College Hockey Attendance and Rankings Correlation

Observing the Correlation Between Men's Collegiate Ice Hockey Programs Attendance to Overall Season Rankings

This Capstone Project is submitted in partial fulfillment of the requirements for the course Data-Driven Decision-Making (MDA 620) during the Fall Semester of 2022.

While writing this Capstone Project, we have not witnessed any wrongdoing, nor have we violated any conditions of the LIU Honor Code.

Anthony Vincent

December 20, 2022

Table of Contents

Figures

Tables

Background

College hockey is an NCAA collegiate sport that began in the late 1890s and has been around for decades. However, not until more recently did college hockey become more competitive and recognized in our society. Every team, on average, is given a 28-35 game schedule where they will play many different opponents. Yet, the ultimate goal is to make the playoffs and ultimately win a championship. So, the NCAA has established a ranking system to keep track of who makes the playoffs and keep track of the highest-ranked teams weekly. This is known as the Pairwise Rankings or PWR. The PWR method compares each team with every other such team, with the winner of each "comparison" earning one PWR point. After all, comparisons are made, the points are totaled up and rankings are listed accordingly. When it comes to the PWR, 3 criteria are combined and form this PWR which include: RPI(Ratings Percentage Index), record against common opponents, and head-to-head competition.[1]

RPI would be the most complex criteria and 3 things tie into how it is calculated. First, the winning percentage is weighted by 25% to the teams' overall record, 21% to the record of a team's opponent, and 54% to the record of the opponent. The second component is where a game is won. Winning on the road counts as 1.2 wins while winning at home only counts as .8 and vice versa for losses. Third, a quality win bonus for beating a top 20 ranked team earns a .050 RPI for beating the #1 ranked team to .0025 which would be if a team beats the 20th ranked team.[2] This math seems complex, but it is computerized and therefore not as complicated. With my data, this RPI information is already calculated and the rankings are more precise.

---

[1] "Pairwise Rankings Explanation - USCHO.com: FAQ: Frequently Asked Questions." *USCHO.com | FAQ | Frequently Asked Questions*, 21 Feb. 2016.

[2] Dilks, Chris. "Explaining the Pairwise Rankings." *SB Nation College Hockey*, SB Nation College Hockey, 25 Feb. 2015.

Problem Scenario/Business Issue

Throughout NCAA Division 1 College Hockey, teams strive to be the best they can be to earn a spot in the playoffs in hopes of winning a championship. However, only a select 16 teams can make the NCAA tournament which ultimately results in one winner. With so many people coming to games and cheering on teams, a question came to my attention. As I looked at the overall data from 2015-2020, I wanted to find out the relationship between Pairwise rankings and Win pct of a NCAA Division 1 College Hockey team on fan attendance.

Figure 1. Denver Rink, average ranking 5th - 1st overall

Figure 2. Alaska Anchorage, average ranking 57th - Last Overall

Objective/Goals of the Project

The purpose of this project is to investigate and understand how Parwise Rankings and WinPct is impacted by fan support and the impacts it has on a team's overall success in NCAA Division 1 Men's Ice Hockey. For example, Figure 1 is a picture of Denver's Rink, and Denver was ranked the highest overall from 2015-2020 cumulatively. Figure 2 is a picture of Alaska Anchorage's Rink, and Alaska Anchorage was ranked last cumulatively in all college hockey from 2015-2020 based on pairwise. Through my analysis, I want to discover if fan support actually impacts a team's overall ranking. Looking at all the NCAA Division 1 Ice Hockey programs from 2015-2020, I will explore whether this is the case or not.

Data Exploration

The NCAA Division 1 Men's Ice Hockey during my data analysis from 2015-2020 consists of 60 NCAA Division 1 hockey teams scattered about the Northeast and Upper Midwest regions of the United States. Each team competes in one of six conferences; Atlantic Hockey, Big 10, ECAC, Hockey East, NCHC, and WCHA.[3] There is one exception in my research for a team that lies outside of these 6 conferences. This is Arizona State, which is considered an independent team, as they are not in any conference. As I mentioned earlier, all 60 of these teams compete in a rigorous schedule of games which can result in a win, loss, tie, overtime loss, overtime win, and sometimes a shutout loss or win depending on the conference.

Through the United States College Hockey Online Database and College Hockey News Website, I was able to collect yearly data on both fan attendance and pairwise rankings. While data exists from 2020-2022, I wanted to remove COVID-19 years which significantly impacted attendance, and the rankings were out of sorts as some teams did not even end up playing during

[3] "Complete List of Hockey Teams: Men's Ice Hockey Colleges." *NCSA College Recruiting*, 22 Nov. 2022.

these years due to shutdowns. Also, it was the most consistent timeline of 60 college hockey teams, as more teams have been added since 2020.

Data Visualization

Table 1 below has the summary statistics (observations, mean, std. Dev., min, and max) for all of the variables I am testing in my project.

Table 1. Summary Statistics

Variables	N	Mean	Std. Dev.	Min	Max
Total Attendance Per Year	60	58,942	42,722	7,419	226,853
Total Attendance Per Game Per Year	60	3,274	2,222	443	11457
Total Capacity % Per Year	60	71.31	31	5.26	211
Total Ranking	60	30	14	5	57
Total Average Win %	60	50	11	31	71

Figure 3 below shows the relationship between “Average Rankings of each team per year” and “Average Attendance Ranking Per year” from 2015-2020 of all 60 NCAA Division 1 Men’s Ice Hockey Teams. From the graph, we can see that there is a positive correlation between the two variables. The lower a team is ranked (closer to #1), the more fan support the team has, and vice versa. Figure 4 shows the relationship between “Average Total Yearly Attendance” and “Average Ranking” from 2015-2020 of all 60 NCAA Division 1 Men’s Ice Hockey Teams. Lastly, Figure 5 shows the relationship between “Average Total Game Attendance Per Year” and “Average Team Ranking” from 2015-2020 of all 60 NCAA Division 1 Men’s Ice Hockey Teams.

Figure 3

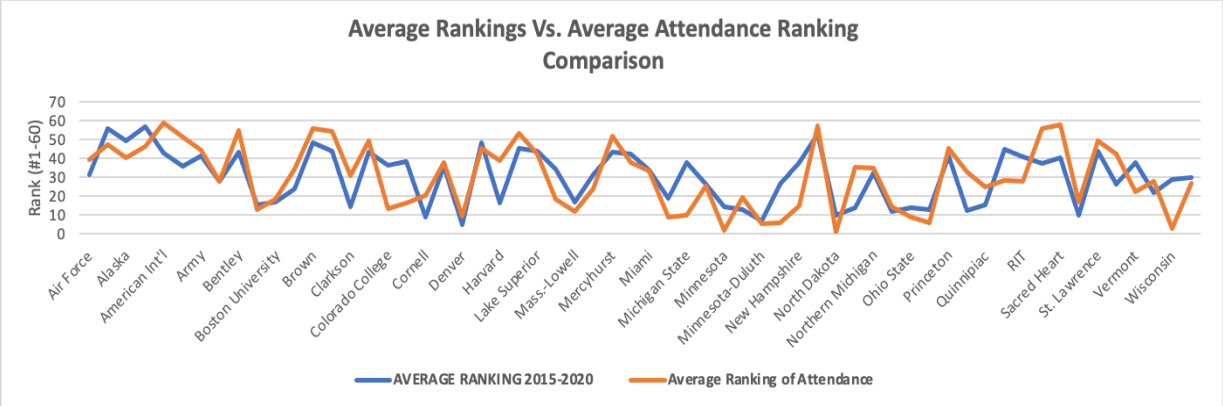


Figure 4

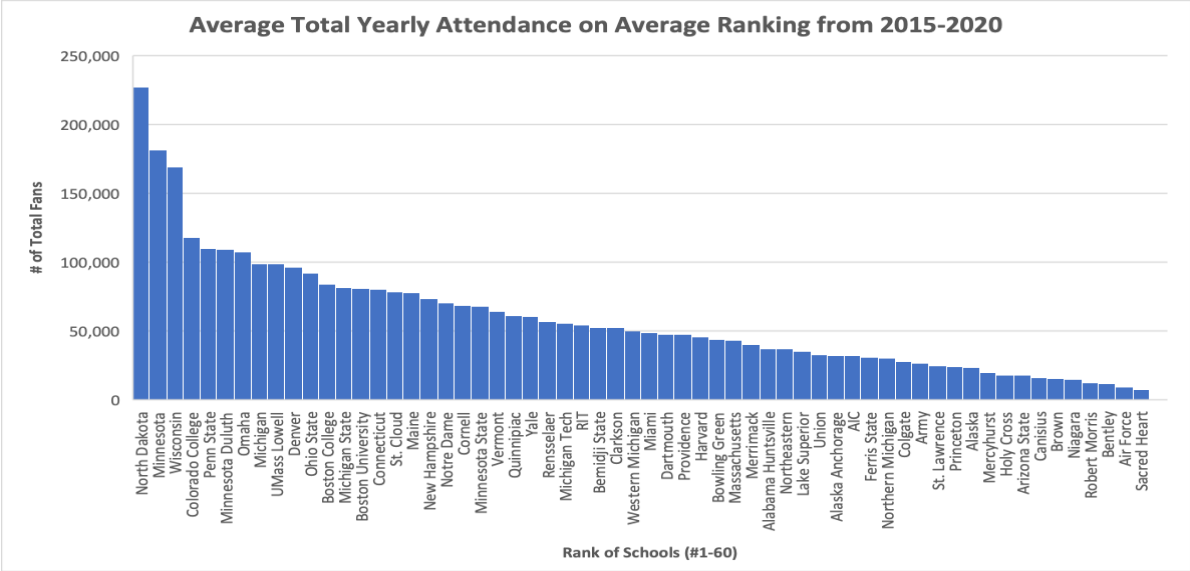
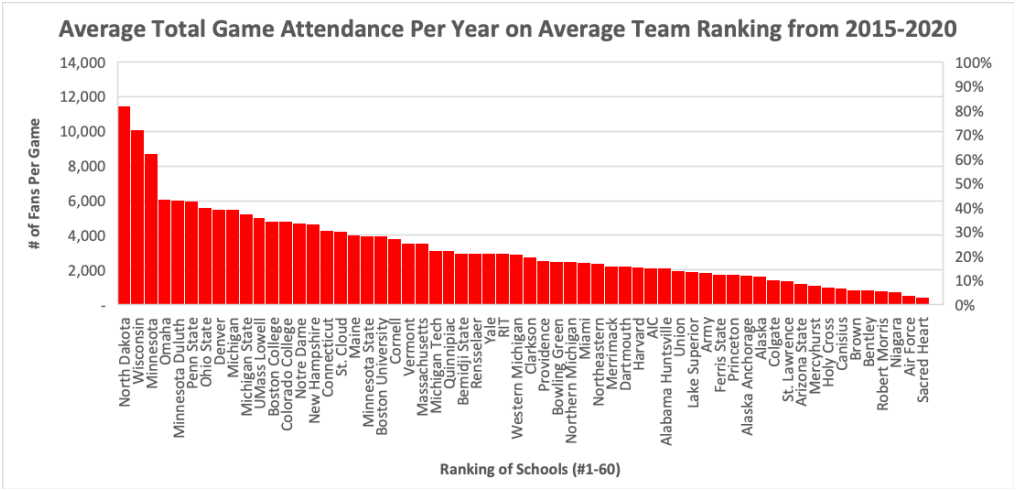


Figure 5



Data Manipulation

My first task in handling the data was to sort through and combine years of 2015-2020 attendance for NCAA Division 1 Men's Ice Hockey Teams from the United States College Hockey Online website.[4] Then my next task was to find the rankings of all 60 NCAA Division 1 Men's Ice hockey teams from 2015-2020 which I was able to find on the College Hockey News Online Website. From there I was able to transfer the data into excel where I could start my data manipulation. Reference Table 2 for the variables I created with my formulas during the process to get a clearer picture of my goal. Through these variables and imported data, I was able to calculate the mean, standard deviation, minimum, maximum, and number of observations for each variable present in Table 1.

Table 2. Created Variables with Formulas

Created Variables	Formula
Average Ranking of Attendance	Mean of "Attendance Ranking"
Average Total Attendance	Sum of Total Attendance (2015-2020) / Total # of Years (5)
Average Attendance Per Game Per Year	Sum of Total Attendance Per Game (2015-2020) / Total # of Years (5)
Avg Capacity %	Sum of Total Capacity % (2015-2020) / Total # of Years (5)

Methodology/Model Building

Model 1

For Model 1, I chose to use a Linear Regression Model. The two independent variables Rank and Win Percentage will be regressed on my dependent variable of Average Attendance.

[4] "Statistics." *College Hockey | USCHO.com*, 4 Mar. 2020.

Below are the equations that I used to generate my Linear Regression Model, and they include Beta(β), which is used to describe the coefficient for Rank and Win Percentage (both independent variables). I also have included Tables 3 and 4 to show my Results of Model 1.

$$\text{Average Attendance} = \beta 0 + \beta 1(\text{Rank})$$
$$\text{Average Attendance} = \beta 0 + \beta 1(\text{WinPct})$$

Table 3. Rank Linear Regression Results	
	Average Attendance
Rank	-116.432*** (3.205)
Cons	6824.641
Observations	300
R-Squared	0.8158

Table 4. WinPct Linear Regression Results	
	Average Attendance
WinPct	2.122*** (8.384)
Cons	3167.425
Observations	300
R-Squared	0.000215

Based on Table 3, the R-squared value was 0.8158. Compared to table 4 of the linear regression models this value was significantly higher. Also for Table 3, Rank has a Beta Coefficient of -116.432 which means that for every increase in 1 in the standings from say #20-#21 in the Pairwise, this would result in a loss of 116 fans attending the game and vice versa. In table 4, the R square value is significantly lower at 0.000215 which indicates that there is no correlation with the dependent variable and WinPct is not significant. However, based on the table and information I concluded that if I were to use this model, an increase in a win percentage of 1 percent would result in an increase in average attendance by 2 (2.122) people.

Model 2

Average Attendance = 7225.691 - 116.943(Rank) - 7.686(WinPct)

Table 5. Multiple Regression Results

	Average Attendance
Rank	-116.943*** (3.195)
WinPct	-7.686* (3.587)
Cons	7225.691
Observations	300
R-Squared	0.8186

After looking at table 5 I observed that the Rank variable barely changed but WinPct coefficient is now negative which was very interesting to me. I cannot pinpoint a reason for this outcome, but after running a multiple regression model, I felt this was important to observe. Also

through table 5, the R-squared value of 0.8186 indicates that this model has a relative goodness of fit with significant coefficients.

Model Selection

After comparing Models 1 and 2, as well as looking over the results and information, I have decided to select Model 2. The fact that this model encompasses all of the independent variables, allows me to interpret how each one impacts the dependent variable rather than assessing them individually. This would be the case for Model 1, a simple linear regression model which only accounts for one independent variable at a time and its relationship to the dependent variable. Lastly, the R-squared value is the highest in Model 2 and is 0.8186. Therefore I can conclude that the relationship between the independent variables and dependent variable in Model 2 has a higher correlation than in Model 1 (especially when looking at the independent variable WinPct).

Analysis of Excel Graphs/Charts/Moving Avg

After analyzing and understanding my initial data, it is clear that a few trends exist. First, using a simple command in excel, I was able to clearly display a line graph that shows the relationship between our manipulated variables. By calculating the moving averages in excel, it is evident that there is a positive relationship between a team's Pairwise Rankings and Total Average attendance. As I mentioned before, Alaska Anchorage was the last ranked team, on average, throughout the 2015-2020 5-year stretch, and Denver was the first ranked team, on average, throughout the 2015-2020 5-year stretch. In order to clearly see the relationships, I felt it was best to pick these two teams and build a model for Future Fan attendance and Future Pairwise Ranking. By creating a moving average for the entire 60 teams, then hand-selecting these two teams, I created two separate projection models. As you can see in Figure 6 and Figure

7, Denver has a significantly greater amount of fans, therefore always finishing way higher in the Pairwise rankings. Conversely, Alaska-Anchorage continues to be projected as the worst-ranked team in NCAA Divison 1 College Hockey, based on my moving average forecasting model. This is positively connected to the fact that they have one of the lowest total average fan supports.

Figure 6. Future Forecast of Fan Attendance Using Moving Average

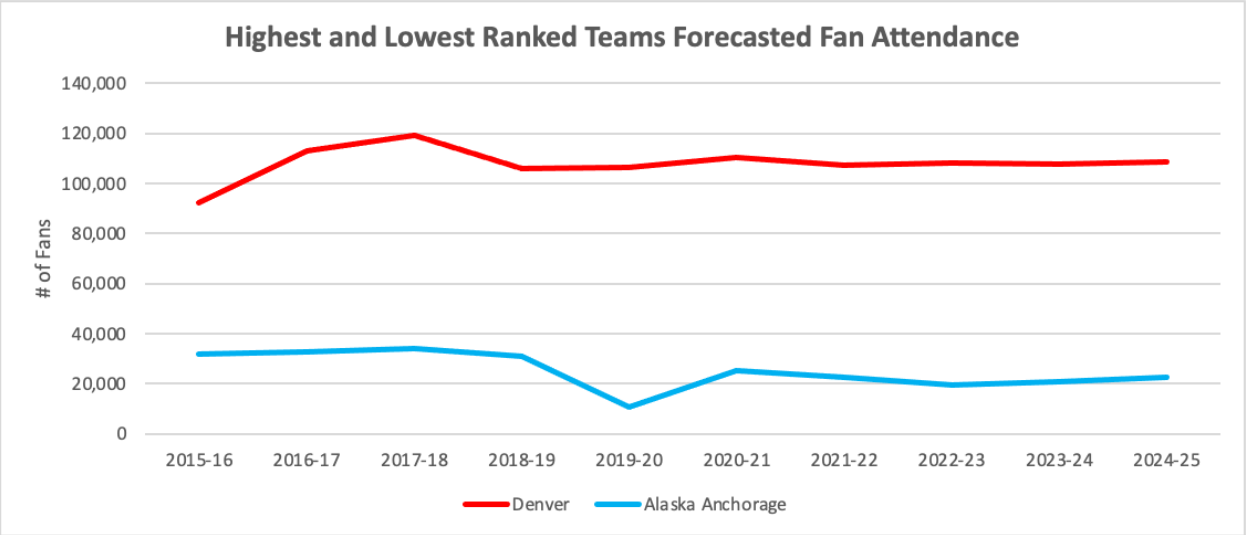
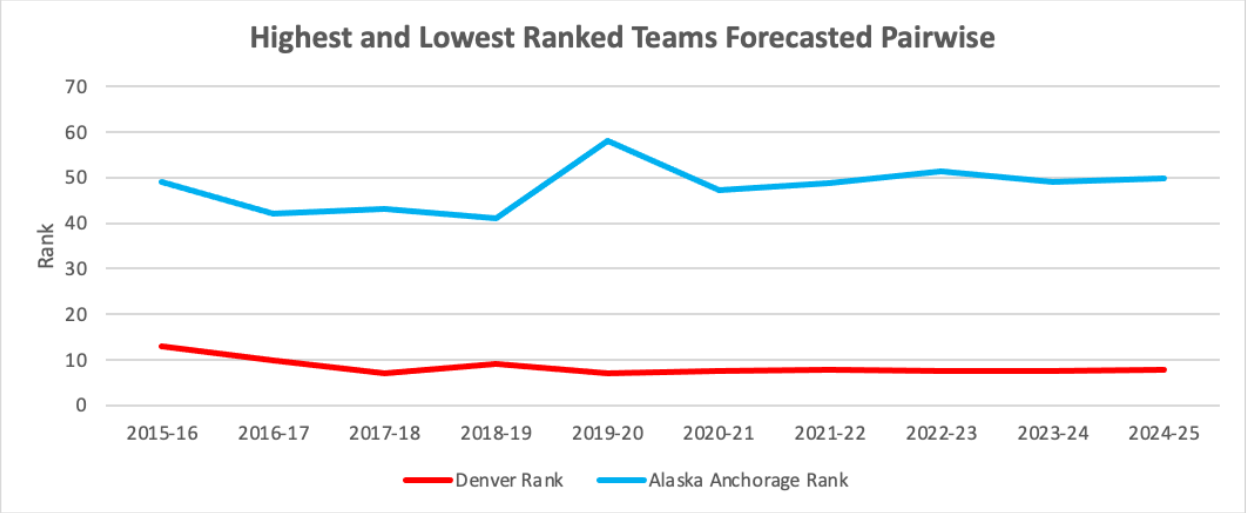


Figure 7. Future Forecast of Pairwise Rankings



Conclusions/Recommendations

Overall, I can conclude that Rank has a positive relationship with attendance giving out an R-square value of 0.8158. The schools that I examined are all NCAA Division 1 Men's Ice hockey programs during the 5 years 2015-2020. Also, as shown in my figures, fan attendance seems to remain about constant throughout the 2025 season, with minimal spikes in an increase. Rink capacity sizes play a factor, yet the overall Pairwise rankings do not seem to fluctuate much at all as well. The top teams continue to remain at the top, while the lower-end teams continue to remain at the bottom. When looking at the larger picture, most of the top-tier programs tend to have an exponentially greater amount of resources at their disposal. This includes better facilities that hold more fans, many more financial resources to generate top-tier programs, and generations of success that bring in top-tier recruits that enhance their programs. This allows for the top teams to stay at the top while the bottom teams seem to continue to struggle for success in the Pairwise rankings.

From my analysis of my datasets, I am confident that these forecasts will not waiver. Success is hard to come by in college hockey, and as the top teams continue to recruit top-tier athletes, it is evident that they will continue to rank higher in the Pairwise year in and year out.

Limitations

Several limitations were present in my study. First, my study only focused on the 60 NCAA Division 1 Men's Ice hockey teams from 2015-2020, which does not include Division 2 or Division 3 NCAA schools, as well as not including any NCAA Division 1, 2, or 3 Women's Ice hockey teams. Therefore, conclusions can only be drawn from these 60 NCAA Division 1 Men's Ice Hockey schools during these 5 years with future forecasts. In addition, I decided to omit COVID-19 data from my data set, as I believe this is not a reliable representation of attendance rates and rankings. Many teams did not even participate in games during this year,

including all of the Ivy League schools. Thus, including this information could have an impact on my analysis. One other notable limitation is rink sizes. When you look at Ohio State University, their rink capacity is 17,500 compared to Alaska Anchorage which has a capacity of 800. Therefore overall attendance may look way less for certain schools when in reality the school is selling out every single game. Nevertheless, this had a minimal impact on overall Pairwise rankings. Overall Pairwise rankings were more affected by the total average attendance.

Works Cited

"2015-16." *College Hockey News*, https://www.collegehockeynews.com/reports/standings.php?s=20152016&conf=0.

"2016-17." *College Hockey News*, https://www.collegehockeynews.com/reports/standings.php?s=20162017&conf=0.

"2017-18." *College Hockey News*, https://www.collegehockeynews.com/reports/standings.php?s=20172018&conf=0.

"2018-19." *College Hockey News*, https://www.collegehockeynews.com/reports/standings.php?s=20182019&conf=0.

"2019-20." *College Hockey News*, https://www.collegehockeynews.com/reports/standings.php?s=20192020&conf=0.

"Complete List of Hockey Teams: Men's Ice Hockey Colleges." *NCSA College Recruiting*, 22 Nov. 2022, https://www.ncsasports.org/mens-hockey/colleges.

Dilks, Chris. "Explaining the Pairwise Rankings." *SB Nation College Hockey*, SB Nation College Hockey, 25 Feb. 2015, https://www.sbncollegehockey.com/2015/2/25/8085357/pairwise-rankings-explained.

"Pairwise Rankings Explanation - USCHO.com: FAQ: Frequently Asked Questions." *USCHO.com | FAQ | Frequently Asked Questions*, 21 Feb. 2016, https://www.uscho.com/faq/pairwise-rankings-explanation/.

"Statistics." *College Hockey | USCHO.com*, 4 Mar. 2020, https://www.uscho.com/stats/attendance/division-i-men/2015-2016/.

"Statistics." *College Hockey | USCHO.com*, 4 Mar. 2020, https://www.uscho.com/stats/attendance/division-i-men/2016-2017/.

"Statistics." *College Hockey | USCHO.com*, 4 Mar. 2020, https://www.uscho.com/stats/attendance/division-i-men/2017-2018/.

"Statistics." *College Hockey | USCHO.com*, 4 Mar. 2020, https://www.uscho.com/stats/attendance/division-i-men/2018-2019/.

"Statistics." *College Hockey | USCHO.com*, 4 Mar. 2020, https://www.uscho.com/stats/attendance/division-i-men/2019-2020/.