**Thesis:**

A player makes a better pass when they control the puck for a longer amount of time immediately before attempting the pass. A player should constantly be processing information as they control the puck, which allows the player to make a better determination on what to do next. When a player with possession is faced with pressure, it forces a decision as losing the puck would be a costly mistake that usually results in a breakdown in player positioning or a quality shot attempt for the other team. When a team has the puck, they have the option to execute a break-out, a practiced sequence of plays that is designed to achieve an outcome, which is typically a positive outcome. On the other hand, just holding on to the puck without skating or assessing the defense will probably decrease the quality of the play. This analysis aims to drill into individual player’s pass decision making compared to various benchmarks with an emphasis on how possession time impacts it.

**Methodology:**

In order to look at an individual player’s performance, we need to compare the stats with a baseline. There can be multiple baselines due to high variance between gender, leagues, and teams. Overall, this analysis will take a high level view, and systematically establish baselines and go into more detail to try to pick out significant factors that seem to affect passing quality. If there are factors that are consistently found across the various levels of detail, a logistic regression model will be made to try to predict “high quality passes”.

In the logistic regression model, we limit our analysis to passing under the following circumstances:

* Possession Time: We will consider if the likelihood of HQP is connected with how much time the passer possesses the puck before making the play.
* Even-strength: When on a power-play or short-handed, the motivations of each team are significantly different from even-strength play, therefore we will not combine passing in those game states due to the positive skew for power-play or negative skew from short-handed. We consider only even strength.
* Score States: We will look into whether or not the score impacts HQP. We expect that a team that is trailing will have lower HQP as they need to take risks to try to score.
* Period: We will consider if fatigue plays a role in HQP score.
* Pass Distance: We will consider if there is a linear relationship between how far a pass travels and the likelihood of the pass being of high quality, which can speak to stretch pass strategy.

Passing quality is defined as a pass that leads to one of the following conditions:

* Two additional possession plays by the passer’s team, or one of: a “Carried” zone entry, any shot, any Goal
* A possession of 3 or more seconds after entering the offensive zone
* A penalty taken by the opposing team

Any other sequence of play will not be considered a high quality pass, and there will be no distinction made between neutral or bad quality passes.

**Findings**

Based on our definition of a high quality pass (HQP), we looked at three leagues that were provided by Stathletes. The teams in the women’s dataset were a mix of national teams and local teams, so the results of this group is not indicative, and the limited number of games also point to skipping interpretation. However, we can see that the Canadian and United States national teams performed better than the average. Surprisingly, Clarkson Golden Knights had the highest HQP score relative to the group baseline, which suggests that quality of opponent is a significant factor for HQP.

The NWHL dataset is a step up in quality as these teams were all members of the same league and played against each other exclusively. However, due to COVID-19 pandemic circumstances, the season was suspended. Some teams in the league had games postponed, one team did not play any games, and others played all their scheduled games. Overall, the HQP scores were lower than in other leagues. Again, due to the lack of games played, there isn’t much point to look at the scores here. However, we do see that teams that won typically make average to above average HQP and vice versa, with the exception of the Boston Pride.

The third group is found in the scouting dataset. The league here is the Ontario Hockey League, which consists of two conferences. In the eastern conference, we see that the standings for the teams closely follow with HQP from top to bottom. However, there is a stark contrast in the western conference, as HQP is not correlated with the season standings.

These observations suggest that passing is a part of the equation, but not able to explain results of games itself. This makes sense as passing only offers the potential to do good things, and doing good things involve concepts such as finishing ability and goaltending, each of which are out of the scope of this analysis. However, we consistently see that the worst performing teams consistently have poor HQP scores. This conclusion encourages the development of a logistic regression model to predict HQP based on all the available information in the dataset.

**Logistic Regression:**

Overall, the model is not robust in determining whether a pass would be a HQP given the available data, referring to adjusted R-squared. Without considering the team that the passer is on, we use an overall logistic regression model to predict if a pass is considered HQP. Using a standard randomized test and train set procedure, the algorithm returns a prediction score of 56.82%. If we include information on which team performed the pass, the model prediction score decreases to 56.12%. What this suggests is that the correlation we saw in the eastern conference HQP scores matching with the standings was merely coincidental.

**Key Points:**

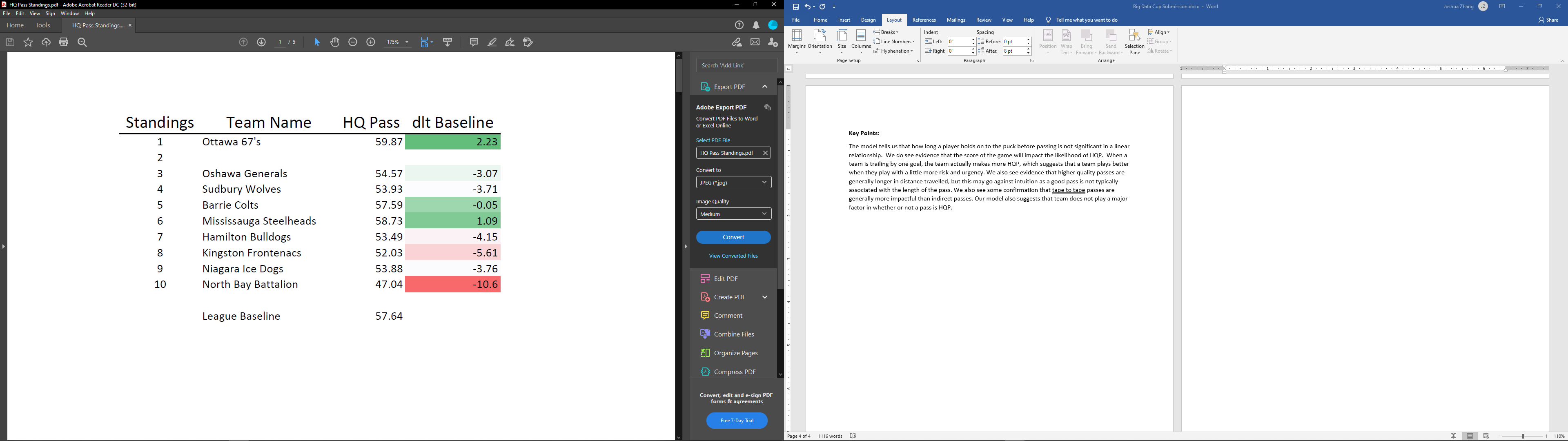
The model tells us that how long a player holds on to the puck before passing is not significant in a linear relationship. We do see evidence that the score of the game will impact the likelihood of HQP. When a team is trailing by one goal, the team actually makes more HQP, which suggests that a team plays better when they play with a little more risk and urgency. We also see evidence that higher quality passes are generally longer in distance travelled, but this may go against intuition as a good pass is not typically associated with the length of the pass. We also see some confirmation that tape to tape passes are generally more impactful than indirect passes. Our model also suggests that team does not play a major factor in whether or not a pass is HQP.

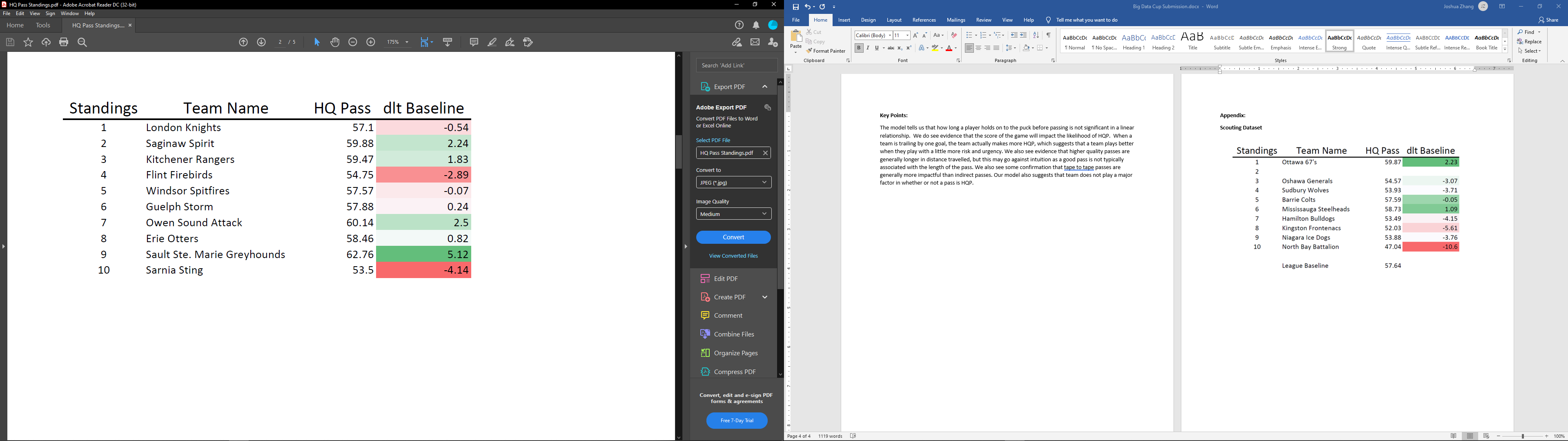
This analysis also has a Tableau public interactive dashboard that allows you to visualize the passing using a variety of filters, which will continue to be developed moving forward.

[https://public.tableau.com/profile/joshua.zhang#!/vizhome/BigDataCup2021/ErieOtters-PasseswithLongPossessions](https://public.tableau.com/profile/joshua.zhang%23!/vizhome/BigDataCup2021/ErieOtters-PasseswithLongPossessions%0c)

**Appendix:**

**Scouting Dataset**





**Logistic Regression Summary**

