Brian_Reppeto_540_week_5_6_Milestone_2

January 21, 2024

0.0.1 DSC 540 Week 2 Data Wrangling with Python:

0.0.2 Project: Milestone 2

0.0.3 Author: Brian Reppeto 1/16/2024

```python
[1]: # import libraries

import pandas as pd
```

```python
[2]: # load the CSV file into a data frame

file_path = '/Users/brianreppeto/DSC540/teams.csv'
teams_df = pd.read_csv(file_path)
```

```python
[3]: # head the new data frame to see file

teams_df.head()
```

```
[3]:    season team  nfl  nfl_team_id espn  pfr  pff pfflabel   fo  \
    0    2002  ARI  ARI         3800  ARI  CRD  NaN      NaN  ARI
    1    2002  ATL  ATL          200  ATL  ATL  NaN      NaN  ATL
    2    2002  BAL  BAL          325  BAL  RAV  NaN      NaN  BAL
    3    2002  BUF  BUF          610  BUF  BUF  NaN      NaN  BUF
    4    2002  CAR  CAR          750  CAR  CAR  NaN      NaN  CAR

                  full   location short_location   nickname        hyphenated  \
    0  Arizona Cardinals    Arizona        Arizona  Cardinals  arizona-cardinals
    1    Atlanta Falcons    Atlanta        Atlanta    Falcons    atlanta-falcons
    2   Baltimore Ravens  Baltimore      Baltimore     Ravens   baltimore-ravens
    3      Buffalo Bills    Buffalo        Buffalo      Bills      buffalo-bills
    4  Carolina Panthers   Carolina       Carolina   Panthers  carolina-panthers

        sbr  sbr_wins sbr_name draft_kings
    0  1549       NaN      NaN         NaN
    1  1546       NaN      NaN         NaN
    2  1521       NaN      NaN         NaN
    3  1526       NaN      NaN         NaN
    4  1545       NaN      NaN         NaN
```

Step 1

```python
[4]:  # Remove unneeded columns

      columns_to_remove = ["sbr", "sbr_wins", "sbr_name", "draft_kings"]
      teams_df = teams_df.drop(columns=columns_to_remove, errors='ignore')
```

```python
[5]:  # head the new data frame to see removed columns

      teams_df.head()
```

```
[5]:    season team  nfl  nfl_team_id espn  pfr  pff pfflabel   fo  \
     0    2002  ARI  ARI         3800  ARI  CRD  NaN      NaN  ARI
     1    2002  ATL  ATL          200  ATL  ATL  NaN      NaN  ATL
     2    2002  BAL  BAL          325  BAL  RAV  NaN      NaN  BAL
     3    2002  BUF  BUF          610  BUF  BUF  NaN      NaN  BUF
     4    2002  CAR  CAR          750  CAR  CAR  NaN      NaN  CAR

                    full    location short_location   nickname          hyphenated
     0  Arizona Cardinals    Arizona        Arizona  Cardinals  arizona-cardinals
     1    Atlanta Falcons    Atlanta        Atlanta    Falcons    atlanta-falcons
     2   Baltimore Ravens  Baltimore      Baltimore     Ravens   baltimore-ravens
     3      Buffalo Bills    Buffalo        Buffalo      Bills      buffalo-bills
     4  Carolina Panthers   Carolina       Carolina   Panthers  carolina-panthers
```

Step 2

```python
[6]:  # Filter rows where the "season" is greater than or equal to 2017

      teams_df = teams_df[teams_df['season'] > 2017]
```

```python
[7]:  # head the data frame to see removed rows

      teams_df.head()
```

```
[7]:      season team  nfl  nfl_team_id espn  pfr  pff pfflabel   fo  \
     512    2018  ARI  ARI         3800  ARI  CRD  1.0      ARZ  ARI
     513    2018  ATL  ATL          200  ATL  ATL  2.0      ATL  ATL
     514    2018  BAL  BAL          325  BAL  RAV  3.0      BLT  BAL
     515    2018  BUF  BUF          610  BUF  BUF  4.0      BUF  BUF
     516    2018  CAR  CAR          750  CAR  CAR  5.0      CAR  CAR

                      full    location short_location   nickname          hyphenated
     512  Arizona Cardinals    Arizona        Arizona  Cardinals  arizona-cardinals
     513    Atlanta Falcons    Atlanta        Atlanta    Falcons    atlanta-falcons
     514   Baltimore Ravens  Baltimore      Baltimore     Ravens   baltimore-ravens
     515      Buffalo Bills    Buffalo        Buffalo      Bills      buffalo-bills
     516  Carolina Panthers   Carolina       Carolina   Panthers  carolina-panthers
```

Step 3

```
[8]: # make all the column headings in upper letters

     teams_df.columns= teams_df.columns.str.upper()
```

```
[9]: # head the data frame to see the changed upper column headings

     teams_df.head()
```

```
[9]:      SEASON TEAM  NFL  NFL_TEAM_ID ESPN  PFR  PFF PFFLABEL   FO  \
     512    2018  ARI  ARI         3800  ARI  CRD  1.0      ARZ  ARI
     513    2018  ATL  ATL          200  ATL  ATL  2.0      ATL  ATL
     514    2018  BAL  BAL          325  BAL  RAV  3.0      BLT  BAL
     515    2018  BUF  BUF          610  BUF  BUF  4.0      BUF  BUF
     516    2018  CAR  CAR          750  CAR  CAR  5.0      CAR  CAR


                       FULL   LOCATION SHORT_LOCATION   NICKNAME          HYPHENATED
     512  Arizona Cardinals    Arizona        Arizona  Cardinals   arizona-cardinals
     513    Atlanta Falcons    Atlanta        Atlanta    Falcons     atlanta-falcons
     514  Baltimore Ravens   Baltimore      Baltimore     Ravens    baltimore-ravens
     515      Buffalo Bills    Buffalo        Buffalo      Bills       buffalo-bills
     516  Carolina Panthers   Carolina       Carolina    Panthers  carolina-panthers
```

Step 4

```
[10]: # Create a new column "TEAM_SEASON" and fill it with the concatenation of␣
      ↪"SEASON" and "TEAM" columns

      teams_df['TEAM_SEASON'] = ''

      for index, row in teams_df.iterrows():
          teams_df.at[index, 'TEAM_SEASON'] = str(row['SEASON']) + '_' +␣
      ↪str(row['TEAM'])
```

```
[11]: # head the data frame to see the updated DF with the new column

      teams_df.head()
```

```
[11]:      SEASON TEAM  NFL  NFL_TEAM_ID ESPN  PFR  PFF PFFLABEL   FO  \
      512    2018  ARI  ARI         3800  ARI  CRD  1.0      ARZ  ARI
      513    2018  ATL  ATL          200  ATL  ATL  2.0      ATL  ATL
      514    2018  BAL  BAL          325  BAL  RAV  3.0      BLT  BAL
      515    2018  BUF  BUF          610  BUF  BUF  4.0      BUF  BUF
      516    2018  CAR  CAR          750  CAR  CAR  5.0      CAR  CAR


                       FULL   LOCATION SHORT_LOCATION   NICKNAME  \
      512  Arizona Cardinals    Arizona        Arizona  Cardinals
      513    Atlanta Falcons    Atlanta        Atlanta    Falcons
```

```
514    Baltimore Ravens   Baltimore       Baltimore       Ravens
515        Buffalo Bills    Buffalo         Buffalo        Bills
516  Carolina Panthers    Carolina        Carolina      Panthers

             HYPHENATED TEAM_SEASON
512  arizona-cardinals     2018_ARI
513    atlanta-falcons     2018_ATL
514   baltimore-ravens     2018_BAL
515      buffalo-bills     2018_BUF
516  carolina-panthers     2018_CAR
```

Step 5

```
[12]: # Capitalize the first letter of each word in the hyphenated column

      teams_df['HYPHENATED'] = teams_df['HYPHENATED'].str.title()
```

```
[13]: # head the data frame to see the updated DF with the new hyphenated upper case
      ↪letter for each word

      teams_df.head()
```

```
[13]:      SEASON TEAM  NFL  NFL_TEAM_ID ESPN  PFR  PFF PFFLABEL   FO  \
      512    2018  ARI  ARI         3800  ARI  CRD  1.0      ARZ  ARI
      513    2018  ATL  ATL          200  ATL  ATL  2.0      ATL  ATL
      514    2018  BAL  BAL          325  BAL  RAV  3.0      BLT  BAL
      515    2018  BUF  BUF          610  BUF  BUF  4.0      BUF  BUF
      516    2018  CAR  CAR          750  CAR  CAR  5.0      CAR  CAR

                       FULL    LOCATION SHORT_LOCATION   NICKNAME  \
      512  Arizona Cardinals   Arizona        Arizona  Cardinals
      513   Atlanta Falcons    Atlanta        Atlanta    Falcons
      514   Baltimore Ravens   Baltimore      Baltimore    Ravens
      515       Buffalo Bills    Buffalo        Buffalo      Bills
      516  Carolina Panthers   Carolina       Carolina   Panthers

                HYPHENATED TEAM_SEASON
      512  Arizona-Cardinals    2018_ARI
      513    Atlanta-Falcons    2018_ATL
      514   Baltimore-Ravens    2018_BAL
      515      Buffalo-Bills    2018_BUF
      516  Carolina-Panthers    2018_CAR
```

0.0.4 Step 6

```python
# Save the cleaned data to a new CSV file

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

```python
# head the data frame to see what the new file will look like

teams_df.head()
```

```
     SEASON TEAM  NFL  NFL_TEAM_ID ESPN  PFR  PFF PFFLABEL   FO  \
512    2018  ARI  ARI         3800  ARI  CRD  1.0      ARZ  ARI
513    2018  ATL  ATL          200  ATL  ATL  2.0      ATL  ATL
514    2018  BAL  BAL          325  BAL  RAV  3.0      BLT  BAL
515    2018  BUF  BUF          610  BUF  BUF  4.0      BUF  BUF
516    2018  CAR  CAR          750  CAR  CAR  5.0      CAR  CAR

                  FULL    LOCATION SHORT_LOCATION   NICKNAME  \
512   Arizona Cardinals    Arizona        Arizona  Cardinals
513     Atlanta Falcons    Atlanta        Atlanta    Falcons
514    Baltimore Ravens  Baltimore      Baltimore     Ravens
515        Buffalo Bills    Buffalo        Buffalo      Bills
516   Carolina Panthers   Carolina       Carolina   Panthers

             HYPHENATED TEAM_SEASON
512  Arizona-Cardinals    2018_ARI
513    Atlanta-Falcons    2018_ATL
514   Baltimore-Ravens    2018_BAL
515       Buffalo-Bills    2018_BUF
516  Carolina-Panthers    2018_CAR
```