Import Section

In this section the dataset and libraries are loaded.

In [1]:
```
import numpy as np
import pandas as pd
import os
import os.path
```

""" Extract all csv files from the dataset """

In [2]:
```
all_dirs = []
for root, dirs, files in os.walk("espana-master"):
    for file in files:
        if file.endswith(".csv"):
            # print(os.path.join(root, file))
            new_dir = os.path.join(root, file)
            all_dirs.append(new_dir)

print('Number of total CSV files: ', len(all_dirs))
```

```
Number of total CSV files:  115
```

''' Merge all CSV files to a single CSV file named: total_table.csv '''

In [3]:
```
fname = 'total_table.csv'  # A single file to save all the records

# Check if file exists, delete it
if os.path.isfile(fname):
    os.remove(fname)

# Add all files to fname csv file
fout = open(fname, "a")
# first file:
for line in open(all_dirs[0]):
    fout.write(line)
# now the rest:
for num in range(len(all_dirs) - 1):
    with open(all_dirs[num + 1], "r+") as f:
        f.readline()  # skip the header
        for line in f:
            fout.write(line)
        f.close()  # not really needed
fout.close()
```

''' Extract data from csv file '''

In [4]:
```
df = pd.read_csv('total_table.csv')
```

""" Delete unused rows """

In [5]:
```
df = df.drop(['Date', 'HT', 'Round'], 1)
print(df)
print('Done 1')
```

```
                    Team 1   FT                    Team 2
0          Arenas Club (1)  2-3       Atlético Madrid (1)
1          RCD Español (1)  3-2       Real Unión Club (1)
2        Real Sociedad (1)  1-1  Athletic Club Bilbao (1)
3       Real Madrid CF (1)  5-0             CE Europa (1)
4         R. Racing C. (1)  0-2          FC Barcelona (1)
...                    ...  ...                      ...
35417   Sporting Gijón (3)  2-0     Albacete Balompié (3)
35418   Rayo Vallecano (3)  3-1         RCD La Coruña (3)
35419        Girona FC (3)  1-0             Málaga CF (3)
35420  SD Ponferradina (3)  4-0           CD Tenerife (3)
35421         Cádiz CF (3)  2-1        Extremadura UD (3)

[35422 rows x 3 columns]
Done 1
```

""" Uniquing the team names """, Because team names contains some numbers like: FC Barcelona (43)

In [6]:
```
df['Team 1'] = df['Team 1'].apply(lambda x: x[0:x.find('(') - 1])
df['Team 2'] = df['Team 2'].apply(lambda x: x[0:x.find('(') - 1])
print(df)
```

```
                Team 1   FT                Team 2
0          Arenas Club  2-3       Atlético Madrid
1          RCD Español  3-2       Real Unión Club
2        Real Sociedad  1-1  Athletic Club Bilbao
3       Real Madrid CF  5-0             CE Europa
4         R. Racing C.  0-2          FC Barcelona
...                ...  ...                   ...
35417   Sporting Gijón  2-0     Albacete Balompié
35418   Rayo Vallecano  3-1         RCD La Coruña
35419        Girona FC  1-0             Málaga CF
35420  SD Ponferradina  4-0           CD Tenerife
35421         Cádiz CF  2-1        Extremadura UD

[35422 rows x 3 columns]
```

""" Create goal difference column """

In [7]:
```
df['Diff point'] = df['FT'].apply(lambda x: int(x[0:x.find('-')]) - int(x[x.fi
nd('-') + 1:len(x)]))
print(df)
print('Done 2')
```

```
                Team 1   FT                Team 2  Diff point
0          Arenas Club  2-3       Atlético Madrid          -1
1          RCD Español  3-2       Real Unión Club           1
2        Real Sociedad  1-1  Athletic Club Bilbao           0
3       Real Madrid CF  5-0             CE Europa           5
4         R. Racing C.  0-2          FC Barcelona          -2
...                ...  ...                   ...         ...
35417   Sporting Gijón  2-0     Albacete Balompié           2
35418   Rayo Vallecano  3-1         RCD La Coruña           2
35419        Girona FC  1-0             Málaga CF           1
35420  SD Ponferradina  4-0           CD Tenerife           4
35421         Cádiz CF  2-1        Extremadura UD           1

[35422 rows x 4 columns]
Done 2
```

""" All teams to Data frame """

In [8]:
```
teams_names = df['Team 1'].unique()
print('Number of teams', len(teams_names))
final_df = pd.DataFrame(teams_names, columns=['Teams'])
final_df['Points'] = float(0)
print(final_df)
teams_names_dict = dict.fromkeys(teams_names, 0)
print(teams_names_dict)
# df.columns = ['Teams', 'points']
# final_df['points'] = df.apply(lambda row: final_df[0].iloc[row['Team 1']] is
3)
print(' ')
print('Working on Final df. PLEASE WAIT ...')
for index in df.iterrows():
    diff_point = index[1]['Diff point']
    team_name_1 = index[1]['Team 1']
    team_name_2 = index[1]['Team 2']
    teams_names_dict[team_name_1] += 1
    teams_names_dict[team_name_2] += 1
    if diff_point > 0:
        a = final_df.index[final_df['Teams'] == team_name_1].tolist()[0]
        final_df.at[a, 'Points'] += 3
    elif diff_point < 0:
        a = final_df.index[final_df['Teams'] == team_name_2].tolist()[0]
        final_df.at[a, 'Points'] += 3
    else:
        a = final_df.index[final_df['Teams'] == team_name_2].tolist()[0]
        final_df.at[a, 'Points'] += 1
        a = final_df.index[final_df['Teams'] == team_name_1].tolist()[0]
        final_df.at[a, 'Points'] += 1
print('Final df finished')
```

```
Number of teams 101
                    Teams  Points
0             Arenas Club     0.0
1             RCD Español     0.0
2          Real Sociedad      0.0
3         Real Madrid CF      0.0
4           R. Racing C.      0.0
..                   ...      ...
96              CF Reus       0.0
97           UCAM Murcia      0.0
98        Extremadura UD      0.0
99   CF Rayo Majadahonda      0.0
100       CF Fuenlabrada      0.0

[101 rows x 2 columns]
{'Arenas Club': 0, 'RCD Español': 0, 'Real Sociedad': 0, 'Real Madrid CF': 0,
'R. Racing C.': 0, 'Athletic Club Bilbao': 0, 'Atlético Madrid': 0, 'FC Barce
lona': 0, 'CE Europa': 0, 'Real Unión Club': 0, 'Deportivo Alavés': 0, 'Valen
cia CF': 0, 'Real Betis': 0, 'Real Oviedo': 0, 'Sevilla FC': 0, 'Hércules C
F': 0, 'CA Osasuna': 0, 'Real Zaragoza': 0, 'RC Celta Vigo': 0, 'Real Murcia
CF': 0, 'RCD La Coruña': 0, 'Granada CF': 0, 'CD Castellón': 0, 'CE Sabadel
l': 0, 'Real Sporting': 0, 'CD Alcoyano': 0, 'Gimnàstic Tarragona': 0, 'Real
Valladolid CF': 0, 'Málaga CF': 0, 'UE Lleida': 0, 'UD Las Palmas': 0, 'Atéti
co Tetuán': 0, 'R. Jaén': 0, 'Cultural Leonesa': 0, 'CD Condal': 0, 'Elche C
F': 0, 'RCD Mallorca': 0, 'CD Tenerife': 0, 'Córdoba CF': 0, 'Pontevedra CF':
0, 'Levante UD': 0, 'Burgos CF': 0, 'UD Salamanca': 0, 'Cádiz CF': 0, 'Rayo V
allecano': 0, 'Recreativo Huelva': 0, 'UD Almería': 0, 'CD Logroñés': 0, 'Rea
l Burgos CF': 0, 'Albacete Balompié': 0, 'Real Racing Santander': 0, 'Real Ma
drid': 0, 'Sporting Gijón': 0, 'SD Compostela': 0, 'RCD Espanyol': 0, 'CP Mér
ida': 0, 'CF Extremadura': 0, 'CD Leganés': 0, 'SD Eibar': 0, 'Real Madrid
B': 0, 'CD Toledo': 0, 'Atlético Madrid B': 0, 'FC Barcelona B': 0, 'CD Badaj
oz': 0, 'Écija Balompié': 0, 'CD Ourense': 0, 'Villarreal CF': 0, 'CD Numanci
a': 0, 'Real Jaén CF': 0, 'Xerez CD': 0, 'RCD Mallorca B': 0, 'Getafe CF': 0,
'Universidad Las Palmas': 0, 'Racing Ferrol': 0, 'Polideportivo Ejido': 0, 'T
errassa FC': 0, 'CF Ciudad de Murcia': 0, 'Málaga CF B': 0, 'Algeciras CF':
0, 'Lorca FC': 0, 'UD Vecindario': 0, 'SD Ponferradina': 0, 'Sevilla FC B':
0, 'Granada 74': 0, 'Alicante CF': 0, 'SD Huesca': 0, 'Girona FC': 0, 'Real U
nión': 0, 'Villarreal CF B': 0, 'FC Cartagena': 0, 'AD Alcorcón': 0, 'CD Guad
alajara': 0, 'CD Mirandés': 0, 'CD Lugo': 0, 'UE Llagostera': 0, 'Athletic Cl
ub Bilbao B': 0, 'CF Reus': 0, 'UCAM Murcia': 0, 'Extremadura UD': 0, 'CF Ray
o Majadahonda': 0, 'CF Fuenlabrada': 0}

Working on Final df. PLEASE WAIT ...
Final df finished
```

''' Sort values by Points '''

In [9]:
```
final_df = final_df.sort_values(['Points'], ascending=False)
print('Sort values by Points')
print(final_df)
print(teams_names_dict)
```

```
Sort values by Points
                      Teams  Points
7               FC Barcelona  5445.0
6            Atlético Madrid  4587.0
5       Athletic Club Bilbao  4325.0
11               Valencia CF  4270.0
14                Sevilla FC  3800.0
..                       ...     ...
95   Athletic Club Bilbao B    32.0
64            Écija Balompié    30.0
34                 CD Condal    29.0
31            Atético Tetuán    26.0
100           CF Fuenlabrada     9.0

[101 rows x 2 columns]
{'Arenas Club': 130, 'RCD Español': 1744, 'Real Sociedad': 2507, 'Real Madrid
CF': 1812, 'R. Racing C.': 734, 'Athletic Club Bilbao': 2811, 'Atlético Madri
d': 2747, 'FC Barcelona': 2811, 'CE Europa': 54, 'Real Unión Club': 72, 'Depo
rtivo Alavés': 875, 'Valencia CF': 2713, 'Real Betis': 1945, 'Real Oviedo': 1
417, 'Sevilla FC': 2583, 'Hércules CF': 1048, 'CA Osasuna': 1621, 'Real Zarag
oza': 2295, 'RC Celta Vigo': 2029, 'Real Murcia CF': 1048, 'RCD La Coruña': 1
667, 'Granada CF': 871, 'CD Castellón': 544, 'CE Sabadell': 594, 'Real Sporti
ng': 1032, 'CD Alcoyano': 150, 'Gimnàstic Tarragona': 620, 'Real Valladolid C
F': 1855, 'Málaga CF': 1434, 'UE Lleida': 358, 'UD Las Palmas': 1805, 'Atétic
o Tetuán': 30, 'R. Jaén': 90, 'Cultural Leonesa': 72, 'CD Condal': 30, 'Elche
CF': 1407, 'RCD Mallorca': 1209, 'CD Tenerife': 1169, 'Córdoba CF': 966, 'Pon
tevedra CF': 222, 'Levante UD': 939, 'Burgos CF': 246, 'UD Salamanca': 924,
'Cádiz CF': 785, 'Rayo Vallecano': 1019, 'Recreativo Huelva': 732, 'UD Almerí
a': 798, 'CD Logroñés': 472, 'Real Burgos CF': 114, 'Albacete Balompié': 983,
'Real Racing Santander': 821, 'Real Madrid': 999, 'Sporting Gijón': 1059, 'SD
Compostela': 328, 'RCD Espanyol': 961, 'CP Mérida': 202, 'CF Extremadura': 24
8, 'CD Leganés': 533, 'SD Eibar': 735, 'Real Madrid B': 206, 'CD Toledo': 16
4, 'Atlético Madrid B': 164, 'FC Barcelona B': 332, 'CD Badajoz': 290, 'Écija
Balompié': 38, 'CD Ourense': 122, 'Villarreal CF': 889, 'CD Numancia': 911,
'Real Jaén CF': 168, 'Xerez CD': 542, 'RCD Mallorca B': 42, 'Getafe CF': 745,
'Universidad Las Palmas': 42, 'Racing Ferrol': 252, 'Polideportivo Ejido': 29
4, 'Terrassa FC': 126, 'CF Ciudad de Murcia': 168, 'Málaga CF B': 126, 'Algec
iras CF': 42, 'Lorca FC': 126, 'UD Vecindario': 42, 'SD Ponferradina': 255,
'Sevilla FC B': 168, 'Granada 74': 42, 'Alicante CF': 42, 'SD Huesca': 377,
'Girona FC': 457, 'Real Unión': 42, 'Villarreal CF B': 126, 'FC Cartagena': 1
26, 'AD Alcorcón': 381, 'CD Guadalajara': 84, 'CD Mirandés': 213, 'CD Lugo':
297, 'UE Llagostera': 84, 'Athletic Club Bilbao B': 42, 'CF Reus': 126, 'UCAM
Murcia': 42, 'Extremadura UD': 45, 'CF Rayo Majadahonda': 42, 'CF Fuenlabrad
a': 3}
```

''' Sort values by Points ratio '''

In [10]:
```
print('Sort values by Points ratio')
final_df['Points ratio'] = final_df['Points']
for key, value in teams_names_dict.items():
    a = final_df.index[final_df['Teams'] == key].tolist()[0]
    final_df.at[a, 'Points ratio'] /= value
final_df = final_df.sort_values(['Points ratio'], ascending=False)
print(final_df)
```

```
Sort values by Points ratio
                      Teams  Points  Points ratio
100          CF Fuenlabrada     9.0      3.000000
51              Real Madrid  2064.0      2.066066
7             FC Barcelona  5445.0      1.937033
3            Real Madrid CF  3506.0      1.934879
6           Atlético Madrid  4587.0      1.669822
..                      ...     ...           ...
84              Alicante CF    35.0      0.833333
80            UD Vecindario    34.0      0.809524
64           Écija Balompié    30.0      0.789474
78             Algeciras CF    33.0      0.785714
95   Athletic Club Bilbao B    32.0      0.761905

[101 rows x 3 columns]
```

In [11]:
```
print('hi')
```

```
hi
```

In [ ]:
```

```