FLORIDA PANTHERS - DATA ENGINEER

INTERVIEW PROJECT

GOAL: Use the publicly accessible portions of the NHL API (https://gitlab.com/dword4/nhlapi/-/blob/master/stats-api.md) to query for historical Florida Panthers statistics, create tables consistent with instructions, and then use SQL to write queries and answer questions using those tables.

PYTHON

1. Make sure you have the necessary packages installed, or similar.

```
pip install pandas
pip install requests
```

2. Create an .ipynb file and import them.

```
import requests
import pandas as pd
```

3. Write one or more function(s) that queries the /teams and the /people endpoints for all Panthers players, starting in the 2014-15 season, and ending in the 2021-22 season.
   a. Should return a pandas data frame where every row is an individual player's season (only seasons where they played for the Panthers), where player id is the first field, season is the second, then descriptive player information (age, height, weight), and finally all of their season averages/totals playing for the Panthers.
   b. Calculate their age during that season using this logic:
      i. https://www.hockey-reference.com/about/glossary.html

```
def create_panthers_players_table():

        return panthers_players_table
```

4. Write one or more function(s) that queries the /schedule and /game endpoints for all Panthers games, starting in the 2014-15 season, and ending in the 2021-22 season.
   a. Should return a data frame where every row is an individual game, with the game id as the first field, whether the Panthers were home or away as the second field, the result for the Panthers as the third (win, loss, overtime loss), and finally the team stats for both the Panthers and their opponent.

```
def create_panthers_game_boxscores_table():

    return panthers_game_boxscores_table
```

5. Write one or more function(s) that queries the /game endpoint for every Panthers player's individual game box score, using the game id's you gathered from above.
   a. Should return a data frame where every row is a Panthers player that played in the game, where the first two fields are player id and game id, and the rest of the fields are all of the player's individual statistics for the game.

```
def create_panthers_player_boxscores_table():

    return panthers_player_boxscores_table
```

6. Write all final data frames as csv's to a local directory to attach later.

SQL

Make sure you have pandas sql installed.

```
pip install pandasql
```

Import sqldf from the pandas sql package.

```
from pandasql import sqldf
```

Create a quick function to facilitate the queries.

```
pysqldf = lambda q: sqldf(q, globals())
```

Load in your three tables and give them clearly defined names. Answer questions in the following format.

```
q = """
    """
q1_answer = pysqldf(q)
print(q1_answer)
```

1. Which five players have played the most full seasons for the Panthers in the time frame we pulled?
2. Which player had the highest plus-minus in an individual game and what was the result for the Panthers?
3. Who were the youngest players to play for the Panthers in this time period, that were also on the team in 2021-22, and what were the first games they played for the Panthers?
4. Who had the most total penalty minutes in a season for the panthers, and what season did this take place in?
5. Rank Panthers seasons by number of total home shutouts.