Brian_Reppeto_540_week_9_10_Milestone_4

February 18, 2024

0.0.1 DSC 540 Week 2 Data Wrangling with Python:

0.0.2 Project: Milestone 4

0.0.3 Author: Brian Reppeto 2/9/2024

```
[19]: # import libraries

      import http.client
      import json
      import pandas as pd
      import time
```

```
[20]: # create empty list to store stadium data

      stad_data = []
```

```
[21]: # base URL and API key parts for the request

      base_url = "/nfl/official/trial/v7/en/games"
      api_key = "k8uqtrmhauf3rmkm6f8t9f54"
```

```
[22]: # create HTTPS connection

      conn = http.client.HTTPSConnection("api.sportradar.us")
```

```
[23]: # Iterate through years from 2017 to 2024

      for year in range(2017, 2024):# API request for each year
          request_url = f"{base_url}/{year}/reg/schedule.json?api_key={api_key}"
          conn.request("GET", request_url)
          res = conn.getresponse()
          data = res.read()
          decoded_data = json.loads(data.decode("utf-8")) # decode JSON

          for week in decoded_data["weeks"]:
              for game in week["games"]:
                  home_id = game["home"]["id"]
                  home_name = game["home"]["name"]
```

```python
            home_alias = game["home"]["alias"]
            venue_id = game["venue"]["id"]
            venue_name = game["venue"]["name"]
            surface = game["venue"]["surface"]
            roof_type = game["venue"]["roof_type"]
            stad_data.append({
                "year": year,
                "home_id": home_id,
                "name": home_name,
                "alias": home_alias,
                "venue_id": venue_id,
                "venue_name": venue_name,
                "surface": surface,
                "roof_type": roof_type
            })

        # pause for 3 seconds before making the next request
        time.sleep(3)
```

[25]:
```python
# create a df

stad_df = pd.DataFrame(stad_data)
```

[26]:
```python
# print the df

print(stad_df)
```

```
      year                               home_id                     name alias  \
0     2017   97354895-8c77-4fd4-a860-32e62ea7382a    New England Patriots    NE
1     2017   4809ecb0-abd3-451d-9c4a-92a90b83ca06          Miami Dolphins   MIA
2     2017   d5a2eb42-8065-4174-ab79-0a6fa820e35e        Cleveland Browns   CLE
3     2017   7b112545-38e6-483c-a55c-96cf6ee49cb8           Chicago Bears   CHI
4     2017   22052ff7-c065-42ee-bc8f-c4691c50e624     Washington Redskins   WAS
...    ...                                   ...                     ...   ...
1838  2023   f0e724b0-4cbf-495a-be47-013907608da9     San Francisco 49ers    SF
1839  2023   7d4fcc64-9cb5-4d1b-8e75-8a906d1e1576        Las Vegas Raiders    LV
1840  2023   1f6dcffb-9823-43cd-9ff4-e7a8466749b5    Los Angeles Chargers   LAC
1841  2023   22052ff7-c065-42ee-bc8f-c4691c50e624  Washington Commanders   WAS
1842  2023   4809ecb0-abd3-451d-9c4a-92a90b83ca06          Miami Dolphins   MIA

                                  venue_id         venue_name     surface  \
0     e43310b1-cb82-4df9-8be5-e9b39637031b   Gillette Stadium   artificial
1     50a5c833-1570-4c38-abc7-7914cf87dbde   Hard Rock Stadium        turf
2     90c38d91-3774-4f5d-82ca-1c806828219f  FirstEnergy Stadium       turf
3     d7866605-5ac6-4b3a-90e8-760cc5a26b75       Soldier Field        turf
4     7c11bb2d-4a53-4842-b842-0f1c63ed78e9           FedExField        turf
...                                    ...                 ...          ...
1838  24c71dbf-6469-46f1-8165-22005c295c8f       Levi's Stadium        turf
```

```
1839   53d42601-5dfb-40a7-bd94-da2c1d3a75e4        Allegiant Stadium        turf
1840   790c1f04-73c6-4f6f-8b1e-78a62260be90            SoFi Stadium  artificial
1841   7c11bb2d-4a53-4842-b842-0f1c63ed78e9              FedExField        turf
1842   50a5c833-1570-4c38-abc7-7914cf87dbde      Hard Rock Stadium        turf

      roof_type
0       outdoor
1       outdoor
2       outdoor
3       outdoor
4       outdoor
…           …
1838    outdoor
1839       dome
1840    outdoor
1841    outdoor
1842    outdoor

[1843 rows x 8 columns]
```

Step 1 drop venue_id & home_id columns

```python
[27]:  # drop the 'venue_id' column from the df

       stad_df = stad_df.drop(columns=['venue_id','home_id'])
```

```python
[28]:  # print the df

       print(stad_df)
```

```
      year                  name alias           venue_name     surface  \
0     2017     New England Patriots   NE     Gillette Stadium  artificial
1     2017          Miami Dolphins  MIA     Hard Rock Stadium        turf
2     2017         Cleveland Browns  CLE  FirstEnergy Stadium        turf
3     2017           Chicago Bears  CHI        Soldier Field        turf
4     2017     Washington Redskins  WAS           FedExField        turf
…      …                     …    …                 …           …
1838  2023      San Francisco 49ers   SF       Levi's Stadium        turf
1839  2023        Las Vegas Raiders   LV     Allegiant Stadium        turf
1840  2023     Los Angeles Chargers  LAC          SoFi Stadium  artificial
1841  2023  Washington Commanders  WAS           FedExField        turf
1842  2023          Miami Dolphins  MIA     Hard Rock Stadium        turf

      roof_type
0       outdoor
1       outdoor
2       outdoor
3       outdoor
4       outdoor
```

```
…        …
1838   outdoor
1839      dome
1840   outdoor
1841   outdoor
1842   outdoor

[1843 rows x 6 columns]
```

Step 2 Capitalize the column headers

```python
[31]:  # capitalize the column headers

       stad_df.columns = [col.upper() for col in stad_df.columns]
```

```python
[32]:  # print the df

       print(stad_df)
```

```
       YEAR                    NAME ALIAS         VENUE_NAME     SURFACE  \
0      2017   New England Patriots    NE     Gillette Stadium  artificial
1      2017         Miami Dolphins   MIA     Hard Rock Stadium        turf
2      2017        Cleveland Browns   CLE  FirstEnergy Stadium        turf
3      2017           Chicago Bears   CHI        Soldier Field        turf
4      2017     Washington Redskins   WAS            FedExField        turf
…      …                        …    …                   …           …
1838   2023    San Francisco 49ers    SF       Levi's Stadium        turf
1839   2023       Las Vegas Raiders    LV     Allegiant Stadium        turf
1840   2023   Los Angeles Chargers   LAC          SoFi Stadium  artificial
1841   2023  Washington Commanders   WAS            FedExField        turf
1842   2023          Miami Dolphins   MIA     Hard Rock Stadium        turf

       ROOF_TYPE
0        outdoor
1        outdoor
2        outdoor
3        outdoor
4        outdoor
…            …
1838     outdoor
1839        dome
1840     outdoor
1841     outdoor
1842     outdoor

[1843 rows x 6 columns]
```

Step 3 Change "turf" to "grass" in the SURFACE column

```
[33]:  # change "turf" to "grass" in the SURFACE column

       stad_df['SURFACE'] = stad_df['SURFACE'].replace('turf', 'grass')

[34]:  # print the df

       print(stad_df)

            YEAR                   NAME ALIAS           VENUE_NAME     SURFACE  \
      0     2017   New England Patriots   NE      Gillette Stadium   artificial
      1     2017         Miami Dolphins  MIA      Hard Rock Stadium       grass
      2     2017       Cleveland Browns  CLE    FirstEnergy Stadium       grass
      3     2017          Chicago Bears  CHI          Soldier Field       grass
      4     2017    Washington Redskins  WAS              FedExField       grass
      ...    ...                    ...  ...                   ...         ...
      1838  2023   San Francisco 49ers   SF         Levi's Stadium       grass
      1839  2023      Las Vegas Raiders   LV      Allegiant Stadium       grass
      1840  2023   Los Angeles Chargers  LAC            SoFi Stadium artificial
      1841  2023   Washington Commanders WAS              FedExField       grass
      1842  2023         Miami Dolphins  MIA      Hard Rock Stadium       grass

           ROOF_TYPE
      0      outdoor
      1      outdoor
      2      outdoor
      3      outdoor
      4      outdoor
      ...        ...
      1838   outdoor
      1839      dome
      1840   outdoor
      1841   outdoor
      1842   outdoor

      [1843 rows x 6 columns]
```

Step 4 create a new column state based on team look-up

```
[40]:  # create look-up for team and state

       team_to_state = {
           'TB': 'FL',   # Tampa Bay Buccaneers to Florida
           'TEN': 'TN',  # Tennessee Titans to Tennessee
           'HOU': 'TX',  # Houston Texans to Texas
           'DAL': 'TX',  # Dallas Cowboys to Texas
           'SF': 'CA',   # San Francisco 49ers to California
           'SEA': 'WA',  # Seattle Seahawks to Washington
           'LAC': 'CA',  # Los Angeles Chargers to California
```

```
    'LAR': 'CA',   # Los Angeles Rams to California
    'ARI': 'AZ',   # Arizona Cardinals to Arizona
    'DEN': 'CO',   # Denver Broncos to Colorado
    'KC': 'MO',   # Kansas City Chiefs to Missouri
    'LV': 'NV',   # Las Vegas Raiders to Nevada
    'MIA': 'FL',   # Miami Dolphins to Florida
    'NE': 'MA',   # New England Patriots to Massachusetts
    'BUF': 'NY',   # Buffalo Bills to New York
    'NYJ': 'NY',   # New York Jets to New York
    'NYG': 'NY',   # New York Giants to New York
    'PHI': 'PA',   # Philadelphia Eagles to Pennsylvania
    'PIT': 'PA',   # Pittsburgh Steelers to Pennsylvania
    'BAL': 'MD',   # Baltimore Ravens to Maryland
    'WAS': 'MD',   # Washington Commanders to Maryland (Note: The team is␣
 ↪associated with Washington D.C., but their stadium is in Maryland)
    'CLE': 'OH',   # Cleveland Browns to Ohio
    'CIN': 'OH',   # Cincinnati Bengals to Ohio
    'IND': 'IN',   # Indianapolis Colts to Indiana
    'JAC': 'FL',   # Jacksonville Jaguars to Florida
    'TEN': 'TN',   # Tennessee Titans to Tennessee
    'ATL': 'GA',   # Atlanta Falcons to Georgia
    'CAR': 'NC',   # Carolina Panthers to North Carolina
    'NO': 'LA',   # New Orleans Saints to Louisiana
    'MIN': 'MN',   # Minnesota Vikings to Minnesota
    'GB': 'WI',   # Green Bay Packers to Wisconsin
    'CHI': 'IL',   # Chicago Bears to Illinois
    'DET': 'MI',   # Detroit Lions to Michigan
    'OAK': 'CA',   # Oakland Raiders to Cal
}
```

[41]: ```
apply the complete team-to-state mapping

stad_df['STATE'] = stad_df['ALIAS'].map(team_to_state)
```

[42]: ```
# print the df

print(stad_df)
```

```
      YEAR                 NAME ALIAS         VENUE_NAME    SURFACE  \
0     2017  New England Patriots    NE    Gillette Stadium  artificial
1     2017        Miami Dolphins   MIA   Hard Rock Stadium      grass
2     2017      Cleveland Browns   CLE  FirstEnergy Stadium      grass
3     2017         Chicago Bears   CHI        Soldier Field      grass
4     2017   Washington Redskins   WAS          FedExField      grass
...    ...                  ...   ...                 ...        ...
1838  2023   San Francisco 49ers    SF     Levi's Stadium      grass
1839  2023     Las Vegas Raiders    LV   Allegiant Stadium      grass
1840  2023   Los Angeles Chargers  LAC        SoFi Stadium  artificial
```

```
1841  2023  Washington Commanders   WAS           FedExField        grass
1842  2023        Miami Dolphins     MIA   Hard Rock Stadium         grass


      ROOF_TYPE STATE
0       outdoor    MA
1       outdoor    FL
2       outdoor    OH
3       outdoor    IL
4       outdoor    MD
...         ...   ...
1838    outdoor    CA
1839       dome    NV
1840    outdoor    CA
1841    outdoor    MD
1842    outdoor    FL

[1843 rows x 7 columns]
```

Step 5 Capitalize the first leter of the surface and roof_type columns

```python
[43]:  # capitalize the first letter of 'SURFACE' and 'ROOF_TYPE' columns

       stad_df['SURFACE'] = stad_df['SURFACE'].str.capitalize()
       stad_df['ROOF_TYPE'] = stad_df['ROOF_TYPE'].str.capitalize()
```

```python
[44]:  # print the df

       print(stad_df)
```

```
      YEAR                    NAME ALIAS           VENUE_NAME     SURFACE  \
0     2017   New England Patriots    NE    Gillette Stadium   Artificial
1     2017         Miami Dolphins   MIA   Hard Rock Stadium        Grass
2     2017       Cleveland Browns   CLE  FirstEnergy Stadium       Grass
3     2017          Chicago Bears   CHI       Soldier Field        Grass
4     2017    Washington Redskins   WAS           FedExField        Grass
...    ...                    ...   ...                 ...          ...
1838  2023    San Francisco 49ers    SF       Levi's Stadium       Grass
1839  2023      Las Vegas Raiders    LV   Allegiant Stadium        Grass
1840  2023   Los Angeles Chargers   LAC          SoFi Stadium   Artificial
1841  2023  Washington Commanders   WAS           FedExField        Grass
1842  2023         Miami Dolphins   MIA   Hard Rock Stadium        Grass


      ROOF_TYPE STATE
0       Outdoor    MA
1       Outdoor    FL
2       Outdoor    OH
3       Outdoor    IL
4       Outdoor    MD
...         ...   ...
```

```
1838    Outdoor    CA
1839       Dome    NV
1840    Outdoor    CA
1841    Outdoor    MD
1842    Outdoor    FL


[1843 rows x 7 columns]
```

[45]:
```python
# create csv

output_file_path = '/Users/brianreppeto/DSC540/api_data.csv'
stad_df.to_csv(output_file_path, index=False)
```

Ethical Implications

The API project involved collecting, transforming, and analyzing NFL data, with a focus on team, venue, and game characteristics such as location, surface type, and roof type. Throughout the process, I adhered to ethical guidelines to ensure the integrity and fairness of the analysis. Key steps included data cleaning, feature engineering, and the creation of a CSV file for the transformed dataset.

1. Privacy: The data utilized in this project was strictly related to publicly available information about NFL teams and games. No personal data about players, staff, or fans was collected or analyzed, thus respecting privacy norms.
2. Bias and Fair Representation: A concerted effort was made to treat all teams and data points equally, avoiding biases towards certain teams.
3. Accuracy: High standards for data accuracy were maintained. The analysis was based on reliable sources and validated methods to ensure the findings were representative and truthful.
4. Use of Data: The project's scope was limited to analysis for informational purposes, avoiding any applications that could lead to ethical concerns, such as gambling.
5. Data Monetization and Long-term Impacts: There was no monetization of the data involved in this project.

[ ]: