Task2

November 9, 2023

Task 2

Based on the historical data, write code to include results of last 5 matches between two teams

Importing python libraries

```
[11]: import pandas as pd
      import numpy as np
      import matplotlib.pyplot as plt
```

Data Explore

```
[2]: df = pd.read_csv("./train.csv")
```

```
/var/folders/r8/__hqjl2n5j5_wf6w5j15gn300000gn/T/ipykernel_5005/1547526635.py:1:
DtypeWarning: Columns (7) have mixed types. Specify dtype option on import or
set low_memory=False.
  df = pd.read_csv("./train.csv")
```

```
[3]: df.head()
```

```
[3]:         id target     home_team_name        away_team_name  \
     0  11906497   away  Newell's Old Boys           River Plate
     1  11984383   home        Real Estelí  Deportivo Las Sabanas
     2  11983301   draw              UPNFM              Marathón
     3  11983471   away               León              Morelia
     4  11883005   home     Cobán Imperial               Iztapa

                 match_date          league_name  league_id is_cup  \
     0  2019-12-01 00:45:00            Superliga        636  False
     1  2019-12-01 01:00:00     Primera Division        752  False
     2  2019-12-01 01:00:00        Liga Nacional        734  False
     3  2019-12-01 01:00:00              Liga MX        743  False
     4  2019-12-01 01:00:00        Liga Nacional        705  False

        home_team_coach_id  away_team_coach_id  … away_team_history_league_id_1  \
     0            468196.0            468200.0  …                        1122.0
     1            516788.0          22169161.0  …                         752.0
     2           2510608.0            456313.0  …                         734.0
     3           1552508.0            465797.0  …                         743.0
```

```
4                 429958.0              426870.0  …                           705.0

   away_team_history_league_id_2 away_team_history_league_id_3  \
0                          642.0                          636.0
1                          752.0                          752.0
2                          734.0                          734.0
3                          743.0                          743.0
4                          705.0                          705.0

   away_team_history_league_id_4 away_team_history_league_id_5  \
0                          636.0                          636.0
1                          752.0                          752.0
2                          734.0                          734.0
3                          743.0                          743.0
4                          705.0                          705.0

   away_team_history_league_id_6 away_team_history_league_id_7  \
0                         1122.0                          636.0
1                          752.0                          752.0
2                          734.0                          734.0
3                          743.0                          743.0
4                          705.0                          705.0

   away_team_history_league_id_8 away_team_history_league_id_9  \
0                          642.0                          636.0
1                          752.0                          752.0
2                          734.0                          734.0
3                          743.0                          746.0
4                          705.0                          705.0

   away_team_history_league_id_10
0                          1122.0
1                           752.0
2                           734.0
3                           743.0
4                           705.0

[5 rows x 190 columns]
```

[4]: `df_=df`

Merging "home_team_name" and "away_team_name" columns into "Combined" column

```python
[5]: df['home_team_name'] = df['home_team_name'].astype(str)
     df['away_team_name'] = df['away_team_name'].astype(str)
     df['Combined'] = df[['home_team_name', 'away_team_name']].agg('_'.join, axis=1)
     df.head()
```

```
[5]:          id target        home_team_name            away_team_name  \
    0  11906497   away  Newell's Old Boys              River Plate
    1  11984383   home        Real Estelí  Deportivo Las Sabanas
    2  11983301   draw              UPNFM                  Marathón
    3  11983471   away               León                   Morelia
    4  11883005   home     Cobán Imperial                    Iztapa

              match_date       league_name  league_id is_cup  \
    0  2019-12-01 00:45:00        Superliga        636  False
    1  2019-12-01 01:00:00  Primera Division        752  False
    2  2019-12-01 01:00:00     Liga Nacional        734  False
    3  2019-12-01 01:00:00          Liga MX        743  False
    4  2019-12-01 01:00:00     Liga Nacional        705  False

       home_team_coach_id  away_team_coach_id  … away_team_history_league_id_2  \
    0            468196.0            468200.0  …                         642.0
    1            516788.0          22169161.0  …                         752.0
    2           2510608.0            456313.0  …                         734.0
    3           1552508.0            465797.0  …                         743.0
    4            429958.0            426870.0  …                         705.0

      away_team_history_league_id_3 away_team_history_league_id_4  \
    0                         636.0                         636.0
    1                         752.0                         752.0
    2                         734.0                         734.0
    3                         743.0                         743.0
    4                         705.0                         705.0

      away_team_history_league_id_5 away_team_history_league_id_6  \
    0                         636.0                        1122.0
    1                         752.0                         752.0
    2                         734.0                         734.0
    3                         743.0                         743.0
    4                         705.0                         705.0

      away_team_history_league_id_7 away_team_history_league_id_8  \
    0                         636.0                         642.0
    1                         752.0                         752.0
    2                         734.0                         734.0
    3                         743.0                         743.0
    4                         705.0                         705.0

      away_team_history_league_id_9 away_team_history_league_id_10  \
    0                         636.0                        1122.0
    1                         752.0                         752.0
    2                         734.0                         734.0
    3                         746.0                         743.0
```

```
4                              705.0                              705.0

                              Combined
0          Newell's Old Boys_River Plate
1    Real Estelí_Deportivo Las Sabanas
2                       UPNFM_Marathón
3                         León_Morelia
4                 Cobán Imperial_Iztapa

[5 rows x 191 columns]
```

Check the number of the same match occurence

```
[6]: df2 = df['Combined'].value_counts()
     home_team_name_counts = df['home_team_name'].value_counts()
     df2_ = pd.DataFrame(df2)
     df2_=df2_.reset_index()
```

```
[15]: combined_count = df2_[df2_['count'] < 5].index
      df2_.drop(combined_count, inplace=True)
      df2_.head()
```

```
[15]:                           Combined  count
      0               Alajuelense_Herediano      7
      1   Beitar Tel Aviv Ramla_Hapoel Ramat Gan      6
      2                       Maribor_Bravo      6
      3                     AEK Larnaca_AEL      6
      4               Herediano_Alajuelense      6
```

Example for index 1 ("Beitar Tel Aviv Ramla_Hapoel Ramat Gan") to show the match date

```
[26]: names = df2_['Combined'][1].split("_")
      home_team_name_split = names[0]
      away_team_name_split = names[1]
      tmp = df.loc[(df['home_team_name']==home_team_name_split)
          & (df['away_team_name']==away_team_name_split), ['id', 'home_team_name',
          'away_team_name','league_name', 'match_date']]
      tmp
```

```
[26]:               id        home_team_name      away_team_name        league_name  \
      18963    11872479  Beitar Tel Aviv Ramla  Hapoel Ramat Gan       Liga Leumit
      28665    14050776  Beitar Tel Aviv Ramla  Hapoel Ramat Gan       Liga Leumit
      33628    16622919  Beitar Tel Aviv Ramla  Hapoel Ramat Gan  Club Friendlies
      92384    16899553  Beitar Tel Aviv Ramla  Hapoel Ramat Gan       Liga Leumit
      96274    17879204  Beitar Tel Aviv Ramla  Hapoel Ramat Gan        State Cup
      104860   17997728  Beitar Tel Aviv Ramla  Hapoel Ramat Gan       Liga Leumit

                    match_date
```

```
18963    2020-02-21 13:00:00
28665    2020-06-18 17:00:00
33628    2020-07-24 06:20:00
92384    2021-03-05 13:00:00
96274    2021-03-15 18:00:00
104860   2021-04-12 16:00:00
```

Running the list of match

```python
[35]: combined_names = df2_['Combined']

for i in range(len(combined_names)):
    names= combined_names[i].split("_")
    home_team_name_split = names[0]
    away_team_name_split = names[1]
    if i ==0:
        df_match=df.loc[(df['home_team_name']==home_team_name_split)
            & (df['away_team_name']==away_team_name_split), ['id',
↪'home_team_name',
            'away_team_name','league_name', 'match_date']].head(5)
    else:
        df_match = pd.concat([df_match, df.
↪loc[(df['home_team_name']==home_team_name_split)
            & (df['away_team_name']==away_team_name_split), ['id',
↪'home_team_name',
            'away_team_name','league_name', 'match_date']].head(5)],
↪ignore_index=True)
```

Saving the results into date_match.csv

```python
[37]: df_match.to_csv("./Task2_submission.csv")
```

```
[ ]:
```