NFL_Playoff_Teams_Scraped

August 9, 2024

[1]:
```
# Import necessary libraries
from bs4 import BeautifulSoup
import requests
import pandas as pd
```

[2]:
```
# Initialize empty list to hold individual DataFrames
df_list = []
```

[3]:
```
# Function to generate URL for scraping
def gen_url(start_year, end_year):
    """
    Generates a Wikipedia URL for NFL playoffs based on the start and end years.

    Parameters:
    start_year (int): The start year of the NFL season.
    end_year (int): The end year of the NFL season.

    Returns:
    str: The URL for the Wikipedia page of the NFL playoffs for that season.
    """
    url_template = 'https://en.wikipedia.org/wiki/
 ↪{start_year}%E2%80%93{end_year}_NFL_playoffs'

    return url_template.format(start_year=start_year, end_year=str(end_year)[-2:
 ↪])
```

[4]:
```
# Loop through year range and scrape data using the gen_url function
for year in range(2011, 2024):
    url = gen_url(year, year + 1)
    print(f'Fetching data for {year}')

    # Send a GET request to fetch the webpage content
    response = requests.get(url)
    webpage = response.content

    # Parse through the webpage content with Beautiful Soup
    soup = BeautifulSoup(webpage, 'html.parser')
```

```
    # Select the target table
    table = soup.find('table', class_ = 'wikitable')

    # Check if table exists
    if table:
        # Locate all rows in the table
        rows = table.find_all('tr')

        # Initialize a list to hold the current year's team names
        year_data = []

        # Iterate over rows starting with index 2
        for row in rows[2:]:
            cells = row.find_all('td') # Find all elements in the row
            cell_texts = [cell.text.strip() for cell in cells] # Extract and␣
↪strip the text from each cell

            # Extract the team names from the text
            for item in cell_texts[1:]: # Skip first element
                full_name = item.split('(')[0].strip() # Extract team name␣
↪before parenthesis
                parts = full_name.split() # Split name by spaces
                team_name = parts[-1] # Extract the last part of team name

                # Append the year and team name to the year's data list
                year_data.append({'year': year, 'team': team_name})

        # Convert the current year's data to a DataFrame and add it to the list␣
↪of DataFrames
        df_list.append(pd.DataFrame(year_data))

    else:
        print(f'No table found for {year}')
```

```
Fetching data for 2011
Fetching data for 2012
Fetching data for 2013
Fetching data for 2014
Fetching data for 2015
Fetching data for 2016
Fetching data for 2017
Fetching data for 2018
Fetching data for 2019
Fetching data for 2020
Fetching data for 2021
Fetching data for 2022
Fetching data for 2023
```

[5]:
```
# Convert the current year's data to a DataFrame and add it to the list of␣
↪DataFrames
df = pd.concat(df_list, ignore_index=True)

print(df)
```

```
     year        team
0    2011    Patriots
1    2011     Packers
2    2011      Ravens
3    2011       49ers
4    2011      Texans
..    …          …
127  2023       Lions
128  2023  Buccaneers
129  2023      Eagles
130  2023        Rams
131  2023     Packers

[132 rows x 2 columns]
```

[6]:
```
# Save DataFrame to CSV file
df.to_csv('nfl_playoff_teams.csv', index=False)
```