Examination of efficient roster design in the National Hockey League (NHL)

Stephanos Tselios, BAcc, MSc,

Submitted in partial fulfillment of the requirements for the degree of

Master of Arts in Applied Health Sciences

(Sport Management)

Faculty of Applied Health Sciences

Brock University

St. Catharines, ON

Abstract

The study estimates the values of NHL roster positions. The analysis was conducted in two phases. First, cluster analysis was used to evaluate and rank players for their overall performance across positions. Second, regression analysis based on aggregated player classifications across team-games estimated the value of roster position and measured diminishing returns to talent across positions. Players were evaluated based on their regular season performance. The clustering of all skaters was administered separately for each position and each year. Standardized regular season-long variables were applied in the analysis. The variables used to cluster all positions were: points per time on ice, goals per time on ice, assists per time on ice, plus/minus per time on ice, shots differential per time on ice, blocks per time on ice, hits per time on ice and penalties per time on ice. Forwards were distributed amongst four lines and defensemen were allocated to three pairings. The linear regression analysis used play-by-play data from the 2010-17 NHL regular seasons. Results indicated that an increase in the quality of centers increased the win probability of a team the most. Teams make player acquisitions decisions based on the talent available and their current composition of players. A team's hockey operations department can use the findings to evaluate their roster composition and identify positions with the greatest marginal benefit from player acquisitions.

Keywords: Roster Design, Cluster Analysis, Player Evaluation, National Hockey League

Acknowledgements

I would like to thank my parents, my brother and my fiancée, Vagia for everything you have done for me. Words cannot describe my gratitude.

I would like to thank all my SPMA colleagues for the wonderful years we spent together: Adam, Chris, Cole, Dan, Eddy, Evan, Jordyn, Josh, Katelyn, Lindsay, Michaela, Michelle, Mike, Mitch, Sean, Stephanie, Tyler and Weller.

A special thank you to my advisor, Dr. Kevin Mongeon, for his guidance and assistance throughout my years at Brock University.

I would also like to thank the Faculty and Staff of the Sport Management Department, especially my committee of Dr. Chris Chard and Dr. Shannon Kerwin for your vital assistance in fulfilling my study.

Table of Contents

CHAPTER I: INTRODUCTION 1

CHAPTER II: THEORETICAL BACKGROUND AND LITERATURE REVIEW 4

Player Performance 4

Marginal Product Theory 4

Player Evaluation 8

CHAPTER III: METHODOLOGY 15

National Hockey League (NHL) 15

Hockey Statistics 18

Cluster Analysis 20

K-means Clustering 22

Roster Design 24

Data 24

CHAPTER IV: RESULTS 29

Cluster Analysis 29

Data Preparation 29

Elbow Analysis 34

Silhouette Analysis 39

Skater Rankings 45

Roster Model Analysis 51

CHAPTER VI: DISCUSSION 62

General Findings 62

Limitations 70

Possible Future Research 71

CHAPTER VII: CONCLUSION 73

REFERENCES 76

APPENDIX A: FIGURES 84

APPENDIX B: TABLES 88

CHAPTER I

INTRODUCTION

This study examines the impact each roster position has on team performance in the National Hockey League (NHL). Roster design is the ability to construct a lineup that will increase the in-game performance of a team, either by developing the quality of existing players or acquiring additional players. It is the effective congregation of forward line combinations and defensive pairings that maximizes the win probability of a team.

To examine the roster efficiency of a given team, individual player performance and effects between teammates must be evaluated. The evaluation of player performance establishes the value of each roster position. The intention of player evaluation is to identify the value of an individual and the assessment of specific skills that improve team performance. Player evaluation can determine the roster position of each player of every team, as it illustrates the value of an individual along with their impact on team success. Finally, it demonstrates the value and effect each roster position has on team performance and success. Roster is the unofficial manifestation of individual athletes consolidated as a group.

The motivation for this research is to identify the key components of in-game team success. Play by play data was used to examine all on ice events to find the optimal combination of variables that would efficiently evaluate and rank players based on their in-game performance. To assist with the assessment, a new statistic that measured shot differential in even strength situations was introduced. Goals, assists, points, plus/minus, even strength shot differential, blocks, hits and penalties, per time on ice, formed the set of variables that weighted the in-game performance of skaters, for the duration of a regular season.

A second goal is to recognize which roster positions have greater contribution to team performance. Play by play data and a linear regression model were used to estimate the relationship between win probability and skater quality per position. The higher the increase in the value of win probability, the greater the contribution of the given roster position. To illustrate the value of an additional top center, winger and defensemen to team success, the natural logarithm (ln) of those variables was calculated to demonstrate the percentage variation in win probability.

A third goal is to establish the significance of roster depth. The use of the mean of each roster position in the win probability linear model also displayed the value of depth in each position. A regression was implemented to measure the relationship between in-game success and the quantity of elite centers, wingers and defensemen. For each roster position, the advancement of the quantity of elite players beyond a specific number had a reduced return in win probability. This indicated the importance of roster depth, as the inclusion of players in lower rankings per position beyond that point, would increase win probability instead of decreasing it.

Previous studies (Macdonald, 2010, 2011, 2012, 2013; Schuckers, D. Lock, Wells, Knickerbocker and R. Lock 2011; Schuckers and Curro, 2013; Thomas, Ventura, Jensen and Ma, 2013) focused on individual player interpretation and not on team assessment. Roster construction research has fixated on team diversity, efficiency and flexibility. Chan, Cho and Novati (2012) studied the contribution of players on team performance based on their style of play and not position. Thus there have been no empirical studies on the impact of roster position on team efficiency. This study will assist in filling the literature gap and aid sport managers to construct a roster that will maximize in-game win probability.

The objective of this study is to establish the value of each NHL roster position on team success. Examination of the NHL occurred for two reasons. First, hockey is a team sport, as unit effort accomplishes in-game success. The nature of the game is physical, accelerated and fluid, making it impossible for an individual athlete to monopolize game outcomes. Rarely will a skater play more than thirty minutes of regulation time. Second, general managers try to compose a high-performing, well-balanced roster with the existence of a strict salary cap. An empirical model will evaluate the overall on-ice performance of all skaters for eight regular seasons (2010-2017). Hockey operations can benefit from using this model to prioritize player acquisitions, given their present composition of players and salary cap restrictions. This study establishes a new statistic and set of variables to evaluate and rank the performance of skaters per season with the use of k-means++ clustering algorithm. Furthermore, a linear regression model was developed for the evaluation of the contribution of each roster position to team success.

The manuscript is structured as follow: Chapter II describes the theoretical background and reviews the literature related to player performance and roster analysis. Marginal product theory in professional sports, alternative methods on evaluating player performance, efficacy theory and diversity are mentioned and explained. Chapter III provides an overview of the research context, empirical approach and data used for analysis. Chapter IV displays and analyses the results from both cluster and regression analysis. Chapter V discusses the implications of the results, limitations of the study and recommendation for future research. Finally, Chapter VI highlights the overall results and considers the contribution of this research in the field of sport analytics.

CHAPTER II

THEORETICAL BACKGROUND AND LITERATURE REVIEW

This chapter will summarize previous studies on player evaluation and roster analysis. Marginal product theory was the initial concept used to examine and associate team performance and winning, to player contribution. Since then, numerous researches (Chan et al., 2012; Macdonald, 2010, 2011, 2012, 2013; Schuckers et al., 2011; Schuckers and Curro, 2013; Thomas et al., 2013) have implemented alternative methods on measuring player performance in the NHL. The in-game contribution of individuals is a fundamental aspect of roster design, as it can establish their roster position and the effect they have on team success.

Player Performance

Marginal Product Theory

Marginal product theory predominated the theory of factor pricing. It was first proposed by T.H. Von Thunen, a German economist, in 1826. Clark (1928) declared that if the allocation of income was equitable, each agent would receive a reward equal to their productivity. Blaug (1997) stated that in equilibrium, each agent would be compensated analogously to their marginal productivity. Liebhafasky (1964) claimed that elements of production would likely earn a pragmatic rate of return equal to their marginal productivity, after a long period of time under optimal competition.

Marginal product indicates the increase in units of output by the addition of one unit of factor of production, while all other factors of production are constant. The increased output is known as marginal product. In consonance with the theory, under perfect competition, the price of services provided by a given factor of production is equal to its marginal product. The organization will employ the optimal number of units of a given factor at which the price of services is equal to its marginal product. From an organizational perspective, it is a factor demand theory while from the industry perspective, it is a theory of factor pricing.

When an organization increases one unit of a factor of production (while keeping the other factors constant), the marginal product escalates to a precise degree of production. After reaching a specific degree, the marginal productivity deteriorates, as the marginal cost is greater than the marginal revenue. However, the organization chooses to hire/use an additional unit of factors, if the marginal cost is lower than the marginal revenue.

The equation to calculate marginal revenue is:

$$MR = \frac{\Delta TR}{\Delta Q}$$

where MR is the marginal revenue, ΔTR is the change in total revenue and ΔQ is the change in quantity of a given factor of production. Profit maximization is achieved at the level where marginal revenue is equal to marginal cost.

The supplementary revenue generated from using one more unit of a given factor of production is known as marginal revenue product. It is the variation in total revenue divided by the change in the number of factors of production. It attributes the marginal product approach with respect to the variation in total revenue. M.J. Ulmer defined marginal revenue productivity as the accretion to total revenue from the usage of one unit of an element of production, while all additional factors being constant. It is an indication of the value of an additional factor of production to an organization, as it illustrates the additional revenue generated.

Marginal revenue product of a factor of production is estimated by the following equation**:**

$$MRP = MP * MR$$

where MRP is marginal revenue product, MP is marginal product and MR is marginal revenue. Marginal revenue product is an indication of the value of an additional factor of production to an organization, as it illustrates the additional revenue generated.

Scully (1974) pioneered the first empirical study on player contribution to team performance and winning. The purpose of his study was to estimate the predicted player marginal revenue product in comparison to the predicted salary of each player. The data set contained team-level data from Major League Baseball (MLB) for the 1968 and 1969 seasons. The margin in the number of wins with and without a given player is an evaluation of his marginal product. Marginal Revenue Product (MRP) of a given player, is the margin in wins multiplied by the margin in gate receipts and broadcasting revenue per win. To evaluate performance, the unit of measurement was player instead of team.

$$MRP(W) = MP(W) * MR(W)$$

where W is the level of team winning percent, $MP(W)$ is marginal product or the player's contribution to team winning percent. In baseball, run scored measures offensive contribution. Prevented runs conceived estimates defensive contribution. Therefore, hitter performance represents offense and pitcher performance defense. For offense, the variable of assessment was slugging average (SA). For defense, the variable of estimation was strikeout-to-walk ratio (SW).

The author detected a linear relationship between team percent wins (PCTWIN), team slugging average (TSA) and team strikeout-to-walk ratio (TSW). Each one point raise in the team slugging average incremented the team win percentage by 0.92 points. A one point raise in the team strikeout-to-walk ratio increased team win percentage by 0.90 points. Scully established a net MRP for every player, after reducing for associated expenses. The salary of each player was then compared to their net MRP. Findings showed that owners lost money by overpaying mediocre players. However, the salaries earned by average and elite players were significantly below their net MRP value.

Despite expanding Scully's original study, Scully (1989) and Zimbalist (1992) were unable to reconstruct or develop a better model to measure player contribution to winning. Fort (1992) improved Scully's 1974 methodology with the use of more skilled statistics and by applying additional independent variables to estimate player performance. Three seasons were used as the data set: 1968, 1977 and 1990. Each player's estimated MRP assisted in the calculation of the predicted salaries. The actual salaries were then correlated to the predicted salaries. For the 1968 season, they payment of most players were significantly below their estimated MRP. For the 1977 season, most players' earnings were still below their estimated MRP, but not as significant as in 1968. For the 1990 season, all players had salaries significantly above their estimated MRP. This highlighted players' capability to apprehend surpluses from team owners.

Marginal product provides a valid framework for computing player contribution to winning. The implementation of Scully's method for player evaluation in hockey is inefficient. The nature of the game is very different to baseball. Hockey has a fluid style of play where all previous on-ice events impact the result of following events. Baseball is a series of independent events. Previous plays have no impact on the outcome of future plays. Hockey has a continuous flow with rapid puck possession alterations, making both teams play offense and defense simultaneously. During a baseball game, teams alternate turns in playing offense and defense. The visitor team starts on offense (at bat), whereas the home team on defense (outfield). When the designated number of players are "out", teams' alternate offensive and defensive positions.

In every on-ice event all skaters from both teams interact, making the outcome of each play unpredictable. The quality of teammates and opponents have an impact on both player performance and the result of each event. In baseball, players that are playing offense are isolated from their teammates; it is a duel between pitcher and outfield versus hitter. There is no significant interaction between teammates and opponents that affects player contribution. For the above reasons, various researchers have designed models with diverse approaches to better fit the nature and style of hockey. Their objective is to consider all aspects of the game, analyze and articulate the overall performance of each player.

Player Evaluation

Over the course of a hockey game, skaters exhibit many skills. Yet, not all skills contribute to team success. The intention of player evaluation is the assessment of specific hockey skills that improve the overall performance of a team. The utmost contribution of a player is his ability to improve the goal differential of his team (i.e. to increase goals scored and decrease goals against). Increased goal differential, increases the win probability of a team. There are a considerable amount of studies measuring players' performance in the NHL. Since the overall performance of each player can determine their roster position and the impact they have on team performance, it is an essential element for roster design. Below are studies that have examined player performance.

Macdonald (2010) formed an adjusted plus-minus statistic for all NHL players. He computed the impact of every player in even strength situations only. The statistic was not affected by the performance of teammates and opponents. To calculate the impact every player had on the success of his team in scoring and preventing goals, a weighted least squares regression was applied. The author used two weighted least square models and then averaged the outcomes of both models. The two types of regression models were a Ilardi-Barzilai-type and a Rosenbaum-type. The second model was implemented as an attempt to enhance estimates and errors.

The data set incorporated all even strength situations of all regular season games from 2007-08 up to 2009-10. Players had to have at least 4,000 shifts throughout those years to be evaluated. Three plus/minus statistics were estimated: offensive, defensive and adjusted plus/minus. Findings indicated that Sidney Crosby was the best offensive forward in the league. His offensive plus/minus was the highest amongst all players. As for defensive plus/minus, all goaltenders were in the top ten rankings. Pekka Rinne had the highest defensive statistic. Pavel Datsyuk was the player who had the highest adjusted plus/minus statistic, balancing both offensive and defensive contribution. In fact, Datsyuk won the Selke trophy (award given to the best defensive forward) all three years (2007-08, 2008-09, 2009-10) illustrating him as the best two-way forward in the league.

Schuckers et al. (2011) introduced a least square model comparable to the adjusted plus/minus proposed by Rosenbaum (2004) and Ilardi (2007) for basketball. The purpose of the model was to weigh, rank and compare the performance of each player, given the plays that occurred during their time on ice. The value of each play was the differential between the observed result and the expected result for that play. The regression model had the ability to adapt for a given series of plays for each player on the ice. All players were then ranked in terms of their average rate contribution to the response.

The data set contained each play of each game from the 2007-08 NHL regular season. The events were: face-off, shot, blocked shot, missed shot, takeaway, giveaway, hit, penalty and goal. Findings showed that all the top players in the league had positive values. The probability of their team scoring a goal increased with their presence on ice. Players with negative values, decreased the probability of their team scoring a goal. For players who switched teams during the season, a small margin between their rating with the different teams was detected. This indicated the balance and independence of player ratings from teammates and opponents.

Macdonald (2011) calculated the effect each player had in even strength and man advantage situations. He created two weighted least square regression models to measure the effect every player had on team performance in terms of scoring and preventing goals. The first model was for even strength situations. The second model was for special teams' situations. The intention was to evaluate the independent impact each player had on team success. The data set included every shift of every game from the 2007-08 to the 2009-10 NHL regular season. Adjusted Plus-Minus (APM) estimated the overall contribution of each player based on goals per season, for man advantage and even strength situations. Findings confirmed common speculations about the value of same top players in the game. For example, Pavel Datsyuk, the best two-way forward in the game, had the highest APM amongst all skaters.

Chan et al. (2012) categorized all players in various categories, based on their on-ice style of play. In detail, forwards were identified and separated into top, second, defensive and physical line. Defensemen into offensive, defensive, average and physical pairing. Goalies into elite, average and bottom. The aspiration was to establish a perceptible relationship between the diagnosed player types and the effect they had on team performance. A multiple linear regression was estimated using the player brands as independent variables and the total points each team acquired throughout the duration of the regular season as the dependent variable. Two models were tested. The first model contained all eleven variables. The second, involved only variables with significant coefficients.

The data set consisted of numerous statistics from six regular seasons (2005-2010) and contract information (salary) from four seasons (2007-10). Results for performance indicated that goaltenders of elite status tend to have provided the greatest when it came to team points. For forwards, as anticipated the top line had a higher impact on team points in comparison to the other forward lines. Similarly, offensive defensemen had a bigger effect on team success in contrast to the other three types. Despite taking into consideration the economic burden of each player, goaltenders still had the most value to their team. For forwards, defensive skaters grant the largest worth for their contribution due to their low cost. Finally, there is a limited amount of high performing defensemen. General managers should be willing and eager to spend more money for their acquisition.

Macdonald (2012) developed an Adjusted Plus Minus (APM) statistic by applying a ridge regression. Fenwick, Corsi, shots, and goals were used as variables. The purpose was to estimate the individual contribution of each player. The statistic was independent of the roster depth of teams and other variables that were uncontrollable. A ridge regression was preferred as it reduces the error bounds in the results and progresses the anticipated performance of the model, if multicollinearity was present in the data set.

The data set involved every shift of every game from the 2007-08 to the 2010-11 NHL regular season, as well as the result of each shift. The data was then broken down to even strength and man advantage situations. Any data that involved empty net observations for both cases were removed. Findings seem to agree with prevalent beliefs about the value of specific players. Pavel Datsyuk, who won the Selke Trophy in 2007-08, 2008-09, and 2009-10, and was a finalist in 2010-11, had the highest APM during those years, making him the most valuable player in the league in the above seasons. As for the top offensive player, Sidney Crosby had the highest rating despite missing a respectful amount of games due to injury.

Macdonald (2013) designed an expected goals model that would analyze player performance. A ridge regression was used to estimate the contribution on the expected goals each player had on their team per game, independent of their zone start, opponents and teammates performance. For adjusted plus/minus, expected goals were used as a variable in preference to goals. The data set contained all NHL regular season games from the 2007-08 up to the 2010-11 season. The first three seasons were used to calculate the restrictions in the model, and the final year for verifying the model. The best offensive player in even strength situations was Sidney Crosby, regardless of not playing a full season during those four years. A notable result for the analysis was teams that had more hits against than hits for, scored more goals.

Schuckers and Curro (2013) came up with an inclusive statistical rating for all NHL players, called Total Hockey Rating (THoR). The probability of an event resulting in a goal, establishes the effect of each play. The statistic measured the influence players had in creating and preventing goals. To cope with multicollinearity, a ridge regression was used. THoR computed the supplementary wins each player contributed to his team over an average player. The data set included all events that occurred in all 2010-11 NHL regular season games. Values were appointed to nine even strength events: face-off, shot on goal, missed shot, blocked shot, penalty, hit, takeaway, giveaway and goal. Based on their analysis and results, elite players in the league provided more than 5 wins. In even strength situations, results revealed that forwards produced more additional wins than defensemen; hence, forwards had greater value in comparison to defensemen.

Thomas et al. (2013) created hazard function models for player ratings by taking under consideration the quality of teammates and opponents. Two diverse shrinkage methods were implemented: full Bayesian hierarchical models that partially pool restrictions per player position, and penalized maximum probability assessment to choose a smaller amount of restrictions that emerge. The data set involved all even strength situations of every NHL regular season games from 2007-08 up to 2011-12 (5 seasons). It should be mentioned that over 98% of the observations were not goal related. Thus, for the analysis, only events that had a goal outcome were considered.

The total contribution of each player was the differential in expected goals, allowed and scored, in relation to an average player on an average team. Results showed that no defenseman was ranked in the top eighty players. The player who had the highest ranked contribution, was Henrik Lundqvist, goaltender for the New York Rangers. The most and least valuable forwards and defenders per team, were also determined. A player was most valuable to his team, when the performance of his team decreased significantly by having a "replacement" in his position. As for the least valuable player, team performance might have increased or remained stable with a "replacement" in his roster spot. Results were not surprising since the best players in the league had the most ice time, making them more valuable to their teams. Centers who specialized in face-offs were identified of being the least valuable players on their team, as they had limited ice time and defensive responsibilities, which restricted their offensive ratings.

The primary variable of focus for evaluating player performance was goals. Macdonald (2010, 2011, 2013) evaluated player performance based on scoring and preventing goals and expected goals (2013). Schuckers et al. (2011) assessed all on-ice events to demonstrate the value of an event based on the variation between the detected outcome and the anticipated outcome for that play. Schuckers and Curro (2013) established the value of an event dependent on the probability of that event resolving to a goal. Thomas et al. (2013) used the differential in expected goals, allowed and scored, in relation to an average player on an average team to discover the contribution of a given player. Chan et al. (2012) used k-means clustering algorithm on a set of standardized variables to weigh and rank forwards and defensemen based on their style of play. Centers and wingers were clustered together while defensemen would categorized into four styles of play. They developed a linear regression model to examine the relationship between the diagnosed styles of play and their impact on team success at the season level.

This study established a new statistic and set of standardized variables to evaluate and rank the performance of skaters per season. K-means++ clustering algorithm was used instead of k-means algorithm, to estimate and classify the in-game performance of centers, wingers and defensemen. Each position was clustered separately, four clusters were formed for centers and wingers while three for defensemen, as the roster of a team contains four lines of forwards and three defensive pairings. No previous study examined the relationship between team success and roster position. By developing a linear regression model, this study managed to evaluate the relationship between win probability and roster position at the game level. Furthermore, a linear model that measured the impact an additional elite center, winger and defenseman had on win probability was developed. Finally, a linear regression was designed to estimate the relationship between win probability and the quantity of top centers, wingers and defensemen.

CHAPTER III

METHODOLOGY

The objective of this chapter is to briefly describe the structure of an NHL team roster and the format of an NHL regular season game. The definition of specific hockey statistics, how they are calculated and their importance in grading the overall performance of a player are interpreted. Cluster analysis is delineated and justified why it was the selected technique to evaluate and rank players dependent of their season-long performance. A regression model was created to measure the contribution of each roster position on team success in terms of winning. This section concludes with an in-depth description of the data used for the analysis.

National Hockey League (NHL)

Each team must have, at any time during the regular season and playoffs, at least twenty players (eighteen skaters, two goaltenders) with a maximum of twenty-three contracts, without exceeding the salary cap limit. The maximum number of players each team can dress for every game, is twenty; eighteen skaters and two goaltenders. For the goaltenders, one is named the starter and usually plays the entire game; the other is named the backup and may enter the game if the starting goaltender performs insufficiently or is injured. Skaters are typically divided into twelve forwards and six defensemen. Every forward line has a left winger, a center and a right winger and every defensive pairing has a left and a right defender. Forwards can be classified into top and bottom six. Top six forwards excel in offensive situations and their objective is to generate goals. Bottom six forwards have more defensive skills and their role is to prevent goals being conceived. Defensemen are categorized into top four and bottom two pairing. Top four defensemen are well rounded players and their responsibilities are to generate offensive opportunities and minimize the offensive production of the opposition. Bottom pairing defensemen are defensively minded and their task is to prohibit the opponent team from scoring a goal. Roster design is the efficient assembling of forward line combinations and defensive pairings that maximizes the win probability of a team.

The definitive quantity of players on ice for each team is six; one goaltender, two defensemen and three forwards. Throughout the duration of a game, on-ice skaters frequently alternate with those on the bench at specific stoppages of play or while the play is active, as long as there is a maximum of six players per team on ice at all times. A team that fails to complete a legal rotation (the player exiting must be completely off the ice prior to his substitute entering) is penalized for having too many men on ice and will have to play with a numerical disadvantage for the duration of two minutes or less (the team with the numerical disadvantage conceives a goal prior to the expiration of the penalty). A team is permitted to withdraw their goaltender at any time and substitute him with a forward. This generally transpires in the final minutes of a game as the team with the less goals scored, may decide to pull their goaltender from the game and add an extra forward to have a numeric advantage and increase the probability of scoring a goal.

A power play occurs when the opposing team is penalized. They must play with a skater or two skaters less (minimum of three skaters required on ice), for a definite amount of time, depending on the penalty. Maximization of numerical advantage will be achieved with the scoring of a goal. The accomplishment of the desired objective for the team with the man advantage is probable with the use of offensively talented players, who consist of a scoring touch and can handle and move the puck efficiently. Thus, the most gifted players of a team will play the majority amount of power play minutes. The less skillful skaters will be on the power play for minimal to zero minutes. Consistently, top six forwards and top four defensemen are on the two power play units of their respectful team.

Penalty kill is the result of a team playing with a numerical disadvantage, for a certain amount of time, due to a penalty or penalties. The intention of penalty kill units is to prohibit a goal to be concede for the duration of time the opposite team hold the man advantage. Bottom six forwards and bottom defensive pairing tend to be defensively talented. They are on the ice for the bulk quantity of penalty kill minutes. Some top six forwards are equally skilled in offensive and defensive styles of play (also known as two way forwards). Hence, they may play a substantial amount of minutes with a numerical disadvantage. When a team is trailing in the final minutes of a game, offensively minded players may be on the penalty kill units, with the anticipation of achieving a goal.

A NHL game has a duration of sixty minutes and is played over three periods, twenty minutes each. If the score is even upon completion of regular time, both teams earn a point and an additional five minutes of sudden-death overtime is played to determine the winning team. If both teams have failed to score a goal prior to the end of overtime, the winning team is decided by a shoot-out competition. There are three rounds of shooters, altering by team. Each team assigns three players, along with the order they will shoot, who will try to score a goal on the opponent goaltender in a one-on-one situation. If the score remains even after three rounds, the shoot-out is extended until one team scores and the other team fails to score in the same round. The winning team is awarded two points in the league standings, indifferent if the win occurred in regulation time, in overtime or in a shoot-out. If a win occurs in regulation time, the losing team will not receive any points. However, if a win occurs in overtime or in a shoot-out, the losing team will receive a point in the league standings.

Hockey Statistics

To assign players to their respectful roster position, top or bottom forward/defensemen, a selected set of statistics were used to evaluate both offensive and defensive performance of each skater. Offensive contribution of a player is measured by their point production, which is the sum of goals and assists for the duration of a regular season. Goals are the total number of goals an individual scored; assists are the number of times a player assisted on a goal scored by a teammate. The last two players who touched the puck prior to the goal scorer are appointed an assist (primary, secondary). The primary assist is assigned to the player who last touched the puck and the secondary assist to the other player. Additional statistic that estimates the offensive contribution of a skater are goals for and shots for. Top six forwards and top pairing defensemen are expected to have a higher offensive contribution than bottom six forwards and bottom pairing defensemen.

Defensive contribution of a player is estimated by his ability to prevent goals being scored. Statistics used to evaluate defensive skills of skaters are blocks, hits and goals against. A block is the prevention of an attempted shot on goal from reaching the net by a player that is not the goaltender. A skater can use his body or equipment to block a shot. A block is assigned to the last player to stop the puck or to the player who significantly altered the motion of the puck. However, if the puck reaches the net despite making contact with a skater, it is considered a shot on goal and a block is not credited to the skater. Hit is the use of physical contact to separate the opponent from the puck and obtain possession. For a hit to be registered, the contact must be legal and intentional with the opposing player possessing the puck, and it must result in possession loss for the player enduring the contact. If puck possession is retrieved by the player himself or his teammate, a hit is still registered. However, if there is illegal physical contact, resulting in a penalty, a hit is not recorded. Blocks and hits are essential defensive attributes for both forwards and defensemen, as less opposing puck possession and less shots on goal decreases the probability of conceiving a goal, which increases the win probability of a given team.

A complex statistic, such as plus-minus is used to assess the overall contribution of a player. Plus/minus is the net contribution of goals a given player was on the ice. It is the difference between goals scored and goals conceived. Goals for, is the sum of goals a team scored in even strength and numerical disadvantage situations while a given player was on the ice. Goals against, is the sum of goals a team conceived in even strength and numerical advantage situations which a given player was on the ice. If a player was on the ice for more goals scored than conceived, plus/minus will have a positive value. However, if a player was on the ice for more goals conceived than goals scored, plus/minus will have a negative value. The greater the value of the plus/minus statistic, the greater the overall contribution to team performance.

An additional composite statistic to appraise the offensive and defensive performance of skaters, is shots differential. It computes the net contribution of shots a player participated in, as the numerical variation of shots for and shots against. Shots for are the sum of shots on net by an individual and his teammates while he was on the ice. Shots against are the total shots on net by the opposition while a given player was on the ice. When a player generates more scoring opportunities than his opponents, the shot differential variable has a positive value. This indicates that a player spent most of his time on ice in the offensive zone creating chances to score a goal. The higher the value of shot differential of a player, the higher offensive contribution to team performance. On the contrary, when a skater produces less offensive opportunities than his competitors, the shot differential variable has a negative value. This shows that a skater spent most of his time on ice in the defensive zone trying to prevent the opposition from developing scoring opportunities instead in the offensive zone. The lower the value of the shot differential of a player, the lesser contribution to team's offense.

To effectively evaluate and compare player performance, all statistics were normalized with the use of time on ice (TOI). This accounted for some skaters playing less minutes per game than others. Since many of these statistics were measured on different scales, it was essential to standardize the data. This was accomplished by subtracting the mean form the data and dividing by the standard deviation. The emerged distribution had a mean value of zero and standard deviation of one. As a result, all statistics were measured on the same scale and were equally weighted, prohibiting a variable or variables to dictate the cluster analysis (Vincent & Eastman, 2009).

Cluster Analysis

Cluster analysis is an exploratory analysis (mining) tool used to classify a multivariate data set, based on a set of selected variables, into a significant number of clusters (groups) in a way that the degree of association amongst these observations is maximal if they belong in the same cluster and minimal alternatively. This tool is used to examine and discover if formerly undefined clusters (groups) exist in the data. The purpose of cluster analysis is to appoint observations to clusters (groups) that are analogous to each other, with respect to the selected variables of interest, and the clusters themselves differ from each other. In other words, the aim is to divide the observations into definite and homogeneous clusters. Cluster analysis seeks to identify the quantity and distribution of the groups whereas classification attempts to forecast the group to which a new observation belongs. Dissimilar to other statistical techniques for classification, such as automatic interaction detection and discriminant analysis, it generates no prior presumptions about essential variations within the data set. However, cluster analysis has no technique for differentiating between significant and insignificant variables. This is vital as the clusters formed can be highly dependent on the selected variables incorporated in the analysis**.**

There are numerous developed clustering techniques in data mining. The selection of a technique is based on the characteristics of the data being clustered and the quantity of clusters. Clustering methods are divided in two fundamental types: hierarchical and non-hierarchical clustering. Hierarchical algorithms assemble analogous observations in a cluster, either by assigning each observation to its own cluster and joining similar clusters until there is only one cluster left (agglomerative method) or by assigning all observations to one cluster and dividing it into similar clusters until each observation has its own cluster (divisive method). The end result of both hierarchical techniques is a dendrogram, which displays how the clusters are formed. Since all observations are compared prior to each clustering phase, the process cost is significantly high.

Non-Hierarchical algorithms attempt to cluster the data directly. With the use of a repetitive algorithm that optimizes a selected criterion, it structures a grouping of a set of observations, into a pre-determined quantity of clusters. Starting from a primary cluster, observations are shifted from one cluster to another or exchanged with observations from other clusters, until there is no additional improvement to the criterion value. Non-hierarchical algorithms typically alter centers until all observations are affiliated to centers. Since the distance of all observations and related centers are estimated successively until all centers are minimized, the cost process is relatively high. However, in terms of estimation resources and time, non-hierarchical method has a lower cost in comparison to hierarchical method.

K-means Clustering

K-means is an unsupervised clustering method that illustrates the “best” partitioning of a data set including k number of clusters. The basic concept of this method is to distinguish k centroids, one for each cluster. Each observation belongs to the cluster with the closet mean. The objective is to minimize the squared error function, or the total intra-cluster variance. An distinctive element of k-means is that the number of clusters is determined in advance. Having a fixed number of clusters is both an advantage and a disadvantage. The benefit of having a fixed number of clusters is that the k-means method does not insert new clusters in the event of an anomaly observation, alternatively it classifies the anomaly observation to its nearest cluster. The deficiency of applying a fixed number of clusters is that the results are sensitive. The use of an inadequate k may possibly lead to insufficient results, which may become impractical.

The efficiency of k-means is highly dependent on the initialization of the centroids. The fundamental concern of k-means is to establish cluster centers that decrease intra-class variance by minimizing the ranges from all clustered data points. Inadequate initialization of centroids will provide poor clustering. K-means++ is a probabilistic algorithm for selecting the primary values (or “seeds”) for the k-means clustering method. It is a technique of averting the occasionally insufficient clustering formed by the standard k-means algorithm. Apart from the first center which is selected erratically, each subsequent center is methodically selected in accordance with its squared distance from the nearest center already selected. The k-means++ algorithm has a inferable approximation guarantee to the optimal solution, in opposition to k-means, which can create clustering arbitrarily inferior than the optimum.

To determine the optimal quantity of clusters for a given data set, elbow method and silhouette analysis are used. Elbow method is a technique of explanation and verification of cohesion within cluster analysis formed to assist in discovering the optimal quantity of clusters in the data set. It examines the percentage of variance interpreted as a function of the quantity of clusters. The optimal number of clusters is identified when by adding an additional cluster, there is no better modeling of the data. Specifically, by plotting the percentage of variance explained across the quantity of clusters, the first clusters will explain a lot of variance but at some point the marginal gain will severely decrease, shaping an angle in the graph. The optimal number of clusters is selected at this point, hence the “elbow criterion”.

Silhouette analysis combines both separation and cohesion. The silhouette value is a measure of how analogous an observation is to its own cluster (cohesion) in contrast to the other clusters formed (separation) by estimating the average distance between clusters. This value has a range of [-1, 1], where a low value (near -1) implies that the observation is inadequately matched to its own cluster and highly matched to neighboring clusters. This means that an observation might have been assigned to the incorrect cluster. A high value (near +1) illustrates that the observation is highly matched with its own cluster and inadequately matched with neighboring clusters. A value of zero signifies that the observation lies between two neighboring clusters. The clustering configuration is applicable, if most observations have a high value. The clustering configuration may have too few or too many clusters, if most observations have a negative or low value.

Roster Design

The formation of every NHL team roster will be finalized with the distribution of all players to their respectful position. The next phase is to generate a linear model that will implement the quantity of additional wins each roster position contributes at the game level.

The linear model to describe additional wins in terms of roster position, can be expressed as:

$$\mathrm{Win}_{s,g,t} = \beta_1 \mathrm{MeanC}_{s,g,t} + \beta_2 \mathrm{MeanW}_{s,g,t} + \beta_3 \mathrm{MeanD}_{s,g,t} + \varepsilon_{s,g,t}$$

where Win is the probability that a given team won the game, MeanC is the mean rank of centers for a given team; MeanW is the mean rank of wingers for a given team; MeanD is the mean rank of defensemen for a given team, s is season, g is game and t is team. The β coefficients symbolize the autonomous contribution of each roster position (independent) to the probability of the dependent variable Win. If all other independent variables stay constant, β_i is the change in Win for each one-unit difference in the relative independent variable. Since players were ranked in ascending order (elite players are ranked as number one), the lower the value of mean a position has, the more skilled players are playing in that position. The intention is to estimate the marginal probability of an additional top and bottom center, winger and defenseman.

Data

A hockey scraper was developed in Python to collect play by play and shift data from the NHL official website. For analysis, play-by-play data from every regular season game of eight seasons (2010-11 up to 2017-18) was used. All eight seasons were applied for the appraisal of player evaluation via cluster analysis and to generate the variables for the roster model. On-ice events, materialized in regulation time only, were enlisted in the data set. Thus, all overtime and shootout play-by-play incidents were not included in the analysis. Each play encompasses a total of twelve observations. Six for the home team and six for the visitor team, along with information related to the outcome of each event. Values are assigned to all on-ice players and are dependent on the volume of influence an individual had on the result of a given play.

Throughout the duration of a game, players participate in numerous on-ice situations; each on-ice event registers an observation. With the completion of a game, players are graded for their complete on-ice performance, which is measured by their impact on team success. Every team during a single regular season plays 82 games, which sums up to a total of 1230 league games. Since the 2012-13 was a half season, a total of 615 league games were played. The addition of a 31st team, increased the total league games to 1271 for the 2017-2018 regular season. In view of the data set for player interpretation consisting of eight seasons, the number of games every team competed in were 615. Cluster analysis estimated the overall performance of a given player, from on-ice events the specific individual took part in, for each season separately.

The 2010-11 regular season was chosen as the starting point due to the rule change that occurred in tie-break procedures. Prior to that season, the number of games won was the variable of separation for teams in the same conference with the equivalent number of points. The new rule divided the number of games won in regulation and in overtime or shootout. It generated a statistic named ROW (Regulation and Overtime Wins). This alteration rewarded in-play team win over a victory gained by individual talent. It proved to be vital during that season. The Philadelphia Flyers and Pittsburgh Penguins had the identical number of points (106). Philadelphia won the Atlantic Division title since their ROW stat was significantly greater. Listed below are additional league committee decisions and adjustments that happened in the years examined. All had an influence on player production and team efficiency.

Prior to the kickoff of the 2011-12 season, the Atlanta Thrashers were sold and relocated to Winnipeg. The Winnipeg Jets took their place in the Southeast Division (Eastern Conference). The unfavorable travelling schedule and time zone differences affected player efficiency for all eastern conference teams. The impact was greater for teams in the same division and significantly higher for the Winnipeg Jets. Both 2010-11 and 2011-12 regular seasons premiered in Europe as part of a marketing strategy by the NHL. The purpose was to broaden its horizons and fan base beyond North America. Six teams (Boston Bruins, Carolina Hurricanes, Columbus Blue Jackets, Minnesota Wild, Phoenix Coyotes and San Jose Sharks) played six games the first year. Four teams (Anaheim Ducks, Buffalo Sabres, Los Angeles Kings and New York Rangers) played four games the following year. These abroad games had an impact on team performance, as skaters were unfamiliar with the hosting arenas and facilities. Teams had no actual home ice advantage, since games happened in a neutral environment. After five straight successful seasons (2007-08 until 2011-12), the NHL decided not to schedule a premiere event for the 2012-13 season.

The 2012-13 regular season started on January 19th, after the National Hockey League Players' Association (NHLPA) was locked out by all team owners due to the termination of the NHL collective bargaining agreement (CBA) on September 15th. This resulted in a short season involving 48 games of intra-conference competition only. Teams in a less competitive conference had a higher probability qualifying for the playoffs in contrast to teams that were contesting in a more challenging conference. Moreover, teams that could perform strongly in the first half of a full season (41 games) had a greater possibility to make the playoffs over teams that would perform better in the second half. The horrific navigation schedule of the Winnipeg Jets the previous year had created issues regarding team distribution. The solution was to reallocate three teams. The Winnipeg Jets joined the Central Division of the Western Conference. The Detroit Red Wings joined the Atlantic Division of the Eastern Conference. The Columbus Blue Jackets joined the Metropolitan Division of the Eastern Conference. Thus, competition in the East has increased by the addition of two more teams.

In February, the 2013-14 NHL regular season was interrupted by the Sochi Winter Olympics. The NHL committee had to adjust the schedule, in anticipation of players' willingness to participate in the Olympics. This resulted in a two-week break. The contradictory influence of the Olympic games on teams, were injuries. Key players were sidelined for a significant period, effecting in a negative way team performance and success. John Tavares of the New York Islanders, who had been out-performing his teammates, suffered a season ending left knee injury. Not only did the Islanders failed to reach the playoffs, they finished last in their Division.

On the first day of the new year, the NHL has programmed an outdoor game, known as the "Winter Classic". For the 2012-13 season the event was canceled due to the lockout. Since the Winter Classic became a success, in terms of revenue and fan base, the 2013-14 season would host five additional outdoor games (six total). Over the course of the remaining years in data set, two outdoor games took place in 2014-15, three in 2015-16, four in 2016-17 and three in 2017-18 regular season. Teams selected to compete in outdoor games were in disadvantage in contrast to the rest of the league. Unfamiliar natural conditions along with the quality of ice, impressed the performance of players. These conditions also impacted the probability of a team winning an outdoor competition.

For the 2015-16 season the New York Islanders played in their new arena, the Barclays Center, in Brooklyn. The piping under the ice generated issues with the ice conditions. Players have characterized the ice "unplayable" and "bouncy". The poor quality of ice had a negative impact of player performance. The team that was affected the most was the Islanders, as all their home games took place in that arena. In-game rule alterations such as hybrid icing, no use of gloves during face-offs etc. had no impact on both player evaluation and roster design models. These changes did not alter the outcome of the on-ice events examined. Modifications made in overtime and shootout situations were disregarded.

Prior to the start of the 2017-2018 regular season announced the addition of a franchise located in Las Vegas, increasing NHL teams to 31, and the Pacific Division to eight. For the formation of the Las Vegas Golden Knights roster, an expansion draft was held. Each team was allowed to protect seven forwards, three defensemen, and one goaltender or, eight skaters and one goaltender. The Golden Knights had to select only one player from each team with a minimum of 14 forwards, nine defensemen, and three goaltenders. A minimum of 20 players had to be under contract for the duration of the 2017-18 season. A minimal cap hit of $43.8 million needed to be reached by Golden Knights via expansion draft.

CHAPTER IV

RESULTS

The objective of this chapter is the display and description of the results from the cluster analysis and the roster design model. Both the elbow and the silhouette method were used for the calculation and presentation of the optimal number of clusters per position and regular season. The quantity of skaters from each formed clusters along with their ranking value were presented. Summary statistics of the data used for the roster model analysis were computed. Estimates from the OLS regression were analyzed at the game level. The impact of an additional elite center, winger, defensemen on the variables of interest were assessed.

Cluster Analysis

Data Preparation

Players were evaluated based on their regular season performance. The clustering of all skaters was administered separately for each year. The objective of conducting cluster analysis per year was to efficiently group and rank players dependent on their contribution to team performance. Since the style of play of an individual may alter from season to season, and his offensive and defensive contribution may had evolved or declined, it was vital to estimate and rate his on-ice performance for the duration of a regular season. Since the data set contained eight regular seasons, clustering was conducted accordingly. Standardized regular season-long variables were applied in the analysis.

Prior to the analysis, goaltenders were excluded from the cluster analysis, leaving forwards and defensemen for evaluation. To determine if a player were to be included in the cluster analysis, his season-long time on ice was calculated. Since there are three defensive pairings in contrast to four forward lines, defensemen average more time on ice than forwards.

For that reason, the data set was subcategorized into forwards and defensemen and two histograms were plotted. Additionally, skaters who played less than a specific quantity of minutes for the whole regular season were excluded from the cluster analysis. These skaters were eliminated to remove outliers whose contribution for a limited time would not have been representative of an entire season performance. These include players who were injured and thus were sidelined for a significant portion of the season, and skaters who play in the American Hockey League (AHL) affiliate team who were called up to participate in a restricted number of NHL games given roster needs. Since game level data was used for the roster model, forwards and defensemen who were excluded from the cluster analysis were assigned to the fourth line and bottom pairing subsequently.

For the 2010-11 regular season, 631 forwards and 334 defensemen participated in at least one regular season game. Skaters who logged less than an explicit portion of ice time for the duration of the regular season, were removed from the cluster analysis. The histogram of the total time on ice distribution for forwards peaked at 187 minutes, indicating that a significantly high quantity of forwards played less or a maximum of 187 minutes for the duration of the regular season (Figure 1). The histogram of the season-long time on ice distribution for defensemen peaked at 223 minutes, implying that a substantial number of defensemen were on the ice for 223 minutes or less (Figure 2). 151 forwards and 68 defencemen (219 total) were discarded from the cluster process as their time on ice failed to exceed the minimum minutes required. The final number of players encompassed in the cluster process for the 2010-11 season were 198 centers, 282 wingers and 267 defensemen (747 total).

Of the 1056 players who played a minimal of one game for the 2011-12 regular season, 269 were centers, 374 were wingers and 324 were defensemen. The accumulated time on ice allocation for forwards spiked at 192 minutes, signifying that forwards had to exceed this time limit to be evaluated. 207 centers and 281 wingers played more than 192 minutes for the duration of the regular season and were included in the clustering process. For defensemen, the overall time on ice appropriation culminated at 230 minutes, revealing that 268 defensemen registered a total time on ice higher than 230 minutes. 62 centers, 93 wingers and 56 defencemen were eliminated from the cluster stage and were directly classified as fourth line forwards and bottom pairing defensemen.

Forwards and defensemen who dressed up for at least one game for the 2012-13 regular season, were 590 and 305. Since the total time on ice allotment for forwards was maximized at 110 minutes, centers and wingers needed to log more than 110 minutes on ice to be involved in the cluster analysis. 48 centers and 60 wingers registered 110 minutes or less for the span of the regular season and were not included in the player ranking procedure. The overall time on ice appropriation for defensemen climaxed at 140 minutes, insinuating that a given defensemen had to be on the ice for more than 140 minutes to be encompassed in the player evaluation phase. A total of 52 defensemen were on the ice for 140 or fewer minutes and were eliminated. The definitive quantity of skaters incorporated in the cluster analysis for the 2012-13 regular season were 197 centers, 285 wingers and 253 defensemen.

For the 2013-14 regular season, 319 defensemen and 628 forwards played a minimal of one regular season game. The graph of the total time on ice distribution for defensemen peaked at 254 minutes, indicating that a significantly high volume of defensemen played less or a maximum of 254 minutes for the duration of the regular season. 253 defensemen surpassed the time frame of 254 minutes and were implicated in the cluster method. The graph of the season-long time on ice dispersion for forwards reached its maximum point at 180 minutes. 208 centers and 279 wingers registered a total time on ice superior to 180 minutes and were ranked. 65 defensemen, 63 centers and 77 wingers were rejected from the cluster stage as their overall time on ice was below the minimum minutes obligated and were ranked as bottom pairing defensemen and fourth line forwards accordingly.

Of the 1064 athletes who participated in at least one game for the 2014-15 regular season, 344 were defensemen, 280 were centers and 342 were wingers. The chart of the accumulated time on ice allotment for defensemen spiked at 243 minutes, specifying that a given defenseman had to have an overall time on ice greater than this time limit to be involved in the evaluation method. 65 defensemen were on the ice for less or a maximum of 243 minutes and therefore were eliminated from the cluster process. The chart of the overall time on ice allocation for forwards climaxed at 174 minutes. For a given center or winger to be involved in the ranking phase, their total time on ice needed to exceed 174 minutes. 55 centers and 78 wingers logged an aggregated time on ice below the required minimal minutes and were eliminated. The concluding number of skaters incorporated in the cluster analysis for the 2014-15 season were 269 defensemen, 225 centers and 264 wingers.

Defensemen and forwards that suited up for a minimal of one game for the 2015-16 regular season, were 330 and 641 accordingly. Since the season-long time on ice distribution for defensemen reached its maximum point at 240 minutes, defensemen were required to log more than 240 minutes on ice to be included in the player ranking stage. 259 defensemen successfully registered more than 240 minutes on ice for the length of the regular season whereas 71 were unsuccessful and therefore excluded. The overall time on ice dispersion for forwards climaxed at 175 minutes, implying that a winger or center had to surpass the playing time of 175 minutes to be involved in the player evaluation procedure. A total of 277 wingers and 222 centers were on the ice for more than 175 minutes whereas 85 wingers and 57 centers were removed.

For the 2016-17 regular season, 633 forwards and 318 defensemen participated in a minimal one regular season game. The histogram of the total time on ice allocation for forwards spiked at 180 minutes, revealing that a significantly high quantity of forwards played less or a maximum of 180 minutes for the duration of the regular season. For a given center or winger to be encompassed in the cluster analysis, his total time on ice needed to be higher than 180 minutes. The histogram of the overall time on ice distribution for defensemen peaked at 240 minutes, suggesting that an essential volume of defensemen played less or a maximum of 240 minutes for the season. 62 centers, 87 wingers and 66 defencemen were eliminated from the cluster process as their overall time on was below the specified minimum time frame.

Defensemen and forwards that dressed up at least once 2017-18 regular season, were 324 and 640 respectively. The season-long time on ice dispersion for forwards reached its climax at 185 minutes, implementing that forwards were required to log more than 185 playing minutes to be included in the player ranking stage. 54 centers and 80 wingers registered less than or equal to 240 total minutes on ice and were discarded. The overall time on ice allotment for defensemen culminated at 225 minutes. For a given defensemen to be involved in the player evaluation procedure, his time on ice had to surpass the time restriction of 225 minutes. 58 defensemen failed to exceed the time limit of 225 minutes and were removed. The quantity of skaters encompassed in the cluster process for the 2017-18 season were 206 centers, 300 wingers and 265 defensemen.

Trades between NHL teams that involve the exchange of players occur throughout the season up until the trade deadline. Since players were being evaluated and ranked on their overall performance, season-long statistics were computed and applied in the clustering process. Whether the performance of a traded player improved or declined after the trade had minimal to insignificant effect on the results from the cluster analysis. An additional issue that transpired with the use of play-by-play data, was skaters changing jersey numbers within the regular season. This concern needed to be addressed adequately to prevent the aggregation of incorrect statistics and the allocation of those season-long statistics to inaccurate players. Imprecise variables would have tampered with the outcome of the cluster analysis, as some skaters would have been ranked higher or lower than their actual seasonal performance.

Centers, wingers and defensemen were clustered separately. The variables used to cluster all positions were: points per time on ice, goals per time on ice, assists per time on ice, plus/minus per time on ice, shots differential per time on ice, blocks per time on ice, hits per time on ice and penalties per time on ice. Time on ice for forwards is distributed amongst four lines whereas time on ice for defensemen is allocated to three pairings. Since all statistics were divided by time on ice, it was vital to cluster forwards and defensemen separately. Despite the use of the same set of variables, each position may exceed in a specific statistic or statistics. Therefore, the reason for implementing cluster analysis independently, was to equally compare and rank skaters who played in the same position.

Elbow Analysis

To determine the optimal number of clusters for k-means clustering, both elbow and silhouette analysis were conducted. As mentioned previously, elbow method explores the percentage of variance demonstrated as a function of the quantity of clusters. The concept is to begin with two clusters (k=2) and to increase the number of clusters each time by one for a range of values of k and to compute the sum of squared errors (SSE) for each value of k. Each time the quantity of clusters increases, SSE decreases towards zero. SSE will equal to zero when k is equal to the total number of observations in the data, as each observation is its own cluster, leaving no error. The objective is to select a small number of k that has a low SSE that reaches a plateau when it is increased further. K-means clustering for all seasons and all positions was conducted for a range of 10 clusters and SSE was calculated for each value of k. Additionally, line charts of the SSE per position were plotted to assist with the interpretation and visualization of the results.

For the 2010-11 regular season, the SSE for each value of k were computed for each position separately (Table 1). The change in SSE for centers when adding a third cluster was relatively higher (165.45) than the change in SSE when increasing the number of clusters from three to four (93.07). Clusters five and six had a variation in the value of the SSE approximately the same, 61.79 and 62.41 accordingly. The line chart displayed that the reduction of SSE decreased significantly after four clusters (Figure 3). For wingers, after advancing the number of clusters from five to six, the depreciation in the value of SSE dropped considerably (6.82% from 10.51%). A graph of the SSE for all 10 clusters was created (Figure 4) to facilitate with the explanation of the outcomes. A substantial drop in the deprecation of the value of SSE was observed beyond five clusters. After that point, increasing the number of clusters had a diminishing return. When the quantity of clusters for defensemen raised from four to five, the deduction in the SSE was significantly low (84.21) in comparison to the reduction in SSE when moving to three clusters four clusters (157.07). This also can be noticed in the line chart for defensemen, as beyond four clusters, reduction of SSE decreased essentially (Figure 5).

The SSE for the 2011-2012 regular season were estimated for centers, wingers and defensemen independently (Table 3). When examining the value of the SSE for centers, it was reduced by 175.07 by increasing the number of clusters from two to three. The volume of reduction when moving from three clusters to four declined significantly (10.32%) while a fifth cluster, had a slightly higher rate of deduction (10.43%). For wingers, the size of reduction in the value of SSE by increasing the quantity of clusters from two to three was 16.87% which was substantially greater than the change in the SSE for four clusters (9.65%). The variation in the SSE for clusters eight and nine were relatively equal with a differential of 0.02 (4.93% and 4.91%, accordingly). As for defensemen, the change in the rate of depreciation of SSE for three and four clusters were both around 10% in contrast to the range after six clusters, which dropped around 6%.

For the 2012-13 regular season, the SSE for ten clusters were estimated for all skater positions individually (Table 5). The inclusion of a third cluster deducted the value of SSE for centers by 158.77 which translated to a change of 13.80%. An efficient slump in the reduction rate was noticed when advancing the number of clusters to four, five and six as it ranged around 10%. When the number of clusters for wingers increased to four, SSE was depreciated by 9.49% which was substantially low in comparison to the decline in the value of SSE when advancing from two clusters to three (16.11%). Additionally, the ninth cluster had a greater impact on SSE than the eighth as it reduced the SSE by 55.45 in contrast to 40.57. For defensemen, extending the amount of clusters to three altered the cluster errors the most (13.23%). The variation in the SSE when advancing the quantity of clusters to six (61.10) was lesser than the variation for seven clusters (61.49).

SSE estimations for the 2013-14 regular season for each value of k were assessed for centers, wingers and defensemen solely (Table 7). For centers, the reduction of the SSE was at 16.24% when increasing the portion of clusters from two to three. By increasing k by one unit (k=4), SSE was decreased by 10.90%. Clusters five and six altered the SSE almost by equal portion, 63.43 and 62.03 subsequently. The variation in the value of SSE for wingers was considerably higher (17.36%) when adding a third cluster in contrast to a fourth (11.09%). A similar change rate was discovered for clusters six and seven, as the SSE was depreciated by 6.39% and 6.22%, accordingly. The SSE defensemen was 1263.08, which was reduced by 15.35% with the addition of another cluster (k=3). After advancing the number of clusters to four, the depreciation in the value of SSE dropped considerably (10.97%). Identically to wingers, the deduction in the SSE for six and seven clusters was 6.38% and 6.23%,.

The SSE for the 2014-2015 regular season were estimated for each position independently (Table 9). As displayed in the table, the shift in SSE for centers when increasing the number of clusters to three was 14.19%. A fourth cluster decreased the cluster errors by 133.70 (12.35%). An essential drop in the depreciation of the value of SSE was noticed with the introduction of a fifth cluster, as it had an impact of 8.62%. The effect a third and fourth cluster had on the SSE for wingers was relatively close, as SSE was reduced by 12.52% and 11.52 respectively. The change in SSE when adding a seven cluster was slightly higher (6.77%) than the change in SSE when increasing the number of clusters from five to six (6.58%). For defensemen, the deduction in the SSE was relatively higher (247.43) than the change in SSE when advancing the number of clusters to four (154.49). The decline rate in the value of SSE for clusters eight (5.54%) and nine (5.37%) were marginally greater than the decline rate of the seventh cluster (5.26%).

For the 2015-16 regular season, the SSE for each value of k were computed for all skater positions separately (Table 11). The variance in SSE for centers when adding a third cluster was 12.68% while the change in SSE when increasing the number of clusters to four was 10.48%.

Clusters six and seven deducted the value of the SSE almost identically by 5.48% and 5.49%, subsequently. By adding an additional cluster (k=8) the change in the SSE was slightly higher than both, with a rate of 5.88%. For wingers, after advancing the number of clusters to three, the switch in the value of the SSE (16.23%) was greater than the switch in SSE when moving from three clusters to four (10.65%). A substantial drop in the deprecation of the value of SSE can be observed beyond five clusters. A ninth cluster had a greater impact on the SSE (5.31%) than a seventh (5.05%) and eighth (3.35%). When the quantity of clusters for defensemen increased to three, the alteration in the SSE was 15.22% .The reduction in SSE when moving for four and five clusters were relatively close and around 10%. Seven clusters decreased the value of the SSE by 5.12% which was lower than the depreciation of the SSE by eight clusters which was 6.22%.

SSE estimations for the 2016-17 regular season for ten clusters were assessed for centers, wingers and defensemen solely (Table 13). For centers, the addition of a third cluster shifted the value of the SSE by 1040 which represented a depreciation volume of 14.89%. When moving from three clusters to four, SSE was deducted by 11.38%. The alteration in the SSE for six and seven clusters were in the same range, around 6%. By increasing the portion of clusters to three, the SSE for wingers was declined by 234.09 (15.49%). An additional cluster (k=4) switched the SSE by 11.33 % which was higher than the effect of a fifth cluster (9.02%). SSE for defensemen was reduced by 13.22% with the inclusion of a third cluster. An efficient drop (4.91%) in the deprecation of the value of SSE was observed for five clusters, as reduction rate declined from 11.94% to 7.03%. The change in the SSE for seven clusters (5.82%) was marginally lower than the change in SSE by eight clusters (5.91%).

The SSE for the 2017-2018 regular season were calculated for each position individually (Table 15). The variation in the SSE for centers when increasing the quantity of clusters to three was significantly higher (14.43%) than the variation in the SSE for four clusters (9.49%). A fifth cluster had a lower effect on the SSE than a sixth, as its reduction rate was 6.96% in comparison to 7.55%. For wingers, advancing the number of clusters to three depreciated the SSE by 263.78 which represented a 15.16% decrease. The SSE was declined by 13.61% and 10.03% when the amount of clusters was increased by one unit (k=4) and two units (k=5), accordingly. A 13.32% depreciation in the SSE for defensemen, was a result of advancing the quantity of clusters from two to three. An extra cluster (k=4) declined the SSE by 13.61 % while the deduction in the SSE from a fifth cluster was lower (10.19%). The change in the SSE for seven and eight clusters were equivalent and slightly above 6%.

Based upon results from the elbow analysis for all regular seasons, the optimal number of clusters for each position was three. Despite the optimal number of clusters for centers and wingers was three, every team has four lines of forwards. Because the rosters of each team were examined in the regression model, the number of selected clusters for forwards was four. In situations where the data is not clustered properly, the elbow method produces incorrect results. Even with the plot of chart lines, it can be unclear to locate and determine the best value of k for each position. To verify that the data was clustered accurately and that the quantity of clusters chosen for every roster position were right, silhouette method was applied.

Silhouette Analysis

Silhouette analysis computes silhouette scores to establish the optimal number of clusters. It is a technique used to estimate the separation span between clusters. The silhouette score assesses the way how dissimilar an observation in a given cluster is to the alternative clusters created and how similar it is to its own cluster. Silhouette score has a range of [-1,1], where a high value (close to +1) signifies that the observation is insufficiently matched with bordering clusters and sufficiently matched with its own cluster. A low value (close to -1) suggests that the observation is sufficiently matched with bordering clusters and insufficiently matched with its own cluster. This implies that an observation might have been appointed to the inaccurate cluster. A value of zero indicates that the observation lies between two bordering clusters. A total number of 10 clusters were created for each roster position per season, for eight seasons.

To select the optimal number of clusters for centers in the 2010-2011 regular season, silhouette scores (Table 2) along with silhouette plots were conducted (Figure 6). The number of clusters with the highest average silhouette score were two, four and five with values 0.2682 0.1907, 0.1982 respectively. Additionally, the thickness of the silhouette plots assisted with the visualization and examination of the formed clusters. The silhouette plots indicated that the number of clusters with the highest silhouette score were two, four and five. The other number of clusters were considered unsuitable as their average silhouette scores were smaller and because of the wide variations in the volume of the silhouette plots.

The same process was followed for wingers in the 2010-2011 regular season. Silhouette scores computed (Table 2) and silhouette graphs were displayed (Figure 7). The quantity of clusters with the greatest average silhouette score were two, three, four, five and six with values 0.3117, 0.2276, 0.1739, 0.1844 and 0.175 accordingly. The silhouette graphs and plots facilitated in the interpretation and observation of the created clusters. The silhouette plot signified that the quantity of clusters with the highest silhouette score were two, three, four. The remaining quantity of clusters (7, 8, 9, 10) were regarded inadequate as a result of the broad variation in the content of the silhouette plots and because their average silhouette scores were lower.

The average silhouette scores were calculated (Table 2) and the silhouette plots were designed (Figure 8) to determine and choose the optimal number of clusters for defensemen in the 2010-2011 regular season. The quantity of clusters with the greatest average silhouette score were two, three, four with values 0.2334, 0.1682, 0.164 correspondingly. The silhouette plots expedited in the interpretation and observation of the created clusters. The silhouette plot illustrated that the number of clusters with the highest silhouette score were two, three, and four. The additional clusters (5, 6, 7, 8, 9, 10) were considered insufficient, as their average silhouette scores were lower and their silhouette plots had a wide variation in size.

For the 2011-2012 regular season, the average silhouette scores for centers, wingers and defensemen were calculated independently (Table 4). The clusters with the highest average silhouette score for centers were two, three, and four with values 0.2510, 0.2121, and 0.1776, respectively. These higher scores indicated that the referred clusters were a good fit for the optimal number of clusters. The additional clusters were considered unsuitable as their average silhouette scores were relatively lower. The average silhouette scores of the first three clusters (k= 2, 3, 4) for wingers displayed a greater value than the other six clusters, which suggested that the ideal quantity of groups was not larger than four. The deduction in the average silhouette scores for defensemen for five clusters and beyond, signified that for the selected variables clusters below five were more suitable for the assembling process.

The average silhouette scores were computed to establish the ideal number of clusters per position for the 2012-2013 regular season (Table 6). Upon completion of the average silhouette scores for centres, the clusters that had the greatest average silhouette scores were two, three, four, six and nine with values 0.2298, 0.1794, 0.1513 and 0.1573 and 0.1583 accordingly. Clusters with a lower score were detected as unfit for the allotment of centers. As for wingers, clusters five, eight, nine and ten had the lowest average silhouette scores (0.1498, 0.1430, 0.1514 and 0.1427) and were therefore regarded as inadequate for the distribution of wingers. The clusters with the highest average silhouette scores for defensemen were two (0.2103), three (0.1847) and four (0.1586). The remaining quantity of clusters were considered unsuitable for the allocation of defensemen due their low average silhouette scores.

To choose the optimal number of clusters for all skater positions, ten average silhouette scores were estimated for the duration of the 2013-14 regular season (Table 8). For centers, the implementation of two, three and four clusters had greater average silhouette scores than the additional clusters with values 0.2438, 0.2400 and 0.1570 correspondingly. These above clusters were considered to categorize centers into groups appropriately. When examining the average silhouette scores for wingers, clusters two, three, four and five (0.2798, 0.2448, 0.1792, 0.1811) had the largest value. Beyond five clusters, the average silhouette scores were diminished, specifying that the ideal portion of clusters were below five. The allocation of defensemen to two, three and four clusters had an average silhouette score of 0.2257, 0.1943 and 0.1760, subsequently. The remaining clusters had an inferior average silhouette score and were determined unfit for the dispersion of defensemen.

For the 2014-2015 regular season, the average silhouette scores for centers, wingers and defensemen were calculated separately (Table 10). Each of the first three average silhouette scores (k=2, 3, 4) had a larger value than the average silhouette score of the remaining six clusters, revealing that the dispersion of centers above five was insufficient. For wingers, the average silhouette scores for clusters two, three, four and five clusters reached a value of 0.2941, 0.2464, 0.1744 and 0.1803, accordingly. The average silhouette score of the additional five clusters were lower, which suggested that the ideal portion of clusters was below six. Clusters two, three and four had the highest average silhouette score for defensemen with a value of 0.2290, 0.1850 and 0.1950, respectively. The deduction in the value of the average silhouette score for five clusters and beyond, implied that for the selected set of variables, clusters below five were more suitable for the distribution process.

The average silhouette score was estimated to establish the ideal number of clusters per position for the 2015-2016 regular season (Table 12). The volume of clusters with the greatest average silhouette score for centers were two, three, four and five with values 0.2633, 0.2282, 0.1598 and 0.1683, subsequently. The rest of the clusters (k>5) had a lower average silhouette and were therefore considered inadequate for the proper distribution of centers. The allocation of wingers into two, three and four clusters generated an average silhouette score of 0.2796, 0.2492 and 0.1762, correspondingly. When increasing the number of cluster above four, a depreciation in the average silhouette score occurred, which illustrated that the ideal number of clusters, based on the given set of variables, were below five. The clusters with the highest average silhouette scores for defensemen were two (0.2106), three (0.2008) and four (0.1707). The remaining six clusters were viewed unfitted for the allocation of defensemen due their low average silhouette scores.

The selection of the optimal number of clusters for all skater positions for the duration of the 2016-17 regular season were based upon the estimated average silhouette scores (Table 14). The application of two, three and four clusters for centers had larger average silhouette scores than the additional clusters, with values 0.2492, 0.2011 and 0.2000, respectively. These high scores revealed that the optimal number of clusters for centers were below five. When examining the average silhouette scores for wingers, clusters two, three, four and five (0.2467, 0.2207, 0.1697, 0.1815) had the highest values. Beyond five clusters, the average silhouette scores were reduced, signifying that the ideal portion of clusters were below five. The distribution of defensemen to two, three and four clusters had an average silhouette score of 0.2222, 0.1909 and

0.1737, subsequently. The remaining clusters had a mediocre average silhouette score and were determined insufficient for the dispersion of defensemen into groups.

The average silhouette score for ten clusters were calculated to establish the optimal number of clusters per position for the 2017-2018 regular season (Table 16). The number of clusters with the highest average silhouette score for centers were two, three and four with values 0.2522, 0.1882 and 0.1680, accordingly. The rest of the clusters (k>4) had an inferior average silhouette score and were therefore viewed unsuitable for the proper allocation of centers. The dispersion of wingers into two, three and four clusters generated an average silhouette score of 0.2528, 0.2006 and 0.1985, respectively. When advancing the number of cluster beyond four, a deduction in the average silhouette score transpired, which implied that the optimal amount of clusters, based on the given set of variables, were below five. The clusters with the greatest average silhouette scores for defensemen were two (0.2138), three (0.1841) and four (0.1769). The additional six quantity of clusters were considered inadequate for the distribution of defensemen as a result of their low average silhouette scores.

Results from the silhouette analysis illustrated that the ideal quantity of clusters for each position was two, since it had the highest average silhouette score for every season. However, the choice of the optimal number of clusters for forwards and defensemen was guided by the assembling of NHL teams roster. Every roster includes four lines of forwards and three pairings of defensemen. The optimal number of clusters chosen for centers and wingers were four: top line, second line, third line and fourth line, while the optimal number of clusters selected for defensemen were three: top two pairing, top four pairing and bottom pairing.

Skater Rankings

Upon completion of k++ clustering method, the optimal number of clusters generated from the selected set of variables, needed to be ranked. The formed clusters along with the value of each variable for each position and regular season were displayed separately. Since all variables were scaled by the total time on ice per season, their value was the amount of standard deviations above the mean for all centers / wingers or defensemen for that given regular season.

The assembled clusters in addition with their value for each variable per position for the 2010-11 regular season were presented in Table 17. The cluster that had the highest value for offensive related variables, contained players with the most offensive skills. For centers, cluster two had a value of 1.1751 for points per time on ice, 0.9984 goals per time on ice and 1.0713 assists per time on ice. This meant that elite forwards had 1.1751 standard deviations above the mean for all centers in terms of points. The cluster with the second highest offensive variables included top six skaters. Cluster one had 0.0548 standard deviations above the mean for points per time on ice, 0.0239 standard deviations above the mean for goals per time on ice and 0.065 standard deviations above the mean for assists per time on ice. The third forward line is assembled by players with defensive skills. Cluster three had the biggest value in blocks per time on ice variable (1.1002). Players on the bottom line are known for their aggressive style of play and are often penalized. Cluster zero had the greatest value in both hits per time on ice and penalties per time on ice, 2.0549 and 2.4305 proportionately.

For wingers, cluster three had the highest standard deviations above the mean for offensive variables with 1.2421 for points per time on ice, 0.9684 goals per time on ice and 1.1338 assists per time on ice, 0.7922 plus/minus per time on ice and 0.3592 even strength shot differential per time one ice. Cluster one had the second highest offensive variables with 0.0705 standard deviations above the mean for points per time on ice, 0.104 standard deviations above the mean for goals per time on ice and 0.0254 standard deviations above the mean for assists per time on ice, and consisted of top six wingers who play on the second line of offense. Cluster zero had the biggest value in blocks per time on ice (0.6816) variable and the third highest value for all offensive variables. The fourth line included skaters that had the lowest offensive contribution but had the greatest standard deviation above the mean for hits and penalties per time on ice with 2.4478 and 3.2484, subsequently.

The cluster with the highest value for offensive related variables contained top pairing defensemen, who had the ability to generate offense. Cluster one had 1.2755 standard deviations above the mean for points per time on ice, 1.1236 goals standard deviations above the mean for goals per time on ice and 1.1509 standard deviations above mean for assists per time on ice, 0.2166 standard deviations above mean for plus/minus per time on ice and 0.1178 standard deviations above the mean for even strength shot differential. The cluster with the second highest offensive variables contained top four defensemen. Cluster zero had 0.2105 standard deviations below the mean for points per time on ice, 0.1953 standard deviations below the mean for goals per time on ice and 0.1861 standard deviations below the mean for assists per time on ice. Bottom pairing defensemen are known for their defensive and psychical skills. Cluster two had the highest value in blocks, hits and penalties per time on ice with values of 0.0938, 0.8531 and 0.6217, accordingly.

Upon finalization of the distribution of all cluster labels to their respectful ranking value, the quantity of players per position and rank were listed. For centers, 53 were rated as first, 79 as second, 49 as third and 17 as fourth line centers. For wingers, 73 were classified as top three, 119 as top six, 73 as bottom six and 17 as bottom three wingers. For defensemen, 72 were regarded as top two, 113 as top four and 82 as bottom defensive pairing. The players that were excluded from the cluster analysis were assigned to the fourth line for forwards and bottom pairing for defensemen. With the addition of those players, centers, wingers and defensemen increased to 84, 101 and 150, correspondingly.

For the 2011-12 regular season, the composed clusters as well as their value for each variable per position were exhibited in Table 18. The cluster consisted of the most offensive centers, was cluster three as all offensive attributes had standard deviations above the mean. This implied that elite centers had 1.2372 standard deviations above the mean for all centers in terms of points per time on ice. Cluster zero contained top six centers, as it had the second largest standard deviations above the mean for offensive variables. Cluster one accommodated fourth line centers as it had the lowest value in offensive contribution and the largest value in hits and penalties per time on ice, 1.644 and 1.4969, respectively. Third line centers were assigned to cluster two, which had the third largest contribution for offensive variables.

Cluster three encompassed top line wingers as it had the highest standard deviations above the mean for offensive variables. Cluster two consisted of top six wingers as it had the second highest standard deviations above the mean for offensive attributes. Cluster one had the largest number in blocks per time on ice (0.577) and the third highest value for all offensive variables. The fourth line wingers were located in cluster zero as offensive contribution were minimum while hits and penalties per time on ice were the largest with 1.7739 and 2.7906, subsequently. Top two pairing defensemen were incorporated in cluster two, as offensive attributes were substantially higher. Top four pairing defensemen were detected in cluster one, as it balanced both defensive and offensive variables. Cluster zero contained bottom pairing defensemen as it had the largest value for hits (1.2203) and penalties (0.8660) per time on ice above the mean and the smallest offensive contribution.

After finishing the dispersion of each cluster label to its appropriate ranking value, the quantity of skaters per position and rank were recorded below. From 207 centers, 57 were ranked as top three, 73 as top six, 49 as bottom six and 28 as bottom three forwards. From 281 wingers, 88 were classified as first, 97 as second, 73 as third and 23 as fourth line wingers. From 268 defensemen, 98 were regarded as top two, 100 as top four and 70 as bottom defensive pairing. With the addition of the skaters that were eliminated from the cluster procedure, centers, wingers and defensemen increased to 90, 116 and 126, respectively.

The created clusters in conjunction with their value for each variable per skater position for the 2012-13 regular season were demonstrated in Table 19. Top line and second line centers were enclosed in clusters zero and two, accordingly, as they had the biggest standard deviations above the mean for all offensive variables. Fourth line centers were involved in cluster one, as it had the biggest value for hits and penalties per time on ice while cluster three contained third line centers. For wingers, cluster three and two had the largest standard deviations above mean for attacking attributes and incorporated first and second line wingers, correspondingly. Cluster zero included third line wingers as it had the highest standard deviations above the mean for blocks per time on ice (0.7735). Fourth line wingers were detected in cluster one with the highest value for hits (1.7902) and penalties per time on ice (2.0783). Top two and four pairing defensemen were incorporated in cluster one and zero, respectively, as offensive attributes were higher. Cluster two contained bottom pairing defensemen as it had the greatest value for hits and penalties per time on ice above the mean along with the smallest offensive contribution.

Upon fulfillment of allocating the proper ranking value to every cluster label, the amount of centers, wingers and defensemen per rank were registered. For centers, 68 were classified as first, 54 as second, 51 as third and 24 as fourth line centers. For wingers, 70 were ranked as top three, 101 as top six, 77 as bottom six and 37 as bottom three forwards. For defensemen, 94 were ranked as top two, 117 as top four and 42 as bottom defensive pairing. When including the players that were removed from the cluster process, fourth lines centers, wingers and bottom pairing defensemen increased to 84, 97 and 94, accordingly.

For the 2013-14 regular season, the formed clusters as well as the value of all variables for centers, wingers and defensemen were featured in Table 20. Clusters two and one had the highest standard deviations above the mean for offensive contribution and incorporated first and second line centers, correspondingly. The lowest value for attacking attributes were registered by cluster three, indicating that it contained fourth line centers. The remaining cluster (zero) involved third line centers. In terms of wingers, clusters zero and three had the largest standard deviations above the mean for offensive variables and enclosed top and second line wingers. Cluster one encompassed fourth line wingers as it had the highest value for hits and penalties per time on ice. Third line wingers were located in cluster two. Clusters two and one had the highest values in attacking attributes and were assigned top two and four pairing defensemen, accordingly. Bottom pairing defensemen were enclosed in cluster zero as it had the lowest offensive contribution and the greatest value for blocks, hits and penalties per time on ice, with 0.7329, 0.9144 and 0.96416, respectively.

Upon completion of the distribution of each cluster label to its applicable ranking value, the number of players per position and rank were specified. For centers, 55 were rated as top three, 72 were as top six, 69 as bottom six and 12 as bottom three forwards. For wingers, 69 were classified as first, 99 as second, 86 as third and 25 as fourth line wingers. For defensemen, 71 were ranked as top two, 115 as top four and 67 as bottom defensive pairing. With the inclusion of the players that were omitted from the cluster analysis, centers, wingers and defensemen increased to 96, 102 and 132, correspondingly.

The structured clusters in addition to their value for every variable per position for the 2014-15 regular season were revealed in Table 21. First and second line centers were contained in clusters one and two correspondingly, as they registered the largest values for every offensive attribute. Cluster three consisted of third line centers while cluster zero included fourth line centers, as it had the highest value for blocks, hits and penalties per time with 0.7522, 2.0866 and 2.6043, accordingly. For wingers, cluster one and three had the largest standard deviations above mean for attacking attributes and encompassed first and second line wingers, correspondingly. Cluster zero enclosed third line wingers as it recorded the largest value for blocks per time on ice (1.0519). Fourth line wingers were located in cluster two at it listed the highest value for hits (1.6464) and penalties per time on ice (1.9322). Top two and four pairing defensemen were assigned to clusters zero and one, as offensive variables were the highest and second highest, respectively. Cluster two contained bottom pairing defensemen as it recorded the lowest contribution to offensive production.

After finishing the allocation of every cluster label to its respectful ranking value, the quantity of skaters per position and rank were listed below. From 225 centers, 51 were ranked as first, 98 as second, 58 as third and 18 as fourth line centers. From 264 wingers, 61 were ranked as top three, 108 as top six, 61 as bottom six and 34 as bottom three forwards. From 269 defensemen, 69 were regarded as top two, 109 as top four and 91 as bottom defensive pairing.

By adding the skaters who were eliminated from the cluster phase, the quantity of fourth line centers, wingers and bottom pairing defensemen raised to 73, 112 and 165, respectively.

For the 2015-16 regular season, the established clusters as well as their value for each variable and position were displayed in Table 22. Cluster zero and three had the highest and second highest values for all offensive variables as they included first and second line centers, correspondingly. Cluster one contained third line centers while fourth line centers were encompassed in cluster two, as it registered the lowest value for each attacking attribute. In terms of wingers, cluster zero and two had considerable higher standard deviations above mean for attacking attributes and consisted of first and second line wingers, accordingly. Cluster three accommodated third line wingers as it recorded the largest value for blocks per time on ice (0.7996). Fourth line wingers were enclosed in cluster one with the highest standard deviations above mean for hits and penalties per time on ice with 1.8133 and 2.4308, respectively. Clusters two and one contained top two and four pairing defensemen, as they registered the highest and second highest values for all offensive variables, accordingly. Cluster zero consisted of bottom pairing defensemen as blocks, hits and penalties per time on ice were above the mean by 0.5345, 15049 and 1.3137 standard deviations, correspondingly.

Upon fulfillment of dispersing the proper ranking value to all cluster labels, the amount of centers, wingers and defensemen per rank were registered. For centers, 58 were classified as top three, 82 as top six, 60 as bottom six and 22 as bottom three forwards. For wingers, 74 were ranked as number one, 105 as number two, 69 as number three and 29 as number four wingers. For defensemen, 63 were rated as top two, 144 as top four and 52 as bottom defensive pairing. When including the skaters that were removed from the cluster stage, fourth lines centers, wingers and bottom pairing defensemen increased to 79, 114 and 123, accordingly.

The assembled clusters as well as their value for each variable per position for the 2016-17 regular season were presented in Table 23. Top line and second line centers were contained in clusters one and two, as they had the largest and second largest standard deviations above the mean for each offensive variable, respectively. Fourth line centers were encompassed in cluster three, as it recorded the largest values for blocks, hits and penalties per time on ice with 0.7518, 2.0485 and 1.9532 standard deviations above the mean, correspondingly. The remaining cluster (zero) enclosed third line centers. For wingers, cluster three and one had the largest and second largest values for attacking attributes and included first and second line wingers, respectively. Cluster zero incorporated third line wingers as it had the highest standard deviations above the mean for blocks per time on ice (0.6495). Fourth line wingers were detected in cluster two with the highest values for hits (2.483) and penalties per time on ice (3.4898). Top pairing defensemen were enclosed in cluster two, as it registered the highest standard deviations above the mean for all offensive attributes while cluster zero incorporated top four pairing defensemen. Bottom pairing defensemen were contained in cluster one, as it had the highest value for hits and penalties per time on ice with 1.3182 and 1.3882 standard deviations above the mean, accordingly.

Upon completion of the distribution of each cluster label to its applicable ranking value, the number of players per position and rank were specified. For centers, 67 were ranked as number one, 85 as number two, 44 as number three and 22 as number four centers. For wingers, 84 were rated as first, 99 as second, 70 as third and 13 as fourth line wingers. For defensemen, 72 were classified as top two, 128 as top four and 52 as bottom defensive pairing. With the inclusion of the players that were omitted from the cluster analysis, centers, wingers and defensemen increased to 84, 100 and 118, correspondingly.

For the 2017-18 regular season, the formed clusters in addition with the value of all variables per position were featured in Table 24. Cluster two and zero had the highest and second highest values for all offensive attributes as they enclosed first and second line centers, respectively. Cluster three encompassed third line centers while fourth line centers were enclosed in cluster one, as it recorded the lowest value for each attacking variable. In terms of wingers, cluster one and zero had substantial higher standard deviations above mean for attacking attributes and contained first and second line wingers, accordingly. Cluster two consisted of third line wingers. Fourth line wingers were included in cluster three with the highest standard deviations above mean for blocks, hits and penalties per time on ice with 0.8292, 2.7171 and 4.3438, correspondingly. Clusters two and zero contained top two and four pairing defensemen, as they illustrated the highest and second highest values for each offensive variable, accordingly. Cluster one enclosed bottom pairing defensemen as hits and penalties per time on ice were above the mean by 1.0945 and 0.8241 standard deviations, respectively.

Upon finalization of the allocation of each cluster label to its respectful ranking value, the quantity of skaters per position and rank were listed. From 206 centers, 50 were regarded as first, 68 as second, 43 as third and 45 as fourth line centers. From 300 wingers, 62 were ranked as top three, 139 as top six, 89 as bottom six and 10 as bottom three forwards. From 268 defensemen, 77 were rated as top two, 125 as top four and 63 as bottom defensive pairing. By including the skaters that were eliminated from the cluster procedure, fourth lines centers, wingers and bottom pairing defensemen increased to 99, 90, and 121, accordingly.

Upon completion of cluster analysis, the estimation of the linear roster model is interpreted below. All forwards and defensemen who were excluded from the cluster analysis were assigned to the fourth line for forwards and bottom pairing for defensemen.

Roster Model Analysis

The linear model to describe additional wins in terms of roster position:

$$\mathrm{Win}_{s,g,t} = \beta_1 \mathrm{MeanC}_{s,g,t} + \beta_2 \mathrm{MeanW}_{s,g,t} + \beta_3 \mathrm{MeanD}_{s,g,t} + \varepsilon_{s,g,t}$$

To estimate the impact each roster position had on team success, game level data was used. Since there were two opponents per game, two observations were created, one for each team. The dependent variable, win probability, has a value of one for the winning team and a value of zero for the losing team. Games with four centers, eight wingers and six defensemen, for both home and visiting team, formed the data set for the linear regression. The quantity of games with the equal amount of players in each position for both teams were: 156 for the 2010-11 regular season; 154 for the 2011-12 regular season; 60 for the 2012-13 regular season; 100 for the 2013-14 regular season; 102 for the 2014-15 regular season; 68 for the 2015-16 regular season; 86 for the 2016-2017 regular season and 102 for the 2017-18 regular season. The data set contained 828 observations, which represented nine percent of the total games played for the duration of eight regular seasons (9266).

Prior to the assessment of the value of each roster position in terms of winning, the mean of every position by team and game were calculated. To compute the mean per position, team and game, results from the cluster analysis were used. Since classification was in ascending order, skaters on the top forward line and top defensive pairing were assigned a value of one; players on the second forward line and second defensive pairing a value of two; forwards on the third line and defensemen on the bottom pairing a value of three; and skaters on the fourth line, a value of four. The mean of each position for every team and game was based on the quality of skaters that were on the roster in a given game. The league mean for centers and wingers was 2.5 and for defensemen two. If the mean of any position for any given team was below the league average, it indicated that in that specific position(s) that team had more quality players. If the mean of a position of a team was above the league mean, it revealed that in that particular position(s) that team had less quantity of elite players. The mean per position and team for eight regular season were displayed (Tables 25-32). Teams that were not listed in the tables did not play a single game were both home and visiting team had four centers, eight wingers and six defensemen.

For the 2017-2018 regular season, the Nashville Predators finished on top of the league with 117 points. The mean of centers and wingers for the regular season were both 1.76 and the mean of defensemen was 1.67 (Table 32). All three positions were below the league average, indicating that the Predators had more high quality players in every position. The Calgary Flames finished the league with 84 points and missed the postseason. The mean of centers, wingers and defensemen was 2.50, 2.33 and 2.00, respectively. This showed that the quality of centers and defensemen for the Flames was equal to the league average, but had better quality of wingers than the league average. The Ottawa Senators finished second last in the league with 67 points. The mean of centers, wingers and defensemen was 3.05, 2.54 and 2.3, accordingly. Results implied that the Senators had significantly less quality of centers and defensemen, and slightly less talented wingers than the league average.

To measure the relationship between win probability and player quality per position, ordinary least square (OLS) regression was applied. OLS minimizes the squared distances of the observation to fit the data. In a linear regression model, the best linear predictor of y given x is the one that has the smallest expected squared prediction error. Magnitude displays the size of the impact a coefficient has on the dependent variable. In linear regression, the predictor values are multiplied by the coefficients. The constant for the linear model had a value of 0.395, implying that win probability would increase by that value when the mean of players for each position were zero (Table 33). By increasing the mean of players in the center position by one, win probability increased by 0.081 with a standard error for the mean of centers was 0.041. Changing the mean of wingers by one unit increased win probability by 0.0753 with a standard error of 0.026. Win probability increased by 0.0694 when the mean of defensemen increased by an additional unit. As it can be seen from the results, by adding quality players in each position, win probability raised. All mean positions were statistically significant at the 5% level. The roster position that had the greatest impact on win probability, for the given data set, was centers.

Additional analysis on the effect each roster position had on in-game performance was conducted. For goal differential as the dependent variable, the constant had a negative value of 0.593, indicating that goal differential would decrease by that value if the mean of players of all positions were zero. By increasing the mean of centers by one, goal differential increased by 0.4052 with a standard error of 0.204. Rising the mean of wingers by one unit increased goal differential by 0.4676 with a standard error of 0.130. Goal differential increased by 0.3505 when the mean of defensemen raised by a single unit. All mean positions were statistically significant at the 5% level. Mean wingers and defensemen were statistically significant at the 1% level. When the quality players increased in each position, goal differential increased. The outcomes can be justified by the fact that talented and more skillful skaters have the ability to generate more offense and minimize the offense of the opposition. The roster position that had the greatest impact on goal differential, for the given data set, was wingers.

When regressing goals for on the mean of players by position, the constant for the linear had a value of 2.587, specifying that goals for increased by that value if the mean of players for every position were zero. An increase by one unit in the mean quality of centers, increased goals for by 0.2763. The standard error for the mean of centers was 0.142. Increasing the mean of wingers by a single unit, increased goals for by 0.2427 with a standard error of 0.090. Goals for increased by 0.0605, when the mean of defensemen increased by an additional unit. Only the mean of wingers was statistically significant at the 5% and 1% level. The mean of centers was significant at the 10% level while the mean of defensemen was not statistically significant. Goals for increased as the mean quality of players in each position increased. This can be supported by the fact that higher ranked players have the skills to produce more offense opportunities and goals in comparison to lower ranked skaters. For the given data set, the roster position with the highest effect on goals for was centers.

For goals against as the dependent variable, the constant had a value of 3.18, signifying that goals against increased by that quantity if the mean of skaters for all positions were zero. Goals against decreased by 0.1288, when the mean rank of centers increased by one, with a standard error of 0.141. By increasing the rank of wingers by one unit, goals against decreased by 0.2249, with a standard error of 0.090. Changing the mean of defensemen by one unit decreased goals against by 0.29, with a standard error of 0.092. The mean of wingers were statistically significant at the 5% level, defensemen were statistically significant at the 5% and 1% level, while the mean of centers was not statistically significant. Results showed that goals against decreased as the mean rank of skaters for all positions increased. This can be substantiated by the fact that higher ranked players have stronger defensive attributes as well as the ability to spend more ice time in the offensive zone, which limits the opportunities and goals generated by the opposition. For the given data set, the roster position with the largest effect on goals against was defensemen.

To prevent the occurrence of multicollinearity, the mean of each position per team per game was computed and used as an independent variable. Variance inflation factor (VIF) was also calculated and inspected (Table 34). VIF estimates how much the variable of the estimated coefficients were increased in comparison to when there was no linear relationship between the predictors. It can verify the level of multicollinearity in a regression model. If VIF is equal to zero, there is no indication of correlation. If VIF is below five, moderate correlation is present and if VIF is above five, there is high correlation. The higher the value of VIF, the less trustworthy the results from the regression analysis. VIF for the constant of the linear regression model had a value of 2.796, which indicated the presence of moderate correlation. VIF had a value of 1.127 for the mean of centers variable, 1.158 for the mean of wingers variable and a value of 1.229 for the mean of defensemen. Since all variables had a VIF below five, there was no evidence of high correlation.

To interpret the value of an additional elite (number one) center, winger and defenseman, the natural logarithm (ln) of those variables was preferred, as estimates showed percentage change in the dependent variable (Table 35). When having win probability as the variable on interest, the constant recorded a value of 0.2056, signifying that win probability increased by that amount if all positions had zero elite skaters. Win probability increased by 3.16%, when elite centers increased by one. Results showed that the addition of an elite player, increased win probability around 3%, regardless of position. The acquisition of an elite winger had the highest impact on win probability. Both t-statistic and p-values implied that all elite positions were significant at the 10% level but only elite centers were significant at the 5% and 1% level.

For goal differential as the dependent variable, the constant had a value of -1.7004, signifying that goal differential decreased by that value if the elite players for all positions were zero. Goal differential increased by 16.33%, when an additional elite center was added to the roster. By increasing the number of elite wingers by one unit, goal differential increased by 20.66%. Increasing elite defensemen by one unit increased goal differential by 28.04%. Outcomes suggested that the acquisition of elite skaters increased goal differential, with centers having the lowest impact whereas defensemen had the greatest. All elite positions were significant at the 5% level, while centers were also statistically significant at the 1% level.

When goals for was used as the dependent variable, the constant recorded a value of 2.1086, specifying that goals for increased by that size if there were no elite players in all positions. Goals for increased by 6.29%, when an additional elite center was added to the roster. By increasing the number of elite wingers by one unit, goals for increased by 16.45%. Increasing elite defensemen by one unit increased goals for by 2.66%. Results indicated that the addition of elite skaters had a positive effect on goal differential. Adding an elite winger had a substantial effect on goals for over centers and defensemen. Only elite wingers were statistically significant at the 10% and 5% level.

For goals against as the dependent variable, the constant registered a value of 3.808, suggesting that goals against increased by that quantity if elite skaters for all positions were zero. Goals against decreased by 10.04%, when an additional elite center was added to the roster. By increasing the number of elite wingers by one unit, goals against decreased by 4.2%. Increasing elite defensemen by one unit decreased goal differential by 15.38%. Estimates showed that the acquisition of elite skaters decreased goals against, with a 5% variation between each position, with elite defensemen having the highest impact while wingers the lowest. Only elite wingers were not statistically significant at the 10% and 5% level.

To measure the relationship between win probability and the quantity of elite centers, wingers and defensemen, a regression was applied. For zero elite centers, win probability was 0.3915 (Table 36). By adding a number one center to a given roster, win probability increased to 0.4837. For two elite centers, win probability increased by 0.0943, reaching 0.5780. The addition of a third top center slight decreased win probability to a value of 0.5745. For four elite centers, win probability was depreciated by 0.0745, reaching a value of 0.5. This indicated that win probability increased the most with the presentence of two elite centers on a given roster. Having more than two elite centers, had a decreased return in win probability.

For zero elite wingers, win probability was 0.42. By adding a top winger to a given roster, win probability decreased to 0.3878 (Table 37). An additional unit of elite wingers, increased win probability to 0.4704. For three first line wingers, win probability increased by 0.0717, reaching 0.5421. When increasing the number of top ranked wingers to four, win probability increased to 0.63. A fifth elite winger deducted win probability by 0.0871. For six top wingers, win probability had a value of 0.7059. Win probability decreased to 0.6, when having seven first line wingers on a given roster. When rising the amount of elite wingers to eight, win probability was depreciated by 0.0286, reaching a value of 0.5714. Findings showed that win probability had the highest value when a given roster contained six top ranked wingers. Interestingly, win probability was larger for no elite wingers than adding an first line winger. The same was noticed when increasing the quantity of top ranked wingers from four to five. The addition of more than six elite wingers, had a reduced return in win probability.

Win probability for a roster with no top pairing defenseman, was 0.4321 (Table 38). The addition of an elite defensemen decreased win probability to 0.3687. By adding a second top ranked defensemen, win probability increased by 0.1677 with a value of 0.5364. Win probability reached 0.5810 for three top pairing defensemen. When increasing the number of elite defensemen to four, win probability hardly increased (0.5816). Having five top pairing defensemen on a given roster, decreased win probability by 0.1271 (0.4545). Outcomes implied that win probability had the greatest value when a given roster consisted of three elite defensemen. Similarly to elite wingers, win probability had a higher value for no top pairing defensemen than one pairing defensemen.

Results indicated that win probability was lower when having no top players in the center position than in any other position. For a roster with no elite players, the acquisition of an elite center would increase win probability more than an elite winger or defensemen. Win probability would increase further with the addition of a second top line center prior to adding an top winger or defensemen. The reduction in win probability was larger for adding a top defensemen than a top winger, by 0.0191. When comparing the addition of a second elite defensemen to the addition of a second elite winger, the increase in win probability was higher for second defensemen by 0.066. Equivalently, the increase in win probability was greater for a third top defensemen than a third top winger by 0.0389. Therefore the increase in win probability would be higher with the acquisition of three top pairing defensemen prior to the acquisition of any elite winger.

CHAPTER VI

DISCUSSION

The objective of this chapter is to review and deliberate the significance of the results from the cluster analysis and regression model and to acknowledge the contribution of this study to new knowledge. Limitations that occurred and impact the methodology and results will be mentioned and briefly discussed. Suggestions in how to improve the analysis for both player evaluation and roster efficiency along with possible future research ideas will be displayed at the end of the chapter.

General Findings

Scully (1974) estimated the MP of a given player in the MLB, as his contribution to team performance and winning. With the use of linear models, this study illustrated the MP of each roster position to team success. The variation in the dependent variable, when increasing the mean of a given position by one unit, showed the MP of that mean position. The MP of an additional elite player by position, was measured as the impact on win probability, goal differential, goals for and goals against. Additionally, the alterations in win probability and the other dependent variables, when increasing the quantity of elite centers, wingers, and defensemen by a unit, demonstrated the MP of the portion of elite skaters per position. Since the salary of skaters was excluded from the data, MR and ergo MRP of each position was not calculated.

Previous researchers (Macdonald, 2010, 2011, 2013; Schuckers et al., 2011; Schuckers and Curro, 2013; Thomas et al., 2013) focused on goals, expected goals and the probability of a goal, as the dependent variable for the evaluation of the in game contribution of a skater. Upon completion of the cluster analysis, this approach can effectively summarize and evaluate the overall performance of players based on a set of variables. The use of one statistic to determine the offensive or defensive contribution of an individual can be deceptive. Clustering has the ability to implement a more inclusive picture of a player compared to using a single variable; thus, it was vital to standardize the data. Doing so, displayed the pragmatic on-ice performance of each player and their contribution to team success. Players that are not on the top two forward lines, or on the power play units, are not often appreciated for their offensive skills. A successful offensive player will continually be productive, even if his time on ice is limited or reduced. Additionally, a defensive minded player will consistently contribute defensively, regardless of his playing time.

Chan et al. (2012) used a different set of variables for the cluster analysis of forwards and defensemen. Forwards were evaluated by a set of six variables while defensemen by five. Forwards had goals per time on ice, assists per time on ice and plus/minus per time on ice to measure their offensive contribution, whereas defensemen had points per time on ice and plus/minus per time on ice. The benefit of this study was that for the cluster analysis of all roster positions, the same set of variables were used. In addition to Chan et al.'s (2012) set of variables, shot differential per time on ice was added. The set of variables chosen in the cluster analysis were: points per time on ice, goals per time on ice, assists per time on ice, plus/minus per time on ice, shot differential per time on ice, blocks per time on ice, hits per time on ice and penalties per time on ice. When using only goals per time on ice and assists per time on ice, players with the identical quantity of points were assigned to different clusters. This occurred as some players had more goals than assists, while others had more assists than goals. The inclusion of points per time on ice minimized the misclassification of skaters.

Plus/minus has the ability to assess the overall contribution of a player, however it focuses only on the net contribution of goals. If a player was on the ice for a goal scored, he was assigned a value of plus one, whereas if he was on the ice for a goal against, he received a minus one. Offensive and defensive performance should not be weighted solely on goals. Shots differential displays the net contribution in shots, as the numerical variation of shots for and shots against. There are games in which teams have been outshooting their opponents but have failed to successfully translate their superiority and opportunities to goals. Players should be evaluated for generating offensive chances, despite if they resulted in a goal or not. For that reason, it was decided to compute the shots differential metric and add it to the set of variables for the cluster analysis.

The intention for implementing cluster analysis for each position independently, was to equally compare and rate skaters who played in the same position. The offensive and defensive production of a given skater may vary from season to season; therefore, it was essential to evaluate and rank the in-game performance for each regular season individually. Standardized regular season-long variables, for eight seasons, were applied in the analysis. Chan et al. (2012) clustered forwards and defensemen separately with the use of k-means algorithm, the efficiency of which is highly dependent on the initialization of the centroids. In this study, k-means++ algorithm was preferred as it has a inferable approximation guarantee to the optimal solution, in opposition to k-means, which can create clustering arbitrarily inferior than the optimum. Elbow and silhouette analysis were conducted to determine the optimal number of clusters per position. Findings from both methods established that the number of clusters for forwards and defensemen were four and three, respectively.

Skaters who played less than a specific amount of minutes for the duration of a regular season were excluded from the analysis. Since the roster model used play by play data, eliminated forwards were automatically assigned to the fourth line, while excluded defensemen to the bottom pairing. Noticeably, 459 centers were ranked as top, 611 as second, 423 as third and 656 as fourth. For wingers, 581 were top three, 867 were top six, 598 were bottom six and 832 were bottom three forwards. For defensemen, 616 were regarded as top two, 951 as top four and 1029 as bottom defensive pairing. As it can be seen from the results, the majority of players were well-balanced as they contributed both offensively and defensively and were ranked as top six and top nine forwards and top four and bottom two pairing, accordingly.

Chan et al. (2012) for the 2008-09 regular season clustered 582 forwards, 303 defensemen and 89 goaltenders. For forwards, 100 were classified as top line, 166 as second line, 128 as defensive and 43 as physical. For defensemen, 70 were ranked as offensive, 71 as defensive, 60 as average and 27 as physical. 18 goaltenders were ranked as elite, 34 as average and 14 as bottom. A difference between Chan et al. (2012) and this study was the quantity of clusters for defensemen. The advantage of this study is that the formed clusters for defensemen (three) represent the pragmatic number of defensive pairings of a given team roster.

Previous studies examined the contribution of individual players, in terms of goals, and the impact of style of play on team success but not the value of each roster position on team performance. To measure the effect each roster position had on team success, in terms of win, the regression model was ran at the game level. The roster of a team alters from game to game and therefore needed to be accounted for. Game level was chosen to examine and measure the impact each position had on the outcome of a single game. Games that both home and visiting team had four centers, eight wingers and six defensemen formed the data set for the linear regression were contained in the analysis to effectively compute the contribution of each roster position. For eight regular seasons, only 828 games (9%) were used as the data set for the regression model. To prevent the occurrence of multicollinearity, the mean of each position per team per game was computed and used as an independent variable.

The coefficients symbolized the sovereign contribution of the mean of each roster position (independent) to the prediction of the dependent variable Win. If all other independent variables stay constant, β_i is the change in Win for each one-unit difference in the mean of the relevant roster position. Win probability increased by 0.0810 when the mean of centers increased by one unit. An increase by one unit in the mean of wingers, increased win probability by 0.0753. Win probability increased by 0.0694, when the mean of defensemen increased by a unit. T-statistic and p-value examined if there was a statistically significant relationship between the mean player quality per position and win probability. All independent variables were statistically significant at the 5% level. These results contribute to new knowledge, as the relationship between win probability and the mean of centers, wingers and defensemen had never been examined prior to this study.

To understand the value of each roster position, additional analysis on in-game performance was administered. The variables of interest were: goal differential, goals for, and goals against. When the quality of players increased, goal differential and goals for increased while goals against decreased. These findings were justified by the fact that higher ranked players are more skillful and talented, creating more offense and minimizing the offense of the opposition. The roster position that had the largest effect on goal differential, for the given data set, was wingers. All mean positions for goal differential were statistically significant at the 5% level. Centers had the biggest impact on goals for. Only the mean of wingers was statistically significant at the 5% level. The mean of centers was significant at the 10% level while the mean of defensemen was not statistically significant. The roster position with the highest effect on goals against was defensemen. The mean of wingers and defensemen were statistically significant at the 5% level while the mean of centers was not significant.

Schuckers et al. (2011) found that the frequent on ice presence of top players increased the probability of scoring a goal and reduced the probability of conceiving a goal. Results from this study were compatible to the findings from Schuckers et al. (2011), as goal differential and goals for increased when the mean of centers, wingers and defensemen improved, while goals against decreased. The addition of an elite center, winger and defensemen had a positive effect on goal differential, goals for and goals against. Findings from this study were different to the results from Thomas et al. (2013), in terms of defensive contribution. According to their findings, centers had a greater effect on defensive performance than defensemen, while the effect of wingers and defensemen was almost equal. This study showed that defensemen had the highest impact on goals against, wingers the second highest, while centers the least.

To cope with multicollinearity, Macdonald (2012, 2013) and Schuckers and Curro (2013) used a ridge regression. Since an OLS regression was applied, the existence or not of multicollinearity needed to be examined. Therefore, VIF was calculated and inspected as it estimated the level of multicollinearity in a regression model. If VIF is equal to zero, there is no indication of correlation. If VIF is below five, moderate correlation is present and if VIF above five there is high correlation. The greater the value of VIF, the less trustworthy the results from the regression analysis. VIF for the constant had a value of 2.796, which implied the presence of moderate correlation. VIF for the mean centers variable was 1.127 which translated to moderate correlation. VIF had a value of 1.158 for the mean wingers variable and a value of 1.129 for the mean defensemen variable. Since all three VIF were below a conservative level of five, there was no evidence of high correlation.

To estimate the value of an additional top center, winger, and defensemen the natural logarithm (ln) was chosen, as the results displayed the percentage change in the dependent variable. The addition of an elite center increased win probability by 3.16%. Increasing elite wingers by one increased win probability by 3.51%. By adding an elite defensemen, win probability increased by 2.9%. All variables were statistically significant at the 10% level. Only elite centers were significant at the 5% level. When goal differential and goals for were used as the dependent variables, the addition of an elite skater, regardless position, increased their values. The position with the highest effect was elite wingers. All variables were statistically significant at the 5% for goal differential but only elite wingers were statistically significant for goals for. Goals against decreased with the addition of an elite skater. Adding an elite defensemen had the largest effect. Only elite wingers were not statistically significant for goals against.

Schuckers and Curro (2013) found that top players contributed five additional wins for their respectful teams, over the course of a regular season. Their data contained only even strength situations for two regular seasons (2010-2012). The coefficients for the acquisition of an additional elite skater in this study were measured at the game level for eight regular seasons (2010 -2017). To compare findings, they were multiplied by 82, which is the total number of games for the duration of a regular season. Results indicated that an additional elite center, winger, and defenseman would contribute 2.59, 2.88, and 2.38 wins per season, accordingly. Both studies implemented a different method to evaluate player performance and a different data set. A benefit of this study was that results represented the overall contribution of a given player, as the data included all on ice situations in regulation time and was not limited to even strength situations.

Chan et al. (2012) indicated that players on the top forward line and top defensive pairing had a greater impact on team success. Thomas et al. (2013) suggested that top players contribute the most and thus had more value to their teams. Results from this study were consistent with the findings of Chan et al. (2012) and Thomas et al. (2013), despite using a different data set, as elite skaters had a higher effect on win probability at the game level. Schuckers and Curro (2013) indicated that forwards had a greater value than defensemen in even strength situations, as they generated more additional wins. This study showed that win probability increased more when the mean of centers and wingers improved in comparison to the mean of defensemen. Similarly, the addition of an elite center or winger had a greater impact on win probability than the addition of an elite defensemen.

To evaluate the relationship between win probability and the quantity of elite centers, wingers and defensemen, a regression was applied. Findings showed that a roster with no elite centers had the lowest win probability. For a roster with no top skaters, the addition of a top center would increase win probability more than a top winger or defensemen. The acquisition of a second elite center would have a higher increase on win probability than the addition of an elite winger or defensemen. Despite the fact that the deduction in win probability was greater for an elite defenseman than an elite winger, the increase in win probability was significantly larger for a second top defenseman than a second top winger. Correspondingly, the increase in win probability was higher for a third top defenseman than a third top winger. For that reason, the increase in win probability would be greater with the addition of three elite defenseman prior to the addition of any elite winger.

K-means++, an improved clustering algorithm, was used as an new approach to divide and rank the overall in game contribution of skaters. A specific set of variables was used to measure both offensive and defensive contribution players. A new statistic, even strength shot differential, was calculated and included in the set of variables that weighed and ranked the performance of centers, wingers and defensemen. This study introduced and implemented a new technique to assess and classify the on ice performance of skaters.

Roster design research has focused on team diversity, efficiency, flexibility and style of play. This study provided new knowledge, as it was the first to analyze the relationship between win probability and roster position; the effect of an additional elite center, winger and defenseman on win probability; and the relationship between win probability and the quantity of top centers, wingers and defensemen. This research can be used as a baseline to examine the impact of roster position on team performance in other leagues or sports, and to further progress the developed models to improve or discover additional findings.

Limitations

Goaltenders were excluded from both cluster analysis and the regression model. The focus of this study was to find the impact of each roster position on team success, in terms of skaters. Goaltenders play a vital role in team success, as they are the only players who are on the ice for a full game. A goaltender can be replaced from the game if he is underperforming or removed if a team is trailing in the last couple of minutes of a game. In those situations, an extra forward is added, in exchange for the goaltender, to increase the probability of scoring a goal.

A possible limitation was the exclusion of players who played limited minutes for the whole regular season from the cluster analysis. Those players were automatically assigned to the bottom line for forwards and the bottom pairing for defensemen for the purpose of the regression game level analysis. If the eliminated players had participated in the clustering process, there was a possibility that their overall performance could have been greater than the position they were ranked and assigned to. Additionally, their inclusion could have regrouped and reranked multiple players.

Another limitation to this study was the low proportion of games included in the roster model analysis. Positions assigned to players by the NHL had a crucial impact on the roster model. Skaters whose primary position is on the wing, had center as their assigned position. For example, Leo Komarov played five seasons on the left wing for the Toronto Maple Leafs and minimal to zero NHL games in the center position; however, the NHL has him listed as a center. Incorrect positions had an impact on both cluster analysis and the roster model. For clustering, players who were listed in the incorrect position were compared and grouped with skaters who had a diverse style of play. As for the roster model, the data only included games that had four centers, eight wingers and six defensemen for both team. An inaccurate assigned position to a given player increasing the number of a given position by one, resulted in excluding the game, as it did not fit the positions criteria. This was the reason for only nine percent of the total regular season games were admitted in the regression model.

Possible Future Research

Future research can include goaltenders and all players despite their limited time of play, in both the cluster analysis and the roster model. The roster model would display the impact of all roster positions on team success and not only skaters. Goaltenders can be evaluated on their season-long performance with a set of variables containing save percentage per time on ice and average goal against per time on ice. The optimal number of clusters for goaltenders can be calculated and each cluster group ranked accordingly. For the roster model, an additional independent variable with the mean of goaltenders per team would be included and analyzed. Goaltenders play the most minutes per game and play a crucial role in team success. As for the players with limited ice time, the formation of a variable that would cluster their on ice performance correctly, would be added to the existing set of variables, to assist with their incorporation.

Another idea for future research would be to implement cluster analysis with the identical set of variables to evaluate player performance for the same data set (2010-2017) for the postseason. Playoff games are considered to be played at a different pace and style in comparison to regular season games. Despite the use of the same set of variables, a given player might be ranked dissimilarly to his regular season-long performance, upon playoff performance evaluation. Additionally, the impact each roster position has on team success might be altered, as the value of specific positions may increase or decrease. A combination of all games (regular season and playoffs) will give the total contribution of a player and his effect on team performance.

An additional idea would be to evaluate player performance in terms of salary to determine a value system for each roster position. The NHL has a salary cap and each team must compose a high-performing and well-balanced roster containing a maximum of twenty-three contracts, without exceeding the salary cap limit. The economic value of a player is important for decision making, as each NHL must have a maximum of 23 contracts at all time. Therefore, players are assessed by combining both on ice performance and salary. Salary in the form of dollars per time on ice can be included in the set of variables for cluster analysis. This way, players will be ranked for their on- and off-ice value.

The final concept for future research would be to find a solution to minimize the incorrect assignment of positions to players. Play by play data has each player listed in numerical value from one to six for each team. The position of each player is also displayed but it is based on how a player is listed in the NHL. For example, in the 2010-2011 regular season the Montreal Canadiens had seven centers playing together in a single game. It would be more helpful and useful to have the position predetermined. There are three forwards position, one center and two wingers, two defensemen and a goaltender. Having visitor center, visitor left winger or visitor winger one, visitor right winger or visitor winger two etc. instead of visitor player one, visitor player two, visitor player three etc. would eliminate teams having too many players per position. Furthermore, more games would be included in the game level data for the roster model analysis and results would be more representative to the whole regular season.

CHAPTER VII

CONCLUSION

This study measured the impact of NHL roster positions on team success. To achieve so, the season-long performance of all NHL skaters were estimated and evaluated. Prior studies on roster construction have concentrated on team diversity, efficiency and flexibility. After a review of the related literature, there was a significant lack of empirical studies on the impact of roster position on team efficiency. The inspiration for this study was to perceive which roster positions have higher contribution to team success and establish the significance of roster depth.

Roster is the unofficial manifest of individual athletes consolidated as a group. Inferior cooperation and interaction amongst teammates are essential for a team to become united and achieve success. Cohesion and collective efficacy influence athletic performance at an individual and team degree. Roster design can be defined as the efficient assembling of forward line combinations and defensive pairings that maximizes the win probability of a team.

The objective of this study was to establish the value of each NHL roster position on team success. To assign players to their respectful roster position, their overall performance was assessed. Cluster analysis, more specifically k-means++ clustering, was selected as the method to evaluate and rank players dependent of their season-long performance. The data was scaled and standardized with the use of the total time on ice per season. The statistics chosen to cluster all positions were: points per time on ice, goals per time on ice, assists per time on ice, plus/minus per time on ice, shots differential per time on ice, blocks per time on ice, hits per time on ice and penalties per time on ice. Centers, wingers and defensemen were clustered separately for each regular season. Elbow method and silhouette analysis were used to find the optimal number of clusters for each roster position. Taking into consideration the formation of a standard hockey roster, results suggested that the optimal quantity of clusters for centers and wingers were four and three for defensemen.

A regression model was created to measure the contribution of each roster position on team success in terms of winning at the game level data. The model consisted of three independent variables, one for each roster position, a constant and a standard error. To restrict the existence of multicollinearity, the mean of each position per team per game was computed and used as an independent variable. To estimate the correlation between win probability and mean player quality per position, ordinary least square (OLS) regression was applied.

To effectively compute the contribution of each roster position to team success, games that had the identical number of skaters were included in the analysis. Games that both home and away team had four centers, eight wingers and six defensemen each, were included in the data. A total of 828 games from eight regular seasons, assembled the data set for the linear regression. The coefficients illustrated the effective contribution of the mean of each roster position to the prediction of win probability. If all other mean roster position variables remain constant, β_i is the variation in win probability for each one-unit variance in the mean of the relevant roster position. The roster position that had the highest impact on win probability, when regressing by mean quality, was centers. Wingers contributed the second most to win probability while defensemen the least. Increase in player quality for all three positions had a positive effect on team success in terms of winning. In terms of elite players only, the position that had the greatest impact on win probability, were elite wingers, followed by centers and defensemen. Results suggested that teams with no elite skaters may what to acquire elite centers first, elite defensemen second and elite wingers last.

Hockey operations can benefit from this study at the individual and team level. Cluster analysis can assist in the evaluation of the players on the current roster of a given team and in the prioritization of player acquisitions either via free agency or trades. The roster model can facilitate in the estimation of the in-game contribution of each newly acquired player, on team success.

The findings of this study contributed to the evaluation of the relationship between team success and roster position; the impact of an additional center, winger and defensemen on win probability; and the assessment of the relationship between win probability and the quantity of elite centers, wingers and defensemen. The methodology in this study can be the starting point for the estimation of the relationship between team success and roster position in other leagues or sports, to be further advanced to enhance or identify additional results, and to potentially progress research within the sphere of roster design.

References

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review*, *84*(2), 191.

Bandura, A. (1997). Self-efficacy: The exercise of control. New York: Freeman.

Berri, D., Schmidt, M., & Brook, S. (2006). The Wages of Wins: Taking Measure of the Many Myths in Modern Sport. Stanford University Press.

Blaug, M. (1997), *Economic Theory in Retrospect*, 5th edn, Cambridge University Press.

Carron, A. V., Brawley, L. R., & Widmeyer, W. N. (1998). The measurement of cohesiveness in sport groups. *Advances in Sport and Exercise Psychology Measurement*, *23*(7), 213-226.

Carron, A. V., Bray, S. R., & Eys, M. A. (2002). Team cohesion and team success in sport. *Journal of Sports Sciences*, *20*(2), 119-126.

Carron, A. V., & Chelladurai, P. (1981). Cohesiveness as a factor in sport performance. *International Review of Sport Sociology*, *16*(2), 21-43.

Carron, A. V., Colman, M. M., Wheeler, J., & Stevens, D. (2002). Cohesion and performance in sport: A meta-analysis. *Journal of Sport and Exercise Psychology*, *24*(2), 168-188.

Chan, T. C., Fearing, D. S., & Fearing, D. (2013). The value of flexibility in baseball roster construction. MIT Sloan Sports Analytics Conference.

Chan, T. C. Y., Cho, J. A., & Novati, D. C. (2012). Quantifying the contribution of NHL player types to team performance. Interfaces, 42(2), 131–145.

Clark, J. M. (1928). Inductive evidence on marginal productivity. *The American Economic Review*, *18*(3), 450-467.

Dawson, P., Dobson, S., & Gerrard, B. (2000). Estimating Coaching Efficiency in Professional Team Sports: Evidence from English Association Football. *Scottish Journal of Political Economy, 47*(4), 399–421.

Dhyani, K., Gualandi, S., & Cremonesi, P. (2010, June). A constraint programming approach for the service consolidation problem. In International Conference on Integration of Artificial Intelligence (AI) and Operations Research (OR) Techniques in Constraint Programming (pp. 97-101). Springer, Berlin, Heidelberg.

Engeler, M., Georgakakis, D., & Ruigrok, W. (n.d.). Newcomers' international experience backgrounds and contributions to team performance.

Fort, R. (1992). Pay and Performance: Is the Field of Dreams Barren?. Diamonds are forever: The business of baseball, 134-162.

Fox, E. C. (1984). Team cohesion, ability and coaches' leadership effectiveness as predictors of success in women's intercollegiate softball (Doctoral dissertation, University of Georgia).

Gammage, K. L., Carron, A. V., & Estabrooks, P. A. (2001). Team cohesion and individual productivity: The influence of the norm for productivity and the identifiability of individual effort. *Small Group Research*, *32*(1), 3-18.

Gramacy, R. B., Jensen, S. T., & Taddy, M. (2013). Estimating player contribution in hockey with regularized logistic regression. *Journal of Quantitative Analysis in Sports, 9*(1), 97-111.

Grieve, F. G., Whelan, J. P., & Meyers, A. W. (2000). An experimental examination of the cohesion-performance relationship in an interactive team sport. *Journal of Applied Sport Psychology, 12*(2), 219-235.

Hall, S., Szymanski, S., & Zimbalist, A. S. (2002). Testing Causality Between Team Performance and Payroll: The Cases of Major League Baseball and English Soccer. *Journal of Sports Economics, 3*(2), 149–168.

Hamilton, B. H., Nickerson, J. A., & Owan, H. (2003). Team Incentives and Worker Heterogeneity: An Empirical Analysis of the Impact of Teams on Productivity and Participation. *Journal of Political Economy, 111*(3), 465–497.

Hofler, R. A., & Payne, J. E. (1996). How close to their offensive potential do national football league teams play?. *Applied Economic Letters, 3*(11), 743–747.

Idson, T. L., and Kahane, L. H. (2000). Team effects on compensation: an application to salary determination in the National Hockey League, *Economic Inquiry, 38*(2), 345–357

Jones, J. C. H., Nadeau, S., & Walsh, W. D. (1997). The wages of sin: Employment and salary effects of violence in the national hockey league. *Atlantic Economic Journal*, *25*(2), 191-206.

Jones, J. C. H., Nadeau, S., & Walsh, W. D. (1999). Ethnicity, productivity and salary: player compensation and discrimination in the National Hockey League. *Applied Economics, 31*(5), 593–608.

Kahane, L. H. (2005). Production efficiency and discriminatory hiring practices in the National Hockey League: A stochastic frontier approach. *Review of Industrial Organization, 27*(1), 47-71.

Kahane, L., Longley, N., & Simmons, R. (2013). The effects of coworker heterogeneity on firm-level output: assessing the impacts of cultural and language diversity in the National Hockey League. *Review of Economics and Statistics, 95*(1), 302–314.

Keidell, R. W. (1987). Team sports models as a generic organizational framework. *Human Relations*, *40*(9), 591-612.

Krautmann, A. C., & Ciecka, J. (2009). The Postseason Value of an Elite Player to a Contending Team. *Journal of Sports Economics*, *10*(2), 168–179.

Krautmann, A. C., Von Allmen, P., & Berri, D. (2009). The underpayment of restricted players in North American sports leagues. *International Journal of Sport Finance, 4*(3), 161.

Kutáč, P., & Sigmund, M. (2015). A comparison of somatic variables of elite ice hockey players from the Czech ELH and Russian KHL. *Journal of Human Kinetics*, *45*(1), 187-195.

Lazear, E.P. (1999). Culure and Language. *Journal of Political Economy, 107*(1), 95-126.

Leonard, J. S., & Levine, D. I. (2003). Diversity, discrimination, and performance.

Liebhafsky, H. H. (1964). Diminishing returns and linear homogeneity: Comment. *The American Economic Review*, 739-744.

Macdonald, B. (2011). A Regression-Based Adjusted Plus-Minus Statistic for NHL Players. *Journal of Quantitative Analysis in Sports*, *7*(3), 1–29.

Macdonald, B. (2012). Adjusted Plus-Minus for NHL Players using Ridge Regression with Goals, Shots, Fenwick, and Corsi. *Journal of Quantitative Analysis in Sports*, *8*(3), 1–24.

Macdonald, B. (2012). An Expected Goals Model for Evaluating NHL Teams and Players. *In Proceedings of the 2012 MIT Sloan Sports Analytics Conference.*

MacLean, D., & Sullivan, P. (2003). A season-long case study investigation of collective efficacy in male intercollegiate basketball. *Online J Sport Psychol*, *5*(3).

Mason, D., & Foster, W. (2007). Putting Moneyball on ice?. *International Journal of Sport Finance*, *2*(4), 206.

Mongeon, K. (2013). A Market Test for Ethnic Discrimination in the National Hockey League? A Game-Level Panel Data Approach. *Journal of Sports Economics*, *16*(5), 460–481.

Mongeon, K. (n.d.). Designing Effective Sport Teams through Physical Diversity.

Myers, N. D., Feltz, D. L., & Short, S. E. (2004). Collective Efficacy and Team Performance: A Longitudinal Study of Collegiate Football Teams. *Group Dynamics: Theory, Research, and Practice*, *8*(2), 126.

Myers, N. D., Payment, C. A., & Feltz, D. L. (2004). Reciprocal relationships between collective efficacy and team performance in women's ice hockey. *Group Dynamics, 8*, 182-195.

Paskevich, D. M. (1997). Conceptual and measurement factors of collective efficacy in its relationship to cohesion and performance outcome.

Paskevich, D. M., Brawley, L. R., Dorsch, K. D., & Widmeyer, W. N. (1999). Relationship between collective efficacy and team cohesion: Conceptual and measurement issues. *Group Dynamics: Theory, Research, and Practice, 3*(3), 210.

Pisarski, A., & Barbour, J. P. (2014). What roles do team climate, roster control, and work life conflict play in shiftworkers' fatigue longitudinally? *Applied Ergonomics, 45*(3), 773–779.

Schuckers, M. E., Lock, D. F., Wells, C., Knickerbocker, C. J., & Lock, R. H. (2011). National Hockey League skater ratings based upon all on-ice events: An adjusted minus/plus probability (AMPP) approach. *Unpublished Manuscript*, *59*.

Schuckers, M., & Curro, J. (2013). Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events. *In 7th annual MIT sloan sports analytics conference.*

Scully, G.W. (1974). Pay and Performance in Major League Baseball. *The American Economic Review*, 64(6), 915-930.

Scully, G.W. (1989). The Business of Major League Baseball. *University of Chicago Press Economics Books.*

Sill, J. (2010). Improved NBA Adjusted +/- Using Regularization and Out-of-Sample Testing. *In Proceedings of the 2010 MIT Sloan Sports Analytics Conference.*

Spink, K. S. (1990). Group cohesion and collective efficacy of volleyball teams. *Journal of Sport and Exercise Psychology*, *12*(3), 301-311.

Szymanski, S. (2000). A market test for discrimination in the English professional soccer leagues. *Journal of political Economy*, *108*(3), 590-603.

Szymanski, S., & Longley, N. (2001). *A Market Test for Discrimination in the National Hockey League*. mimeo.

Thomas, A. C., Ventura, S. L., Jensen, S., & Ma, S. (2012). Competing Process Hazard Function Models for Player Ratings in Ice Hockey. *The Annals of Applied Statistics*, 1497-1524.

Thomas, A. C. (2007). Inter-arrival Times of Goals in Ice Hockey. *Journal of Quantitative Analysis in Sports, 3*(3).

Thomas, A. C. (2006). The impact of puck possession and location on ice hockey strategy. *Journal of Quantitative Analysis in Sports*, *2*(1).

Timmerman, T. A. (2000). Racial diversity, age diversity, interdependence, and team performance. *Small Group Research*, *31*(5), 592-606.

Ulmer, M. J. (1971). The Limitations of Revenue Sharing. *The ANNALS of the American Academy of Political and Social Science, 397*(1), 48-59.

Vincent, C., & Eastman, B. (2009). Determinants of pay in the NHL: a quantile regression approach. *Journal of Sports Economics*, *10*(3), 256-277.

Von Thünen, J. (1826), *The Isolated State*, English edn, London, Pergamon Press.

Watson, C. B., Chemers, M. M., & Preiser, N. (2001). Collective efficacy: A multilevel analysis. *Personality and Social Psychology Bulletin*, *27*(8), 1057-1068.

Wiseman, F., & Chatterjee, S. (2003). Team payroll and team performance in major league baseball: 1985–2002. *Economics Bulletin, 1*(2), 1-10.

Zaccaro, S. J., Blair, V., Peterson, C., & Zazanis, M. (1995). Collective efficacy. In J. E. Maddux (Ed.), *Self-efficacy, adaptation, and adjustment: Theory research, and application* (pp. 305-328). New York: Plenum Press.

Zak, T. A., Huang, C. J., & Siegfried, J. J. (1979). Production efficiency: the case of professional basketball. *Journal of Business*, 379-392.

Zimbalist, A. S. (1992). *Baseball and billions: A probing look inside the big business of our national pastime*. Basic Books.

Appendix A: Figures

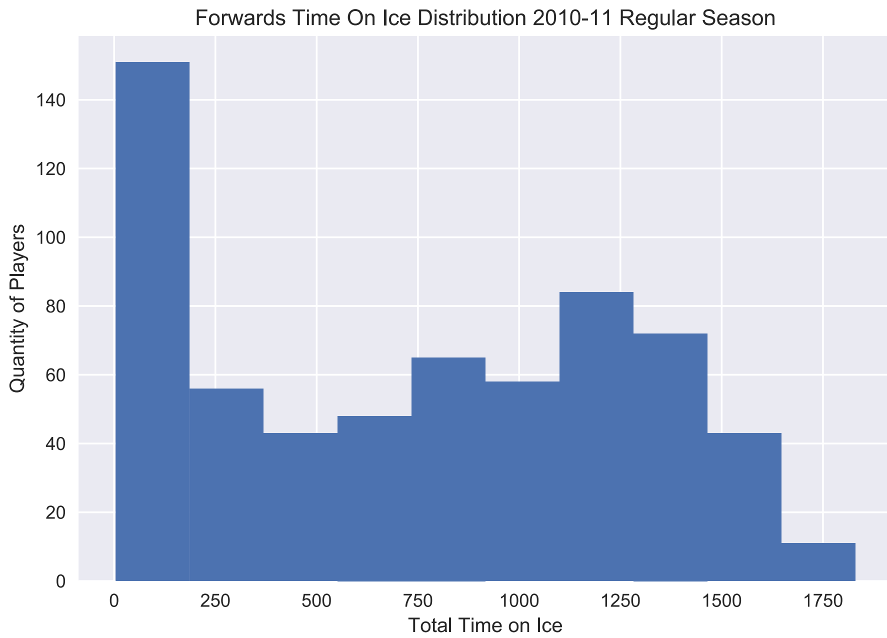

Figure 1. Forwards Time On Ice Distribution For The 2010-11 Regular Season.

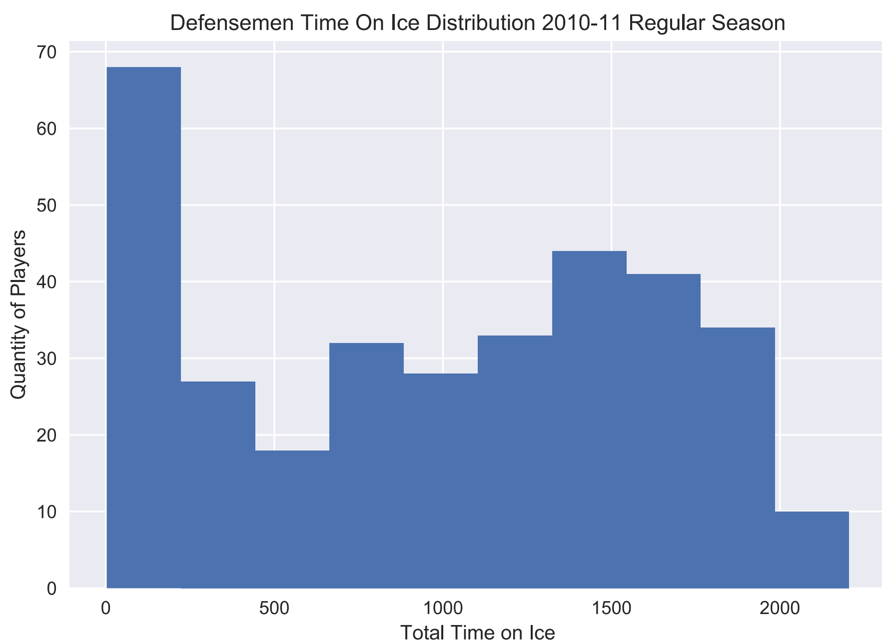

Figure 2. Defensemen Time On Ice Distribution For The 2010-11 Regular Season.

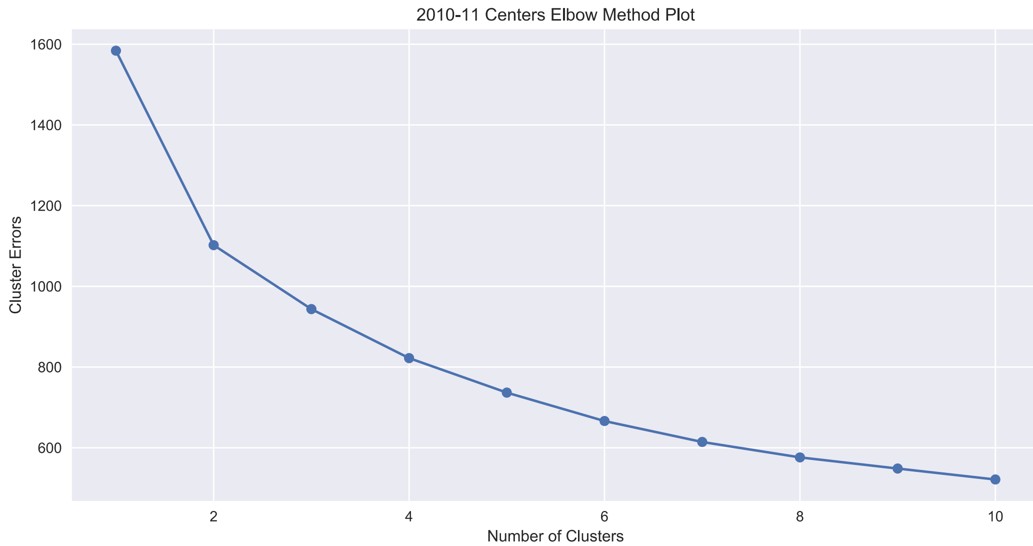

Figure 3. Centers Elbow Method Plot For The 2010-11 Regular Season.

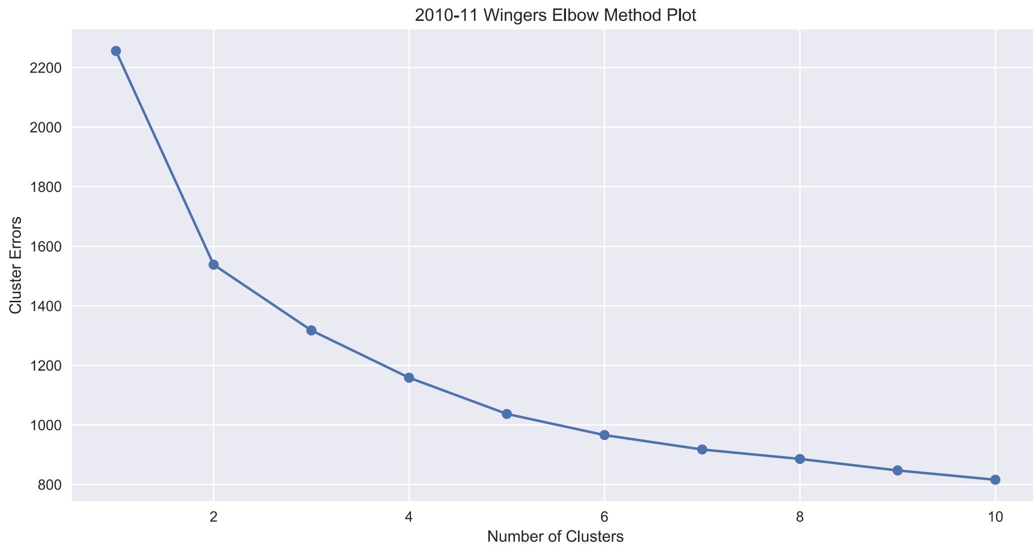

Figure 4. Wingers Elbow Method Plot For The 2010-11 Regular Season.

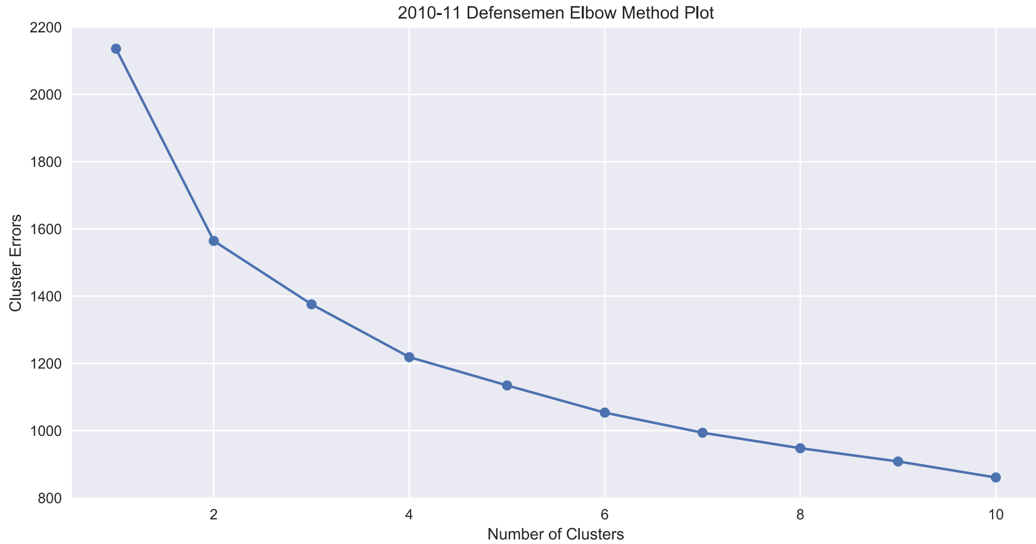

Figure 5. Defensemen Elbow Method Plot For The 2010-11 Regular Season.

```
For n_clusters = 2 The average silhouette_score is : 0.268174274769
For n_clusters = 3 The average silhouette_score is : 0.183160214367
For n_clusters = 4 The average silhouette_score is : 0.190659922622
For n_clusters = 5 The average silhouette_score is : 0.198241966931
For n_clusters = 6 The average silhouette_score is : 0.187881852035
For n_clusters = 7 The average silhouette_score is : 0.189854428786
For n_clusters = 8 The average silhouette_score is : 0.188541722968
For n_clusters = 9 The average silhouette_score is : 0.158983632516
For n_clusters = 10 The average silhouette_score is : 0.184467631752
```

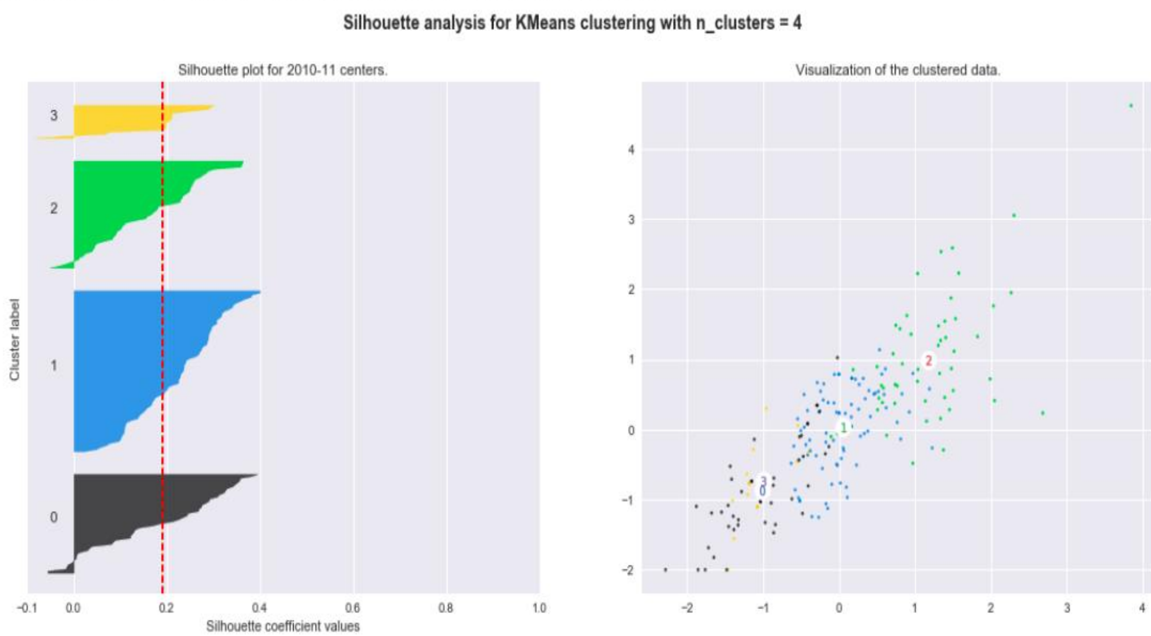

Figure 6. Centers Silhouette Scores and Plot For The 2010-11 Regular Season.

```
For n_clusters = 2 The average silhouette_score is : 0.311716614401
For n_clusters = 3 The average silhouette_score is : 0.227563543748
For n_clusters = 4 The average silhouette_score is : 0.173948452061
For n_clusters = 5 The average silhouette_score is : 0.184394822055
For n_clusters = 6 The average silhouette_score is : 0.174987002022
For n_clusters = 7 The average silhouette_score is : 0.163371412899
For n_clusters = 8 The average silhouette_score is : 0.147387210222
For n_clusters = 9 The average silhouette_score is : 0.155575125568
For n_clusters = 10 The average silhouette_score is : 0.152954579265
```

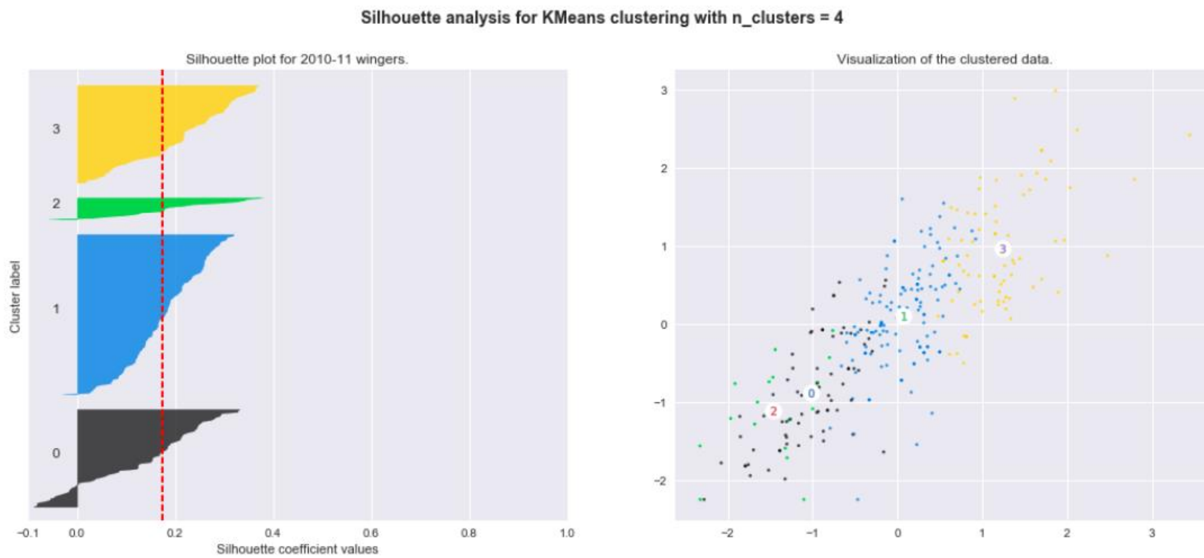

Figure 7. Wingers Silhouette Scores and Plot For The 2010-11 Regular Season.

```
For n_clusters = 2 The average silhouette_score is : 0.233412525181
For n_clusters = 3 The average silhouette_score is : 0.168220883198
For n_clusters = 4 The average silhouette_score is : 0.163988321301
For n_clusters = 5 The average silhouette_score is : 0.149905061097
For n_clusters = 6 The average silhouette_score is : 0.160456677627
For n_clusters = 7 The average silhouette_score is : 0.158129254916
For n_clusters = 8 The average silhouette_score is : 0.161691751851
For n_clusters = 9 The average silhouette_score is : 0.147109978696
For n_clusters = 10 The average silhouette_score is : 0.16092166979
```

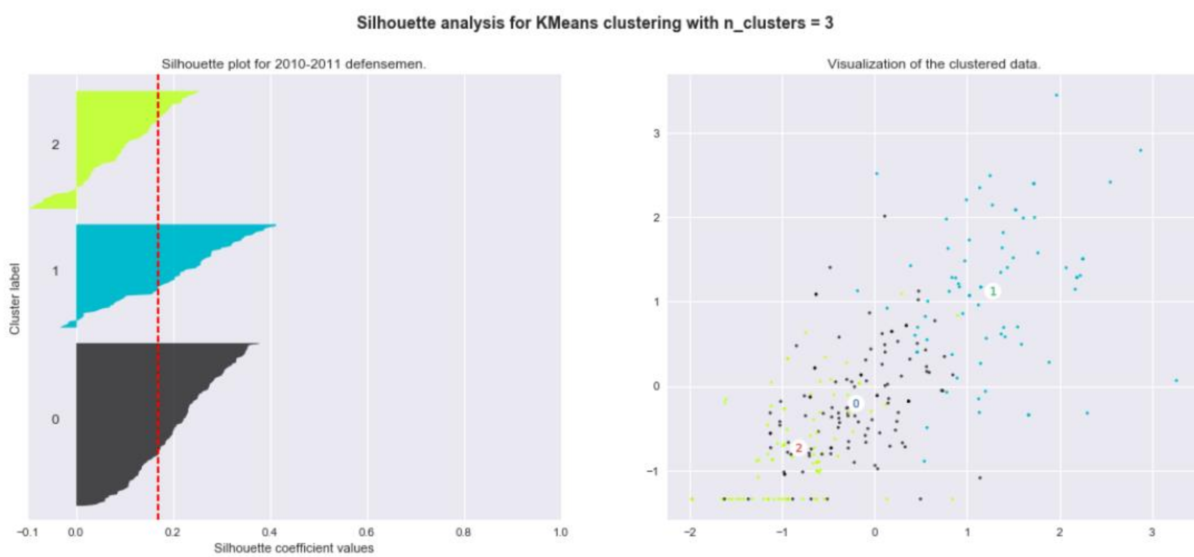

Figure 8. Defensemen Silhouette Scores and Plot For The 2010-11 Regular Season.

Appendix B: Tables

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1648.0000	2256.0000	2136.0000
2	1146.4737	1538.2604	1564.4791
3	981.0154	1317.6353	1375.8018
4	887.9426	1158.8458	1218.7314
5	826.1492	1037.0584	1134.5137
6	763.7392	966.2859	1053.8008
7	720.7542	917.8949	993.8912
8	686.8841	885.9656	947.8466
9	664.7453	847.4030	908.4384
10	628.8622	816.1988	860.8761

Table 1. Elbow analysis for 2010-11 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2682	0.3117	0.2334
3	0.1832	0.2276	0.1682
4	0.1907	0.1739	0.1640
5	0.1982	0.1844	0.1499
6	0.1879	0.1750	0.1605
7	0.1899	0.1634	0.1581
8	0.1885	0.1474	0.1617
9	0.1590	0.1556	0.1471
10	0.1845	0.1530	0.1609

Table 2. Silhouette analysis for 2010-11 regular season.

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1656.0000	2248.0000	2144.0000
2	1176.1268	1495.7824	1589.1786
3	1001.0553	1243.3755	1418.0328
4	897.7881	1123.4020	1276.4281
5	804.1739	1027.4628	1172.2680
6	726.4886	954.9538	1098.8983
7	669.8752	910.9492	1035.5217
8	631.2018	866.0443	979.2357
9	598.5298	823.5331	928.7899
10	569.7525	788.5738	885.9342

Table 3. Elbow analysis for 2011-12 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2510	0.2819	0.2325
3	0.2121	0.2463	0.1510
4	0.1776	0.1730	0.1529
5	0.1719	0.1697	0.1482
6	0.1775	0.1627	0.1481
7	0.1777	0.1491	0.1486
8	0.1662	0.1463	0.1446
9	0.1545	0.1517	0.1466
10	0.1631	0.1556	0.1424

Table 4. Silhouette analysis for 2011-12 regular season.

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1576.0000	2280.0000	2024.0000
2	1150.4326	1632.1872	1537.8658
3	991.6546	1369.1607	1334.4348
4	898.7642	1239.1651	1215.0155
5	813.5690	1144.3284	1119.5478
6	739.6850	1063.2469	1058.4466
7	697.4491	999.5591	996.9490
8	657.9087	958.9812	953.7342
9	617.7200	903.5296	932.4692
10	599.2178	870.5734	891.2438

Table 5. Elbow analysis for 2012-13 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2298	0.2516	0.2103
3	0.1794	0.2093	0.1847
4	0.1513	0.1528	0.1586
5	0.1448	0.1498	0.1438
6	0.1573	0.1572	0.1421
7	0.1418	0.1618	0.1383
8	0.1471	0.1430	0.1352
9	0.1583	0.1514	0.1294
10	0.1426	0.1427	0.1271

Table 6. Silhouette analysis for 2012-13 regular season.

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1664.0000	2232.0000	2024.0000
2	1186.2710	1500.2739	1492.0765
3	993.6040	1239.8170	1263.0892
4	885.2980	1102.3662	1124.5143
5	821.8644	1005.8257	1035.3803
6	759.8309	941.5719	969.3705
7	717.7981	883.0258	908.9900
8	677.0363	843.3054	866.8190
9	635.2187	804.6818	826.1274
10	605.8666	759.0591	802.0385

Table 7. Elbow analysis for 2013-14 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2438	0.2798	0.2257
3	0.2400	0.2448	0.1943
4	0.1570	0.1792	0.1760
5	0.1513	0.1811	0.1534
6	0.1479	0.1704	0.1653
7	0.1431	0.1746	0.1537
8	0.1517	0.1764	0.1443
9	0.1358	0.1571	0.1455
10	0.1503	0.1642	0.1418

Table 8. Silhouette analysis for 2013-14 regular season.

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1800.0000	2112.0000	2152.0000
2	1261.4209	1419.0407	1575.6110
3	1082.4512	1241.4294	1328.1745
4	948.7477	1098.3968	1173.6842
5	866.9733	997.0707	1087.1445
6	799.7259	931.4941	1021.1584
7	749.5254	868.3939	967.4004
8	717.4418	818.6340	913.8163
9	678.3954	778.7674	864.7844
10	655.8522	745.1517	824.2369

Table 9. Elbow analysis for 2014-15 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2731	0.2941	0.2290
3	0.1842	0.2464	0.1850
4	0.1787	0.1744	0.1950
5	0.1640	0.1803	0.1675
6	0.1654	0.1708	0.1756
7	0.1712	0.1647	0.1514
8	0.1469	0.1728	0.1691
9	0.1581	0.1667	0.1610
10	0.1473	0.1592	0.1576

Table 10. Silhouette analysis for 2014-15 regular season.

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1776.0000	2216.0000	2072.0000
2	1235.0073	1517.0740	1556.2210
3	1078.4287	1270.9187	1319.3482
4	965.4112	1135.5316	1186.7439
5	878.6998	1057.5477	1079.1117
6	830.5387	985.4188	1004.2205
7	784.9011	935.6537	952.8394
8	738.7515	904.3145	893.5621
9	699.8445	856.2943	852.5061
10	653.7416	832.6475	823.4783

Table 11. Elbow analysis for 2015-16 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2633	0.2796	0.2106
3	0.2282	0.2492	0.2008
4	0.1598	0.1762	0.1707
5	0.1683	0.1637	0.1654
6	0.1571	0.1605	0.1673
7	0.1520	0.1457	0.1594
8	0.1567	0.1346	0.1598
9	0.1573	0.1488	0.1690
10	0.1586	0.1620	0.1604

Table 12. Silhouette analysis for 2015-16 regular season.

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1744.0000	2128.0000	2016.0000
2	1222.5311	1511.2485	1545.2600
3	1040.5559	1277.1568	1340.9281
4	922.1537	1132.4910	1180.8439
5	832.4571	1030.3505	1097.8427
6	777.3578	955.5697	1027.3795
7	729.3710	894.2038	967.5564
8	696.9966	851.0954	910.3509
9	657.8034	802.6501	866.5741
10	640.6097	770.3963	834.5966

Table 13. Elbow analysis for 2016-17 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2492	0.2467	0.2222
3	0.2011	0.2207	0.1909
4	0.2000	0.1697	0.1737
5	0.1798	0.1815	0.1489
6	0.1686	0.1676	0.1417
7	0.1677	0.1624	0.1579
8	0.1545	0.1594	0.1602
9	0.1513	0.1494	0.1416
10	0.1521	0.1693	0.1454

Table 14. Silhouette analysis for 2016-17 regular season.

Number of clusters	Cluster errors for centers	Cluster errors for wingers	Cluster errors for defensemen
1	1648.0000	2400.0000	2120.0000
2	1146.4737	1739.7949	1600.3314
3	981.0154	1476.0139	1387.1444
4	887.9426	1275.1856	1245.7814
5	826.1492	1147.3286	1139.9626
6	763.7392	1059.9093	1057.3627
7	720.7542	996.5809	989.2442
8	686.8841	941.5902	927.5167
9	664.7453	897.1673	898.8134
10	628.8622	868.6342	877.2288

Table 15. Elbow analysis for 2017-18 regular season.

Number of clusters	Silhouette scores for centers	Silhouette scores for wingers	Silhouette scores for defensemen
2	0.2522	0.2528	0.2138
3	0.1882	0.2006	0.1841
4	0.1680	0.1985	0.1769
5	0.1446	0.1854	0.1500
6	0.1527	0.1894	0.1548
7	0.1422	0.1741	0.1568
8	0.1453	0.1797	0.1595
9	0.1542	0.1666	0.1488
10	0.1372	0.1512	0.1313

Table 16. Silhouette analysis for 2017-18 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	-1.0140	-0.8609	-0.9248	-0.6313	-0.1494	1.1002	0.4168	-0.1669
	1	0.0548	0.0239	0.0650	-0.1234	-0.1946	-0.4277	-0.3827	-0.2047
	2	1.1751	0.9984	1.0713	0.9589	0.6221	-0.4075	-0.4740	-0.3202
	3	-0.9957	-0.7422	-0.9763	-0.5967	-0.6042	0.0868	2.0549	2.4305
W	0	-1.0172	-0.8779	-0.8612	-0.6392	-0.3589	0.6816	0.4978	0.1835
	1	0.0705	0.1040	0.0254	0.0289	0.1334	-0.1970	-0.3428	-0.3628
	2	-1.4592	-1.1170	-1.3484	-0.8597	-0.9351	0.5939	2.4478	3.2484
	3	1.2421	0.9684	1.1338	0.7922	0.3592	-0.4988	-0.5091	-0.3487
D	0	-0.2105	-0.1953	-0.1861	0.4328	0.3981	0.2878	-0.2584	-0.2297
	1	1.2755	1.1236	1.1509	0.2166	0.1178	-0.5586	-0.5661	-0.3477
	2	-0.8299	-0.7174	-0.7541	-0.7865	-0.6520	0.0938	0.8531	0.6217

Table 17. Cluster labels for 2010-11 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	-0.1467	-0.0617	-0.1714	0.4327	0.4141	0.3048	0.2282	-0.1337
	1	-1.1351	-1.0209	-0.9808	-0.8218	-0.5280	0.9244	1.6440	1.4969
	2	-0.5721	-0.5461	-0.4743	-0.8857	-0.7404	-0.1674	-0.4281	-0.2526
	3	1.2372	1.0499	1.1091	0.6108	0.3655	-0.7006	-0.7318	-0.3470
W	0	-1.2883	-1.0335	-1.2422	-0.7430	-0.7949	0.2498	1.7739	2.7906
	1	-0.9811	-0.8724	-0.8743	-0.7852	-0.5176	0.5770	0.4556	-0.1334
	2	0.0131	0.0062	0.0162	-0.0733	0.1096	-0.1692	-0.2338	-0.2632
	3	1.1361	0.9870	1.0320	0.9264	0.5163	-0.3575	-0.5838	-0.3286
D	0	-0.8480	-0.7279	-0.7338	-0.3164	-0.3657	0.9117	1.1220	0.8660
	1	-0.3654	-0.2457	-0.3440	-0.2314	-0.1628	-0.1269	-0.3167	-0.3835
	2	0.9786	0.7707	0.8752	0.4621	0.4274	-0.5217	-0.4783	-0.2272

Table 18. Cluster labels for 2011-12 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	1.0801	0.8076	0.9836	0.6302	0.4330	-0.5006	-0.3955	-0.1864
	1	-1.1181	-0.8832	-0.9887	-0.5841	-0.9115	0.7644	1.8471	1.5465
	2	-0.1637	-0.0172	-0.2148	0.2348	-0.1380	0.7699	-0.1406	-0.1908
	3	-0.7407	-0.6429	-0.6188	-0.8140	-0.0022	-0.5074	-0.1929	-0.2772
W	0	-0.6960	-0.6278	-0.5424	-0.2499	-0.1701	0.7735	0.2361	-0.2193
	1	-1.1557	-0.7180	-1.1476	-0.7102	-0.9012	0.3385	1.7902	2.0783
	2	0.0542	0.1050	-0.0005	-0.1249	0.1560	-0.5086	-0.4697	-0.3616
	3	1.2983	0.9186	1.2040	0.8305	0.4383	-0.2959	-0.5283	-0.3356
D	0	-0.5442	-0.5369	-0.4156	-0.1972	-0.2359	0.1246	-0.2175	-0.3529
	1	0.9827	0.8091	0.8339	0.3585	0.3408	-0.4981	-0.4052	-0.2897
	2	-0.6835	-0.3152	-0.7085	-0.2530	-0.1055	0.7676	1.5128	1.6313

Table 19. Cluster labels for 2012-13 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	-0.9011	-0.7723	-0.7523	-0.5202	-0.0912	0.3884	0.2373	-0.0707
	1	0.2027	0.2756	0.0964	-0.1089	-0.1834	-0.2944	-0.1458	-0.1795
	2	1.1693	0.8408	1.0917	1.0816	0.6373	-0.4375	-0.5651	-0.3299
	3	-1.3943	-1.0666	-1.2558	-1.3128	-1.2960	1.5382	2.1000	2.9955
W	0	1.3300	1.1403	1.1433	0.9971	0.4654	-0.4785	-0.6008	-0.3334
	1	-1.2445	-0.7947	-1.2819	-1.0407	-0.9876	0.0909	1.9583	2.5489
	2	-0.8395	-0.7377	-0.7076	-0.5395	-0.3272	0.7516	0.3553	-0.0871
	3	0.1165	0.0467	0.1416	0.0365	0.2092	-0.3423	-0.3845	-0.3356
D	0	-0.6687	-0.4372	-0.6529	-0.6903	-0.8915	0.7329	0.9144	0.9641
	1	-0.4058	-0.3710	-0.3440	0.1883	0.4127	-0.0187	-0.2039	-0.3071
	2	1.2883	1.0134	1.1733	0.3465	0.1728	-0.6613	-0.5327	-0.4124

Table 20. Cluster labels for 2013-14 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	-1.0486	-0.9625	-0.8574	-0.5336	-0.6384	0.7522	2.0866	2.6043
	1	1.3328	0.9423	1.2897	0.8927	0.6718	-0.6742	-0.5661	-0.2607
	2	0.1225	0.2097	0.0310	0.0697	0.1350	-0.1197	-0.1295	-0.2875
	3	-1.0536	-0.8843	-0.9203	-0.7371	-0.6207	0.5617	0.0691	-0.0932
W	0	-1.0050	-0.8723	-0.8345	-0.8989	-0.6335	1.0519	0.2778	-0.1939
	1	1.2446	1.1139	1.0064	0.9990	0.6938	-0.2249	-0.6017	-0.3942
	2	-0.9824	-0.6930	-0.9433	-0.3698	-0.2637	-0.0141	1.6464	1.9322
	3	0.1740	0.0817	0.1999	0.0599	0.0489	-0.4627	-0.3354	-0.2761
D	0	1.3850	1.2267	1.1865	0.4869	0.2782	-0.4489	-0.4059	-0.2692
	1	-0.2748	-0.3037	-0.2064	0.2460	0.4081	-0.2160	-0.5072	-0.3508
	2	-0.7210	-0.5664	-0.6525	-0.6639	-0.6999	0.5992	0.9153	0.6243

Table 21. Cluster labels for 2014-15 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	1.2445	0.8845	1.1941	0.7047	0.2547	-0.6047	-0.6440	-0.3750
	1	-0.9384	-0.8386	-0.7763	-0.7849	-0.5409	0.3502	0.0429	-0.2455
	2	-1.1739	-0.9606	-1.0351	-0.8504	-0.4426	0.6342	1.9784	2.0850
	3	0.1213	0.2457	0.0012	0.3041	0.3344	0.0013	-0.1067	-0.1145
W	0	1.2527	1.0201	1.1029	0.7571	0.2799	-0.5369	-0.6622	-0.3512
	1	-0.9612	-0.5523	-1.0362	-0.8301	-0.7597	0.2105	1.8133	2.4308
	2	0.0476	-0.0057	0.0786	0.0889	0.4122	-0.2052	-0.2344	-0.2720
	3	-1.0120	-0.8533	-0.8669	-0.5984	-0.6082	0.7996	0.3048	-0.2310
D	0	-0.8214	-0.5722	-0.7839	-0.1972	0.1633	0.5345	1.5049	1.3137
	1	-0.2678	-0.2560	-0.2237	-0.2617	-0.3250	0.0284	-0.3015	-0.3204
	2	1.2901	1.0575	1.1583	0.7609	0.6080	-0.5060	-0.5529	-0.3519

Table 22. Cluster labels for 2015-16 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	-0.8877	-0.6425	-0.8531	-1.1292	-1.1209	-0.0604	0.0150	0.0003
	1	1.2026	0.9194	1.1218	0.5324	0.2714	-0.5863	-0.7090	-0.3675
	2	-0.2300	-0.2142	-0.1880	0.2532	0.4847	0.2988	0.0209	-0.2160
	3	-0.9986	-0.6875	-0.9840	-0.3413	-0.4575	0.7518	2.0485	1.9532
W	0	-0.9943	-0.6877	-0.9507	-0.9934	-0.4832	0.6495	0.1977	-0.0917
	1	-0.1327	-0.1471	-0.0830	0.2231	0.2227	-0.2451	0.0004	-0.1200
	2	-1.2946	-1.1794	-1.0126	-0.6532	-1.1798	0.4773	2.4830	3.4898
	3	1.1853	0.9290	1.0468	0.6661	0.3228	-0.3262	-0.5495	-0.3223
D	0	-0.4494	-0.3668	-0.4044	-0.0406	-0.1060	0.2919	-0.2428	-0.3677
	1	-0.6068	-0.3485	-0.6082	-0.4216	-0.1335	-0.0210	1.3182	1.3882
	2	1.2372	0.9038	1.1583	0.3766	0.2849	-0.5037	-0.5204	-0.3489

Table 23. Cluster labels for 2016-17 regular season.

Position	Cluster label	TOIPoints	TOIGoals	TOIAssists	TOI+/-	TOIEVDShots	TOIBlocks	TOIHits	TOIPenalties
C	0	0.0001	0.0184	-0.0123	-0.0891	-0.2651	-0.1544	-0.5900	-0.5418
	1	-1.0502	-1.0462	-0.8255	-0.7827	-0.8952	0.8122	1.0588	0.4985
	2	1.3946	1.1165	1.2802	0.8058	0.2678	-0.6570	-0.6652	-0.1202
	3	-0.5227	-0.2325	-0.6054	0.0231	1.0447	0.1582	0.5984	0.4748
W	0	0.0343	0.0551	0.0080	-0.0361	-0.1323	-0.2676	-0.4337	-0.2572
	1	1.4673	1.1179	1.2728	0.9644	0.4262	-0.5955	-0.6203	-0.3098
	2	-0.9772	-0.7622	-0.8345	-0.4986	0.0512	0.7396	0.8042	0.1294
	3	-0.8767	-0.9136	-0.5769	-1.0403	-1.2589	0.8292	2.7171	4.3438
D	0	-0.3462	-0.2011	-0.3402	0.2986	0.2869	0.4266	-0.2415	-0.1997
	1	-0.7987	-0.6782	-0.7011	-0.7948	-0.6851	0.1380	1.0945	0.8241
	2	1.2156	0.8814	1.1259	0.1656	0.0948	-0.8055	-0.5035	-0.3501

Table 24. Cluster labels for 2017-18 regular season.

Team	MeanC	MeanW	MeanD
Anaheim Ducks	2.32	2.21	2.17
Columbus Blue Jackets	2.14	2.40	2.05
Calgary Flames	1.70	1.81	1.92
Chicago Blackhawks	2.28	1.73	1.73
Colorado Avalanche	2.25	2.50	2.00
Dallas Stars	2.19	2.41	2.38
Detroit Red Wings	1.70	2.01	1.62
Edmonton Oilers	2.36	2.51	2.46
Florida Panthers	2.33	2.68	2.15
Los Angeles Kings	2.00	2.25	1.83
Minnesota Wild	1.83	2.88	2.22
Nashville Predators	2.19	1.92	1.68
New York Islanders	2.05	2.62	2.13
Ottawa Senators	2.55	2.68	2.57
Philadelphia Flyers	2.03	1.72	1.94
Phoenix Coyotes	2.30	1.83	1.96
Pittsburgh Penguins	2.25	2.38	1.83
Toronto Maple Leafs	2.28	2.33	2.25
Vancouver Canucks	2.28	1.86	1.88
Washington Capitals	2.25	2.00	2.00

Table 25. Mean position per team for the 2010-2011 regular season.

Team	MeanC	MeanW	MeanD
Anaheim Ducks	2.40	2.17	2.47
Buffalo Sabres	2.50	1.78	2.07
Calgary Flames	2.33	2.50	1.78
Chicago Blackhawks	1.60	1.98	1.42
Colorado Avalanche	1.62	2.53	1.71
Dallas Stars	2.44	2.06	2.04
Detroit Red Wings	1.50	2.00	1.56
Edmonton Oilers	2.29	2.38	2.43
Florida Panthers	2.05	2.15	1.77
Los Angeles Kings	2.12	2.56	1.67
Minnesota Wild	3.50	2.75	2.33
New Jersey Devils	2.14	2.08	2.13
Nashville Predators	1.35	1.99	1.61
New York Islanders	2.30	2.23	2.47
Ottawa Senators	1.79	2.09	1.63
Philadelphia Flyers	2.50	1.17	2.08
Phoenix Coyotes	2.12	2.12	1.87
Pittsburgh Penguins	1.80	1.68	1.92
St. Louis Blues	2.29	2.12	1.48
Tampa Bay Lightning	2.25	2.38	2.17
Toronto Maple Leafs	1.83	2.58	1.61
Vancouver Canucks	2.38	1.88	1.58

Table 26. Mean position per team for the 2011-2012 regular season.

Team	MeanC	MeanW	MeanD
Carolina Hurricanes	1.50	2.75	1.67
Calgary Flames	1.81	2.31	1.67
Chicago Blackhawks	2.00	1.50	1.50
Colorado Avalanche	1.81	2.62	2.12
Edmonton Oilers	2.30	2.20	2.07
Montreal Canadiens	1.00	2.25	1.67
New Jersey Devils	2.67	2.22	1.74
Nashville Predators	2.33	2.83	1.67
New York Rangers	1.75	2.38	1.50
Ottawa Senators	2.25	2.25	1.50
Philadelphia Flyers	2.50	1.90	2.30
Phoenix Coyotes	1.69	2.00	1.96
Pittsburgh Penguins	2.00	2.04	1.83
Tampa Bay Lightning	1.50	2.12	1.83
Vancouver Canucks	2.62	2.33	1.42
Washington Capitals	1.75	2.00	1.92

Table 27. Mean position per team for the 2012-2013 regular season.

Team	MeanC	MeanW	MeanD
Boston Bruins	2.00	1.62	1.75
Buffalo Sabres	2.75	2.88	2.17
Carolina Hurricanes	2.38	2.62	2.04
Chicago Blackhawks	1.95	1.58	1.53
Detroit Red Wings	2.42	2.02	1.33
Edmonton Oilers	2.65	2.50	2.60
Minnesota Wild	2.50	2.25	1.83
Montreal Canadiens	2.00	2.62	2.00
New Jersey Devils	2.09	2.17	1.27
New York Islanders	2.00	2.50	1.94
Ottawa Senators	2.00	2.50	1.83
Philadelphia Flyers	2.05	2.15	2.17
Pittsburgh Penguins	2.25	2.09	2.25
San Jose Sharks	1.75	2.25	1.33
St. Louis Blues	1.81	1.85	1.70
Toronto Maple Leafs	2.56	2.62	2.56
Vancouver Canucks	2.00	2.59	1.58
Winnipeg Jets	2.00	2.46	1.58
Washington Capitals	2.04	2.12	2.10

Table 28. Mean position per team for the 2013-2014 regular season.

Team	MeanC	MeanW	MeanD
Anaheim Ducks	1.76	2.10	1.76
Arizona Coyotes	2.81	2.85	2.31
Boston Bruins	1.50	2.25	2.17
Columbus Blue Jackets	2.25	2.54	1.56
Chicago Blackhawks	2.12	1.84	1.81
Detroit Red Wings	2.38	1.62	1.67
Edmonton Oilers	2.33	2.42	2.55
Florida Panthers	2.50	2.24	2.22
Montreal Canadiens	1.62	2.31	1.88
New Jersey Devils	2.38	2.69	2.08
New York Rangers	1.50	1.56	1.62
Ottawa Senators	1.88	2.06	2.38
Vancouver Canucks	1.75	2.19	2.26
Winnipeg Jets	2.25	2.33	2.19
Washington Capitals	1.80	2.02	2.03

Table 29. Mean position per team for the 2014-2015 regular season.

Team	MeanC	MeanW	MeanD
Arizona Coyotes	2.56	1.83	2.04
Buffalo Sabres	2.25	2.75	2.33
Columbus Blue Jackets	2.50	2.00	2.17
Chicago Blackhawks	1.83	2.08	1.56
Colorado Avalanche	2.50	2.50	1.83
Edmonton Oilers	2.04	2.41	2.36
Florida Panthers	2.12	2.09	1.92
Minnesota Wild	1.88	2.03	1.67
Montreal Canadiens	1.75	2.38	2.17
New Jersey Devils	2.50	2.44	2.17
Nashville Predators	2.50	2.12	1.50
New York Islanders	1.75	2.18	1.79
New York Rangers	2.00	1.88	1.83
Ottawa Senators	1.85	2.30	2.03
Pittsburgh Penguins	1.79	1.96	1.72
St. Louis Blues	2.00	1.88	1.83
Vancouver Canucks	2.45	2.48	2.13
Winnipeg Jets	1.88	2.25	2.00
Washington Capitals	2.15	1.50	1.90

Table 30. Mean position per team for the 2015-2016 regular season.

Team	MeanC	MeanW	MeanD
Arizona Coyotes	2.75	2.25	2.23
Buffalo Sabres	1.50	2.38	2.08
Carolina Hurricanes	1.92	2.38	1.56
Calgary Flames	1.75	2.00	1.67
Chicago Blackhawks	1.50	1.78	1.73
Dallas Stars	2.00	1.88	2.17
Edmonton Oilers	1.50	2.11	1.93
Montreal Canadiens	2.33	1.75	1.72
New Jersey Devils	2.85	2.33	1.80
Nashville Predators	1.92	2.01	1.60
New York Islanders	1.68	1.88	2.00
New York Rangers	1.25	1.64	1.67
Pittsburgh Penguins	1.35	1.65	1.63
Winnipeg Jets	1.75	2.25	1.67
Washington Capitals	1.50	1.67	1.40

Table 31. Mean position per team for the 2016-2017 regular season.

Team	MeanC	MeanW	MeanD
Arizona Coyotes	2.72	2.20	1.68
Carolina Hurricanes	2.50	2.12	1.79
Columbus Blue Jackets	3.25	1.72	1.58
Calgary Flames	2.50	2.33	2.00
Colorado Avalanche	1.75	2.12	2.17
Edmonton Oilers	1.25	2.57	2.24
Florida Panthers	1.75	2.50	1.83
Minnesota Wild	1.79	1.88	1.51
New Jersey Devils	2.17	1.98	1.58
Nashville Predators	1.75	1.75	1.67
New York Islanders	2.50	2.16	1.94
New York Rangers	3.00	2.03	2.25
Ottawa Senators	3.08	2.54	2.33
Pittsburgh Penguins	2.00	2.12	1.67
St. Louis Blues	2.46	2.41	1.69
Vegas Golden Knights	1.75	2.08	1.45
Winnipeg Jets	2.33	1.96	1.67
Washington Capitals	1.58	2.15	1.93

Table 32. Mean position per team for the 2017-2018 regular season.

	Win	GD	GF	GA
Constant	0.3954***	-0.5932***	2.5879***	3.8110***
	(0.0290)	(0.2410)	(0.0980)	(0.0980)
Mean of Centers	0.0810**	0.4052**	0.2763*	-0.1288
	(0.0410)	(0.2040)	(0.1420)	(0.1410)
Mean of Wingers	0.0753***	0.4676***	0.2427***	-0.2249**
	(0.0260)	(0.1300)	(0.0900)	(0.9000)
Mean of Defensemen	0.0694**	0.3505***	0.0605	-0.2900***
	(0.0270)	(0.1340)	(0.0930)	(0.002)
Number of Observations	828	828	828	828

Table 33. OLS in relation to the mean position and the variables of interest.
Notes: ***, **, *, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses.

	VIF
Constant	2.7960
Mean of Centers	1.1270
Mean of Wingers	1.1580
Mean of Defensemen	1.2290
Number of Observations	828

Table 34 Variance Inflation Factor (VIF).
Notes: VIF = 1, no correlation; VIF < 5, moderate correlation; VIF > 5, high correlation.

	Win	GD	GF	GA
Constant	0.2056***	-1.7004***	2.1086***	3.8089***
	(0.0720)	(0.3560)	(0.2460)	(0.2450)
Elite Centers	0.0316***	0.1633***	0.0629	-0.1004**
	(0.0120)	(0.0580)	(0.0400)	(0.0400)
Elite Wingers	0.0351*	0.2066**	0.1645**	-0.0420
	(0.0190)	(0.0960)	(0.0660)	(0.0660)
Elite Defensemen	0.0290*	0.2804**	0.0266	-0.1540***
	(0.0150)	(0.0770)	(0.0530)	(0.0530)
Number of Observations	828	828	828	828

Table 35. OLS in relation to the addition of an elite player and the variables of interest.
Notes: ***, **, *, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses.

	Win	GD	GF	GA
No Elite Centers	0.3915***	-0.5608***	2.6931***	3.2540***
	(0.0360)	(0.1790)	(0.1230)	(0.1230)
One Elite Center	0.4837***	-0.1471	2.7222***	2.8693***
	(0.0280)	(0.1410)	(0.0970)	(0.0970)
Two Elite Centers	0.5780***	0.4362***	3.1844***	2.7482***
	(0.0300)	(0.1470)	(0.1010)	(0.1010)
Three Elite Centers	0.5745***	0.5319	3.0851***	2.5532***
	(0.0150)	(0.3600)	(0.2470)	(0.2480)
Four Elite Centers	0.5000***	0.7500	3.7500***	3.0000***
	(0.2480)	(1.2330)	(0.8470)	(0.8490)
Number of Observations	828	828	828	828

Table 36. OLS in relation to the quantity of elite centers and the variables of interest.
Notes: ***, **, *, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses.

	Win	GD	GF	GA
No Elite Winger	0.4200***	-0.3800	2.5000***	2.8800***
	(0.0700)	(0.3480)	(0.2400)	(0.2390)
One Elite Winger	0.3878***	-0.6871***	2.6259***	3.3129***
	(0.0410)	(0.2030)	(0.1400)	(0.1400)
Two Elite Wingers	0.4704***	-0.1344	2.7945***	2.9289***
	(0.0310)	(0.1550)	(0.1070)	(0.1060)
Three Elite Wingers	0.5421***	0.1495	3.0794***	2.9299***
	(0.0340)	(0.1680)	(0.1160)	(0.1160)
Four Elite Wingers	0.6300***	0.6500***	3.0900***	2.4400***
	(0.0500)	(1.2460)	(0.1700)	(0.1690)
Five Elite Wingers	0.5429***	0.6286	3.2286***	2.600***
	(0.0840)	(1.4160)	(0.2870)	(0.2860)
Six Elite Wingers	0.7059***	1.5294**	3.7059***	2.1765***
	(0.1200)	(0.5960)	(0.4110)	(0.4100)
Seven Elite Wingers	0.6000***	0.4000	2.6000***	2.200***
	(0.2220)	(1.0990)	(0.7590)	(0.7570)
Eight Elite Wingers	0.5714***	1.0000	3.5714***	2.5714***
	(0.1880)	(0.9290)	(0.6410)	(0.6400)
Number of Observations	828	828	828	828

Table 37. OLS in relation to the quantity of elite wingers and the variables of interest.
Notes: ***, **, *, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses.

	Win	GD	GF	GA
No Elite Defenseman	0.4321***	-0.5679**	2.7407***	3.3086***
	(0.0550)	(0.2740)	(0.1890)	(0.1880)
One Elite Defenseman	0.3687***	-0.5707***	2.6616***	3.2323***
	(0.0350)	(0.1750)	(0.1210)	(0.1200)
Two Elite Defensemen	0.5364***	0.2529*	3.0728***	2.8199
	(0.0310)	(0.1530)	(0.1050)	(0.1050)
Three Elite Defensemen	0.5810***	0.3296*	3.0391***	2.7095***
	(0.0370)	(0.1840)	(0.1270)	(0.1260)
Four Elite Defensemen	0.5816***	0.3980	2.8469***	2.4490***
	(0.0500)	(1.249)	(0.1720)	(0.1710)
Five Elite Defensemen	0.4545***	-0.4545	2.3636***	2.8182***
	(0.1490)	(0.7430)	(0.5130)	(0.5090)
Six Elite Defensemen	0.0000	0.0000	0.0000	0.0000
	0.0000	0.0000	0.0000	0.0000
Number of Observations	828	828	828	828

Table 38. OLS in relation to the quantity of elite defensemen and the variables of interest.
Notes: ***, **, *, indicates significance at 1%, 5%, and 10% levels, respectively. Standard errors are in parentheses.