For my project, I analyzed the RIT men's hockey team statistics, both individually and on the team as a whole. I wanted to look into the statistics of the team roughly half way into the season (specifically during the semester/winter break) and see where the statistics were at that point. I chose to do an analysis of them because I am big into college hockey and they are my favorite team along with their sister team, the RIT women's hockey team, and have been following them since the 2021-2022 season.

The way I took in the statistics was by web scraping the team's statistics page online and saving them as a data frame. The tools I used to do this was Jupyter notebook and python to get the statistics and I used Gemini specifically for the line of code to make it easier to get since there were multiple tables on the page. Other than those two lines of code, I did not use any AI. Instead of using BeautifulSoup, I just simply used "tables = pd.read_html(url)" and "stats = tables[1]" to get the table I needed for this analysis. On top of that, I used pandas to help me handle the data as a data frame and used the python packages "seaborn" and "matplotlib.pyplot" to help me make plots of the data.

I initially had five questions that I asked about the data but answered 4 of them. The 4 questions were "How many and what percentage of the goals scored were from special teams? (from short handed and power play goals)", "Based on the shooting/scoring percentage, who needs the most help to improve on scoring/getting the puck in the back of the net?", "Who are the strongest players defensively, in terms of blocks?", and "Based on the points per game average, who is projected to be the top scorer at the end of the season?".

Some of the insights I got were when I looked at the goals scored, 46 total goals were scored and of the 46, 6 were power play goals (making up 13.04% of the goals) and 1 short handed goal (making up 2.17% of goals scored). Additionally, I have learned that Tristan Allen,

Gustav Blom, Brock Reinhart, Adam Jeffery, and Kevin Scott have the lowest shooting and scoring percentage (which means of all the shots they have taken, they have the fewest of their shots going into the net). So those players need some help and improvement of shooting placement and quality shots to help them score goals. Some other insights I got were that Kevin Scott leads the team in blocks with 24 with Tristan Allen and Mason Croucher following with 17 and 15 respectively. This is a good measure to see because it can look in to see if players are making blocks of course but also if there are a lot of blocks then it might be a problem since the team is spending too much time in the defensive zone. I also looked into the points per game of each player to see who is leading and is projected to potentially be the point leader at the end of the season.

Some recommendations I have are to work on the power play since there are very few power play goals scored. Although there is only one short handed goal scored, not scoring a lot of them is a big deal. Another recommendation I have is to help out the few players with a 0% scoring percentage, specifically with placement on net and getting the right shots.

Some more questions I would like to ask and look into are “What is the overall save percentage and goals against average for the goalies on the team?” and “What is the attendance per game average? And how much does it fluctuate?”. I want to look into these since I focus a lot on the players/skaters more than the goalies and don’t look into their statistics too much. I also want to look into the attendance to see how much it fluctuates during the season and see if there’s a correlation between that number and anything else in connection. I would also like to look into the women’s hockey team as well to see some of the similar statistics too and see any similarities.