

```
In [1]: import numpy as np
import pandas as pd
from datetime import datetime
```

```
In [2]: print('Numpy version :',np.__version__)
print('Pandas Version :',pd.__version__)
```

Numpy version : 1.16.4
Pandas Version : 0.23.4

```
In [3]: df = pd.read_csv(r"C:\Users\HP\Desktop\Datasets\Fifa.csv",
encoding='UTF-8', sep=',', skiprows=0, index_col=False)
data = pd.DataFrame(df)
df.head()
```

Out[3]:

	Number	ID	Name	Age	Photo	Nationality	Flag	Over
0	0	158023	L. Messi	31	https://cdn.sofifa.org/players/4/19/158023.png	Argentina	https://cdn.sofifa.org/flags/52.png	
1	1	20801	Cristiano Ronaldo	33	https://cdn.sofifa.org/players/4/19/20801.png	Portugal	https://cdn.sofifa.org/flags/38.png	
2	2	190871	Neymar Jr	26	https://cdn.sofifa.org/players/4/19/190871.png	Brazil	https://cdn.sofifa.org/flags/54.png	
3	3	193080	De Gea	27	https://cdn.sofifa.org/players/4/19/193080.png	Spain	https://cdn.sofifa.org/flags/45.png	
4	4	192985	K. De Bruyne	27	https://cdn.sofifa.org/players/4/19/192985.png	Belgium	https://cdn.sofifa.org/flags/7.png	

5 rows × 89 columns

```
In [4]: #Copying Data Frame
df1=df.copy()
```

```
In [5]: #Removing Unecessary Column
df1.drop(['Joined','Contract Valid Until',
"Photo","Flag","Club Logo","Special","ID","Preferred Foot','Real Face',
'Jersey Number','Loaned From','LS','ST','RS','LW','LF','CF',
'RF','RW','LAM','CAM','RAM','LM', 'LB','LCB','CB','RCB',
'RB','LCM','CM','RCM','RM','LNB','LDM','CDM','RDM','RWB',
'StandingTackle','SlidingTackle','GKDivng','GKHandling','GKKicking','GKPositioning','Numb
er','Release_Clause' ],
axis=1, inplace=True)
```

```
In [6]: ## Converting to numeric Format

def convert_to_numeric(df_value):
    try:
        value = float(df_value[1:-1])
        suffix = df_value[-1:]
        if suffix == 'M':
            value = value * 1000000
        elif suffix == 'K':
            value = value * 1000
        except ValueError:
            value = 0
        return value
```

```
In [7]: ## Applying The Function

df1['Value'] = df1['Value'].apply(convert_to_numeric)
df1['Wage'] = df1['Wage'].apply(convert_to_numeric)
df1.Value = df1.Value.replace(0, np.nan)
df1.Wage = df1.Wage.replace(0, np.nan)
df1.head(3)
```

Out[7]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	International Reputation	Weak Foot	...	Strength	LongShr
0	L. Messi	31	Argentina	94	94	FC Barcelona	110500000.0	565000.0	5.0	4.0	...	59.0	94.0
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000000.0	405000.0	5.0	4.0	...	79.0	93.0
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500000.0	290000.0	5.0	5.0	...	49.0	88.0

3 rows × 44 columns

```
In [8]: ##Separating Work Rate
df1['Work_Rate'] = df1['Work_Rate'].astype(str)
df1['Work_Rate'] = df1['Work_Rate'].str.split('/')
df1['Attacking_rates']=df1['Work_Rate'].str.get(0)
df1['Attacking_rates'] = df1['Attacking_rates'].str.strip()
df1['Defensive_rates'] = df1['Work_Rate'].str.get(1)
df1['Defensive_rates'] = df1['Defensive_rates'].str.strip()
```

In []:

```
In [9]: df1.dtypes
df1.head(10)
```

Out[9]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	International Reputation	Weak Foot	...	Aggression	Inte
0	L. Messi	31	Argentina	94	94	FC Barcelona	110500000.0	565000.0	5.0	4.0	...	48.0	
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000000.0	405000.0	5.0	4.0	...	63.0	
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500000.0	290000.0	5.0	5.0	...	56.0	
3	De Gea	27	Spain	91	93	Manchester United	72000000.0	260000.0	4.0	3.0	...	38.0	
4	K. De Bruyne	27	Belgium	91	92	Manchester City	102000000.0	355000.0	4.0	5.0	...	76.0	
5	E. Hazard	27	Belgium	91	91	Chelsea	93000000.0	340000.0	4.0	4.0	...	54.0	
6	L. Modrić	32	Croatia	91	91	Real Madrid	67000000.0	420000.0	4.0	4.0	...	62.0	
7	L. Suárez	31	Uruguay	91	91	FC Barcelona	80000000.0	455000.0	5.0	4.0	...	87.0	
8	Sergio Ramos	32	Spain	91	91	Real Madrid	51000000.0	380000.0	4.0	3.0	...	88.0	
9	J. Oblak	25	Slovenia	90	93	Atlético Madrid	68000000.0	94000.0	3.0	3.0	...	34.0	

10 rows × 46 columns

```
In [10]: #Fixing Body Type
df1.loc[df['Body Type'] == "Lean", 'Body Type'] = 1
df1.loc[df['Body Type'] == "Normal", 'Body Type'] = 2
df1.loc[df['Body Type'] == "Stocky", 'Body Type'] = 3
df1.loc[df['Body Type'] == "Messi", 'Body Type'] = 2
df1.loc[df['Body Type'] == "C. Ronaldo", 'Body Type'] = 2
df1.loc[df['Body Type'] == "Neymar", 'Body Type'] = 1
df1.loc[df['Body Type'] == "Courtois", 'Body Type'] = 2
df1.loc[df['Body Type'] == "Shaqiri", 'Body Type'] = 3
df1.loc[df['Body Type'] == "Akinfenwa", 'Body Type'] = 3
df1.loc[df['Body Type'] == "PLAYER_BODY_TYPE_25", 'Body Type'] = np.nan
```

```
In [11]: ##Dropping Body Type that have "nan" values and changing Body Type column into int
df1.dropna(subset=['Body Type'], inplace=True)
df1['Body Type'].astype('int64')
```

Out[11]:

0	2
1	2
2	1
3	1
4	2
5	2
6	1
7	2
8	2
9	2
10	2
11	2
12	1
13	2
14	1
15	2
16	2
17	1
18	2
19	2
20	1
21	1
22	2
23	3
24	2
25	1
27	2
28	2
29	2
30	2
...	...
18177	1
18178	1
18179	1
18180	1
18181	1
18182	1
18183	2
18184	2
18185	2
18186	2
18187	1
18188	1
18189	1
18190	2
18191	1
18192	1
18193	1
18194	2
18195	2
18196	1
18197	2
18198	1
18199	1
18200	1
18201	1
18202	1
18203	2
18204	2
18205	1
18206	1

Name: Body Type, Length: 18158, dtype: int64

```
In [12]: df1.head(5)
```

Out[12]:

	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	International Reputation	Weak Foot	...	Aggression	Inte
0	L. Messi	31	Argentina	94	94	FC Barcelona	110500000.0	565000.0	5.0	4.0	...	48.0	
1	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000000.0	405000.0	5.0	4.0	...	63.0	
2	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500000.0	290000.0	5.0	5.0	...	56.0	
3	De Gea	27	Spain	91	93	Manchester United	72000000.0	260000.0	4.0	3.0	...	38.0	
4	K. De Bruyne	27	Belgium	91	92	Manchester City	102000000.0	355000.0	4.0	5.0	...	76.0	

5 rows × 46 columns

In []:

```
In [14]: #Make Variable and column name lower case and export in into clean csv
df1.columns = map(str.lower, df1.columns)
df1.to_csv(r"C:\Users\HP\Desktop\Datasets\Fifa_cleaned.csv")
```