```
In [48]: # checking for incorrect datatype
         fifa_Data.applymap(lambda x: isinstance(x,(int,float,str)))
```

Out[48]:

	ID	Name	LongName	photoUrl	playerUrl	Nationality	Age	↓OVA	POT	Club	...	A/W	D/W	IR	PAC	SHO	PAS	DRI	DEF	PHY	Hit
0	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
1	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
2	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
3	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
4	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	.
18974	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
18975	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
18976	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
18977	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru
18978	True	True	True	True	True	True	True	True	True	True	...	True	True	True	True	True	True	True	True	True	Tru

18979 rows × 77 columns

```
In [50]: # Correcting wrong datatype
         fifa_Data['Joined'] = pd.to_datetime(fifa_Data['Joined'])

         #notice that the dtype of 'Joined' which was initially object has change to datetime
```

```
In [45]: fifa_Data.info()

         <class 'pandas.core.frame.DataFrame'>
         RangeIndex: 18979 entries, 0 to 18978
         Data columns (total 77 columns):
          #   Column           Non-Null Count  Dtype
         ---  ------           --------------  -----
          0   ID               18979 non-null  int64
          1   Name             18979 non-null  object
          2   LongName         18979 non-null  object
          3   photoUrl         18979 non-null  object
          4   playerUrl        18979 non-null  object
          5   Nationality      18979 non-null  object
          6   Age              18979 non-null  int64
          7   ↓OVA             18979 non-null  int64
          8   POT              18979 non-null  int64
          9   Club             18979 non-null  object
          10  Contract         18979 non-null  object
          11  Positions        18979 non-null  object
          12  Height           18979 non-null  object
          13  Weight           18979 non-null  object
          14  Preferred Foot   18979 non-null  object
          15  BOV              18979 non-null  int64
          16  Best Position    18979 non-null  object
          17  Joined           18979 non-null  datetime64[ns]
          18  Loan Date End    18979 non-null  object
          19  Value            18979 non-null  object
          20  Wage             18979 non-null  object
          21  Release Clause   18979 non-null  object
          22  Attacking        18979 non-null  int64
          23  Crossing         18979 non-null  int64
          24  Finishing        18979 non-null  int64
          25  Heading Accuracy 18979 non-null  int64
          26  Short Passing    18979 non-null  int64
```
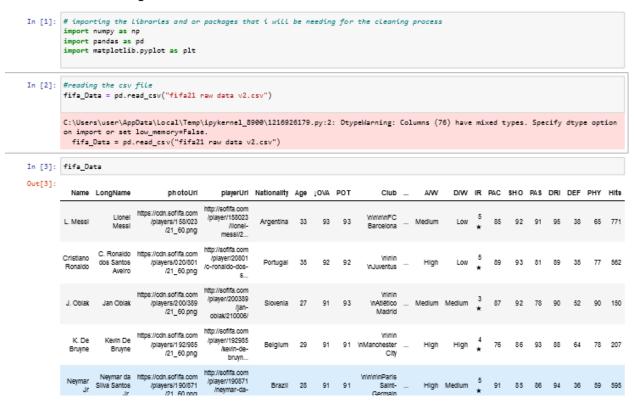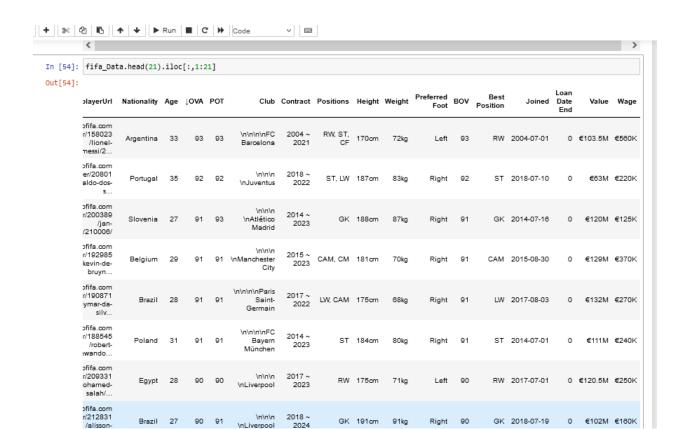
In [36]: `fifa_Data.iloc[:,11:20]`

Out[36]:

	Positions	Height	Weight	Preferred Foot	BOV	Best Position	Joined	Loan Date End	Value
0	RW, ST, CF	170cm	72kg	Left	93	RW	Jul 1, 2004	0	€103.5M
1	ST, LW	187cm	83kg	Right	92	ST	Jul 10, 2018	0	€63M
2	GK	188cm	87kg	Right	91	GK	Jul 16, 2014	0	€120M
3	CAM, CM	181cm	70kg	Right	91	CAM	Aug 30, 2015	0	€129M
4	LW, CAM	175cm	68kg	Right	91	LW	Aug 3, 2017	0	€132M
...	...	...	...	...	...	...	...	...	...
18974	CB	178cm	66kg	Right	49	CB	Jul 13, 2018	0	€100K
18975	CM	175cm	65kg	Right	51	CAM	Aug 1, 2020	0	€130K
18976	CM	179cm	74kg	Right	49	CAM	Mar 8, 2019	0	€120K
18977	RW	175cm	69kg	Right	48	ST	Sep 22, 2020	0	€100K
18978	CB, LB	188cm	75kg	Left	50	LB	Jul 29, 2019	0	€100K

18979 rows × 9 columns

In [37]: `fifa_Data.iloc[:,21:30]`

Out[37]:

	Release Clause	Attacking	Crossing	Finishing	Heading Accuracy	Short Passing	Volleys	Skill	Dribbling
0	€138.4M	429	85	95	70	91	88	470	96
1	€75.9M	437	84	95	90	82	86	414	88
2	€159.4M	95	13	11	15	43	13	109	12
3	€161M	407	94	82	55	94	82	441	88

```
10   Contract          18979 non-null  object
11   Positions         18979 non-null  object
12   Height            18979 non-null  object
13   Weight            18979 non-null  object
14   Preferred Foot    18979 non-null  object
15   BOV               18979 non-null  int64
16   Best Position     18979 non-null  object
17   Joined            18979 non-null  object
18   Loan Date End     18979 non-null  object
19   Value             18979 non-null  object
20   Wage              18979 non-null  object
21   Release Clause    18979 non-null  object
22   Attacking         18979 non-null  int64
23   Crossing          18979 non-null  int64
24   Finishing         18979 non-null  int64
25   Heading Accuracy  18979 non-null  int64
26   Short Passing     18979 non-null  int64
27   Volleys           18979 non-null  int64
28   Skill             18979 non-null  int64
29   Dribbling         18979 non-null  int64
30   Curve             18979 non-null  int64
31   FK Accuracy       18979 non-null  int64
32   Long Passing      18979 non-null  int64
33   Ball Control      18979 non-null  int64
34   Movement          18979 non-null  int64
35   Acceleration      18979 non-null  int64
36   Sprint Speed      18979 non-null  int64
37   Agility           18979 non-null  int64
38   Reactions         18979 non-null  int64
39   Balance           18979 non-null  int64
40   Power             18979 non-null  int64
41   Shot Power        18979 non-null  int64
42   Jumping           18979 non-null  int64
43   Stamina           18979 non-null  int64
44   Strength          18979 non-null  int64
45   Long Shots        18979 non-null  int64
46   Mentality         18979 non-null  int64
47   Aggression        18979 non-null  int64
48   Interceptions     18979 non-null  int64
```

In [34]: `fifa_Data.columns`

Out[34]:
```
Index(['ID', 'Name', 'LongName', 'photoUrl', 'playerUrl', 'Nationality', 'Age',
       '↓OVA', 'POT', 'Club', 'Contract', 'Positions', 'Height', 'Weight',
       'Preferred Foot', 'BOV', 'Best Position', 'Joined', 'Loan Date End',
       'Value', 'Wage', 'Release Clause', 'Attacking', 'Crossing', 'Finishing',
       'Heading Accuracy', 'Short Passing', 'Volleys', 'Skill', 'Dribbling',
       'Curve', 'FK Accuracy', 'Long Passing', 'Ball Control', 'Movement',
       'Acceleration', 'Sprint Speed', 'Agility', 'Reactions', 'Balance',
       'Power', 'Shot Power', 'Jumping', 'Stamina', 'Strength', 'Long Shots',
       'Mentality', 'Aggression', 'Interceptions', 'Positioning', 'Vision',
       'Penalties', 'Composure', 'Defending', 'Marking', 'Standing Tackle',
       'Sliding Tackle', 'Goalkeeping', 'GK Diving', 'GK Handling',
       'GK Kicking', 'GK Positioning', 'GK Reflexes', 'Total Stats',
       'Base Stats', 'W/F', 'SM', 'A/W', 'D/W', 'IR', 'PAC', 'SHO', 'PAS',
       'DRI', 'DEF', 'PHY', 'Hits'],
      dtype='object')
```

In [35]: `fifa_Data.info()`

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18979 entries, 0 to 18978
Data columns (total 77 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   ID              18979 non-null  int64
 1   Name            18979 non-null  object
 2   LongName        18979 non-null  object
 3   photoUrl        18979 non-null  object
 4   playerUrl       18979 non-null  object
 5   Nationality     18979 non-null  object
 6   Age             18979 non-null  int64
 7   ↓OVA            18979 non-null  int64
 8   POT             18979 non-null  int64
 9   Club            18979 non-null  object
 10  Contract        18979 non-null  object
 11  Positions       18979 non-null  object
 12  Height          18979 non-null  object
 13  Weight          18979 non-null  object
 14  Preferred Foot  18979 non-null  object
```

```python
In [25]:   # Filling the missing values
           fifa_Data.fillna(value = 0, inplace = True)
```

```python
In [26]:   fifa_Data.isnull().sum()
           # notice that 'Hit' has 0 null or missing values now
```

```
Out[26]:   ID           0
           Name         0
           LongName     0
           photoUrl     0
           playerUrl    0
                       ..
           PAS          0
           DRI          0
           DEF          0
           PHY          0
           Hits         0
           Length: 77, dtype: int64
```

```python
In [28]:   #checking for duplicate values
           fifa_Data.duplicated().sum()
```

```
Out[28]:   0
```

```python
In [29]:   # There are no duplicate values
```

```python
In [31]:   # ALL datatypes are correct
```

```python
In [10]:   # Looking out for null values
           fifa_Data.isnull()
```

Out[10]:

	ID	Name	LongName	photoUrl	playerUrl	Nationality	Age	↓OVA	POT	Club	...	A/W	D/W	IR	PAC	SHO	PAS	DRI	DEF	PHY	Hits
0	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
1	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
2	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
3	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
4	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
18974	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
18975	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
18976	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
18977	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
18978	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True

18979 rows × 77 columns

```python
In [ ]:    # only Hits has null values among the columns.The total null values can be seeing below
```

```python
In [8]:    fifa_Data.isnull().sum()
```

```
Out[8]:    ID           0
           Name         0
           LongName     0
           photoUrl     0
           playerUrl    0
                       ...
           PAS          0
           DRI          0
           DEF          0
           PHY          0
           Hits      2595
           Length: 77, dtype: int64
```

```
In [10]:  # Looking out for null values
          fifa_Data.isnull()
```

Out[10]:

	ID	Name	LongName	photoUrl	playerUrl	Nationality	Age	↓OVA	POT	Club	...	A/W	D/W	IR	PAC	SHO	PAS	DRI	DEF	PHY	Hits
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	False
	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True
	False	False	False	False	False	False	False	False	False	False	...	False	False	False	False	False	False	False	False	False	True

vs × 77 columns

```
In [ ]:  # only Hits has null values among the columns.The total null values can be seeing below
```

```
In [8]:  fifa_Data.isnull().sum()
```

Out[8]:
```
ID           0
Name         0
LongName     0
photoUrl     0
playerUrl    0
```

Working on fifa21 raw data v2.csv using jupyter notebook

Data cleaning

```
In [1]:  # importing the libraries and or packages that i will be needing for the cleaning process
         import numpy as np
         import pandas as pd
         import matplotlib.pyplot as plt
```

```
In [2]:  #reading the csv file
         fifa_Data = pd.read_csv("fifa21 raw data v2.csv")

         C:\Users\user\AppData\Local\Temp\ipykernel_8900\1216926179.py:2: DtypeWarning: Columns (76) have mixed types. Specify dtype option
         on import or set low_memory=False.
           fifa_Data = pd.read_csv("fifa21 raw data v2.csv")
```

```
In [3]:  fifa_Data
```

Out[3]:

	Name	LongName	photoUrl	playerUrl	Nationality	Age	↓OVA	POT	Club	...	A/W	D/W	IR	PAC	SHO	PAS	DRI	DEF	PHY	Hits
	L. Messi	Lionel Messi	https://cdn.sofifa.com /players/158/023 /21_60.png	http://sofifa.com /player/158023 /lionel-messi/2...	Argentina	33	93	93	\n\n\n\nFC Barcelona	...	Medium	Low	5★	85	92	91	95	38	65	771
	Cristiano Ronaldo	C. Ronaldo dos Santos Aveiro	https://cdn.sofifa.com /players/020/801 /21_60.png	http://sofifa.com /player/20801 /c-ronaldo-dos-s...	Portugal	35	92	92	\n\n\n \nJuventus	...	High	Low	5★	89	93	81	89	35	77	562
	J. Oblak	Jan Oblak	https://cdn.sofifa.com /players/200/389 /21_60.png	http://sofifa.com /player/200389 /jan-oblak/210006/	Slovenia	27	91	93	\n\n\n \nAtlético Madrid	...	Medium	Medium	3★	87	92	78	90	52	90	150
	K. De Bruyne	Kevin De Bruyne	https://cdn.sofifa.com /players/192/985 /21_60.png	http://sofifa.com /player/192985 /kevin-de-bruyn...	Belgium	29	91	91	\n\n\n \nManchester City	...	High	High	4★	76	86	93	88	64	78	207
	Neymar Jr	Neymar da Silva Santos Jr	https://cdn.sofifa.com /players/190/871 /21_60.png	http://sofifa.com /player/190871 /neymar-da-	Brazil	28	91	91	\n\n\n\nParis Saint- Germain	...	High	Medium	5★	91	85	86	94	36	59	595

PlayerUrl	Nationality	Age	↓OVA	POT	Club	Contract	Positions	Height	Weight	Preferred Foot	BOV	Best Position	Joined	Loan Date End	Value	Wage
ofifa.com r/158023 /lionel-messi/2...	Argentina	33	93	93	\n\n\n\nFC Barcelona	2004 ~ 2021	RW, ST, CF	170cm	72kg	Left	93	RW	2004-07-01	0	€103.5M	€560K
ofifa.com er/20801 aldo-dos-s...	Portugal	35	92	92	\n\n\n \nJuventus	2018 ~ 2022	ST, LW	187cm	83kg	Right	92	ST	2018-07-10	0	€63M	€220K
ofifa.com r/200389 /jan-/210006/	Slovenia	27	91	93	\n\n\n \nAtlético Madrid	2014 ~ 2023	GK	188cm	87kg	Right	91	GK	2014-07-16	0	€120M	€125K
ofifa.com r/192985 kevin-de-bruyn...	Belgium	29	91	91	\n\n\n \nManchester City	2015 ~ 2023	CAM, CM	181cm	70kg	Right	91	CAM	2015-08-30	0	€129M	€370K
ofifa.com r/190871 ymar-da-silv...	Brazil	28	91	91	\n\n\n\nParis Saint-Germain	2017 ~ 2022	LW, CAM	175cm	68kg	Right	91	LW	2017-08-03	0	€132M	€270K
ofifa.com r/188545 /robert-ewando...	Poland	31	91	91	\n\n\n\nFC Bayern München	2014 ~ 2023	ST	184cm	80kg	Right	91	ST	2014-07-01	0	€111M	€240K
ofifa.com r/209331 ohamed-salah/...	Egypt	28	90	90	\n\n\n \nLiverpool	2017 ~ 2023	RW	175cm	71kg	Left	90	RW	2017-07-01	0	€120.5M	€250K
ofifa.com r/212831 /alisson-	Brazil	27	90	91	\n\n\n \nLiverpool	2018 ~ 2024	GK	191cm	91kg	Right	90	GK	2018-07-19	0	€102M	€160K