

Construcción de un Dataset de los Ranking de la Liga Mexicana de Fútbol

a partir de la información de Wikipedia

https://en.wikipedia.org/wiki/2021-22_Liga_MX_season
(https://en.wikipedia.org/wiki/2021%E2%80%9322_Liga_MX_season)

Instrucciones:

1. De la página de wikipedia busque la tabla de resultados
2. Copie y pegue en una Hoja de Cálculo de Google Drive
3. Descargue en formato .CSV
4. Renombre el archivo con las siguientes características FMF_T(A|C)_(año4-año2).csv.
Ejemplo: FMF_TA_2021-22.csv
5. Copie el archivo en la carpeta /ligas/data (importante cree las carpetas antes de este paso)

```
In [14]: %matplotlib inline
%config InlineBackend.figure_format = 'retina'
```

```
In [15]: import os
import warnings
warnings.filterwarnings('ignore')

import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import penaltyblog as pb
```

```
In [16]: DATA_DIR = os.path.join(os.getcwd(), 'data/')
CHART_DIR = os.path.join(os.getcwd(), 'charts/')
```

```
In [17]: DATA_DIR
```

Out[17]: '/Users/rubenrodriguez/Documents/anaconda/premiereLeague/data/'

```
In [18]: #data_file = './data/FMF_TA_2021.csv'
data_file = './data/lmf-ac-2021-22.csv'
df = pd.read_csv(data_file, index_col=0)
df.head()
```

Out[18]:

	AMÉ	ATL	ASL	CAZ	GUA	JUÁ	LEÓ	MAZ	MON	NEC	PAC	PUE	QUE	SAI
Home \ Away														
América	—	0-2	2-3	0-0	0-0	3-0	2-0	2-0	0-0	2-1	1-3	2-0	1-1	2-
Atlas	0-1	—	1-0	0-0	1-1	2-0	2-0	1-2	2-1	2-1	0-1	0-1	2-0	2-
Atlético San Luis	0-1	2-6	—	0-0	2-2	0-1	2-0	1-0	1-1	0-2	0-2	2-1	1-1	1-
Cruz Azul	2-1	1-0	0-1	—	0-1	1-0	0-1	0-2	1-1	1-2	1-1	1-3	2-0	1-
Guadalajara	0-0	0-1	1-2	1-1	—	2-2	0-3	3-0	1-3	2-1	1-0	2-3	1-1	1-

```
In [19]: print(df.columns)

Index(['AMÉ', 'ATL', 'ASL', 'CAZ', 'GUA', 'JUÁ', 'LEÓ', 'MAZ', 'MON',
      'NEC',
      'PAC', 'PUE', 'QUE', 'SAN', 'TIJ', 'TOL', 'UNL', 'UNM'],
      dtype='object')
```

```
In [20]: x = df.columns
[]
n df.index:
c in df.columns:
if i == c: continue
score = df.loc[i, c]
if score == '-': continue
ssplit = score.split('-')
rows.append([i, c, ssplit[0], ssplit[1]])
DataFrame(rows, columns = ['home', 'away', 'home_score', 'away_score']
())
```

Out[20]:

	home	away	home_score	away_score
0	AMÉ	ATL	0	2
1	AMÉ	ASL	2	3
2	AMÉ	CAZ	0	0
3	AMÉ	GUA	0	0
4	AMÉ	JUÁ	3	0

```
In [21]: df.dtypes
```

Out[21]:

```
home          object
away          object
home_score    object
away_score    object
dtype: object
```

In [22]:

```
df.home_score = df['home_score'].astype('int')
df.away_score = df['away_score'].astype('int')
```

In [23]:

```
colley = pb.ratings.Colley(df["home_score"], df["away_score"], df["home_team"], df["away_team"])
colley.get_ratings()
```

Out[23]:

	team	rating
0	UNL	1.907895
1	AMÉ	1.907895
2	ATL	1.875
3	PAC	1.842105
4	LEÓ	1.822368
5	PUE	1.822368
6	CAZ	1.815789
7	GUA	1.815789
8	MON	1.815789
9	SAN	1.802632
10	UNM	1.75
11	ASL	1.75
12	TOL	1.75
13	MAZ	1.710526
14	QUE	1.697368
15	NEC	1.684211
16	TIJ	1.664474
17	JUÁ	1.565789

Out[24]:

	team	rating
0	AMÉ	0.631579
1	UNL	0.631579
2	ATL	0.592105
3	PAC	0.578947
4	LEÓ	0.539474
5	PUE	0.539474
6	CAZ	0.526316
7	MON	0.526316
8	GUA	0.526316
9	SAN	0.5
10	TOL	0.473684
11	UNM	0.473684
12	ASL	0.473684
13	MAZ	0.447368
14	NEC	0.447368
15	QUE	0.394737
16	TIJ	0.381579
17	JUÁ	0.315789

In []: