Abby Hustler

Completed Module

May 6th, 2023

|  |  |
| --- | --- |
| Powerplay Data | NHL powerplay stats during 2022 season |

### Description

The data can be found from the [MoneyPuck](https://moneypuck.com/moneypuck/playerData/seasonSummary/2022/regular/teams.csv). This data shows the statistics for all powerplays by each team.

### 

### Format

A data frame has observations from all 32 teams for the following variables.

|  |  |
| --- | --- |
| team | Name of the team |
| goals\_per\_2\_min | how many goals scored in 2 minute increments. |
| SOG\_per\_2\_min | how many shots on goal were taken on a 2 minute average. |
| hits\_per\_2\_min | average hits recorded for 2 minutes |
| League | what league the team plays in (east vs west) |
| goal\_differential | combinations of goals for and against. |
| Icetime | how much time spent on the powerplay |

### 

### Source

[Money Puck](https://moneypuck.com/moneypuck/playerData/seasonSummary/2022/regular/teams.csv) (moneypuck.com/moneypuck/playerData/seasonSummary/2022/regular/teams)

1. Create a histogram that will show us the distribution of goals per 2 minutes. Is the histogram normally distributed?

2. Look over the data set. Give one example each of a categorical and quantitative variable.

3. Find the mean for the goals\_per\_2\_min variable.

4. Create a scatterplot that shows us the shots on goal compared to the goals per 2 minutes and shots on goal per 2 minutes. Does the graph you created appear to have a strong correlation? First make an initial guess of the correlation's value. Now compute it using the cor function and compare it to your estimate?

5. Create a hypothesis test to see if there is a correlation between goals per 2 minutes and shots on goal per 2 minutes. Find the T value which will help you make your conclusion.

6. Create a boxplot that shows us goals\_per\_2min vs hits\_per\_2min for each conference (hint: add facet\_wrap(~League) to your ggplot to separate the east vs west). Once you have created the plot tell us about the differences between leagues. Are there any apparent outliers?

7. Create a scatterplot that will compare goals\_per\_2min with goal differential. Do you see any outliers? If so, do it follow the linear pattern of the other data points?

8. Look at the output below.

What is the best individual predictor for goals\_per2min? Is this predictor significant?

mod1= lm(goals\_per\_2min ~ SOG\_per\_2min + faceoffs\_won\_per2min + hits\_per\_2min, data = powerplay\_data)

Table

Description automatically generated

9. Based on the same output, does the entire model appear to be significant? If so, explain how you know this.