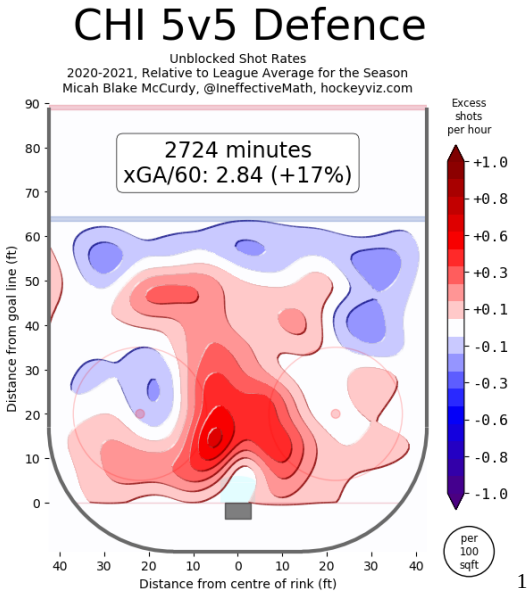
Final Project Proposal

I am writing to you today to propose my final project idea for your class. I plan on centering my final project around creating a model that will allow me to analyze historical NHL hockey data to be able to generate predictions about how players and prospects will perform in the future. One main reason for wanting to do this is that I would like to go into sports analytics as a career, specifically in hockey, and creating my own proprietary model that I could continue to refine over time would be a great steppingstone to hopefully one day achieving that goal. I plan on, throughout the course of the proposal, giving you the background, potential benefits, feasibility, what results might look like, my planned method, as well as a few concluding statements.

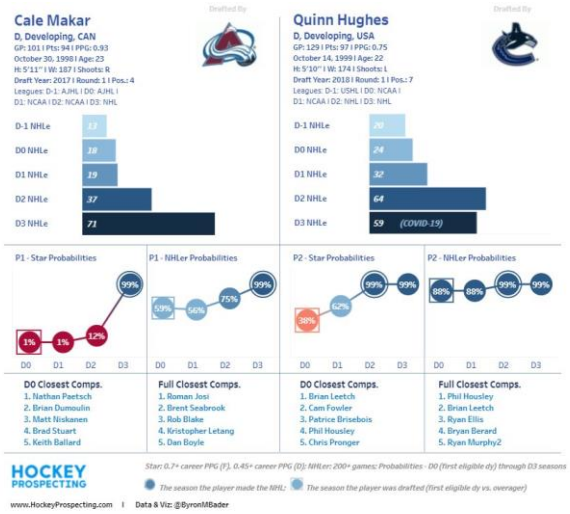
Hockey, being a sport that isn't nearly as easy to analyze numerically as say baseball, still has a somewhat relatively small analytics scene when compared to sports like football and basketball. Luckily, the NHL is also very forthcoming with their data, and it is easily accessed to amateurs such as me. What models are already out there online all have their strengths and weaknesses, and they all still have certain holes in them when it comes to assessing certain types of players. Ideally, I'd be able to find at least one new strength, if not filling several holes in one model. Of course, most teams have their own in-house proprietary models as well, but the contents of those may never be known so I only have what is publicly available to base this claim off.



[1]

[1] McCurdy, M. B. (2021, October 5). *Commentator on Sportsnet just described the Chicago system as "ultra-defensive" and I remember now why I mostly watch games on mute these days.* Twitter. Retrieved October 8, 2021, from https://twitter.com/IneffectiveMath/status/1445194091382378500?s=20.

Seeing as how I am nowhere near the first person to have this idea, this project likely wouldn't revolutionize the NHL analytics industry. Instead, this project would mostly have large effects for my personal life. It would give me a great piece to include on my resume as well as an excellent project to use to market myself in the social media space. On the off chance that I can stumble upon something that others haven't already found, it could, in some small way, impact the way my favorite sport is played and managed.



[2]

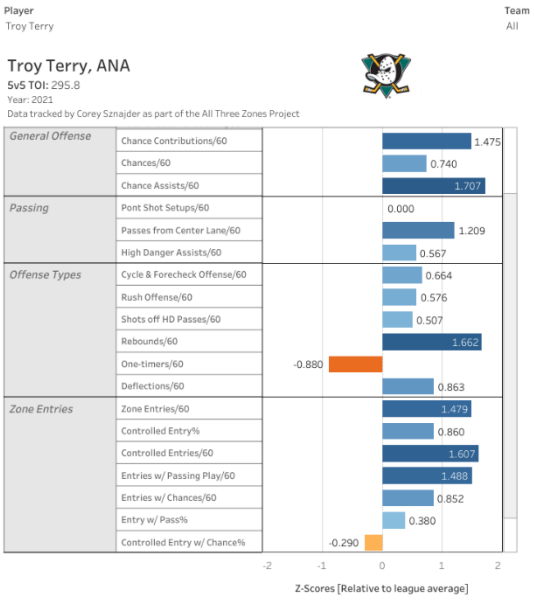
For this project, I would be using the NHL's official REST API.[3] I would theoretically, take in data from a wide variety of topics, including player zone time, player shooting times, player possession times, player linemates, and seasonal scoring environments to analyze the entirety of a player's game more totally. I would likely be able to produce a final write-up for this project. However, a more important end-goal would be a constantly flexible database which could at some point be connected to some sort of live visual presentation which could be accessed by fellow fans.

---

[2] Bader, B. M. (2021, October 7). *When Makar and Hughes cross over 200 games, they will be part of the very elite and very tiny D group that meet the star requirements for both a defenseman and a forward.* Twitter. Retrieved October 7, 2021, from https://twitter.com/ByronMBader/status/1446151379395899400?s=20.

[3] Pastor, K. (2020, April 24). *NHL Analytics with Python*. Medium. Retrieved October 7, 2021, from https://towardsdatascience.com/nhl-analytics-with-python-6390c5d3206d.

Theoretically, I would at first poke around with the dataset to see what all it includes, as it is quite expansive.[4] Once that is completed, I would identify which data points are the most important to what I am trying to accomplish and would begin attempting to train a model based on the data to produce a player score from 0-100 for simplicity. I am also the type of data scientist who likes to play around with a variety of different model-building methods to determine which ones best fits the data. This is because many times I've found that prior to working with data, I may understand it in one way, but instead come to find it works better in a different situation. For that reason, I do not wish to commit to one specific model at the current time. I imagine that such a model itself wouldn't be the most complicated thing out there. Instead, this project is a lot more likely to rely on innate knowledge of the game itself as well as being able to sift through a massive amount of data to best fine-tune the model.



[5]

To conclude, I imagine it is not difficult to understand what the project's end-goal will look like as well as what impact it will have on my career opportunities as well as the potential overall results. I plan on working on this project alone as I don't know anyone else, especially not in this class, with enough of an in-depth understanding of the game of hockey and the analytics involved.

---

[4] Hynes, D. (2021, August 31). *nhlapi*. GitLab. Retrieved October 7, 2021, from https://gitlab.com/dword4/nhlapi.

[5] Sznajder, C. (2021, October 5). *I am always super high on Troy Terry because he is one of the only ducks young forwards who isn't a black hole in most of my stats (him and Zegras).* Twitter. Retrieved October 8, 2021, from https://twitter.com/ShutdownLine/status/1445457698128994317?s=20.