Abstract

Salary Cap modelling has become an increasingly relevant application of analytics in all major sports, as projecting the current and future value of players helps teams forecast and manage future payroll constraints and retain talent. Front offices across the National Hockey League (NHL) have been employing such analyses for decades, but the recent data revolution experienced within the sport has increased the amount of data available for player evaluation. Public blogs have begun parsing data published by the NHL to create consumable data sets for the league’s armchair analysts. Among these data sets are corpi of information to measure on-ice player performance as well as off-ice contract details. Measures of player similarity can be calculated and player contract valuations can be estimated to enable us to ponder:

*If Player A is 95% similar to Player B, can we assume that Player A’s contract will reasonably resemble the terms of Player B?*

To further explore this question, we’ll combine historical player performance data with historical contract valuation data to establish a database of precedent examples of varying player qualities and requisite paydays. Once this database is established, we’ll intake player performance data from the recently-begun 2020-21 season to estimate a player’s value based on their season output thus far.

Team Members

*Self-identified skills and planned contributions*

**Andrew Napolitano** is a second-term MSDS Master’s student with proficiency in data pre-processing, statistical concepts, and visualization/concept delivery. While possessing a foundational pre-processing skillset, Andrew hopes to gain experience managing and combining data sets for analytic use during this project. Additionally, he appreciates the opportunity to practice clean workflow creation and repeatability practices.

**Abdulaziz Alquzi** has a background in Electronics and Communication Engineering. My focus for this project will be obtaining experience on how to handle dataset files with large sizes and analyzing, combining . Also, I'm looking forward to obtaining experience in pre-processing, exploring, and analyzing. Even though I have gained knowledge scraping data and exploring and analyzing on the previous project . I'll appreciate the opportunity to gain more experience and practice on this project.

**Evan Falkowski** is a first year master’s in data science student. I obtained my bachelor’s in computer science, and therefore have a strong background in programming. While also obtaining and having experience in statistical analysis and pre-processing. I hope to further gain experience in the pre-processing of data in a new domain, as well as being able to predict valuable insights from that data. While practicing healthy data science habits throughout our projects lifecycle.

**Wynton Britton** is a first year MSDS student at Drexel University. I also have a computer science background, having just graduated this past spring. I have experience wrangling data and pre-processing, as well as using applications like Excel and Tableau to visualize data. I hope to learn to utilize various statistical analysis and machine learning techniques so I may help to implement them in our project and in the future. I also hope to foster collaborative work habits throughout the course of the project.

Analysis Description

This analysis is intended to produce an estimated contract value for players in the NHL that can be used to manage current and future salary cap constraints. A thorough exploration of this topic is deeply valuable to front offices across the NHL and any other similarly established sports leagues.

There are a few notable constraints of the publicly available data being used for this analysis:

* PuckPedia is only able to provide contract valuation data from 2018-onward. The current NHL Collective Bargaining Agreement with the National Hockey League Players Association (NHLPA) has been active since 2013. Theoretically, the salary cap constraints have not changed since then, and data from that period would strengthen our ability to produce accurate estimates.
* Player valuation metrics are limited to publicly available data points, such as scoring rates and shot production rates, yet there are far more advanced and accurate methods of measuring player value in the private domain.

Given these and other constraints, this analysis may not be as robust as theoretically possible, but a proof of concept may be just as valuable in the short term.

**Dataset Selection**

This analysis will require the following datasets:

* Player Contract Valuation Data provided by [PuckPedia.com](https://puckpedia.com/)
  + Free contract data from 2018-onward
* Player Performance Data provided by [PuckPedia.com](https://puckpedia.com/)
  + Free on-ice impact data from 2018-onward

Initial exploratory analysis was conducted to ensure cooperation between both datasets. The details of this analysis are presented in these [slides](https://docs.google.com/presentation/d/1y888bcBDiavXlwR0rNGFSMBA91PflR-YoT_QY5NlaiY/edit?usp=sharing) and confirm that the ending population will contain about 4,500 distinct players with contract and performance.

**Analysis Methodology**

***Step By Step Planned Methodology:***

1. Merge historical player performance data with historical player contract valuation to create table of historical player data
2. Intake Current Season Player performance data
3. Conduct similarity analyses with methods such as cosine similarity and perhaps some regression/neural network concepts covered later in the course.
4. Estimate contract term and value based on preceding contract terms for similar players

Once completed, the code and results of this analysis can be published to GitHub and socialized within the hockey analytic community and sport analytic community as a whole.