OWEN HEASLIP

2842 Coach House Way, Naples, FL

(239) 216 - 2832

oheaslip@friars.providence.edu

www.linkedin.com/in/owenheaslip

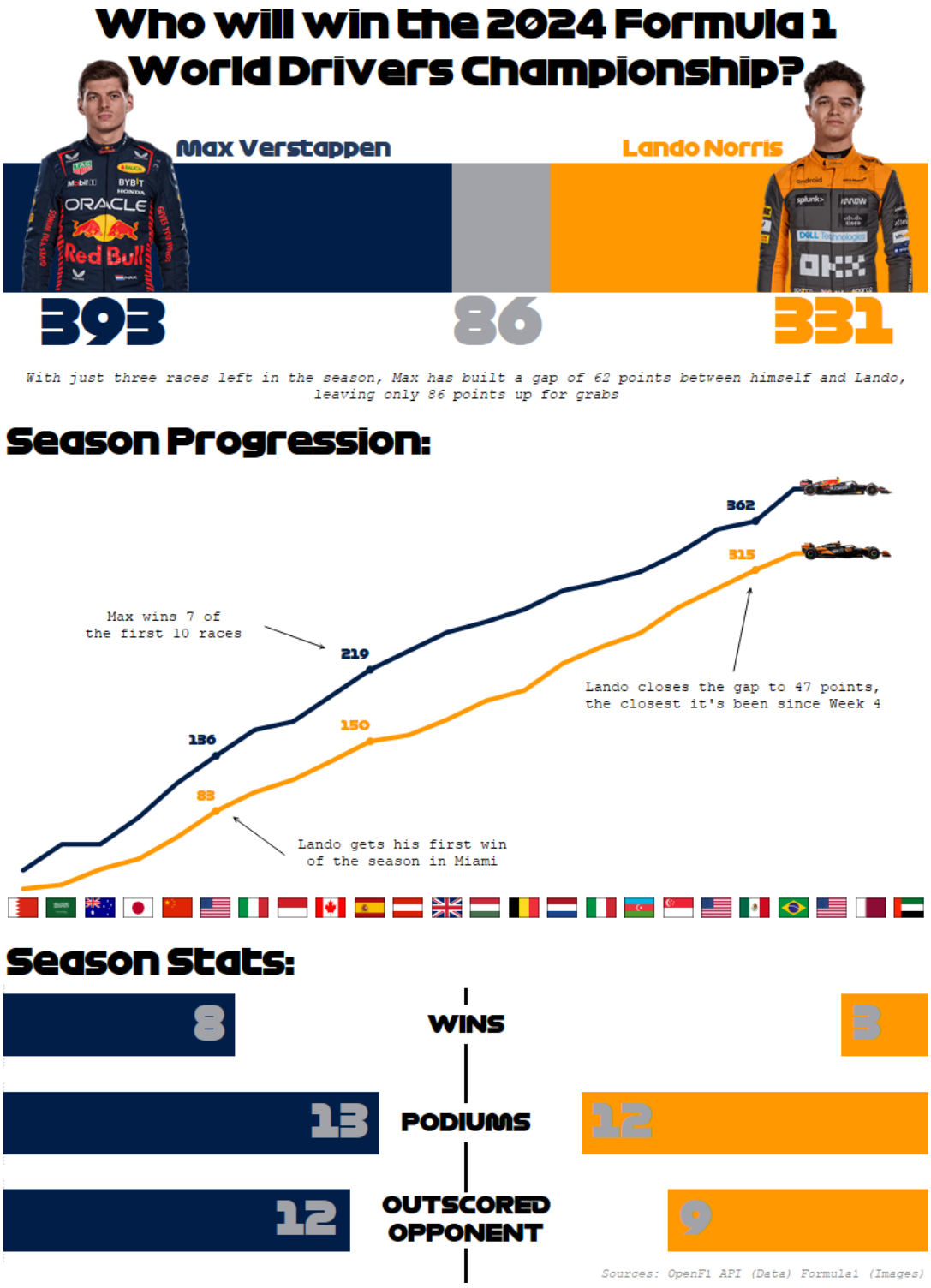
FOR A DETAILED VIEW OF ANY PROJECT, VISIT MY GIT HUB REPOSITORY HERE:

https://github.com/OwenHeaslip/Portfolio

TABLE OF CONTENTS

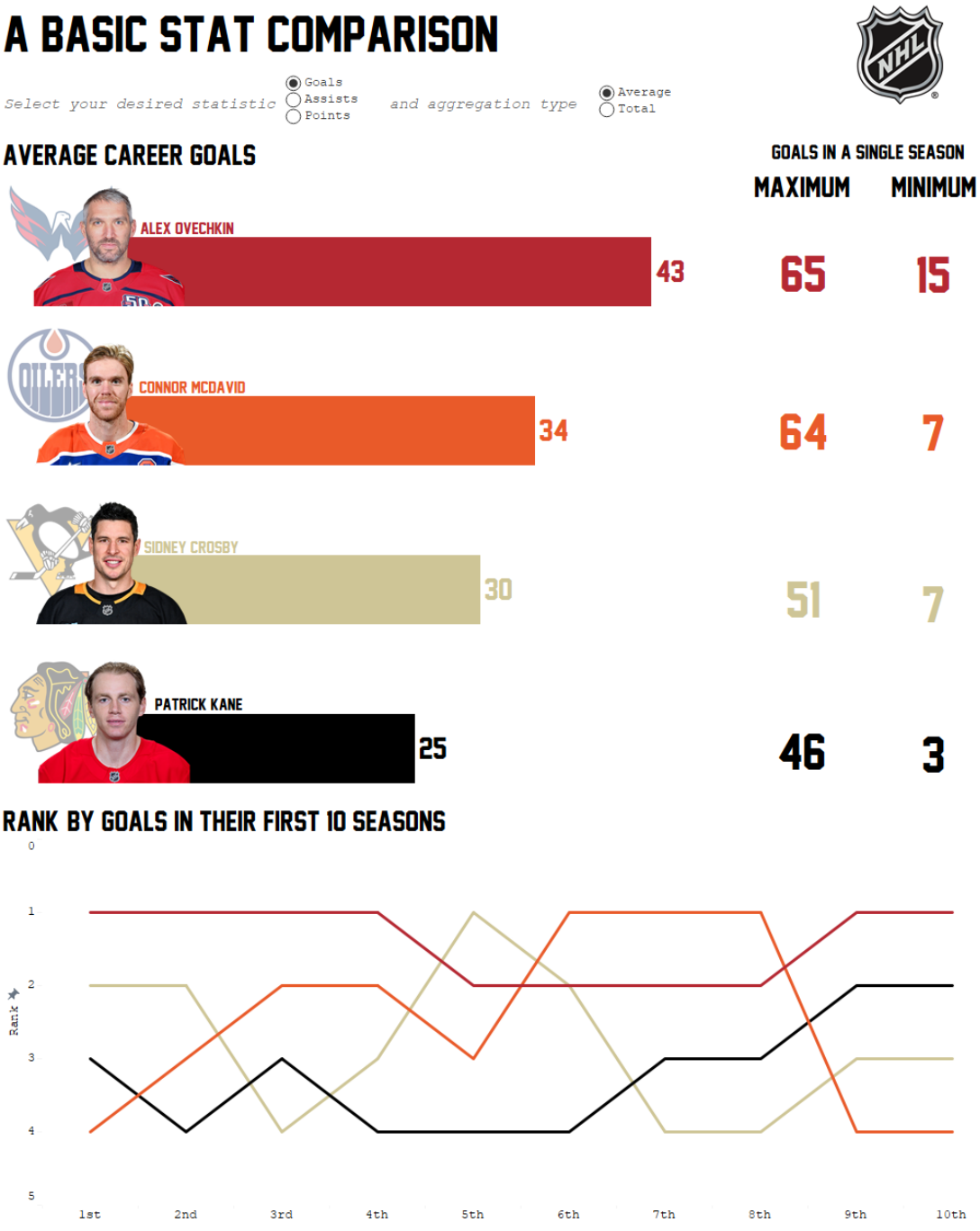
WHO WILL WIN THE 2024 FORMULA 1 DRIVERS CHAMPIONSHIP?

In this project, I was tasked to create an infographic. I chose to create a snapshot of the final stretch of the 2024 F1 World Driver's Championship. This included what points were left up for grabs and how the final two contenders matched up against each other throughout the season thus far.



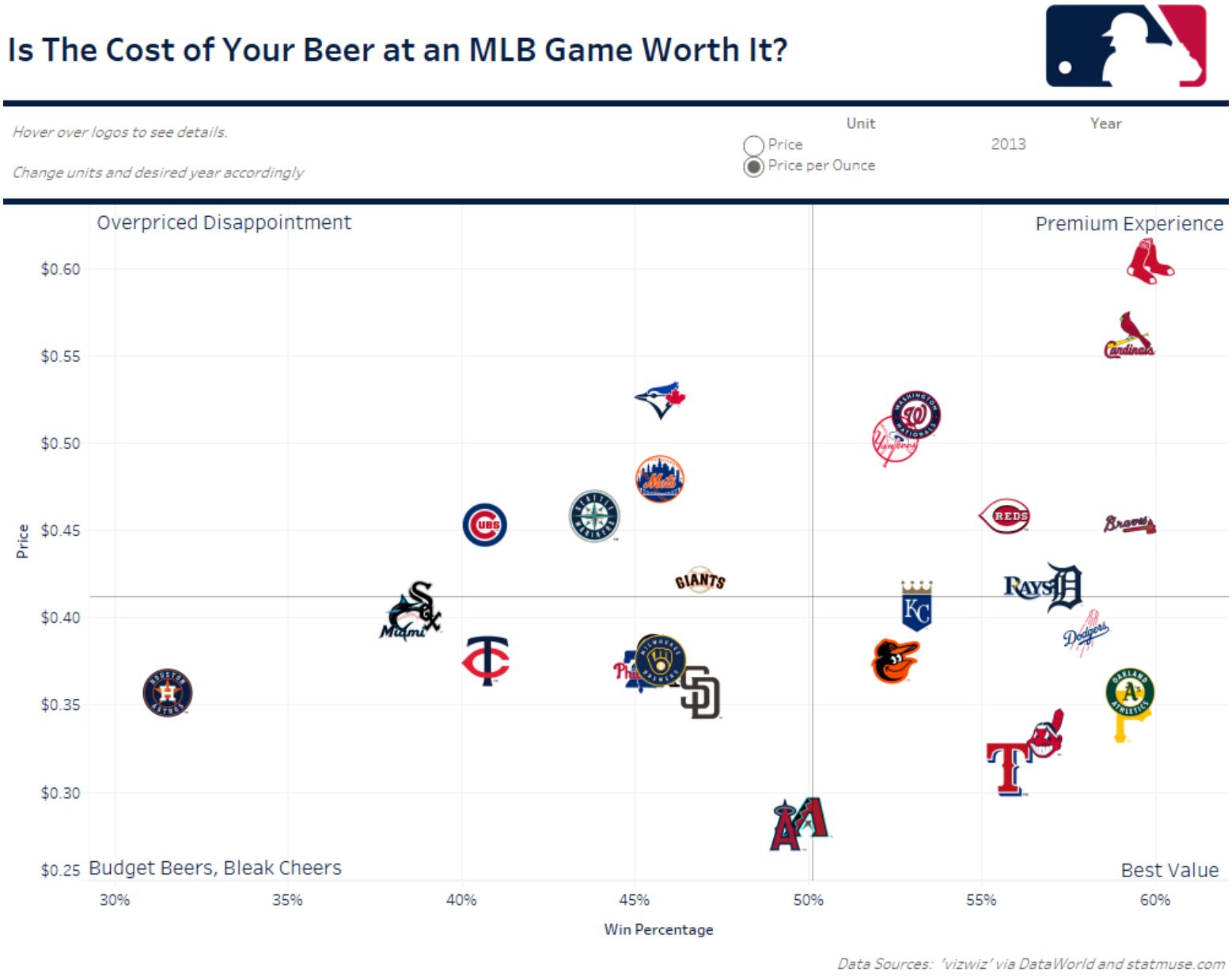
NHL BASIC STAT COMPARISON

In this project, my goal was to provide a simple, yet intriguing comparison of some of the best player in the NHL by comparing basic statistical measures. Parameters allow users to flip between metric and aggregation types. While the middle charts give numerical and bar chart comparisons, the bottom chart ranks the players in the selected stat over the first ten years of their career.



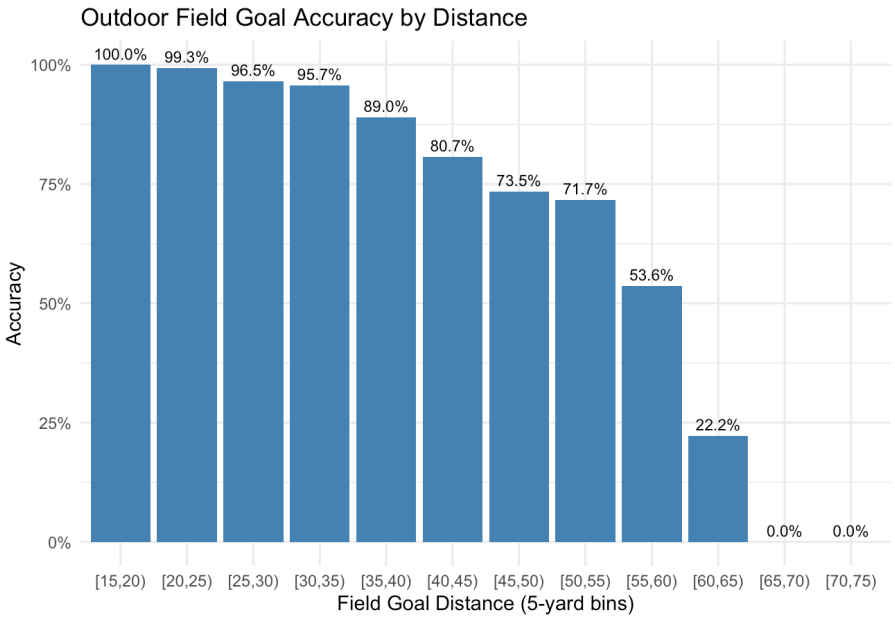
IS THE COST OF YOUR BEER AT AN MLB GAME WORTH IT?

In this project, I was tasked to recreate a visualization. The original viz displayed the cost of a beer across all MLB teams. However, to improve the design, I transformed the narrative to not only include price but also price per ounce. I also included team win percentage to see if you're getting a bang for your buck. By scrolling through the seasons, users can see trends in their desired team over time.



NFL FIELD GOAL ANALYSIS

In this project, I worked on filtering, wrangling, and summarizing the NFL's field goal data. The goal of the project was to utilize a logistic regression to predict the probability of making a field goal in outdoor games.



To begin the project, I filtered all plays to those that were field goals located in outdoor fields, removed N/As, and created bins for distances.

The resulting histogram shows a severe drop off in success rate starting at distances greater than 60 yds.

ORIGINAL MODEL

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    6.238202   0.345087  18.077   <2e-16 ***
kick_distance -0.110092   0.006251 -17.611   <2e-16 ***
temp           0.003903   0.003127   1.248    0.212
wind          -0.015574   0.010389  -1.499    0.134
---
```

By running a regression using all three independent variables, I can see there is not significant evidence that temperature and wind will have a meaningful impact on the outcome, so, they will not be included in the final model.

FINAL MODEL

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)    6.284849   0.286687   21.92   <2e-16 ***
kick_distance -0.109181   0.006212  -17.57   <2e-16 ***
---
```

Therefore, kick distance will be the sole contributor in predicting accuracy.

CONFUSION MATRIX

```
          Reference
Prediction    0    1
         0   36   19
         1  453 2561
```

According to the confusion matrix, the final model produced 19 false negatives (type II errors) and 453 false positives (type I errors), but for the remaining 2597 observations, the model predicted the correct outcome.

WHO ARE THE MOST AND LEAST LUCKY TEAMS IN MLB HISTORY?

In this project, I used Lahman's database to analyze historical trends in multiple variables for Major League Baseball. This predictive model aimed to predict the number of runs a team would score in a season based on runs, walks, hits, doubles, homeruns, stolen bases, and caught stealing. The model was exceptionally accurate with a R-squared of .92, and all variables were significant (except for triples).

LINEAR REGRESSION SUMMARY

new_model <- lm(R ~ Walks + H + X2B + HR + SB + CS, data = Teams)

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept).	*-320.35056*	*39.46058*	*-8.118*	*1.99e-13 ****
Walks	*0.32654*	*0.03424*	*9.536*	*< 2e-16 ****
H	*0.39373*	*0.03915*	*10.057*	*< 2e-16 ****
X2B	*0.50091*	*0.09571*	*5.234*	*5.80e-07 ****
HR	*0.96032*	*0.06776*	*14.172*	*< 2e-16 ****
SB	*0.28728*	*0.07260*	*3.957*	*0.000119 ****
CS	*-0.65646*	*0.28803*	*-2.279*	*0.024139 **

Residual standard error: ***22.85*** *on* ***143*** *degrees of freedom*

Multiple R-squared: ***0.9267****, Adjusted R-squared:* ***0.9237***

F-statistic: 301.4 on 6 and 143 DF, p-value: < 2.2e-16

To find the most and least lucky teams, I found the difference in predicted runs and actual runs. Teams with a large positive difference were lucky, while those with a large negative difference are the least lucky.

MOST LUCKY

Actual > Predicted

Team	Year	Runs	Predicted Runs	Difference
TBA	2021	857	775	82
BAL	2023	807	760	47
TBA	2023	860	823	36
CHN	2023	819	783	36
ARI	2022	702	666	36

LEAST LUCKY

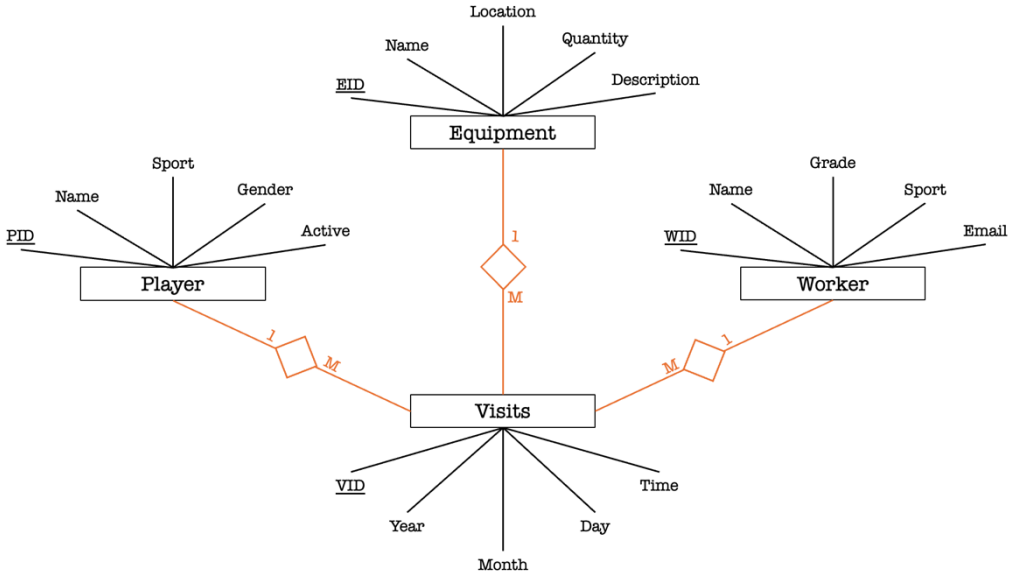
Actual < Predicted

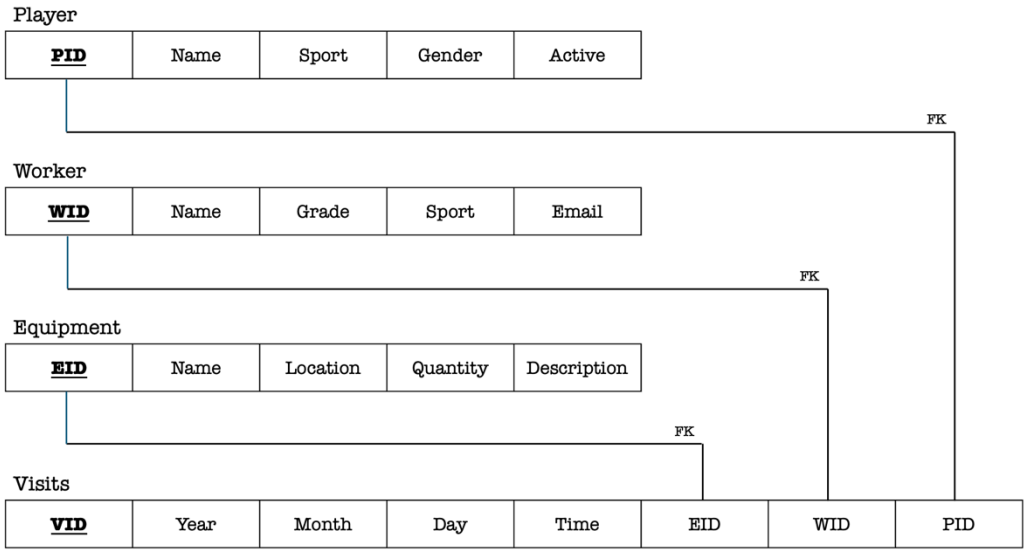
Team	Year	Runs	Predicted Runs	Difference
MIL	2019	769	839	-70
DET	2019	582	640	-58
SLN	2023	719	773	-54
OAK	2023	585	637	-52
BAL	2018	622	665	-43

INNOVATION LAB RELATIONAL DATABASE

In this project, I have built a multi-table relational database to store data on the usage of the Innovation Lab. The Innovation Lab is a room dedicated to rest and recovery for athletes at Providence College, and by storing data, coaches and other athletic staff can monitor and analyze usage patterns.

DESIGN





IMPLEMENTATION

In the implementation stage, the goal is to create each table, define the variables used within each table, identify the data types for each variable, and establish the primary and foreign keys to relate the tables to one another.

```
CREATE TABLE Player
(
   PID INT,
   Name TEXT NOT NULL,
   Sport VARCHAR(10) NOT NULL,
   Gender CHAR(1) NOT NULL,
   Active BOOLEAN NOT NULL,
   PRIMARY KEY (PID)
);
```

```
CREATE TABLE Equipment
(
   EID INT,
   Name TEXT NOT NULL,
   Location TEXT NOT NULL,
   Quantity INT NOT NULL,
   Description TEXT,
   PRIMARY KEY (EID)
);
```

```
CREATE TABLE Worker
(
   WID INT,
   Name TEXT NOT NULL,
   Grade VARCHAR(10) NOT NULL,
   Sport VARCHAR(10) NOT NULL,
   Email TEXT NOT NULL,
   PRIMARY KEY (WID)
);
```

```
CREATE TABLE Visits
(
   VID INT,
   Year INT NOT NULL,
   Month VARCHAR(12) NOT NULL,
   Day INT NOT NULL,
   Time DECIMAL NOT NULL,
   PID INT NOT NULL,
   EID INT NOT NULL,
   WID INT NOT NULL,
   PRIMARY KEY (VID),
   FOREIGN KEY (PID) REFERENCES Player(PID),
   FOREIGN KEY (EID) REFERENCES Equipment(EID),
   FOREIGN KEY (WID) REFERENCES Worker(WID)
);
```

QUERY RETRIEVAL

In this retrieval, the goal is to provide the top ten most visited players and how many times they have visited.

```
SELECT P.Name AS PlayerName, COUNT(V.VID) AS VisitCount
FROM Visits V, Player P
WHERE V.PID = P.PID
GROUP BY P.Name
ORDER BY VisitCount DESC
LIMIT 10;
```

PlayerName	VisitCount
Ryan Welch	70
Macie Pennisi	50
Stone Evans	43
Keira Kelly	32
Emily Bush	30
Lukas Burns	30
Daniel Prescott	27
Harry Scarborough	27
Emily Archibald	25
Ryan Sharp	24

In this retrieval, the goal is to use an exists function to find the most used equipment given that the visits were by females.

```
SELECT E.Name, COUNT(*) AS UsageCount
FROM Equipment E JOIN Visits V ON E.EID = V.EID
WHERE EXISTS ( SELECT *
                FROM Visits V2 JOIN Player P ON V2.PID = P.PID
                WHERE V2.EID = E.EID AND P.Gender = "F")
GROUP BY E.Name
ORDER BY UsageCount DESC
LIMIT 1;
```

Name	UsageCount
Compression Boots	1561

More example queries can be found in Git Hub Repository