

# datapp

December 4, 2021

```
[1]: import os
import pandas as pd
```

```
[2]: df = pd.read_csv("english (1).csv")
```

C:\Users\spoon\Anaconda3\lib\site-packages\IPython\core\interactiveshell.py:3165: DtypeWarning: Columns (3,12) have mixed types.Specify dtype option on import or set low\_memory=False.  
has\_raised = await self.run\_ast\_nodes(code\_ast.body, cell\_name,

```
[4]: df.in
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10878 entries, 0 to 10877
Columns: 182 entries, Unnamed: 0 to Unnamed: 27
dtypes: float64(172), int64(2), object(8)
memory usage: 15.1+ MB
```

```
[38]: df.info(verbose=True, show_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10878 entries, 0 to 10877
Data columns (total 182 columns):
#   Column          Non-Null Count  Dtype
---  -
0   Unnamed: 0      10878 non-null  int64
1   Div             10181 non-null  object
2   Date            10181 non-null  object
3   Time            897 non-null    object
4   HomeTeam        10181 non-null  object
5   AwayTeam        10181 non-null  object
6   FTHG            10181 non-null  float64
7   FTAG            10181 non-null  float64
8   FTR             10181 non-null  object
9   HTHG            9257 non-null   float64
10  HTAG            9257 non-null   float64
11  HTR             9257 non-null   object
12  Referee         7357 non-null   object
13  HS              7357 non-null   float64
```

14	AS	7357 non-null	float64
15	HST	7357 non-null	float64
16	AST	7357 non-null	float64
17	HF	7357 non-null	float64
18	AF	7357 non-null	float64
19	HC	7357 non-null	float64
20	AC	7357 non-null	float64
21	HY	7357 non-null	float64
22	AY	7357 non-null	float64
23	HR	7357 non-null	float64
24	AR	7357 non-null	float64
25	B365H	6597 non-null	float64
26	B365D	6597 non-null	float64
27	B365A	6597 non-null	float64
28	BWH	6216 non-null	float64
29	BWD	6216 non-null	float64
30	BWA	6216 non-null	float64
31	IWH	7341 non-null	float64
32	IWD	7341 non-null	float64
33	IWA	7341 non-null	float64
34	PSH	3557 non-null	float64
35	PSD	3557 non-null	float64
36	PSA	3557 non-null	float64
37	WHH	7243 non-null	float64
38	WHD	7243 non-null	float64
39	WHA	7243 non-null	float64
40	VCH	6217 non-null	float64
41	VCD	6217 non-null	float64
42	VCA	6217 non-null	float64
43	MaxH	897 non-null	float64
44	MaxD	897 non-null	float64
45	MaxA	897 non-null	float64
46	AvgH	897 non-null	float64
47	AvgD	897 non-null	float64
48	AvgA	897 non-null	float64
49	B365>2.5	1148 non-null	float64
50	B365<2.5	1148 non-null	float64
51	P>2.5	897 non-null	float64
52	P<2.5	897 non-null	float64
53	Max>2.5	897 non-null	float64
54	Max<2.5	897 non-null	float64
55	Avg>2.5	897 non-null	float64
56	Avg<2.5	897 non-null	float64
57	AHh	897 non-null	float64
58	B365AHH	897 non-null	float64
59	B365AHA	897 non-null	float64
60	PAHH	897 non-null	float64
61	PAHA	897 non-null	float64

62	MaxAHH	897 non-null	float64
63	MaxAHA	897 non-null	float64
64	AvgAHH	897 non-null	float64
65	AvgAHA	897 non-null	float64
66	B365CH	897 non-null	float64
67	B365CD	897 non-null	float64
68	B365CA	897 non-null	float64
69	BWCH	897 non-null	float64
70	BWCD	897 non-null	float64
71	BWCA	897 non-null	float64
72	IWCH	897 non-null	float64
73	IWCD	897 non-null	float64
74	IWCA	897 non-null	float64
75	PSCH	3557 non-null	float64
76	PSCD	3557 non-null	float64
77	PSCA	3557 non-null	float64
78	WHCH	897 non-null	float64
79	WHCD	897 non-null	float64
80	WHCA	897 non-null	float64
81	VCCH	897 non-null	float64
82	VCCD	897 non-null	float64
83	VCCA	897 non-null	float64
84	MaxCH	897 non-null	float64
85	MaxCD	897 non-null	float64
86	MaxCA	897 non-null	float64
87	AvgCH	897 non-null	float64
88	AvgCD	897 non-null	float64
89	AvgCA	897 non-null	float64
90	B365C>2.5	897 non-null	float64
91	B365C<2.5	897 non-null	float64
92	PC>2.5	897 non-null	float64
93	PC<2.5	897 non-null	float64
94	MaxC>2.5	897 non-null	float64
95	MaxC<2.5	897 non-null	float64
96	AvgC>2.5	897 non-null	float64
97	AvgC<2.5	897 non-null	float64
98	AHCh	897 non-null	float64
99	B365CAHH	897 non-null	float64
100	B365CAHA	897 non-null	float64
101	PCAHH	897 non-null	float64
102	PCAHA	897 non-null	float64
103	MaxCAHH	897 non-null	float64
104	MaxCAHA	897 non-null	float64
105	AvgCAHH	897 non-null	float64
106	AvgCAHA	897 non-null	float64
107	season	10878 non-null	int64
108	Bb1X2	5320 non-null	float64
109	BbMxH	5320 non-null	float64

110	BbAvH	5320 non-null	float64
111	BbMxD	5320 non-null	float64
112	BbAvD	5320 non-null	float64
113	BbMxA	5320 non-null	float64
114	BbAvA	5320 non-null	float64
115	BbOU	5320 non-null	float64
116	BbMx>2.5	5320 non-null	float64
117	BbAv>2.5	5320 non-null	float64
118	BbMx<2.5	5320 non-null	float64
119	BbAv<2.5	5320 non-null	float64
120	BbAH	5310 non-null	float64
121	BbAHh	5310 non-null	float64
122	BbMxAHH	5310 non-null	float64
123	BbAvAHH	5310 non-null	float64
124	BbMxAHA	5310 non-null	float64
125	BbAvAHA	5310 non-null	float64
126	LBH	6019 non-null	float64
127	LBD	6019 non-null	float64
128	LBA	6019 non-null	float64
129	SJH	3459 non-null	float64
130	SJD	3459 non-null	float64
131	SJA	3459 non-null	float64
132	GBH	4096 non-null	float64
133	GBD	4096 non-null	float64
134	GBA	4096 non-null	float64
135	BSH	2280 non-null	float64
136	BSD	2280 non-null	float64
137	BSA	2280 non-null	float64
138	SBH	3744 non-null	float64
139	SBD	3744 non-null	float64
140	SBA	3744 non-null	float64
141	SOH	380 non-null	float64
142	SOD	380 non-null	float64
143	SOA	380 non-null	float64
144	GB>2.5	379 non-null	float64
145	GB<2.5	379 non-null	float64
146	Unnamed: 48	0 non-null	float64
147	Unnamed: 49	0 non-null	float64
148	Unnamed: 50	0 non-null	float64
149	Unnamed: 51	0 non-null	float64
150	Unnamed: 52	0 non-null	float64
151	Attendance	759 non-null	float64
152	HHW	760 non-null	float64
153	AHW	760 non-null	float64
154	HO	760 non-null	float64
155	AO	760 non-null	float64
156	HBP	760 non-null	float64
157	ABP	760 non-null	float64

```

158 SYH          373 non-null    float64
159 SYD          373 non-null    float64
160 SYA          373 non-null    float64
161 Unnamed: 10  0 non-null      float64
162 Unnamed: 11  0 non-null      float64
163 Unnamed: 12  0 non-null      float64
164 Unnamed: 13  0 non-null      float64
165 Unnamed: 14  0 non-null      float64
166 Unnamed: 15  0 non-null      float64
167 Unnamed: 16  0 non-null      float64
168 Unnamed: 17  0 non-null      float64
169 Unnamed: 18  0 non-null      float64
170 Unnamed: 19  0 non-null      float64
171 Unnamed: 20  0 non-null      float64
172 Unnamed: 21  0 non-null      float64
173 Unnamed: 7   0 non-null      float64
174 Unnamed: 8   0 non-null      float64
175 Unnamed: 9   0 non-null      float64
176 Unnamed: 22  0 non-null      float64
177 Unnamed: 23  0 non-null      float64
178 Unnamed: 24  0 non-null      float64
179 Unnamed: 25  0 non-null      float64
180 Unnamed: 26  0 non-null      float64
181 Unnamed: 27  0 non-null      float64
dtypes: float64(172), int64(2), object(8)
memory usage: 15.1+ MB

```

```
[5]: df2 = df.dropna(subset=["B365H", "B365D", "B365A", "BWH", "BWD", "BWA", "IWH",
↪ "IWD", "IWA"])
```

```
[10]: df.groupby("season").count()
```

```
[10]:
```

	Unnamed: 0	Div	Date	Time	HomeTeam	AwayTeam	FTHG	FTAG	FTR	\
season										
1	380	380	380	0	380	380	380	380	380	
102	380	380	380	0	380	380	380	380	380	
203	380	380	380	0	380	380	380	380	380	
506	380	380	380	0	380	380	380	380	380	
607	380	380	380	0	380	380	380	380	380	
708	380	380	380	0	380	380	380	380	380	
809	380	380	380	0	380	380	380	380	380	
910	380	380	380	0	380	380	380	380	380	
1011	380	380	380	0	380	380	380	380	380	
1112	380	380	380	0	380	380	380	380	380	
1213	380	380	380	0	380	380	380	380	380	
1314	380	380	380	0	380	380	380	380	380	
1415	381	380	380	0	380	380	380	380	380	

1516	380	380	380	0	380	380	380	380	380
1617	380	380	380	0	380	380	380	380	380
1718	380	380	380	0	380	380	380	380	380
1819	380	380	380	0	380	380	380	380	380
1920	380	380	380	380	380	380	380	380	380
2021	380	380	380	380	380	380	380	380	380
2122	137	137	137	137	137	137	137	137	137
9394	552	462	462	0	462	462	462	462	462
9495	552	462	462	0	462	462	462	462	462
9596	552	380	380	0	380	380	380	380	380
9697	552	380	380	0	380	380	380	380	380
9798	380	380	380	0	380	380	380	380	380
9899	380	380	380	0	380	380	380	380	380
9900	552	380	380	0	380	380	380	380	380

	HTHG	...	Unnamed: 21	Unnamed: 7	Unnamed: 8	Unnamed: 9	\
season		...					
1	380	...	0	0	0	0	
102	380	...	0	0	0	0	
203	380	...	0	0	0	0	
506	380	...	0	0	0	0	
607	380	...	0	0	0	0	
708	380	...	0	0	0	0	
809	380	...	0	0	0	0	
910	380	...	0	0	0	0	
1011	380	...	0	0	0	0	
1112	380	...	0	0	0	0	
1213	380	...	0	0	0	0	
1314	380	...	0	0	0	0	
1415	380	...	0	0	0	0	
1516	380	...	0	0	0	0	
1617	380	...	0	0	0	0	
1718	380	...	0	0	0	0	
1819	380	...	0	0	0	0	
1920	380	...	0	0	0	0	
2021	380	...	0	0	0	0	
2122	137	...	0	0	0	0	
9394	0	...	0	0	0	0	
9495	0	...	0	0	0	0	
9596	380	...	0	0	0	0	
9697	380	...	0	0	0	0	
9798	380	...	0	0	0	0	
9899	380	...	0	0	0	0	
9900	380	...	0	0	0	0	

	Unnamed: 22	Unnamed: 23	Unnamed: 24	Unnamed: 25	Unnamed: 26	\
season						

1	0	0	0	0	0
102	0	0	0	0	0
203	0	0	0	0	0
506	0	0	0	0	0
607	0	0	0	0	0
708	0	0	0	0	0
809	0	0	0	0	0
910	0	0	0	0	0
1011	0	0	0	0	0
1112	0	0	0	0	0
1213	0	0	0	0	0
1314	0	0	0	0	0
1415	0	0	0	0	0
1516	0	0	0	0	0
1617	0	0	0	0	0
1718	0	0	0	0	0
1819	0	0	0	0	0
1920	0	0	0	0	0
2021	0	0	0	0	0
2122	0	0	0	0	0
9394	0	0	0	0	0
9495	0	0	0	0	0
9596	0	0	0	0	0
9697	0	0	0	0	0
9798	0	0	0	0	0
9899	0	0	0	0	0
9900	0	0	0	0	0

Unnamed: 27

season

1	0
102	0
203	0
506	0
607	0
708	0
809	0
910	0
1011	0
1112	0
1213	0
1314	0
1415	0
1516	0
1617	0
1718	0
1819	0

1920	0
2021	0
2122	0
9394	0
9495	0
9596	0
9697	0
9798	0
9899	0
9900	0

[27 rows x 181 columns]

```
[9]: df2.info(verbose=True, show_counts=True)
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 6214 entries, 0 to 6217
Data columns (total 182 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Unnamed: 0   6214 non-null   int64
1   Div          6214 non-null   object
2   Date         6214 non-null   object
3   Time         897 non-null    object
4   HomeTeam     6214 non-null   object
5   AwayTeam     6214 non-null   object
6   FTHG         6214 non-null   float64
7   FTAG         6214 non-null   float64
8   FTR          6214 non-null   object
9   HTHG         6214 non-null   float64
10  HTAG         6214 non-null   float64
11  HTR          6214 non-null   object
12  Referee      6214 non-null   object
13  HS           6214 non-null   float64
14  AS           6214 non-null   float64
15  HST          6214 non-null   float64
16  AST          6214 non-null   float64
17  HF           6214 non-null   float64
18  AF           6214 non-null   float64
19  HC           6214 non-null   float64
20  AC           6214 non-null   float64
21  HY           6214 non-null   float64
22  AY           6214 non-null   float64
23  HR           6214 non-null   float64
24  AR           6214 non-null   float64
25  B365H        6214 non-null   float64
26  B365D        6214 non-null   float64
27  B365A        6214 non-null   float64
```



28	BWH	6214 non-null	float64
29	BWD	6214 non-null	float64
30	BWA	6214 non-null	float64
31	IWH	6214 non-null	float64
32	IWD	6214 non-null	float64
33	IWA	6214 non-null	float64
34	PSH	3555 non-null	float64
35	PSD	3555 non-null	float64
36	PSA	3555 non-null	float64
37	WHH	6107 non-null	float64
38	WHD	6107 non-null	float64
39	WHA	6107 non-null	float64
40	VCH	6214 non-null	float64
41	VCD	6214 non-null	float64
42	VCA	6214 non-null	float64
43	MaxH	897 non-null	float64
44	MaxD	897 non-null	float64
45	MaxA	897 non-null	float64
46	AvgH	897 non-null	float64
47	AvgD	897 non-null	float64
48	AvgA	897 non-null	float64
49	B365>2.5	897 non-null	float64
50	B365<2.5	897 non-null	float64
51	P>2.5	897 non-null	float64
52	P<2.5	897 non-null	float64
53	Max>2.5	897 non-null	float64
54	Max<2.5	897 non-null	float64
55	Avg>2.5	897 non-null	float64
56	Avg<2.5	897 non-null	float64
57	AHh	897 non-null	float64
58	B365AHH	897 non-null	float64
59	B365AHA	897 non-null	float64
60	PAHH	897 non-null	float64
61	PAHA	897 non-null	float64
62	MaxAHH	897 non-null	float64
63	MaxAHA	897 non-null	float64
64	AvgAHH	897 non-null	float64
65	AvgAHA	897 non-null	float64
66	B365CH	897 non-null	float64
67	B365CD	897 non-null	float64
68	B365CA	897 non-null	float64
69	BWCH	897 non-null	float64
70	BWCD	897 non-null	float64
71	BWCA	897 non-null	float64
72	IWCH	897 non-null	float64
73	IWCD	897 non-null	float64
74	IWCA	897 non-null	float64
75	PSCH	3555 non-null	float64

76	PSCD	3555 non-null	float64
77	PSCA	3555 non-null	float64
78	WHCH	897 non-null	float64
79	WHCD	897 non-null	float64
80	WHCA	897 non-null	float64
81	VCCH	897 non-null	float64
82	VCCD	897 non-null	float64
83	VCCA	897 non-null	float64
84	MaxCH	897 non-null	float64
85	MaxCD	897 non-null	float64
86	MaxCA	897 non-null	float64
87	AvgCH	897 non-null	float64
88	AvgCD	897 non-null	float64
89	AvgCA	897 non-null	float64
90	B365C>2.5	897 non-null	float64
91	B365C<2.5	897 non-null	float64
92	PC>2.5	897 non-null	float64
93	PC<2.5	897 non-null	float64
94	MaxC>2.5	897 non-null	float64
95	MaxC<2.5	897 non-null	float64
96	AvgC>2.5	897 non-null	float64
97	AvgC<2.5	897 non-null	float64
98	AHCh	897 non-null	float64
99	B365CAHH	897 non-null	float64
100	B365CAHA	897 non-null	float64
101	PCAAH	897 non-null	float64
102	PCAAH	897 non-null	float64
103	MaxCAHH	897 non-null	float64
104	MaxCAHA	897 non-null	float64
105	AvgCAHH	897 non-null	float64
106	AvgCAHA	897 non-null	float64
107	season	6214 non-null	int64
108	Bb1X2	5317 non-null	float64
109	BbMxH	5317 non-null	float64
110	BbAvH	5317 non-null	float64
111	BbMxD	5317 non-null	float64
112	BbAvD	5317 non-null	float64
113	BbMxA	5317 non-null	float64
114	BbAvA	5317 non-null	float64
115	BbOU	5317 non-null	float64
116	BbMx>2.5	5317 non-null	float64
117	BbAv>2.5	5317 non-null	float64
118	BbMx<2.5	5317 non-null	float64
119	BbAv<2.5	5317 non-null	float64
120	BbAH	5307 non-null	float64
121	BbAHh	5307 non-null	float64
122	BbMxAHH	5307 non-null	float64
123	BbAvAHH	5307 non-null	float64

124	BbMxAHA	5307 non-null	float64
125	BbAvAHA	5307 non-null	float64
126	LBH	4936 non-null	float64
127	LBD	4936 non-null	float64
128	LBA	4936 non-null	float64
129	SJH	3458 non-null	float64
130	SJD	3458 non-null	float64
131	SJA	3458 non-null	float64
132	GBH	3038 non-null	float64
133	GBD	3038 non-null	float64
134	GBA	3038 non-null	float64
135	BSH	2279 non-null	float64
136	BSD	2279 non-null	float64
137	BSA	2279 non-null	float64
138	SBH	2659 non-null	float64
139	SBD	2659 non-null	float64
140	SBA	2659 non-null	float64
141	SOH	0 non-null	float64
142	SOD	0 non-null	float64
143	SOA	0 non-null	float64
144	GB>2.5	0 non-null	float64
145	GB<2.5	0 non-null	float64
146	Unnamed: 48	0 non-null	float64
147	Unnamed: 49	0 non-null	float64
148	Unnamed: 50	0 non-null	float64
149	Unnamed: 51	0 non-null	float64
150	Unnamed: 52	0 non-null	float64
151	Attendance	0 non-null	float64
152	HHW	0 non-null	float64
153	AHW	0 non-null	float64
154	HO	0 non-null	float64
155	AO	0 non-null	float64
156	HBP	0 non-null	float64
157	ABP	0 non-null	float64
158	SYH	0 non-null	float64
159	SYD	0 non-null	float64
160	SYA	0 non-null	float64
161	Unnamed: 10	0 non-null	float64
162	Unnamed: 11	0 non-null	float64
163	Unnamed: 12	0 non-null	float64
164	Unnamed: 13	0 non-null	float64
165	Unnamed: 14	0 non-null	float64
166	Unnamed: 15	0 non-null	float64
167	Unnamed: 16	0 non-null	float64
168	Unnamed: 17	0 non-null	float64
169	Unnamed: 18	0 non-null	float64
170	Unnamed: 19	0 non-null	float64
171	Unnamed: 20	0 non-null	float64

```

172 Unnamed: 21  0 non-null    float64
173 Unnamed: 7   0 non-null    float64
174 Unnamed: 8   0 non-null    float64
175 Unnamed: 9   0 non-null    float64
176 Unnamed: 22  0 non-null    float64
177 Unnamed: 23  0 non-null    float64
178 Unnamed: 24  0 non-null    float64
179 Unnamed: 25  0 non-null    float64
180 Unnamed: 26  0 non-null    float64
181 Unnamed: 27  0 non-null    float64
dtypes: float64(172), int64(2), object(8)
memory usage: 8.7+ MB

```

```

[151]: from datetime import datetime, timedelta
date_parser = lambda x: datetime.strptime(x, "%Y-%m-%d")

df3 = pd.read_csv("processed_data.csv", date_parser=date_parser,
↳ parse_dates=["Date"])

```

```

[61]: len(df3.groupby("HomeTeam").count())

```

```

[61]: 41

```

```

[58]: columns_req = ['Date', 'HomeTeam', 'AwayTeam', 'FTHG', 'FTAG', 'FTR']

df4 = df3[columns_req]
df4.info()

```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6214 entries, 0 to 6213
Data columns (total 6 columns):
#   Column      Non-Null Count  Dtype
---  -
0   Date        6214 non-null   datetime64[ns]
1   HomeTeam    6214 non-null   object
2   AwayTeam    6214 non-null   object
3   FTHG        6214 non-null   int64
4   FTAG        6214 non-null   int64
5   FTR         6214 non-null   object
dtypes: datetime64[ns](1), int64(2), object(3)
memory usage: 291.4+ KB

```

```

[98]: def find_time_interval(time_series):
      return abs(min(time_series) - max(time_series)) / 7

```

```

[101]: min(df4["Date"])

```

```

[101]: Timestamp('2005-08-13 00:00:00')

```

```
[122]: import math
```

```
[266]: mins = abs(min(df4["Date"]) - max(df4["Date"])).total_seconds() / 60
hour = mins / 60
days = hour / 24
weeks = days / 7
math.ceil(weeks)
```

```
[266]: 851
```

```
[267]: dummy_teams = df4.HomeTeam.unique()
```

```
[268]: dummy_teams_dict = dict()

for i in range(len(dummy_teams)):
    dummy_teams_dict[dummy_teams[i]] = [i for i in range(math.ceil(weeks) + 1)]

dummy_df = pd.DataFrame(dummy_teams_dict)

df5 = dummy_df.T

for col in df5.columns:
    df5[col].values[:] = 0

dates = list(dummy_df.T.columns)
initial_date_str = min(df4.Date).strftime("%Y-%m-%d").split('-')
initial = datetime(int(initial_date_str[0]),
                   int(initial_date_str[1]),
                   int(initial_date_str[2]))# date

dates_t = []

for i in range(len(dates)):
    delta = i * timedelta(days = 7)
    dates_t.append(initial + delta)
```

```
[269]: df5.columns = dates_t
df5
```

```
[269]:
```

	2005-08-13	2005-08-20	2005-08-27	2005-09-03	2005-09-10	\
Aston Villa	0	0	0	0	0	
Everton	0	0	0	0	0	
Fulham	0	0	0	0	0	
Man City	0	0	0	0	0	
Middlesbrough	0	0	0	0	0	
Portsmouth	0	0	0	0	0	
Sunderland	0	0	0	0	0	
West Ham	0	0	0	0	0	

Arsenal	0	0	0	0	0
Wigan	0	0	0	0	0
Birmingham	0	0	0	0	0
Blackburn	0	0	0	0	0
Charlton	0	0	0	0	0
Liverpool	0	0	0	0	0
Man United	0	0	0	0	0
Newcastle	0	0	0	0	0
Tottenham	0	0	0	0	0
West Brom	0	0	0	0	0
Bolton	0	0	0	0	0
Chelsea	0	0	0	0	0
Reading	0	0	0	0	0
Sheffield United	0	0	0	0	0
Watford	0	0	0	0	0
Derby	0	0	0	0	0
Hull	0	0	0	0	0
Stoke	0	0	0	0	0
Wolves	0	0	0	0	0
Burnley	0	0	0	0	0
Blackpool	0	0	0	0	0
QPR	0	0	0	0	0
Swansea	0	0	0	0	0
Norwich	0	0	0	0	0
Southampton	0	0	0	0	0
Crystal Palace	0	0	0	0	0
Cardiff	0	0	0	0	0
Leicester	0	0	0	0	0
Bournemouth	0	0	0	0	0
Brighton	0	0	0	0	0
Huddersfield	0	0	0	0	0
Leeds	0	0	0	0	0
Brentford	0	0	0	0	0

	2005-09-17	2005-09-24	2005-10-01	2005-10-08	2005-10-15 \
Aston Villa	0	0	0	0	0
Everton	0	0	0	0	0
Fulham	0	0	0	0	0
Man City	0	0	0	0	0
Middlesbrough	0	0	0	0	0
Portsmouth	0	0	0	0	0
Sunderland	0	0	0	0	0
West Ham	0	0	0	0	0
Arsenal	0	0	0	0	0
Wigan	0	0	0	0	0
Birmingham	0	0	0	0	0
Blackburn	0	0	0	0	0

Charlton	0	0	0	0	0
Liverpool	0	0	0	0	0
Man United	0	0	0	0	0
Newcastle	0	0	0	0	0
Tottenham	0	0	0	0	0
West Brom	0	0	0	0	0
Bolton	0	0	0	0	0
Chelsea	0	0	0	0	0
Reading	0	0	0	0	0
Sheffield United	0	0	0	0	0
Watford	0	0	0	0	0
Derby	0	0	0	0	0
Hull	0	0	0	0	0
Stoke	0	0	0	0	0
Wolves	0	0	0	0	0
Burnley	0	0	0	0	0
Blackpool	0	0	0	0	0
QPR	0	0	0	0	0
Swansea	0	0	0	0	0
Norwich	0	0	0	0	0
Southampton	0	0	0	0	0
Crystal Palace	0	0	0	0	0
Cardiff	0	0	0	0	0
Leicester	0	0	0	0	0
Bournemouth	0	0	0	0	0
Brighton	0	0	0	0	0
Huddersfield	0	0	0	0	0
Leeds	0	0	0	0	0
Brentford	0	0	0	0	0

	...	2021-10-02	2021-10-09	2021-10-16	2021-10-23	\
Aston Villa	...	0	0	0	0	
Everton	...	0	0	0	0	
Fulham	...	0	0	0	0	
Man City	...	0	0	0	0	
Middlesbrough	...	0	0	0	0	
Portsmouth	...	0	0	0	0	
Sunderland	...	0	0	0	0	
West Ham	...	0	0	0	0	
Arsenal	...	0	0	0	0	
Wigan	...	0	0	0	0	
Birmingham	...	0	0	0	0	
Blackburn	...	0	0	0	0	
Charlton	...	0	0	0	0	
Liverpool	...	0	0	0	0	
Man United	...	0	0	0	0	
Newcastle	...	0	0	0	0	

Tottenham	...	0	0	0	0
West Brom	...	0	0	0	0
Bolton	...	0	0	0	0
Chelsea	...	0	0	0	0
Reading	...	0	0	0	0
Sheffield United	...	0	0	0	0
Watford	...	0	0	0	0
Derby	...	0	0	0	0
Hull	...	0	0	0	0
Stoke	...	0	0	0	0
Wolves	...	0	0	0	0
Burnley	...	0	0	0	0
Blackpool	...	0	0	0	0
QPR	...	0	0	0	0
Swansea	...	0	0	0	0
Norwich	...	0	0	0	0
Southampton	...	0	0	0	0
Crystal Palace	...	0	0	0	0
Cardiff	...	0	0	0	0
Leicester	...	0	0	0	0
Bournemouth	...	0	0	0	0
Brighton	...	0	0	0	0
Huddersfield	...	0	0	0	0
Leeds	...	0	0	0	0
Brentford	...	0	0	0	0

	2021-10-30	2021-11-06	2021-11-13	2021-11-20	2021-11-27 \
Aston Villa	0	0	0	0	0
Everton	0	0	0	0	0
Fulham	0	0	0	0	0
Man City	0	0	0	0	0
Middlesbrough	0	0	0	0	0
Portsmouth	0	0	0	0	0
Sunderland	0	0	0	0	0
West Ham	0	0	0	0	0
Arsenal	0	0	0	0	0
Wigan	0	0	0	0	0
Birmingham	0	0	0	0	0
Blackburn	0	0	0	0	0
Charlton	0	0	0	0	0
Liverpool	0	0	0	0	0
Man United	0	0	0	0	0
Newcastle	0	0	0	0	0
Tottenham	0	0	0	0	0
West Brom	0	0	0	0	0
Bolton	0	0	0	0	0
Chelsea	0	0	0	0	0



Reading	0	0	0	0	0
Sheffield United	0	0	0	0	0
Watford	0	0	0	0	0
Derby	0	0	0	0	0
Hull	0	0	0	0	0
Stoke	0	0	0	0	0
Wolves	0	0	0	0	0
Burnley	0	0	0	0	0
Blackpool	0	0	0	0	0
QPR	0	0	0	0	0
Swansea	0	0	0	0	0
Norwich	0	0	0	0	0
Southampton	0	0	0	0	0
Crystal Palace	0	0	0	0	0
Cardiff	0	0	0	0	0
Leicester	0	0	0	0	0
Bournemouth	0	0	0	0	0
Brighton	0	0	0	0	0
Huddersfield	0	0	0	0	0
Leeds	0	0	0	0	0
Brentford	0	0	0	0	0

2021-12-04

Aston Villa	0
Everton	0
Fulham	0
Man City	0
Middlesbrough	0
Portsmouth	0
Sunderland	0
West Ham	0
Arsenal	0
Wigan	0
Birmingham	0
Blackburn	0
Charlton	0
Liverpool	0
Man United	0
Newcastle	0
Tottenham	0
West Brom	0
Bolton	0
Chelsea	0
Reading	0
Sheffield United	0
Watford	0
Derby	0

Hull	0
Stoke	0
Wolves	0
Burnley	0
Blackpool	0
QPR	0
Swansea	0
Norwich	0
Southampton	0
Crystal Palace	0
Cardiff	0
Leicester	0
Bournemouth	0
Brighton	0
Huddersfield	0
Leeds	0
Brentford	0

[41 rows x 852 columns]

```
[292]: from tqdm import tqdm

def find_time_interval(date, series):
    for time_i in series:
        if date <= time_i:
            return time_i.strftime("%Y-%m-%d")

def find_team_idx(team, dfIndex):
    return list(dfIndex.index(team))

for i in tqdm(range(len(dummy_teams))):
    home_teams = df4.loc[(df4["HomeTeam"] == dummy_teams[i])].
    ↪reset_index(drop=True)
    away_teams = df4.loc[(df4["AwayTeam"] == dummy_teams[i])].
    ↪reset_index(drop=True)

    scores_per_matchweek = dict()

    for idx, game in home_teams.iterrows():
        time_i = find_time_interval(game["Date"], df5.columns)
        team_idx = find_team_idx(game["HomeTeam"], df5.index)
        goals_scored = game["FTHG"]
        df5.iloc[team_idx][time_i] += goals_scored

    for idx, game in away_teams.iterrows():
        time_i = find_time_interval(game["Date"], df5.columns)
        team_idx = find_team_idx(game["AwayTeam"], df5.index)
```

```
goals_scored = game["FTAG"]
df5.iloc[team_idx][time_i] += goals_scored
```

```
100%|
| 41/41 [00:14<00:00, 2.90it/s]
```

```
[293]: df5
```

```
[293]:
```

	2005-08-13	2005-08-20	2005-08-27	2005-09-03	2005-09-10	\
Aston Villa	12	0	9	0	0	
Everton	0	0	2	0	0	
Fulham	0	2	6	0	2	
Man City	0	4	12	0	2	
Middlesbrough	0	0	6	0	8	
Portsmouth	0	2	6	0	2	
Sunderland	4	0	4	0	0	
West Ham	12	0	4	0	0	
Arsenal	0	8	16	0	2	
Wigan	0	0	4	0	4	
Birmingham	0	4	6	0	0	
Blackburn	2	8	0	0	0	
Charlton	6	4	0	6	2	
Liverpool	0	4	0	0	0	
Man United	4	4	0	4	4	
Newcastle	0	0	0	0	4	
Tottenham	4	8	0	0	0	
West Brom	0	8	8	0	4	
Bolton	4	0	12	0	0	
Chelsea	0	2	24	0	8	
Reading	0	0	0	0	0	
Sheffield United	0	0	0	0	0	
Watford	0	0	0	0	0	
Derby	0	0	0	0	0	
Hull	0	0	0	0	0	
Stoke	0	0	0	0	0	
Wolves	0	0	0	0	0	
Burnley	0	0	0	0	0	
Blackpool	0	0	0	0	0	
QPR	0	0	0	0	0	
Swansea	0	0	0	0	0	
Norwich	0	0	0	0	0	
Southampton	0	0	0	0	0	
Crystal Palace	0	0	0	0	0	
Cardiff	0	0	0	0	0	
Leicester	0	0	0	0	0	
Bournemouth	0	0	0	0	0	
Brighton	0	0	0	0	0	

Huddersfield	0	0	0	0	0
Leeds	0	0	0	0	0
Brentford	0	0	0	0	0
	2005-09-17	2005-09-24	2005-10-01	2005-10-08	2005-10-15 \
Aston Villa	7	2	0	12	0
Everton	0	0	0	0	0
Fulham	4	0	8	0	0
Man City	0	0	0	8	0
Middlesbrough	0	2	0	6	4
Portsmouth	4	0	0	0	2
Sunderland	4	0	8	0	4
West Ham	20	0	2	0	0
Arsenal	0	8	0	4	2
Wigan	0	6	0	8	4
Birmingham	2	8	0	0	0
Blackburn	0	4	8	0	0
Charlton	0	4	8	0	0
Liverpool	0	4	0	4	4
Man United	0	4	6	0	6
Newcastle	0	10	0	0	0
Tottenham	2	0	10	0	8
West Brom	2	4	0	0	8
Bolton	0	6	0	2	2
Chelsea	4	8	0	8	20
Reading	0	0	0	0	0
Sheffield United	0	0	0	0	0
Watford	0	0	0	0	0
Derby	0	0	0	0	0
Hull	0	0	0	0	0
Stoke	0	0	0	0	0
Wolves	0	0	0	0	0
Burnley	0	0	0	0	0
Blackpool	0	0	0	0	0
QPR	0	0	0	0	0
Swansea	0	0	0	0	0
Norwich	0	0	0	0	0
Southampton	0	0	0	0	0
Crystal Palace	0	0	0	0	0
Cardiff	0	0	0	0	0
Leicester	0	0	0	0	0
Bournemouth	0	0	0	0	0
Brighton	0	0	0	0	0
Huddersfield	0	0	0	0	0
Leeds	0	0	0	0	0
Brentford	0	0	0	0	0

	...	2021-10-02	2021-10-09	2021-10-16	2021-10-23	\
Aston Villa	...	0	2	12	2	
Everton	...	2	0	0	8	
Fulham	...	0	0	0	0	
Man City	...	0	4	8	8	
Middlesbrough	...	0	0	0	0	
Portsmouth	...	0	0	0	0	
Sunderland	...	0	0	0	0	
West Ham	...	0	4	0	2	
Arsenal	...	12	0	0	20	
Wigan	...	0	0	0	0	
Birmingham	...	0	0	0	0	
Blackburn	...	0	0	0	0	
Charlton	...	0	0	0	0	
Liverpool	...	0	8	10	0	
Man United	...	4	0	4	0	
Newcastle	...	2	0	0	10	
Tottenham	...	2	8	0	6	
West Brom	...	0	0	0	0	
Bolton	...	0	0	0	0	
Chelsea	...	9	0	2	21	
Reading	...	0	0	0	0	
Sheffield United	...	0	0	0	0	
Watford	...	0	0	0	10	
Derby	...	0	0	0	0	
Hull	...	0	0	0	0	
Stoke	...	0	0	0	0	
Wolves	...	8	0	6	2	
Burnley	...	0	0	0	4	
Blackpool	...	0	0	0	0	
QPR	...	0	0	0	0	
Swansea	...	0	0	0	0	
Norwich	...	0	0	0	0	
Southampton	...	2	0	3	6	
Crystal Palace	...	3	6	0	7	
Cardiff	...	0	0	0	0	
Leicester	...	0	4	12	0	
Bournemouth	...	0	0	0	0	
Brighton	...	1	0	0	2	
Huddersfield	...	0	0	0	0	
Leeds	...	2	0	0	2	
Brentford	...	0	2	0	0	
		2021-10-30	2021-11-06	2021-11-13	2021-11-20	2021-11-27 \
Aston Villa		0	7	0	12	4
Everton		0	2	0	0	0
Fulham		0	0	0	0	0

Man City	0	4	0	0	12
Middlesbrough	0	0	0	0	0
Portsmouth	0	0	0	0	0
Sunderland	0	0	0	0	0
West Ham	4	8	12	0	0
Arsenal	4	0	4	0	8
Wigan	0	0	0	0	0
Birmingham	0	0	0	0	0
Blackburn	0	0	0	0	0
Charlton	0	0	0	0	0
Liverpool	18	0	4	16	16
Man United	6	0	0	2	0
Newcastle	0	2	0	12	0
Tottenham	0	0	0	0	8
West Brom	0	0	0	0	0
Bolton	0	0	0	0	0
Chelsea	6	3	0	6	0
Reading	0	0	0	0	0
Sheffield United	0	0	0	0	0
Watford	0	0	0	12	0
Derby	0	0	0	0	0
Hull	0	0	0	0	0
Stoke	0	0	0	0	0
Wolves	0	6	0	3	0
Burnley	9	2	0	9	0
Blackpool	0	0	0	0	0
QPR	0	0	0	0	0
Swansea	0	0	0	0	0
Norwich	0	7	0	6	0
Southampton	2	3	0	2	0
Crystal Palace	4	6	0	6	3
Cardiff	0	0	0	0	0
Leicester	4	0	2	0	0
Bournemouth	0	0	0	0	0
Brighton	2	2	0	0	0
Huddersfield	0	0	0	0	0
Leeds	0	2	2	0	1
Brentford	3	2	0	3	0

2021-12-04

Aston Villa	7
Everton	4
Fulham	0
Man City	12
Middlesbrough	0
Portsmouth	0
Sunderland	0

West Ham	6
Arsenal	0
Wigan	0
Birmingham	0
Blackburn	0
Charlton	0
Liverpool	8
Man United	2
Newcastle	4
Tottenham	0
West Brom	0
Bolton	0
Chelsea	7
Reading	0
Sheffield United	0
Watford	7
Derby	0
Hull	0
Stoke	0
Wolves	0
Burnley	0
Blackpool	0
QPR	0
Swansea	0
Norwich	2
Southampton	6
Crystal Palace	0
Cardiff	0
Leicester	16
Bournemouth	0
Brighton	1
Huddersfield	0
Leeds	2
Brentford	2

[41 rows x 852 columns]

[ ]: