Aubreigh DeFriece

Hockey Regression

DS 325

The regression shows the relationship between time on ice per game and points per game for the NHL over 10 seasons (2016-17 to 2024-25). I first gathered the data from Hockey-Reference.com for the 10 seasons. It showed statistics such as games played, time on ice, goals, assists, etc. To clean the dataset, I removed any missing data and converted them to N/A to make the data easier to handle the data. The time on ice column showed the data in minutes and seconds, so I converted it to only minutes to make it easier to compare. Then, found an average for each player by dividing the points column by the games played column. I also filtered out players who played less than 20 games during the season to prevent skewing the results. I used a linear regression to show relationships. The validation was done through R^2 and P-values. The regression showed a positive correlation. However, as time increased at about 22 minutes, the number of goals scored started to decrease.