Capstone - Sprint 0

Problem Area

- Area of Interest: National Hockey League (NHL)
- Research Question: When a player is traded midseason (prior to the trade deadline) does the significance of that player's prestige (role - player versus all- star- caliber) affect a team's playoff success? By analyzing 50+ years of NHL transaction data, can we predict whether a team will have a more positive impact from signing role players rather than all stars for their playoff run?

- The User: Every season NHL teams that are well positioned to make the playoffs 'buy' before the trade deadline. Many try to swing big for all - star - caliber players, while others add more modestly with role players to round out their current rosters. If this model can accurately show significance favoring one strategy or another, it could help optimize which trade assets to prioritize.
    - Additionally, while many factors will differ in other sports - this could possibly be a study altered for situations like Baseball, Football, or Basketball.

- The Big Idea - With supervised machine learning, this model will attempt a categorical prediction on which type of trades will lead to the best outcomes for a team's playoff success and attempt at the Stanley Cup.

- The Impact - Beyond the potential awards of playoff success, a franchise and it's players can see significant benefits from a deep playoff run. There is financial benefit from both ticket, merchandise, sponsorship, and TV revenues ; future marketing and notoriety benefits which help increase the overall value of the franchise ; player bonuses, which often come with increasing the desirability of top players to want to play for the team in the future. Additionally, playoff runs can create economic impact beyond just the team itself, as it can often energize and rally a community and its fans.

- The Data - From Kaggle 'NHL Transactions 1918 - 2020' - 12,500+ transactions. I plan to focus on trades after 1967 (~10,000 rows of data)
    - Additionally, I have found suggestions on reddit, and google for various APIs and scraping across other websites containing more recent NHL data. Including "capfriendly', 'spotrac', 'NHL.com' and 'ESPN'.

- The Alternative
    - NHL - if this data does not prove sufficient - I would likely transition to a project based on advanced analytics and on-ice player / puck tracking.
    - Alcohol (Wine) - Climate change effects on wine production.
    - Jeopardy - I found a cool data set on 'Jeopardy!'. It could be interesting to do something on game theory within jeopardy (i.e. betting on 'Daily Doubles' or 'Final Jeopardy'), but unsure if the data gets into player scores within an episode as much as questions / answers / values.