*2018-2019 Season NHL Database*

Martin Knuth   
University of PittsburghPittsburgh, PA, United States  
MartyKnuth@pitt.edu

Table of Contents

Title Page………………………...…………………………………………..……………………1

Table of Contents………………...………………………………………………………..………2

Introduction and Abstract………..………………………………………………………………...3

ER Diagram……………………..………………………………………………………………....4

Business Rules…………………..………………………………………………………………...5

Entity and Attribute Descriptions..………………………………………………………………...6

List of Questions………………….……………………………………………………………...13

Closing Section………………….……………………………………………….........................14

My database uses data from the 2018-19 NHL season. I retrieved the data from *Hockey-Reference.com* so all queries run on my database will return actual results from this season. I began to insert data manually, but quickly realized that it would take me too long to insert what ended up being about thirty-eight hundred insert statements. What I came up with was as follows: download the data into a comma separated file, change some of the rows to make insertion easier, and use Jupyter Notebooks to print out the insertion statements using Python. My Python scripts would take in the CSV files I downloaded and run loops to print out the insertion statements I needed. I copied the insertion statements into my SQL script. This was the biggest breakthrough in my entire project, saving me hours of time.

If my database was available to the public, my target audience would be fans of the NHL or the management of the teams. As useful as *Hockey-Reference* is, my database would be more beneficial for the 2018-19 season because of the user’s ability to query the data. Users of *Hockey-Reference* cannot run queries to see specific statistics and data. This gives my database a purpose for die-hard fans of the NHL or future sports historian who want to know specific data you couldn’t easily obtain from a record keeping website like *Hockey-Reference*. The other benefit of my database would be the gathering of more advanced statistics that *Hockey-Reference* does not offer. Statistics like *Goals Saved Above Average, Points Per 60,* and *Adjusted Power Play Percentage* could be calculated with queries I’ve already written. In no way does my database replace the vast amounts of data that *Hockey-Reference* offers, but it does allow fans and front offices to gather statistics not readily available to the public.

A screenshot of a cell phone

Description automatically generated

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Business Rules** | | | | |
| **Entity 1** | **Entity 2** | **Cardinality on Entity 1 side** | **Cardinality on Entity 2 side** | **Business Rule(s)** |
| Arenas | Teams | 1 | 1 | An arena only has one home team. Each team only has one arena. |
| Coaches | Teams | 1…\* | 1…\* | A coach can coach multiple teams in a season. A team can have multiple coaches for a season. |
| Draft Picks | Teams | 1…\* | 1 | A team can draft multiple players, but each player can only be drafted by one team. |
| Games | Teams | 1…\* | 1 | A team can play in multiple games, but a game can only have one home team. |
| Games | Teams | 1…\* | 1 | A team can play in multiple games, but a game can only have one away team. |
| General Manager | Teams | 1 | 1 | In 2018-19, each team only had one general manger, so this is a one-to-one relationship. A team can have one general manager and one general manager only managed one team. |
| Goalies | Teams | 1…\* | 1…\* | A goalie can play for multiple teams and a team can have multiple goalies. |
| Players | Teams | 1…\* | 1…\* | A player can play on multiple teams and a team can have multiple players. |

|  |  |  |
| --- | --- | --- |
| **Arenas** | | |
| PK Arena ID | Integer | Primary Key |
| Arena Name | Varchar | Name of the arena |
| Capacity | Integer | How many people can sit in the arena |
| Year Opened | Integer | Year the arena opened |

|  |  |  |
| --- | --- | --- |
| **Awards** | | |
| PK Award ID | Integer | Primary Key |
| Name of Award | Varchar | Name of the award |
| Winner | Varchar | Name of the winner |
| Description of Award | Varchar | Description of what the award is |

|  |  |  |
| --- | --- | --- |
| **Coaches** | | |
| PK Coach ID | Integer | Primary Key |
| Coach Name | Varchar | Name of the coach |
| Games Coached | Integer | Number of games they’ve coached in their entire career |
| Wins | Integer | Number of wins in their career |
| Losses | Integer | Career losses |
| First Year Coaching | Integer | First year coaching in the NHL |
| Latest Year Coaching | Integer | Latest year coaching in the NHL |
| Stanley Cups Won | Integer | Number of Stanley Cups the coach has won |

|  |  |  |
| --- | --- | --- |
| **Coaches on Teams** | | |
| PK Coach on Team ID | Integer | Primary Key |
| FK Coach ID | Integer | Foreign Key |
| FK Team ID | Integer | Foreign Key |
| Games Coached | Integer | Number of games they’ve coached in the 2018-19 season |
| Wins | Integer | Wins this season |
| Losses | Integer | Losses this season |
| OTL | Integer | Overtime Losses this season |

|  |  |  |
| --- | --- | --- |
| **Draft Picks** | | |
| PK Draft ID | Integer | Primary Key |
| Overall | Integer | The pick this player was selected with |
| FK Team ID | Integer | Foreign Key |
| Draftee Name | Varchar | Name of the person getting drafted |
| Nationality | Varchar | Nationality of the draft pick |
| Pos | Varchar | Position of the player |
| Age | Integer | Age of the person |
| Amateur team | Varchar | The team they were drafted off of |

|  |  |  |
| --- | --- | --- |
| **Games** | | |
| PK Game ID | Integer | Primary Key |
| Game Date | Datetime | Date of the game |
| FK Home ID | Integer | Foreign Key of the home team |
| Home Goal Total | Integer | Number of goals the home team scored |
| FK Visitor ID | Integer | Foreign Key of the visiting team |
| Visitor Goal Total | Integer | Number of goals the visiting team scored |
| Type of Win | Varchar | Type of win, Regulation, Overtime |
| Attendance | Integer | Number of people at the game |
| Length of Game | Integer | Number of minutes the game lasted |

|  |  |  |
| --- | --- | --- |
| **General Manager** | | |
| PK GM ID | Integer | Primary Key |
| GM Name | Varchar | Name of the general manager |
| Tendered Since | Datetime | The day they began as the general manager |
| FK Team ID | Integer | Foreign Key to the team they manage |

|  |  |  |
| --- | --- | --- |
| **Goalies** | | |
| PK Goalie ID | Integer | Primary Key |
| Goalie Name | Varchar | Name of the goalie |
| Age | Integer | Age of the goalie |
| Games Played | Integer | Number of games played that season |
| Games Started | Integer | Number of games they began the game as the starter |
| Wins | Integer | Wins that season |
| Losses | Integer | Losses that season |
| Goals Against | Integer | Goals they’ve allowed |
| Shots Against | Integer | Shots they’ve had |
| Saves | Integer | Shots they’ve saved |
| Shut Outs | Integer | Games they’ve had they didn’t allow a goal |
| Minutes | Integer | Minutes of ice time they’ve had that season |

|  |  |  |
| --- | --- | --- |
| **Goalies on Teams** | | |
| PK Goalie on Team ID | Integer | Primary Key |
| FK Goalie ID | Integer | Foreign Key to their goalie ID |
| FK Team ID | Integer | Foreign Key to their team ID |
| Games Played | Integer | Games they played for this team |
| Games Started | Integer | Games they started for this team |
| Wins | Integer | Wins on this team |
| Losses | Integer | Losses on this team |
| Goals Against | Integer | Goals they’ve allowed on this team |
| Shots Against | Integer | Shots they’ve had on this team |
| Saves | Integer | Shots they’ve saved on this team |
| Shut Outs | Integer | Games they’ve had they didn’t allow a goal on this team |
| Minutes | Integer | Minutes of ice time they’ve for this team |

|  |  |  |
| --- | --- | --- |
| **Players** | | |
| PK Player ID | Integer | Primary Key |
| Player Name | Integer | Name of player |
| Age | Integer | Age of player |
| Position | Varchar | Position of the player |
| Games Played | Integer | Games they’ve played this season |
| Goals | Integer | Goals they’ve scored on the season |
| Assists | Integer | Assists on the season |
| Points | Integer | Points on the season |
| Plus Minus | Integer | On ice goal differential for the season |
| Penalties in Minutes | Integer | Number of minutes spent in penalty box |
| Shots | Integer | Number of shots for the season |
| Time on Ice | Integer | Number of minutes they’ve played for the season on the ice |
| Blocked Shots | Integer | Blocked shots for the season |
| Hits | Integer | Body checks for the season |

|  |  |  |
| --- | --- | --- |
| **Players on Teams** | | |
| PK Player on Team ID | Integer | Primary Key |
| FK Player ID | Integer | Foreign Key for player ID |
| FK Team ID | Integer | Foreign Key for team ID |
| Games Played | Integer | Games they’ve played this team |
| Goals | Integer | Goals they’ve scored on this team |
| Assists | Integer | Assists on this team |
| Points | Integer | Points on this team |
| Plus Minus | Integer | On ice goal differential for this team |
| Penalties in Minutes | Integer | Number of minutes spent in penalty box on this team |
| Shots | Integer | Number of shots for the season on this team |
| Time on Ice | Integer | Number of minutes they’ve played for this team on the ice |
| Blocked Shots | Integer | Blocked shots for this team |
| Hits | Integer | Body checks for this team |

|  |  |  |
| --- | --- | --- |
| **Teams** | | |
| PK Team ID | Integer | Primary Key |
| Team Name | Varchar | Name of the team |
| Wins | Integer | Number of wins |
| Losses | Integer | Number of losses |
| Overtime Losses | Integer | Number of losses in overtime |
| Points | Integer | Number of points in the standings the team has |
| Goals For | Integer | Number of goals the team scored |
| Goals Against | Integer | Number of goals the team allowed |
| Power Plays | Integer | Number of power plays they’ve had |
| PP Goals | Integer | Goals scored on the power play |
| Power Plays Against | Integer | Power plays their opponent has had |
| PPA Goals | Integer | Goals their opponents have scored on the power play |
| SH Goals | Integer | Shorthanded goals |
| SH Goals Against | Integer | Shorthanded goals the opponent has scored |
| Shots | Integer | Shots on net for |
| Shots Against | Integer | Shots against |
| FK Arena ID | Integer | Foreign Key for their arena |

QUESTIONS:

The SQL scripts to answer these questions and more can be found in the file named “NHL queries.sql.” That file is in the zip file along with the creation script.

1. Can the database provide the standings of the teams? This is important because the league and the fans need to see what teams finished where to determine who made the playoffs and draft position.
2. Can the database provide all the players that played for a specific team? This is important so fans and the league can see who played on that team.
3. Can the database provide a list of all the games a team played? This adds value by being able to look at a team’s schedule and what the results of each game were.
4. Can the database calculate the average age of all players in the NHL? This adds value by giving the user the ability to compare the average age of an NHL player to the average age of a player on their team, players from previous seasons, or players from other leagues.
5. Can the database provide a list of players that played for the most teams in this season? This would be important for users that want to track roster moves.
6. Can the database calculate the average number of goals scored in each month? Each season the opening games feel much higher scoring. Can I check this theory by looking at the average goals in each game over time?
7. Can the database provide a list of what players average the most ice time? This shows what workload players get in the season. This could be important to show how good a player is if they are trusted to play a lot of ice time.
8. Can the database provide a list of the top ten goal tenders based on save percentage? Save percentage is the most used stat to determine a goalies’ ability. The top ten goalies in the league might be considered for awards.
9. Can the database provide a list of the highest scoring players on each team? This is important so fans know what players to look out for when they go to games.
10. Can the database provide a list of the players who won the Stanley Cup? This would add value because the league needs to know what players’ names to engrave onto the Stanley Cup.

CLOSING SECTION

The biggest thing I learned in this project was going through the process of designing my own database and the obstacles to implement my design. The biggest breakthrough in creating the database was to automate the insertion statements. Without coming up with the idea to print the insertion statements in a Python script, I simply would not have been able to insert all the data I would have needed to build a fully functioning database. I have gained a whole new respect for websites with large databases thousands of times bigger than mine, especially *Hockey-Reference* which has season wide data going back to 1917.