Stat468 Final Project

Trevor Steunenberg

2025-07-12

Table of contents

1 Index

2 Abstract

In the days leading up to and days of the 2025 NHL Entry Draft there were 18 trades which only included draft picks. This report aims to use player contribution data to determine the relative value of selections in the NHL Entry Draft. Knowing the relative value of picks allows NHL teams to both determine whether they should accept trade offers they have received as well as propose favourable trades to other teams.

Depending on the context, there can be countless approaches one could take to quantify the value of a draft pick. As one example, one could estimate the fairness of a trade by comparing the assets given up and acquired to previous trades. In contrast, this report will estimate value of pick n by utilizing the points contributed and games played by previous players selected at pick n along with a k-nearest neighbours algorithm.

The data used by this report is imported from HockeyReference, which has data on the NHL Draft and player games played and point share counts dating back to 1963, though we will only use a subset of this data as will be explained later.

Some work in this area has been done before, such as:

- Valuation of NHL Draft Picks using Functional Data Analysis
- Examining the value of NHL Draft picks
- NHL draft: What does it cost to trade up?

As an interesting aside, Eric Tulsky, who wrote the last article listed above in 2013, was hired as General Manager of the Carolina Hurricanes in 2024. One critical point that it is common knowledge in ice hockey circles and confirmed by the resources above, along with the Model chapter of this report is that NHL draft picks do **not** decrease in value linearly. In particular, the difference in value pick 1 and 30 is much greater than between pick 101 and 130.

Note that if picks did decrease linearly in value linearly then it would be very easy to create a model of draft pick value since we would have

$$v_1 = v_2 + c = v_3 + 2c = ... = v_{224} + 223c$$

where $c > 0$ and v_i is the value of the i^{th} selection, meaning we would only have to find the value of c.

3 Question

This report will estimate the relative value of selections in the NHL Entry Draft. I am not sure what else I am supposed to put here.

4 Import

5 Introduction

As mentioned before, we will be importing data from Hockey Reference. Before we import anything, it's important to consider *which* and *how many* years we want to include in this analysis. Since the NHL has changed dramatically over the years, care must be taken to ensure we do not include drafts from too long ago. The primary concern with including data from too many years ago is that teams have likely changed their drafting approach over time. For example, teams may have become better at evaluating prospects as more advanced statistics have been developed, meaning that there are likely fewer late round draft "steals" in the 2020s than there were in the 1980s. Thus including drafts from the 1980s would skew our calculations because it would overestimate contributions by players who were drafted in the later rounds, since those players would potentially have been drafted sooner if the teams of the 1980s had the resources available to teams today. This would make our model a poor estimator of draft pick value for drafts occurring in the 2020s. That being said, players drafted in recent years have not had sufficient time to contribute to their teams, so we should not include drafts from too recently either. Ideally, we would wait until all players from a draft class have retired before including it in our analysis . Practically speaking, this is not feasible since players can have very long careers (for example, Alex Ovechkin was drafted in 2004 and is still playing) which would force us to include older drafts to maintain the same sample size, which is also not ideal as explained above.

Having considered this, we make the somewhat arbitrary decision to use the 25 drafts between and 1996 and 2020 (inclusive). Note that a significant portion of the players in our dataset are still active, so we will have to make an adjustment to account for this. Additionally, it makes sense to give more recent drafts more weight for the reasons described above. We will make both of these adjustments in the Transform chapter.

6 Setup

```
install.packages("rvest")
```

```
The following package(s) will be installed:
- rvest [1.0.4]
These packages will be installed into "~/Downloads/Stat468_Final/renv/library/windows/R-4.5/
```

```
# Installing packages --------------------------------------------------------
- Installing rvest ...                          OK [linked from cache]
Successfully installed 1 package in 15 milliseconds.
```

```
install.packages("tidyverse")
```

```
The following package(s) will be installed:
- tidyverse [2.0.0]
These packages will be installed into "~/Downloads/Stat468_Final/renv/library/windows/R-4.5/
```

```
# Installing packages --------------------------------------------------------
- Installing tidyverse ...                      OK [linked from cache]
Successfully installed 1 package in 16 milliseconds.
```

```
install.packages("janitor")
```

```
The following package(s) will be installed:
- janitor [2.2.1]
These packages will be installed into "~/Downloads/Stat468_Final/renv/library/windows/R-4.5/
```

```
# Installing packages --------------------------------------------------------
- Installing janitor ...                        OK [linked from cache]
Successfully installed 1 package in 11 milliseconds.
```

```
library(rvest)
library(tidyverse)
library(janitor)
```

7 Code

We start off by creating a function to import data from Hockey Reference.

```
start_year <- 1996
end_year <- 2020

import_draft <- function(year){
  url <- str_glue("https://www.hockey-reference.com/draft/NHL_{year}_entry.html")
  html <- read_html(url)
  draft_year__table <- html |>
    html_element("table") |>
    html_table() |>
    janitor::row_to_names(1) |>
    janitor::clean_names()
  draft_year__table
}

head(import_draft(start_year), 10)
```

```
Warning: Row 1 does not provide unique names. Consider running clean_names()
after row_to_names().
```

```
# A tibble: 10 x 21
   overall team    player nat   pos   age   to    amateur_team gp    g     a
   <chr>   <chr>   <chr>  <chr> <chr> <chr> <chr> <chr>        <chr> <chr> <chr>
 1 1       Ottawa~ Chris~ CA    D     18    2015  Prince Albe~ 1179  71    217
 2 2       San Jo~ Andre~ RU    D     18    2008  Salavat Yul~ 496   38    82
 3 3       New Yo~ J.P. ~ CA    RW    18    2011  Val-d'Or Fo~ 822   214   309
 4 4       Washin~ Alexa~ RU    C     18    2000  Barrie Colt~ 3     0     0
 5 5       Dallas~ Ric J~ CA    D     18    2007  Soo Greyhou~ 231   19    58
 6 6       Edmont~ Boyd ~ CA    C     18    2009  Kitchener R~ 627   67    112
 7 7       Buffal~ Erik ~ US    LW/C  19    2007  Minnesota (~ 545   52    76
 8 8       Boston~ Johna~ CA    D     18    2004  Medicine Ha~ 44    0     1
 9 9       Anahei~ Rusla~ BY    D     21    2011  Las Vegas T~ 917   45    159
10 10      New Je~ Lance~ CA    D     18    2004  Red Deer Re~ 209   4     12
```

```
# i 10 more variables: pts <chr>, x <chr>, pim <chr>, gp_2 <chr>, w <chr>,
#   l <chr>, t_o <chr>, sv_percent <chr>, gaa <chr>, ps <chr>
```

8 Tidy

This is a placeholder for Ch2 – Tidy.

9 Transform

This is a placeholder for Ch3 – Transform.

10 Visualize

This is a placeholder for Ch4 – Visualize.

11 Model

This is a placeholder for Ch5 – Model.

12 Communicate

This is a placeholder for Ch6 – Communicate.