# Analyzing NCAA - D1 Players' First Season in the NHL

---

Ever since the NHL debut of Cale Makar during the 2019-20 NHL Playoffs, he has dominated every conversation discussing the top defensemen in the league. He won the Calder Trophy as the NHL's most impressive Rookie, then was nominated for the Norris Trophy the next year, won it the year after that, and then won the Conn Smythe Trophy as the NHL playoff's MVP when the Avalanche won the cup in 2022. On top of that, he won the Hobey Baker award as a college sophomore at UMass and led the Hockey East in scoring as defenseman.

All of these accomplishments led me to the question:

---

# *Does success in Division 1 hockey lead to success in the NHL?*

---

First, we will import all the required packages. The main package to note here is the TopDownHockey Web Scraper. This package can be used to scrape NHL data as well as Elite Prospects data, and can be installed with `pip install TopDownHockey-Scraper`

In [ ]:
```
import TopDownHockey_Scraper.TopDownHockey_EliteProspects_Scraper as tdhepscrap
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

```
Welcome to the TopDownHockey EliteProspects Scraper, built by Patrick Bacon.
This scraper is built strictly for personal use. For commercial or professiona
l use, please look into the EliteProspects API.
If you enjoy the scraper and would like to support my work, feel free to follo
w me on Twitter @TopDownHockey. Have fun!
```

We would like to get all the players that played NCAA D1 Men's Hockey in the 2020-21 season and then played in the NHL for the 2021-22 seaon. This will likely take about a minute to run.

In [ ]:
```
leagues = ["ncaa", "nhl"]
seasons = ["2020-2021", "2021-2022"]

d1_nhl_skaters_20to22 = tdhepscrape.get_skaters(leagues, seasons)
```

```
Your scrape request is skater data from the following leagues:
ncaa and nhl
In the following seasons:
2020-2021 and 2021-2022
Beginning scrape of ncaa skater data from 2020-2021.
Successfully scraped all ncaa skater data from 2020-2021.
Beginning scrape of ncaa skater data from 2021-2022.
Successfully scraped all ncaa skater data from 2021-2022.
Scraping ncaa data is complete. You scraped skater data from 2020-2021 and 202
1-2022.
Beginning scrape of nhl skater data from 2020-2021.
Successfully scraped all nhl skater data from 2020-2021.
Beginning scrape of nhl skater data from 2021-2022.
Successfully scraped all nhl skater data from 2021-2022.
Scraping nhl data is complete. You scraped skater data from 2020-2021 and 2021
-2022.
Completed scraping skater data from the following leagues:
nhl and ncaa
Incorporating the following seasons:
2020-2021 and 2021-2022
```

In [ ]:
```
d1_nhl_skaters_20to22
```

Out[ ]:

| | player | team | gp | g | a | tp | ppg | pim | +/- | |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Cole Caufield (LW/RW) | Univ. of Wisconsin | 31 | 30 | 22 | 52 | 1.68 | 4 | 9 | https://www.eliteprospects.com/pla |
| **1** | Odeen Tufto (C) | Quinnipiac Univ. | 29 | 8 | 39 | 47 | 1.62 | 22 | 8 | https://www.eliteprospects.com/play |
| **2** | Linus Weissbach (LW) | Univ. of Wisconsin | 31 | 12 | 29 | 41 | 1.32 | 12 | 7 | https://www.eliteprospects.com/pla |
| **3** | Jordan Kawaguchi (C) | Univ. of North Dakota | 28 | 10 | 26 | 36 | 1.29 | 24 | 19 | https://www.eliteprospects.com/pl |
| **4** | Dylan Holloway (LW/C) | Univ. of Wisconsin | 23 | 11 | 24 | 35 | 1.52 | 19 | 10 | https://www.eliteprospects.com/pla |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | ... | |
| **4969** | Victor Söderström (D) | Arizona Coyotes | 16 | 0 | 0 | 0 | 0.00 | 6 | -7 | https://www.eliteprospects.com/play |
| **4970** | Austin Poganski (RW) | Winnipeg Jets | 16 | 0 | 0 | 0 | 0.00 | 7 | -3 | https://www.eliteprospects.com/pla |
| **4971** | Mikhail Maltsev (LW) | Colorado Avalanche | 18 | 0 | 0 | 0 | 0.00 | 2 | -5 | https://www.eliteprospects.com/play |
| **4972** | Olli Juolevi (D) | totals | 18 | 0 | 0 | 0 | 0.00 | 6 | -1 | https://www.eliteprospects.com/pla |
| **4973** | Mason Geertsen (LW/D) | New Jersey Devils | 25 | 0 | 0 | 0 | 0.00 | 77 | -6 | https://www.eliteprospects.com/play |

4974 rows × 14 columns

This is a dataframe with 4,974 rows because it is a compiled list of all players that played NCAA D1 or NHL between 2020 and 2022. We will now clean it up a bit and filter it down even more.

In [ ]:

```
# The player and playername columns are the same, so we can drop one of them

d1_nhl_skaters_20to22 = d1_nhl_skaters_20to22.drop(columns=["player"])

d1_nhl_skaters_20to22
```

Out[ ]:

| | team | gp | g | a | tp | ppg | pim | +/- | lin |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Univ. of Wisconsin | 31 | 30 | 22 | 52 | 1.68 | 4 | 9 | https://www.eliteprospects.com/player/316168/c. |
| **1** | Quinnipiac Univ. | 29 | 8 | 39 | 47 | 1.62 | 22 | 8 | https://www.eliteprospects.com/player/248460/o. |
| **2** | Univ. of Wisconsin | 31 | 12 | 29 | 41 | 1.32 | 12 | 7 | https://www.eliteprospects.com/player/230472/l. |
| **3** | Univ. of North Dakota | 28 | 10 | 26 | 36 | 1.29 | 24 | 19 | https://www.eliteprospects.com/player/199911/j. |
| **4** | Univ. of Wisconsin | 23 | 11 | 24 | 35 | 1.52 | 19 | 10 | https://www.eliteprospects.com/player/322111/d. |
| **...** | ... | ... | ... | ... | ... | ... | ... | ... | . |
| **4969** | Arizona Coyotes | 16 | 0 | 0 | 0 | 0.00 | 6 | -7 | https://www.eliteprospects.com/player/344555/v. |
| **4970** | Winnipeg Jets | 16 | 0 | 0 | 0 | 0.00 | 7 | -3 | https://www.eliteprospects.com/player/231724/a. |
| **4971** | Colorado Avalanche | 18 | 0 | 0 | 0 | 0.00 | 2 | -5 | https://www.eliteprospects.com/player/312218/m. |
| **4972** | totals | 18 | 0 | 0 | 0 | 0.00 | 6 | -1 | https://www.eliteprospects.com/player/196391/o. |
| **4973** | New Jersey Devils | 25 | 0 | 0 | 0 | 0.00 | 77 | -6 | https://www.eliteprospects.com/player/64567/ma. |

4974 rows × 13 columns

Now to test something out, lets see if Cole Caufield is in the dataframe three times. He should be in there for the University of Wisconsin and then for the Montreal Canadiens twice.

In [ ]:
```
d1_nhl_skaters_20to22.loc[(d1_nhl_skaters_20to22.playername=="Cole Caufield")]
```

Out[ ]:

| team | gp | g | a | tp | ppg | pim | +/- | link | season | league | playername | position |
|---|---|---|---|---|---|---|---|---|---|---|---|---|

Nothing showed up because after inspection, we see there is whitespace at the end of the names in `playername` so we can use `str.strip()` to fix that.

In [ ]:
```
d1_nhl_skaters_20to22.playername = d1_nhl_skaters_20to22.playername.str.strip()

# Now we can see if it worked
d1_nhl_skaters_20to22.loc[(d1_nhl_skaters_20to22.playername=="Cole Caufield")]
```

Out[ ]:

| | team | gp | g | a | tp | ppg | pim | +/- | link |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Univ. of Wisconsin | 31 | 30 | 22 | 52 | 1.68 | 4 | 9 | https://www.eliteprospects.com/player/316168/c.. |
| **3599** | Montréal Canadiens | 10 | 4 | 1 | 5 | 0.50 | 2 | -1 | https://www.eliteprospects.com/player/316168/c.. |
| **4124** | Montréal Canadiens | 67 | 23 | 20 | 43 | 0.64 | 10 | -24 | https://www.eliteprospects.com/player/316168/c.. |

Now we will rearrange the columns, rename some, and then create 2 different dataframes; one for NCAA 2020-21 season and one for NHL 2021-22 season.

In [ ]:
```
d1_nhl_skaters_20to22 = d1_nhl_skaters_20to22.loc[:, ['playername', 'team', 'se

d1_nhl_skaters_20to22 = d1_nhl_skaters_20to22.rename(columns={'playername':'pla

# Now to create the 2 dataframes for the 2 leagues/seasons
d1_skaters_2020_21 = d1_nhl_skaters_20to22.loc[(d1_nhl_skaters_20to22.season=='

nhl_skaters_2021_22 = d1_nhl_skaters_20to22.loc[(d1_nhl_skaters_20to22.season==
```

Now we will merge these two dataframes we just created so that we can have one main dataframe. To ensure we are only keeping D1 hockey players from 2020-21 that played in the NHL in 2021-22, we will use an inner merge. We will use 'link' as the foreign key since each player has their own link no matter what team they played on.

In [ ]:
```
ncaa_2020_21_nhl_2021_22 = d1_skaters_2020_21.merge(nhl_skaters_2021_22, on = '

ncaa_2020_21_nhl_2021_22
```

Out[ ]:

| | player_x | team_x | season_x | league_x | position_x | gp_x | g_x | a_x | tp_x | ppg_x |
|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Cole Caufield | Univ. of Wisconsin | 2020-2021 | ncaa | LW/RW | 31 | 30 | 22 | 52 | 1.68 |
| **1** | Dylan Holloway | Univ. of Wisconsin | 2020-2021 | ncaa | LW/C | 23 | 11 | 24 | 35 | 1.52 |
| **2** | Shane Pinto | Univ. of North Dakota | 2020-2021 | ncaa | C | 28 | 15 | 17 | 32 | 1.14 |
| **3** | Alex Steeves | Univ. of Notre Dame | 2020-2021 | ncaa | F | 29 | 15 | 17 | 32 | 1.10 |
| **4** | Sampo Ranta | Univ. of Minnesota | 2020-2021 | ncaa | LW/RW | 31 | 19 | 12 | 31 | 1.00 |
| **5** | Matt Boldy | Boston College | 2020-2021 | ncaa | LW/RW | 22 | 11 | 20 | 31 | 1.41 |
| **6** | Thomas Bordeleau | Univ. of Michigan | 2020-2021 | ncaa | C | 24 | 8 | 22 | 30 | 1.25 |
| **7** | Ben Meyers | Univ. of Minnesota | 2020-2021 | ncaa | C/LW | 31 | 12 | 16 | 28 | 0.90 |
| **8** | Scott Reedy | Univ. of Minnesota | 2020-2021 | ncaa | C | 28 | 11 | 17 | 28 | 1.00 |
| **9** | Jackson Cates | Univ. of Minnesota-Duluth | 2020-2021 | ncaa | C | 28 | 11 | 16 | 27 | 0.96 |
| **10** | Kent Johnson | Univ. of Michigan | 2020-2021 | ncaa | LW/C | 26 | 9 | 18 | 27 | 1.04 |
| **11** | Tyce Thompson | Providence College | 2020-2021 | ncaa | RW | 25 | 11 | 14 | 25 | 1.00 |
| **12** | Nathan Smith | Minnesota State Univ. (Mankato) | 2020-2021 | ncaa | C | 28 | 9 | 16 | 25 | 0.89 |
| **13** | Jasper Weatherby | Univ. of North Dakota | 2020-2021 | ncaa | C/LW | 29 | 14 | 10 | 24 | 0.83 |
| **14** | Matty Beniers | Univ. of Michigan | 2020-2021 | ncaa | C | 24 | 10 | 14 | 24 | 1.00 |
| **15** | Marc McLaughlin | Boston College | 2020-2021 | ncaa | C | 24 | 10 | 14 | 24 | 1.00 |
| **16** | Zac Jones | UMass | 2020-2021 | ncaa | D | 29 | 9 | 15 | 24 | 0.83 |
| **17** | Ronnie Attard | Western Michigan Univ. | 2020-2021 | ncaa | D | 25 | 8 | 14 | 22 | 0.88 |
| **18** | Matt Kiersted | Univ. of North Dakota | 2020-2021 | ncaa | D | 29 | 3 | 19 | 22 | 0.76 |
| **19** | Cam York | Univ. of Michigan | 2020-2021 | ncaa | D | 24 | 4 | 16 | 20 | 0.83 |

| | player_x | team_x | season_x | league_x | position_x | gp_x | g_x | a_x | tp_x | ppg_x |
|---|---|---|---|---|---|---|---|---|---|---|
| **20** | Mike Hardman | Boston College | 2020-2021 | ncaa | W/C | 24 | 10 | 9 | 19 | 0.79 |
| **21** | Jordan Harris | Northeastern Univ. | 2020-2021 | ncaa | D | 19 | 6 | 13 | 19 | 1.00 |
| **22** | Jack McBain | Boston College | 2020-2021 | ncaa | C | 24 | 6 | 13 | 19 | 0.79 |
| **23** | Noah Cates | Univ. of Minnesota-Duluth | 2020-2021 | ncaa | LW | 28 | 5 | 14 | 19 | 0.68 |
| **24** | Jacob Bernard-Docker | Univ. of North Dakota | 2020-2021 | ncaa | D | 27 | 3 | 15 | 18 | 0.67 |
| **25** | Walker Duehr | Minnesota State Univ. (Mankato) | 2020-2021 | ncaa | RW | 28 | 10 | 7 | 17 | 0.61 |
| **26** | Alex Newhook | Boston College | 2020-2021 | ncaa | C/LW | 12 | 7 | 9 | 16 | 1.33 |
| **27** | Owen Power | Univ. of Michigan | 2020-2021 | ncaa | D | 26 | 3 | 13 | 16 | 0.62 |
| **28** | Josiah Slavin | Colorado College | 2020-2021 | ncaa | LW | 22 | 5 | 8 | 13 | 0.59 |
| **29** | Nick Blankenburg | Univ. of Michigan | 2020-2021 | ncaa | D | 26 | 5 | 8 | 13 | 0.50 |
| **30** | Bobby Brink | Univ. of Denver | 2020-2021 | ncaa | RW | 15 | 2 | 9 | 11 | 0.73 |
| **31** | Alex Vlasic | Boston Univ. | 2020-2021 | ncaa | D | 16 | 3 | 5 | 8 | 0.50 |
| **32** | Nicholas Abruzzese | Harvard Univ. | 2020-2021 | ncaa | C/LW | 0 | 0 | 0 | 0 | 0.00 |

33 rows × 25 columns

Unfortunately our sample size is very minimal now (33 players). But that was expected. We will carry on with the analysis.

We can drop irrelevant columns now though because some were duplicated in the merge.

In [ ]:
```
ncaa_2020_21_nhl_2021_22 = ncaa_2020_21_nhl_2021_22.drop(columns = ['season_x',

ncaa_2020_21_nhl_2021_22
```

Out[ ]:

| | player_x | team_x | position_x | gp_x | g_x | a_x | tp_x | ppg_x | pim_x | +/-_x | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **0** | Cole Caufield | Univ. of Wisconsin | LW/RW | 31 | 30 | 22 | 52 | 1.68 | 4 | 9 | https://ww |
| **1** | Dylan Holloway | Univ. of Wisconsin | LW/C | 23 | 11 | 24 | 35 | 1.52 | 19 | 10 | https://w |
| **2** | Shane Pinto | Univ. of North Dakota | C | 28 | 15 | 17 | 32 | 1.14 | 4 | 19 | https://ww |
| **3** | Alex Steeves | Univ. of Notre Dame | F | 29 | 15 | 17 | 32 | 1.10 | 8 | 2 | https://ww |
| **4** | Sampo Ranta | Univ. of Minnesota | LW/RW | 31 | 19 | 12 | 31 | 1.00 | 10 | 19 | https://ww |
| **5** | Matt Boldy | Boston College | LW/RW | 22 | 11 | 20 | 31 | 1.41 | 4 | 15 | https://ww |
| **6** | Thomas Bordeleau | Univ. of Michigan | C | 24 | 8 | 22 | 30 | 1.25 | 12 | 18 | https://ww |
| **7** | Ben Meyers | Univ. of Minnesota | C/LW | 31 | 12 | 16 | 28 | 0.90 | 12 | 15 | https://ww |
| **8** | Scott Reedy | Univ. of Minnesota | C | 28 | 11 | 17 | 28 | 1.00 | 6 | 20 | https://ww |
| **9** | Jackson Cates | Univ. of Minnesota-Duluth | C | 28 | 11 | 16 | 27 | 0.96 | 4 | 8 | https://w |
| **10** | Kent Johnson | Univ. of Michigan | LW/C | 26 | 9 | 18 | 27 | 1.04 | 4 | 16 | https://w |
| **11** | Tyce Thompson | Providence College | RW | 25 | 11 | 14 | 25 | 1.00 | 16 | -2 | https://w |
| **12** | Nathan Smith | Minnesota State Univ. (Mankato) | C | 28 | 9 | 16 | 25 | 0.89 | 16 | 7 | https://ww |
| **13** | Jasper Weatherby | Univ. of North Dakota | C/LW | 29 | 14 | 10 | 24 | 0.83 | 8 | 7 | https://w |
| **14** | Matty Beniers | Univ. of Michigan | C | 24 | 10 | 14 | 24 | 1.00 | 0 | 21 | https://ww |
| **15** | Marc McLaughlin | Boston College | C | 24 | 10 | 14 | 24 | 1.00 | 6 | 16 | https://ww |
| **16** | Zac Jones | UMass | D | 29 | 9 | 15 | 24 | 0.83 | 8 | 17 | https://ww |
| **17** | Ronnie Attard | Western Michigan Univ. | D | 25 | 8 | 14 | 22 | 0.88 | 8 | -3 | https://w |
| **18** | Matt Kiersted | Univ. of North Dakota | D | 29 | 3 | 19 | 22 | 0.76 | 28 | 15 | https://ww |

| | player_x | team_x | position_x | gp_x | g_x | a_x | tp_x | ppg_x | pim_x | +/-_x | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| **19** | Cam York | Univ. of Michigan | D | 24 | 4 | 16 | 20 | 0.83 | 4 | 13 | https://ww |
| **20** | Mike Hardman | Boston College | W/C | 24 | 10 | 9 | 19 | 0.79 | 14 | 5 | https://ww |
| **21** | Jordan Harris | Northeastern Univ. | D | 19 | 6 | 13 | 19 | 1.00 | 8 | -5 | https://w |
| **22** | Jack McBain | Boston College | C | 24 | 6 | 13 | 19 | 0.79 | 12 | 14 | https://w |
| **23** | Noah Cates | Univ. of Minnesota-Duluth | LW | 28 | 5 | 14 | 19 | 0.68 | 25 | 6 | https://w |
| **24** | Jacob Bernard-Docker | Univ. of North Dakota | D | 27 | 3 | 15 | 18 | 0.67 | 20 | 17 | https://w |
| **25** | Walker Duehr | Minnesota State Univ. (Mankato) | RW | 28 | 10 | 7 | 17 | 0.61 | 26 | 9 | https://ww |
| **26** | Alex Newhook | Boston College | C/LW | 12 | 7 | 9 | 16 | 1.33 | 8 | -1 | https://w |
| **27** | Owen Power | Univ. of Michigan | D | 26 | 3 | 13 | 16 | 0.62 | 6 | 18 | https://ww |
| **28** | Josiah Slavin | Colorado College | LW | 22 | 5 | 8 | 13 | 0.59 | 4 | -5 | https://w |
| **29** | Nick Blankenburg | Univ. of Michigan | D | 26 | 5 | 8 | 13 | 0.50 | 14 | 19 | https://ww |
| **30** | Bobby Brink | Univ. of Denver | RW | 15 | 2 | 9 | 11 | 0.73 | 4 | -4 | https://ww |
| **31** | Alex Vlasic | Boston Univ. | D | 16 | 3 | 5 | 8 | 0.50 | 10 | 3 | https://w |
| **32** | Nicholas Abruzzese | Harvard Univ. | C/LW | 0 | 0 | 0 | 0 | 0.00 | 0 | - | https://w |

Time to rearrange some columns.

In [ ]:
```
my_columns = list(ncaa_2020_21_nhl_2021_22.columns)

my_columns.remove('link')

my_columns.append('link')
```

In [ ]:
```
ncaa_2020_21_nhl_2021_22 = ncaa_2020_21_nhl_2021_22.loc[:, my_columns]

ncaa_2020_21_nhl_2021_22 = ncaa_2020_21_nhl_2021_22.rename(columns = {'player_x
                                                                      'g_x':'ncaa_g', 'a_x':'
                                                                      '+/-_x':'ncaa_+/-', 'te
                                                                      'tp_y':'nhl_p', 'ppg_y'
```

```
ncaa_2020_21_nhl_2021_22
```

Out[ ]:

| | player | ncaa_team | position | ncaa_gp | ncaa_g | ncaa_a | ncaa_p | ncaa_ppg | ncaa_pi |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Cole Caufield | Univ. of Wisconsin | LW/RW | 31 | 30 | 22 | 52 | 1.68 | |
| **1** | Dylan Holloway | Univ. of Wisconsin | LW/C | 23 | 11 | 24 | 35 | 1.52 | 1 |
| **2** | Shane Pinto | Univ. of North Dakota | C | 28 | 15 | 17 | 32 | 1.14 | |
| **3** | Alex Steeves | Univ. of Notre Dame | F | 29 | 15 | 17 | 32 | 1.10 | |
| **4** | Sampo Ranta | Univ. of Minnesota | LW/RW | 31 | 19 | 12 | 31 | 1.00 | 1 |
| **5** | Matt Boldy | Boston College | LW/RW | 22 | 11 | 20 | 31 | 1.41 | |
| **6** | Thomas Bordeleau | Univ. of Michigan | C | 24 | 8 | 22 | 30 | 1.25 | 1 |
| **7** | Ben Meyers | Univ. of Minnesota | C/LW | 31 | 12 | 16 | 28 | 0.90 | 1 |
| **8** | Scott Reedy | Univ. of Minnesota | C | 28 | 11 | 17 | 28 | 1.00 | |
| **9** | Jackson Cates | Univ. of Minnesota-Duluth | C | 28 | 11 | 16 | 27 | 0.96 | |
| **10** | Kent Johnson | Univ. of Michigan | LW/C | 26 | 9 | 18 | 27 | 1.04 | |
| **11** | Tyce Thompson | Providence College | RW | 25 | 11 | 14 | 25 | 1.00 | 1 |
| **12** | Nathan Smith | Minnesota State Univ. (Mankato) | C | 28 | 9 | 16 | 25 | 0.89 | 1 |
| **13** | Jasper Weatherby | Univ. of North Dakota | C/LW | 29 | 14 | 10 | 24 | 0.83 | |
| **14** | Matty Beniers | Univ. of Michigan | C | 24 | 10 | 14 | 24 | 1.00 | |
| **15** | Marc McLaughlin | Boston College | C | 24 | 10 | 14 | 24 | 1.00 | |
| **16** | Zac Jones | UMass | D | 29 | 9 | 15 | 24 | 0.83 | |
| **17** | Ronnie Attard | Western Michigan Univ. | D | 25 | 8 | 14 | 22 | 0.88 | |
| **18** | Matt Kiersted | Univ. of North Dakota | D | 29 | 3 | 19 | 22 | 0.76 | 2 |
| **19** | Cam York | Univ. of Michigan | D | 24 | 4 | 16 | 20 | 0.83 | |

| | player | ncaa_team | position | ncaa_gp | ncaa_g | ncaa_a | ncaa_p | ncaa_ppg | ncaa_pi |
|---|---|---|---|---|---|---|---|---|---|
| **20** | Mike Hardman | Boston College | W/C | 24 | 10 | 9 | 19 | 0.79 | 1 |
| **21** | Jordan Harris | Northeastern Univ. | D | 19 | 6 | 13 | 19 | 1.00 | |
| **22** | Jack McBain | Boston College | C | 24 | 6 | 13 | 19 | 0.79 | 1 |
| **23** | Noah Cates | Univ. of Minnesota-Duluth | LW | 28 | 5 | 14 | 19 | 0.68 | 2 |
| **24** | Jacob Bernard-Docker | Univ. of North Dakota | D | 27 | 3 | 15 | 18 | 0.67 | 2 |
| **25** | Walker Duehr | Minnesota State Univ. (Mankato) | RW | 28 | 10 | 7 | 17 | 0.61 | 2 |
| **26** | Alex Newhook | Boston College | C/LW | 12 | 7 | 9 | 16 | 1.33 | |
| **27** | Owen Power | Univ. of Michigan | D | 26 | 3 | 13 | 16 | 0.62 | |
| **28** | Josiah Slavin | Colorado College | LW | 22 | 5 | 8 | 13 | 0.59 | |
| **29** | Nick Blankenburg | Univ. of Michigan | D | 26 | 5 | 8 | 13 | 0.50 | 1 |
| **30** | Bobby Brink | Univ. of Denver | RW | 15 | 2 | 9 | 11 | 0.73 | |
| **31** | Alex Vlasic | Boston Univ. | D | 16 | 3 | 5 | 8 | 0.50 | 1 |
| **32** | Nicholas Abruzzese | Harvard Univ. | C/LW | 0 | 0 | 0 | 0 | 0.00 | |

We have to do a little extra munging first before we move on to analysis. For example, the numbers in the dataframe right now are objects, not floats or integers. Let's fix that. While we're at it, we can also filter out the players who have not played any NHL games (it looks like there's only 1).

In [ ]:
```
ncaa_2020_21_nhl_2021_22.ncaa_ppg = ncaa_2020_21_nhl_2021_22.ncaa_ppg.astype(fl
ncaa_2020_21_nhl_2021_22.nhl_ppg = ncaa_2020_21_nhl_2021_22.nhl_ppg.astype(floa
ncaa_2020_21_nhl_2021_22.ncaa_gp = ncaa_2020_21_nhl_2021_22.ncaa_gp.astype(floa
ncaa_2020_21_nhl_2021_22.nhl_gp = ncaa_2020_21_nhl_2021_22.nhl_gp.astype(float)
ncaa_2020_21_nhl_2021_22 = ncaa_2020_21_nhl_2021_22.loc[(ncaa_2020_21_nhl_2021_
ncaa_2020_21_nhl_2021_22
```

Out[ ]:

| | player | ncaa_team | position | ncaa_gp | ncaa_g | ncaa_a | ncaa_p | ncaa_ppg | ncaa_pi |
|---|---|---|---|---|---|---|---|---|---|
| **0** | Cole Caufield | Univ. of Wisconsin | LW/RW | 31.0 | 30 | 22 | 52 | 1.68 | |
| **2** | Shane Pinto | Univ. of North Dakota | C | 28.0 | 15 | 17 | 32 | 1.14 | |
| **3** | Alex Steeves | Univ. of Notre Dame | F | 29.0 | 15 | 17 | 32 | 1.10 | |
| **4** | Sampo Ranta | Univ. of Minnesota | LW/RW | 31.0 | 19 | 12 | 31 | 1.00 | 1 |
| **5** | Matt Boldy | Boston College | LW/RW | 22.0 | 11 | 20 | 31 | 1.41 | |
| **6** | Thomas Bordeleau | Univ. of Michigan | C | 24.0 | 8 | 22 | 30 | 1.25 | 1 |
| **7** | Ben Meyers | Univ. of Minnesota | C/LW | 31.0 | 12 | 16 | 28 | 0.90 | 1 |
| **8** | Scott Reedy | Univ. of Minnesota | C | 28.0 | 11 | 17 | 28 | 1.00 | |
| **9** | Jackson Cates | Univ. of Minnesota-Duluth | C | 28.0 | 11 | 16 | 27 | 0.96 | |
| **10** | Kent Johnson | Univ. of Michigan | LW/C | 26.0 | 9 | 18 | 27 | 1.04 | |
| **11** | Tyce Thompson | Providence College | RW | 25.0 | 11 | 14 | 25 | 1.00 | 1 |
| **12** | Nathan Smith | Minnesota State Univ. (Mankato) | C | 28.0 | 9 | 16 | 25 | 0.89 | 1 |
| **13** | Jasper Weatherby | Univ. of North Dakota | C/LW | 29.0 | 14 | 10 | 24 | 0.83 | |
| **14** | Matty Beniers | Univ. of Michigan | C | 24.0 | 10 | 14 | 24 | 1.00 | |
| **15** | Marc McLaughlin | Boston College | C | 24.0 | 10 | 14 | 24 | 1.00 | |
| **16** | Zac Jones | UMass | D | 29.0 | 9 | 15 | 24 | 0.83 | |
| **17** | Ronnie Attard | Western Michigan Univ. | D | 25.0 | 8 | 14 | 22 | 0.88 | |
| **18** | Matt Kiersted | Univ. of North Dakota | D | 29.0 | 3 | 19 | 22 | 0.76 | 2 |
| **19** | Cam York | Univ. of Michigan | D | 24.0 | 4 | 16 | 20 | 0.83 | |
| **20** | Mike Hardman | Boston College | W/C | 24.0 | 10 | 9 | 19 | 0.79 | 1 |

| | player | ncaa_team | position | ncaa_gp | ncaa_g | ncaa_a | ncaa_p | ncaa_ppg | ncaa_pi |
|---|---|---|---|---|---|---|---|---|---|
| **21** | Jordan Harris | Northeastern Univ. | D | 19.0 | 6 | 13 | 19 | 1.00 | |
| **22** | Jack McBain | Boston College | C | 24.0 | 6 | 13 | 19 | 0.79 | 1 |
| **23** | Noah Cates | Univ. of Minnesota-Duluth | LW | 28.0 | 5 | 14 | 19 | 0.68 | 2 |
| **24** | Jacob Bernard-Docker | Univ. of North Dakota | D | 27.0 | 3 | 15 | 18 | 0.67 | 2 |
| **25** | Walker Duehr | Minnesota State Univ. (Mankato) | RW | 28.0 | 10 | 7 | 17 | 0.61 | 2 |
| **26** | Alex Newhook | Boston College | C/LW | 12.0 | 7 | 9 | 16 | 1.33 | |
| **27** | Owen Power | Univ. of Michigan | D | 26.0 | 3 | 13 | 16 | 0.62 | |
| **28** | Josiah Slavin | Colorado College | LW | 22.0 | 5 | 8 | 13 | 0.59 | |
| **29** | Nick Blankenburg | Univ. of Michigan | D | 26.0 | 5 | 8 | 13 | 0.50 | 1 |
| **30** | Bobby Brink | Univ. of Denver | RW | 15.0 | 2 | 9 | 11 | 0.73 | |
| **31** | Alex Vlasic | Boston Univ. | D | 16.0 | 3 | 5 | 8 | 0.50 | 1 |

---

## Finally, Some Analysis:

Let's see how many players from each 2020-21 NCAA team played in the NHL in the 2021-22 season.

In [ ]:
```
ncaa_counts = ncaa_2020_21_nhl_2021_22.groupby('ncaa_team').player.count()

ncaa_counts = pd.DataFrame(ncaa_counts)

ncaa_counts.sort_values(by = 'player', ascending = False)
```

Out[ ]:

| ncaa_team | player |
|---|---|
| **Univ. of Michigan** | 6 |
| **Boston College** | 5 |
| **Univ. of North Dakota** | 4 |
| **Univ. of Minnesota** | 3 |
| **Minnesota State Univ. (Mankato)** | 2 |
| **Univ. of Minnesota-Duluth** | 2 |
| **Boston Univ.** | 1 |
| **Colorado College** | 1 |
| **Northeastern Univ.** | 1 |
| **Providence College** | 1 |
| **UMass** | 1 |
| **Univ. of Denver** | 1 |
| **Univ. of Notre Dame** | 1 |
| **Univ. of Wisconsin** | 1 |
| **Western Michigan Univ.** | 1 |

Michigan had the most players on theior team that played at least 1 game in the NHL the next year. They did not, however, make it to the Frozen Four. This was likely due to their 6 players being called up to the NHL before the end of their NCAA seasons.

Lets see how many NHL teams had NCAA players playing for them their first year out of college.

In [ ]:
```
nhl_d1_counts = ncaa_2020_21_nhl_2021_22.groupby('nhl_team').player.count()

nhl_d1_counts = pd.DataFrame(nhl_d1_counts)

nhl_d1_counts.sort_values(by = 'player', ascending = False)
```

Out[ ]:

| nhl_team | player |
|---|---|
| Philadelphia Flyers | 5 |
| San Jose Sharks | 3 |
| Chicago Blackhawks | 3 |
| Colorado Avalanche | 3 |
| Arizona Coyotes | 2 |
| Columbus Blue Jackets | 2 |
| Ottawa Senators | 2 |
| Montréal Canadiens | 2 |
| New York Rangers | 1 |
| Seattle Kraken | 1 |
| Minnesota Wild | 1 |
| New Jersey Devils | 1 |
| Boston Bruins | 1 |
| Florida Panthers | 1 |
| Calgary Flames | 1 |
| Buffalo Sabres | 1 |
| Toronto Maple Leafs | 1 |

17 NHL teams had first year players from college play on their team at least once in the 2021-22 season! The NCAA certainly has a strong presence in the NHL.

This is all cool to know but lets go a little deeper:

### *How well does scoring in NCAA D1 predict scoring in the NHL?*

Lets find the correlation coefficient between NCAA points per game and NHL points per game.

In [ ]:
```
np.corrcoef(ncaa_2020_21_nhl_2021_22.ncaa_ppg, ncaa_2020_21_nhl_2021_22.nhl_ppg
```

Out[ ]:
```
0.4579755515984456
```

0.46 correlation coefficient. Lets find R-squared though. That tells us how much of the variance in one variable can be explained by the other.

In [ ]:
```
correlation_coefficient = (np.corrcoef(ncaa_2020_21_nhl_2021_22.ncaa_ppg, ncaa_

RSQ = correlation_coefficient**2

RSQ = round(RSQ, 2)
```

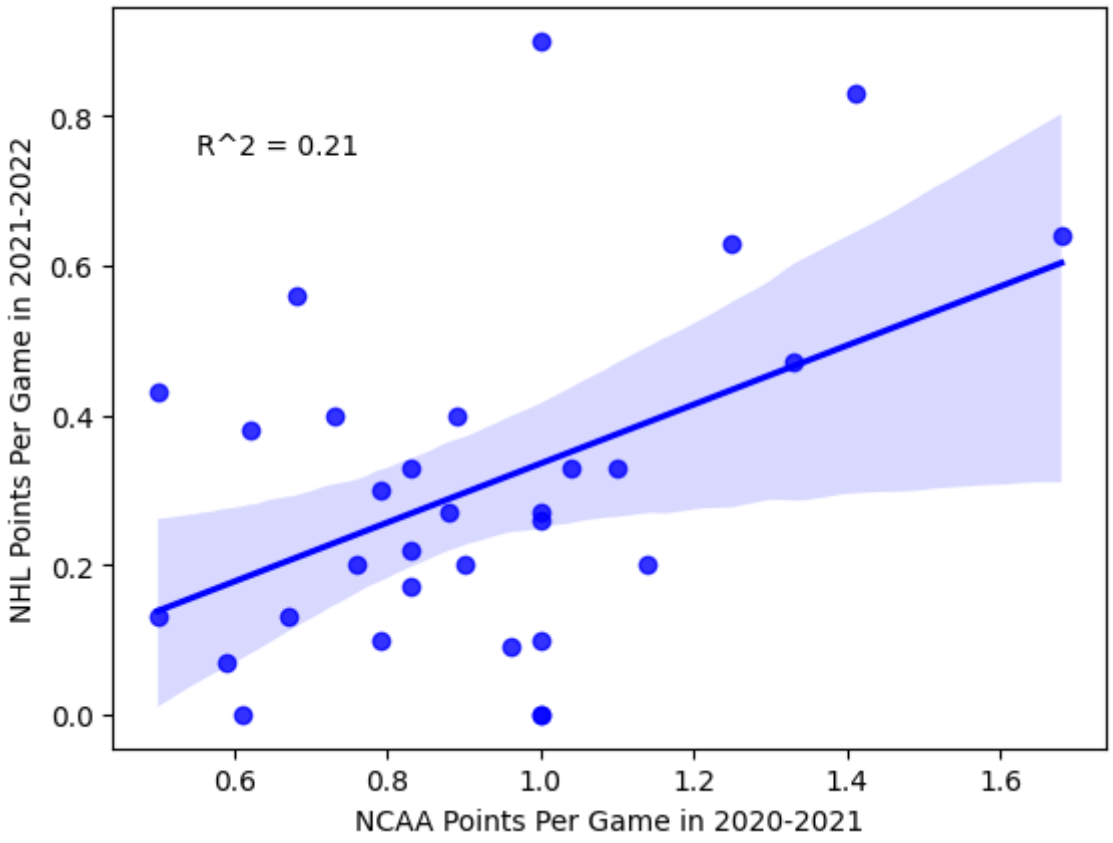
```
RSQ
```

Out[ ]: 0.21

An R-squared value of 0.21 means there is low correlation between scoring in the NCAA and the NHL. However, we can still visualize this with a seaborn plot and dig even deeper afterwards.

In [ ]:
```
sns.regplot(x = ncaa_2020_21_nhl_2021_22.ncaa_ppg, y = ncaa_2020_21_nhl_2021_22
plt.xlabel("NCAA Points Per Game in 2020-2021")
plt.ylabel("NHL Points Per Game in 2021-2022")

# Lets add R^2 to the plot
RSQString = "R^2 = " + str(RSQ)
plt.text(0.55, 0.75, RSQString)
plt.show()
```




It is important to note in the plot above that the axes are not the same, so it could be misconstrued that someone scored more in the NHL than in the NCAA, but that is likely not the case.

The values are not grouped closely to the line and there is some sort of loose correlation, but not enough to be significant. With a larger sample size, we could dial in the results.

---

Now we will create an NHL equivalency model using the methodology laid out by Gabriel Desjardins in League Equivalencies. This is his formula:

- Quality of League x = (Average PPG in NHL in Year 2) / (Average PPG in league x in Year 1)

In [ ]:
```
nhl_ppg_average = np.mean(ncaa_2020_21_nhl_2021_22.nhl_ppg)

ncaa_ppg_average = np.mean(ncaa_2020_21_nhl_2021_22.ncaa_ppg)

ncaa_nhl_equivalency = nhl_ppg_average/ncaa_ppg_average

ncaa_nhl_equivalency
```

Out[ ]:
```
0.32991875662310144
```

Our calculation of 0.329919 is exactly the same as Desjardins' calculation of 0.33 in League Equivalencies. This means that one point in NCAA D1 Men's Hockey is worth 0.33 points in the NHL, or one third of a point.

Now we will finish up by creating a function that calculates the predicted points per game for an NHL season given a points per game metric from the NCAA.

In [ ]:
```
def obtain_nhle_given_ncaa_ppg(ncaa_ppg):
    nhle = ncaa_ppg * 82 * ncaa_nhl_equivalency
    nhle = round(nhle, 2)
    print(nhle)
```

If you pass the NCAA points per game into the function, you receive the NHL equivalent points in an 82 game season.

In [ ]:
```
obtain_nhle_given_ncaa_ppg(52/30)
```

```
46.89
```

Result: A 1 point per game season in the NCAA translates to a 27 point season in the NHL. For perspective, an NCAA season is typically 34 games.

In 2020-21, Cole Caufield had 52 points in 30 games, so based off this model he should have had a 47 point season in the NHL the next year. He had a 43 point season in 67 games in 2021-22 in the NHL. I'd say this model is spot on.

---

This project was inspired by the TopDownHockey Tutorial by Patrick Bacon