scrape football matches data from [https://www.football-data.co.uk/data.php" (https://www.football-data.co.uk/data.php")](https://www.football-data.co.uk/data.php)

In [47]:
```python
import pandas as pd
from bs4 import BeautifulSoup as bs
import requests
import re
```

In [48]:
```python
url="https://www.football-data.co.uk/data.php"
soup=bs(requests.get(url).content,"html.parser")
```

In [49]:
```python
data=pd.DataFrame()
links=soup.select("td[valign='top'] table[cellspacing='2'] tr td:nth-child(2) a")
for link in links:
    if link["href"]!="belgiumm.php":

        final_url="https://www.football-data.co.uk/"+link["href"]
        soup=bs(requests.get(final_url).content,"html.parser")
        csvs=soup.select("td[valign='top'] a[href]")
        for csv in csvs:
            if "csv" in csv["href"]:
                df=pd.read_csv("https://www.football-data.co.uk/"+csv["href"],on_bad_lines='skip',encoding_errors="ignore")
                df.insert(1,"Country",link["href"].split(".")[0][:-1])
                df.insert(2,"League",csv.text)
                data=pd.concat([data,df],ignore_index=True)

    else:
        break
```

In [55]:
```python
pd.set_option("display.max_columns",100)
```

In [56]:
```python
df=data.iloc[:,1:15]
```

In [57]:
```python
df.dropna(subset=["HomeTeam"],inplace=True)
```

In [58]:
```python
df=df.reset_index()
df=df.iloc[:,df.columns.isin(["Time","index"])==False]
```

In [60]:
```python
df
```

Out[60]:

	Country	League	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	Referee	HS
0	england	Premier League	13/08/2021	Brentford	Arsenal	2.0	0.0	H	1.0	0.0	H	M Oliver	8.0
1	england	Premier League	14/08/2021	Man United	Leeds	5.0	1.0	H	1.0	0.0	H	P Tierney	16.0
2	england	Premier League	14/08/2021	Burnley	Brighton	1.0	2.0	A	1.0	0.0	H	D Coote	14.0
3	england	Premier League	14/08/2021	Chelsea	Crystal Palace	3.0	0.0	H	2.0	0.0	H	J Moss	13.0
4	england	Premier League	14/08/2021	Everton	Southampton	3.0	1.0	H	0.0	1.0	A	A Madley	14.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
173108	netherlands	Eredivisie	08/05/94	NAC Breda	Utrecht	2.0	2.0	D	NaN	NaN	NaN	NaN	NaN
173109	netherlands	Eredivisie	08/05/94	Twente	Cambuur	4.0	1.0	H	NaN	NaN	NaN	NaN	NaN
173110	netherlands	Eredivisie	08/05/94	Volendam	PSV Eindhoven	3.0	2.0	H	NaN	NaN	NaN	NaN	NaN
173111	netherlands	Eredivisie	08/05/94	VVV Venlo	Waalwijk	0.0	2.0	A	NaN	NaN	NaN	NaN	NaN
173112	netherlands	Eredivisie	08/05/94	Willem II	Ajax	2.0	1.0	H	NaN	NaN	NaN	NaN	NaN

173113 rows × 13 columns

Create connection

In [61]:
```python
import mysql.connector

mydb=mysql.connector.connect(host="localhost",user="root",password="gigaberosql",database="international")
mycursor=mydb.cursor()
```

Create table

In [62]:
```python
columns1 = ' varchar(255), '.join([str(elem) for elem in df.columns])
create_table="create table club_matches(ID int auto_increment primary key, "+columns1+" varchar(255))"

mycursor.execute("drop table if exists club_matches")
mycursor.execute(create_table)
```

Insert data into table ¶

In [63]:
```python
columns2 = ', '.join([str(elem) for elem in df.columns])
values=len(df.columns)*"%s,"
insert_into="insert into club_matches("+columns2+")"+" values("+values[:-1]+")"
columns3=[]
for i in df.columns:
    i="row."+str(i)
    columns3.append(i)
```

In [68]:
```python
import numpy as np

df2 = df.replace({np.nan: None})

#df2 = df.where(pd.notnull(df), None)


for row in df2.itertuples():
    mycursor.execute(insert_into,(row.Country, row.League, row.Date, row.HomeTeam, row.AwayTeam, row.FTHG, row.FTAG, row.FTR, row.HTHG, row.HTAG, row.HTR, row.Referee, row.HS))
    mydb.commit()
```

In [65]:
```python
print(*columns3,sep=", ")
```

row.Country, row.League, row.Date, row.HomeTeam, row.AwayTeam, row.FTHG, row.FTAG, row.FTR, row.HTHG, row.HTAG, row.HTR, row.Referee, row.HS

select everything from club_matches

In [72]: `pd.read_sql("select * from club_matches",mydb)`

Out[72]:

	ID	Country	League	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	Referee	HS
0	1	england	Premier League	13/08/2021	Brentford	Arsenal	2.0	0.0	H	1.0	0.0	H	M Oliver	8.0
1	2	england	Premier League	14/08/2021	Man United	Leeds	5.0	1.0	H	1.0	0.0	H	P Tierney	16.0
2	3	england	Premier League	14/08/2021	Burnley	Brighton	1.0	2.0	A	1.0	0.0	H	D Coote	14.0
3	4	england	Premier League	14/08/2021	Chelsea	Crystal Palace	3.0	0.0	H	2.0	0.0	H	J Moss	13.0
4	5	england	Premier League	14/08/2021	Everton	Southampton	3.0	1.0	H	0.0	1.0	A	A Madley	14.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
173108	173109	netherlands	Eredivisie	08/05/94	NAC Breda	Utrecht	2.0	2.0	D	None	None	None	None	None
173109	173110	netherlands	Eredivisie	08/05/94	Twente	Cambuur	4.0	1.0	H	None	None	None	None	None
173110	173111	netherlands	Eredivisie	08/05/94	Volendam	PSV Eindhoven	3.0	2.0	H	None	None	None	None	None
173111	173112	netherlands	Eredivisie	08/05/94	VVV Venlo	Waalwijk	0.0	2.0	A	None	None	None	None	None
173112	173113	netherlands	Eredivisie	08/05/94	Willem II	Ajax	2.0	1.0	H	None	None	None	None	None

173113 rows × 14 columns

In [ ]: