event. Used 4, 5, 6, etc. for overtime periods.
b. Strength: 5v5, 5v4, etc. All situations were preloaded into the cell.
c. Team: The team that generated the shot attempt. Abbreviations are loaded into the cell.
d. Shooter: Jersey number of the shooter. On tip/deflected plays by offensive player, last player to touch the puck is the shooter.
e. Shot Type: One-Timer, "o"; Slap Shot, "s", Wrist/Snapshot, "w", Back hand, "b", Wraparound, "a", Tip, "t", Rebound "r".
f. A1: Player number of primary passer.
g. A2: Player number of secondary passer.
h. A1 Zone: Zone on the ice in which the primary pass originated from.
i. A2 Zone: Zone on the ice in which the secondary pass originated from.
j. SC: Scoring Chance Passes. Primary pass leading to a shot attempt from within the home plate area.
k. SOG: Pass leading to a Shot on Goal. Accounted for if the shot forces a save or results in a goal.
l. Oddman: Situations during the game when a team gets a 2 on 1, 3 on 2, etc.…
m. G: Goal was scored.

The primary purpose of this research is to utilize AAV and passing metrics from a variety of NHL games to identify undervalued "playmakers" in the NHL. Regarding this research, "playmakers" are the players who create a shot on goal or scoring chance by passing to one of their teammates but generate a low number of total points. The AAV is found by taking a player's total contract not accounting for signing bonuses and dividing it by the length of the contract. For example, if a player signs a $3 million contract over a 3-year period, their AAV is $1 million. The AAV is how teams and organization negotiate the value of players (Tollerud, 2017).

For the scoring chance (SC) metric, Sznajder marked any pass that led to a shot attempt within the home plate area. The home plate area is also referred to as "The House" by other hockey analysts. Figure 1 displays a visual representation of where the home plate area is located within the offensive zone.

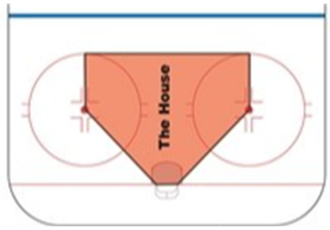

Figure 1. Visual Representation of "The House." (Evenson, 2023)

To start, all 966 different excel sheets (each game collected by Sznajder) were transferred into one single dataset in R Studio. All empty ("NA") values for the Team, A1, and A2 variables were filtered out of the dataset. Using variables A1 and Team, a unique player identifier was created for each row as shown in Table 1.

Table 1. Depiction of the initial dataset in R Studio after scoping down to only 5v5 and creating the new variables with their respective player identifier.

Player-Identifier	Assists	Goals	Scoring Chance Ratio	Shot on Goal Ratio
ANA-11	7	5	0.3294	0.5882
ANA-14	2	6	0.3415	0.6341

Given the absence of player names in the original datasets, an assumption made in this research was that the unique identifier accounted for a single player throughout the entirety of the season. In other words, this research assumed no player in the data was traded mid-season. New variables were established after creating the player identifier: assists, goals, scoring chance ratio, and shot on goal ratio. In this research, the assists statistic was created by taking the total number of primary passes (A1) leading to goals (G). Thus, only primary assists were accounted for in the research. The goals statistic is strictly the number of times a player made a direct pass leading to a 5v5 goal. To scale down the data, only even strength opportunities (5v5) were accounted for. It is important to scale the data to only even strength opportunities because not everyone plays during odd man situations such as 5v4 or 4v4. The equations for calculating the Scoring Chance Ratio and Shot on Goal Ratio are displayed in equations 7 and 8 below.

$$Scoring\ Chance\ Ratio = \frac{The\ number\ of\ times\ a\ player\ made\ a\ pass\ into\ the\ house\ that\ directly\ lead\ to\ a\ scoring\ chance}{Total\ nmber\ of\ times\ the\ player\ made\ a\ pass\ that\ was\ accounted\ for\ in\ the\ dataset} \tag{7}$$

$$Shot\ on\ Goal\ Ratio = \frac{The\ number\ of\ times\ a\ player\ made\ a\ pass\ that\ directly\ led\ to\ a\ shot\ on\ goal}{Total\ nmber\ of\ times\ the\ player\ made\ a\ pass\ that\ was\ accounted\ for\ in\ the\ dataset} \tag{8}$$

3.2 K-means clustering

The next step in determining undervalued playmakers was utilizing k-means clustering to identify different clusters of players in the data. K-means clustering is an iterative algorithm that seeks to partition the data into distinct clusters that do not overlap (Dabbura, 2018). Since K-means clustering analyzes and compares each data point using distance-based measurements, it is important to scale down the data before utilizing the k-means clustering (Dabbura, 2018). This allows the data to have a mean of 0 and standard deviation of 1 (Dabbura, 2018). After scaling down the data, k-means clustering was used to identify 5 different centers in the data. Each of the five centers represents a different type of player. For the purposes of this analysis, these clusters were named (1) Playmaker, (2) Franchise Player, (3) Energy Player (4) Sniper (5) Power Forward as shown in Table 2.

Table 2. Cluster Output for the Five Different Clusters of Players

Cluster	Assists	Scoring Chance Ratio	Shot on Goal Ratio	Goals
1	-0.132	1.125	1.047	-0.224
2	2.157	0.927	0.336	2.225
3	0.320	0.186	-0.142	0.320
4	-0.526	-0.558	0.380	-0.563

5	-0.576	-0.969	-1.339	-0.483

The players within each cluster were consistent with that respective cluster's player type. For example, the players in cluster 2 ("Franchise Players") were McDavid, Crosby, Ovechkin, Matthews, etc.… whereas cluster 1 ("Playmaker") identified high level playmakers such as Ivan Provorov, Brendan Gallagher, Alex DeBrincat, and Brayden Schenn. After identifying and naming the clusters to a specific player type, the clusters were graphed together. Graphing the clusters was important for visually understanding where each cluster fit within the data and gaining an idea where each player is located within the clusters.

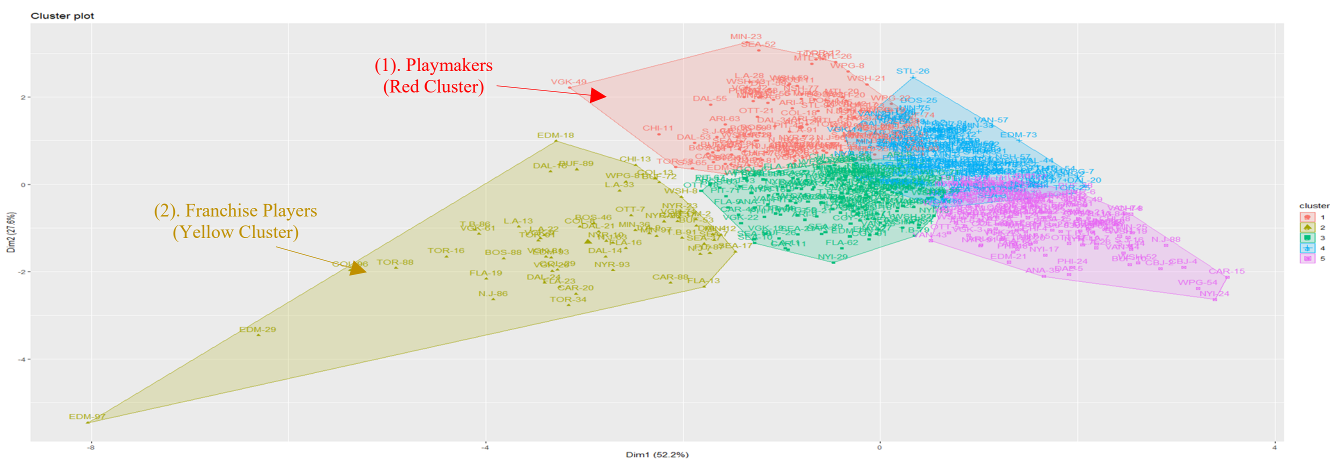

Figure 2. Graph of K-means clustering identifying the five different clusters in the data.

The cluster results in the R Studio file were exported to an excel csv file to filter the data only to the players in cluster 1. The reason the data was filtered to strictly cluster 1 players was because the research is solely focused on players who are contributing a high value to their team in respect to their Scoring Chance Ratio and Shot on Goal Ratio, but lack in the number of goals and assists they produce for their team. Additionally, the dataset was strengthened from exporting the R Studio file back into excel where each playmaker's AAV could be inputted from Cap Friendly. Once the AAV's were collected, each player's total number of games played in the 2022-23 season were added to identify and filter out the players who were traded mid-year. Filtering out players who got traded mid-year enhanced the data by validating each player's unique identifier throughout the entirety of the 2022-23 season.

Table 3. Top 6 Playmakers with the highest scoring chance ratio.

Unique Identifier	Cluster	Player Type	Player Name	Points	Scoring Chance Ratio	Shot on Goal Ratio	AAV	Games Played
L.A-28	1	Playmaker	Jaret Anderson-Dolan	3	0.55	0.73	$750,000	46
MIN-23	1	Playmaker	Marco Rossi	1	0.55	0.82	$1,713,333	19
MTL-20	1	Playmaker	Juraj Slafkovsky	0	0.54	0.62	$4,450,000	39
PHI-55	1	Playmaker	Rasmus Ristolainen	4	0.54	0.69	$5,100,000	74
MTL-56	1	Playmaker	Jesse Ylonen	3	0.53	0.67	$925,000	37
TOR-58	1	Playmaker	Michael Bunting	11	0.52	0.62	$950,000	82

From here, the excel sheet above was exported back into R Studio to plot the data in an interactive graph using the Plotly graphing command. As shown in figure 3, each player within cluster 1 was plotted to compare each player's "Shot on Goal Ratio" and "Scoring Chance Ratio" to their respective AAV. The size and color of each point is consistent with the players AAV. For example, the players with a small blue dot have a small AAV, whereas players with large yellow dot have a large AAV. This helps visually depict which players are undervalued based off the amount of scoring chances they create each game.

4. Results

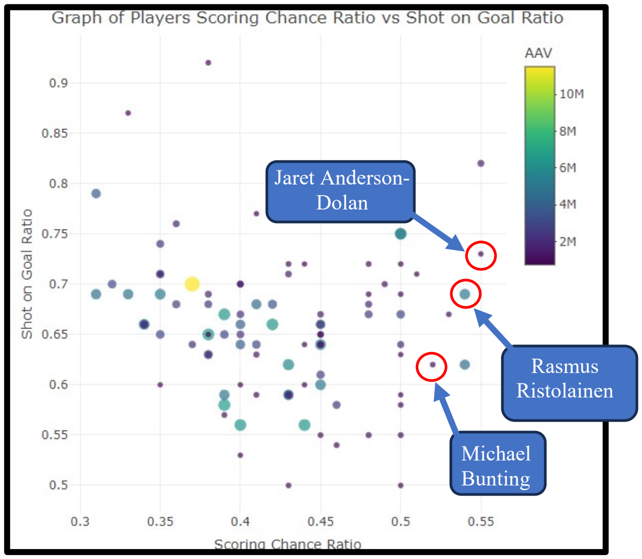

Figure 3. Graph comparing each "Playmakers" Scoring Chance Ratio, Shot on Goal Ratio, and AAV

Analyzing the players within cluster 1 gave insight to specific players who are being undervalued based off newly established passing metrics and their AVV. When analyzing Table 3 and Figure 3 above, players such as Jaret Anderson-Dolan had an AAV of $750,000 with a scoring chance ratio of 0.55, shots on goal ratio of 0.73, and tallied 3 points within the dataset. Additionally, Michael Bunting had an AAV of $950,000 with a scoring chance ratio of 0.52, shots on goal ratio of 0.62, and tallied 11 points for Toronto. These players are undervalued compared to players such as Rasmus Ristolainen. Ristolainen received an AAV five times larger ($5,100,000) than the previous two players but received relatively similar statistics. Ristolainen only tallied 4 points with a scoring chance ratio of 0.54, and shots on goal ratio of 0.69. Players such as Jaret Anderson-Dolan and Michael Bunting are just two of many identified as undervalued based off the scoring chance ratio, shot on goal ratio, points, and AAV. That is, these players bring a high value to their team, but their success is not captured in other statistics and metrics. Thus, this methodology offers another insight to identifying undervalued playmakers.

5. Conclusion

In conclusion, this research proposes a new methodology for identifying undervalued playmakers. The methodology displayed above incorporates a shot on goal ratio, scoring chance ratio, total points, and AAV from the 2022-23 season. The complexity of hockey data stems from a high volume of plays occurring in rapid sequence within seconds of each other, making for a difficult analytical challenge (OpenAI, 2023). While it is important to study the macro-level statistics such as goals, assists, blocked shots, and missed shots, it is imperative to study the micro-level statistics as well. By using these micro-level statistics, analysts can gain a better perspective on various plays that may go unnoticed otherwise. Hockey analytics lacks an approach to determine solid playmakers. It is important to understand that there are multiple approaches to identifying the true value of a hockey player; however, utilizing this methodology offers a unique lens to identifying undervalued players based on their passing abilities. These undervalued players offer teams the same value as other playmakers but for a lower cost. In the Jaret Anderson-Dolan versus Rasmus Ristolainen example above, teams and organizations now have a means of identifying players who bring the same value to a team but for five times less money.

Using the clustering approach to group different player styles can be easily replicated for any number of games and seasons. As technological advancements continue to arise in the hockey analytics world, the tracking of raw data at both macro and micro-levels is steadily evolving into a more efficient process. If teams have the means of retrieving raw data for passes that lead to scoring chances and passes that lead to a shot on goal, they can easily replicate this methodology to identify clusters of players within an organization. From this research, teams and organizations now possess the methodology for comparing players within the same cluster of data. In the future, it would be beneficial to identify who the player is playing with regarding forward lines or defensive pairings. This would lead to a more comprehensive understanding of a player's value and ability by analyzing their performance with players from other clusters (OpenAI, 2023). By seeking to match various players from different clusters, teams and organizations have the capability to maximize the value on each offensive and defensive line pairing.

6. References

Alamar, Benjamin C. (2023). Sports Analytics: A Guide for Coaches, Managers, and other Decision Makers. *Colombia University Press, New York*. West Point, NY. Accessed 19 September 2023.

Bond, Phillip L, (2024), assistance given via verbal and written assistance. Mr. Bond read through my paper and gave me feedback on areas I could improve my grammar. Additionally, Mr. Bond pointed out areas in the paper where I could explain in more detail so the reader could gain a better understanding of the details within the game of hockey. West Point, NY. Accessed 15 April 2024.

CapFriendly, (2024), https://www.capfriendly.com/. West Point, NY. Accessed 10 March 2024.

Dabbura, Imad. (2018). K-means Clustering: Algorithm, Applications, Evaluations Methods, and Drawbacks. *Medium*: Towards Data Science. https://towardsdatascience.com/k-means-clustering-algorithm-applications-evaluation-methods-and-drawbacks-aa03e644b48a. West Point, NY. Accessed 16 March 2024.

Dennis, Declan, CDT, Cl 24', Co. H1, (2024), assitance given via verbal and written assitance. CDT Dennis gave me feedback on areas I could improve my grammar and sentence structure throughout the paper. West Point, NY. Accessed 15 April 2024.

Evenson, Justin, 2LT. (2023). Establishing a Model to Estimate Goal Differential for Hockey Goalies. *Department of Systems Engineering*. West Point, NY. Accessed 25 October 2023.

Geisler, Trent, LTC, (2023), assistance given via verbal and typed assistance. LTC Geisler helped me understand how to code portions of my data cleaning in R Studio, especially when I needed to work on scaling the data. LTC Geisler also explained how to use k-means clustering effectively. West Point, NY. 27 September 2023.

Kloo, Ian, Dr., (2024), assistance given via verbal and typed assistance. Dr. Kloo helped me organize my code so it would run without errors. Additonally, Dr. Kloo helped me effectively pull all 966 excel sheets into R Studio and make them into one big list. This helped me put the list into a large dataframe. Additonally, Dr. Kloo helped me develop the total points statistic, shots on goal ratio, and scoring chance ratio. Lastly, Dr. Kloo helped me code an interactive plot using plotly. West Point, NY. 07 November 2023.

Krueger, M., & Enos, J. (2021). Fenwick and Corsi in Army Hockey, Incorporating "Shot-Zones" and Rebounds into Weighted Shots. *Department of Systems Engineering, United States Military Academy*. West Point, NY. Accessed 25 September 2023.

McAdoo, Timothy. (2024) How to cite ChatGPT. *American Phychological Association*, https://apastyle.apa.org/blog/how-to-cite-chatgpt. West Point, NY. Accessed 19 March 2024.

Nandakumar, N., & Jensen, S. (2019). Historical Perspecttvies and Current Directions in Hockey Analytics. *Annual Review of Statistics and Its Application*. West Point, NY. Accessed 07 October 2023.

OpenAI, (2023). *ChatGPT* (Mar 14 version), https://chat.openai.com/, assistance in restructuring a few of my sentences to make them more fluid and grammatical correct. West Point, NY. Accessed 17 March 2024.

Radke, D., Lu, J., Woloschuk, T., Radke, D., Liu, C., & Brecht, T. "Analyzing Passing Metrics in Ice Hockey using Puck and Player Tracking Data." (2024) *Cheriton School of Computer Science, University of Waterloo and University of Chicago.* West Point, NY. Accessed 15 October 2023.

"Salary and Contract Glossary," *SportsNet*. (2024) https://www.sportsnet.ca/hockey/nhl/salary-contract-glossary/. West Point, NY. Accessed 18 April 2024.

Shea, S., & Baker, C. (2017). Hockey Analytics: A Game-Changing Perspective. *Advanced Metrics LLC*. West Point, NY. Accessed 19 September 2023.

Sznajder, Corey. (2023). "All Three Zones Project." *All Three Zones*. https://www.allthreezones.com/. West Point, NY. Accessed 28 August. 2024.

Tollerud, Jon. (2017). Average Annual Value and Guarantees." *MileHighReport*. https://www.milehighreport.com/2017/3/2/14794314/average-annual-value-and-guarantees. West Point, NY. Accessed 16 March 2024.

Vollman, Rob. (2016). Stat Shot: The Ultimate Guide to Hockey Analytics. *ECW Press*. West Point, NY. Accessed 25 September 2023.

(2022). Zibanejad scores 2, Rangers beat Lightning 3-1 in Opener. *ESPN*. https://www.espn.com/nhl/boxscore/_/gameId/401458593. West Point, NY. Accessed 27NOV2023.

(2024). 3D Scatter Plots in R. *Plotly*. https://plotly.com/r/3d-scatter-plots/. West Point, NY. Accessed 17 November 2023.