Módulo 17: Data Massaging Part 2

Generar un archivo tipo Notebook de Python que contenga el código fuente de varios ejercicios aplicados a los conceptos vistos:

- Leer archivo fifa_eda.csv
- Visualización de una muestra de 20 líneas
- De ser necesario, utiliza el código del ejercicio anterior para no tener que repetir refinamientos de datos
- Generar la media, mínimo y máximo de los jugadores en total y por país
- Generar 2 jugadores ficticios con todos los campos del dataset y añarirlos al mismo. Visualizarlos con .tail() y correr nuevamente las estadísticas y compararlas
- Generar una lambda que añada la zona horaria y la capital el país de nacionalidad
- Generar una lambda que en base al potencial, añada una columna "Candidato" o "Next Year" para el premio para el balón de oro. Jugadores con más de 85 son candidatos

```
In [ ]:  import pandas as pd
         import pytz
         from countryinfo import CountryInfo
```

```
In [ ]:  # Load df
         df = pd.read_csv('d:/Documentos/DataAnalysis/EBAC/Python/Modulo14/fifa_eda.csv')
         df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 18207 entries, 0 to 18206
Data columns (total 18 columns):
 #   Column                   Non-Null Count  Dtype
---  ------                   --------------  -----
 0   ID                       18207 non-null  int64
 1   Name                     18207 non-null  object
 2   Age                      18207 non-null  int64
 3   Nationality              18207 non-null  object
 4   Overall                  18207 non-null  int64
 5   Potential                18207 non-null  int64
 6   Club                     17966 non-null  object
 7   Value                    17955 non-null  float64
 8   Wage                     18207 non-null  float64
 9   Preferred Foot           18207 non-null  object
 10  International Reputation 18159 non-null  float64
 11  Skill Moves              18159 non-null  float64
 12  Position                 18207 non-null  object
 13  Joined                   18207 non-null  int64
 14  Contract Valid Until     17918 non-null  object
 15  Height                   18207 non-null  float64
 16  Weight                   18207 non-null  float64
 17  Release Clause           18207 non-null  float64
dtypes: float64(7), int64(5), object(6)
memory usage: 2.5+ MB
```

In [ ]:
```python
# Muestra de 20 líneas
df.sample(20)
```

Out[ ]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Pre
2752	225859	M. Niakhaté	22	France	73	80	1. FSV Mainz 05	4800.0	17.0	
10653	223925	A. Brunst	22	Germany	65	74	1. FC Magdeburg	625.0	3.0	
13872	212973	M. Marić	22	Croatia	62	69	Lillestrøm SK	325.0	3.0	
4539	166080	Evandro	31	Brazil	71	71	Hull City	2100.0	18.0	
7836	227370	M. Zetterer	22	Germany	67	75	SV Werder Bremen	850.0	5.0	
7575	241135	J. Collins	23	Nigeria	68	73	SC Paderborn 07	1000.0	4.0	
11303	172872	M. Yıldırım	31	Turkey	64	64	Yeni Malatyaspor	300.0	3.0	
18164	243876	M. Frøkjær-Jensen	18	Denmark	48	63	Odense Boldklub	60.0	1.0	
9976	214028	C. Amaya	27	Colombia	65	65	Once Caldas	575.0	1.0	
5533	232926	K. Watanabe	31	Japan	70	70	Gamba Osaka	1600.0	8.0	
3660	210321	R. Quioto	26	Honduras	72	72	Houston Dynamo	3300.0	7.0	
9327	207741	M. Lieder	28	Netherlands	66	66	SønderjyskE	675.0	5.0	
11663	235171	A. Bongiovanni	18	Belgium	64	79	KSV Cercle Brugge	925.0	6.0	
17544	239893	C. Coxe	19	Wales	53	69	Cardiff City	110.0	2.0	
12384	236883	M. Anderson	20	Iceland	63	76	Excelsior	700.0	3.0	
11326	138830	D. Lewington	34	England	64	64	Milton Keynes Dons	130.0	3.0	
427	226226	G. Lo Celso	22	Argentina	80	86	Real Betis	20000.0	72.0	
352	204884	B. Mendy	23	France	81	86	Manchester City	20000.0	105.0	
11204	135218	D. Perquis	34	Poland	64	64	GFC Ajaccio	140.0	1.0	
279	192984	K. Casteels	26	Belgium	82	85	VfL Wolfsburg	19500.0	46.0	

In [ ]:
```python
# Media, mínimo y máximo de los jugadores
print(df.mean(numeric_only=True), '\n\n---\n')
print(df.min(numeric_only=True), '\n\n---\n')
print(df.max(numeric_only=True))
```

```
ID                        214298.338606
Age                           25.122206
Overall                       66.238699
Potential                     71.307299
Value                       2444.530214
Wage                           9.731312
International Reputation       1.113222
Skill Moves                    2.361308
Joined                      2016.420607
Height                         5.946771
Weight                       165.979129
Release Clause              4585.060971
dtype: float64

---

ID                          16.000000
Age                         16.000000
Overall                     46.000000
Potential                   48.000000
Value                       10.000000
Wage                         0.000000
International Reputation     1.000000
Skill Moves                  1.000000
Joined                    1991.000000
Height                       5.083333
Weight                     110.000000
Release Clause              13.000000
dtype: float64

---

ID                        246620.00
Age                           45.00
Overall                       94.00
Potential                     95.00
Value                     118500.00
Wage                         565.00
International Reputation        5.00
Skill Moves                     5.00
Joined                      2018.00
Height                         6.75
Weight                       243.00
Release Clause            228100.00
dtype: float64
```

In [ ]:
```python
# Media, mínimo y máximo de los jugadores por país
print('MEDIA\n',df.groupby('Nationality').mean(numeric_only=True),'\n\n---\n')
print('MAXIMO\n',df.groupby('Nationality').max(numeric_only=True),'\n\n---\n')
print('MÍNIMO\n',df.groupby('Nationality').min(numeric_only=True))
```

MEDIA

Nationality	ID	Age	Overall	Potential	Value
Afghanistan	234743.500000	22.500000	61.000000	67.750000	342.500000
Albania	221682.450000	24.025000	65.925000	71.700000	1920.250000
Algeria	206355.716667	27.050000	70.633333	72.983333	4670.916667
Andorra	192937.000000	28.000000	62.000000	64.000000	290.000000
Angola	206676.000000	25.866667	67.600000	71.533333	2018.333333
...	...	...	...	...	...
Uzbekistan	189585.000000	29.500000	67.500000	67.500000	3100.000000
Venezuela	224879.641791	24.313433	67.268657	73.149254	2498.333333
Wales	201690.651163	24.736434	64.139535	69.968992	1769.257812
Zambia	236750.555556	22.222222	65.222222	73.777778	1129.444444
Zimbabwe	211857.769231	26.846154	69.769231	72.461538	2780.769231

Nationality	Wage	International Reputation	Skill Moves	Joined
Afghanistan	1.250000	1.000000	2.500000	2017.250000
Albania	7.100000	1.102564	2.487179	2016.025000
Algeria	18.083333	1.300000	2.816667	2016.916667
Andorra	1.000000	1.000000	2.000000	2018.000000
Angola	7.666667	1.133333	2.933333	2017.000000
...	...	...	...	...
Uzbekistan	9.000000	1.000000	2.500000	2017.000000
Venezuela	6.194030	1.106061	2.530303	2017.119403
Wales	12.093023	1.093023	2.248062	2016.193798
Zambia	3.444444	1.000000	2.333333	2017.555556
Zimbabwe	6.846154	1.000000	2.692308	2016.461538

Nationality	Height	Weight	Release Clause	Candidato
Afghanistan	5.729167	149.250000	617.250000	0.000000
Albania	5.950753	164.424478	3694.029560	0.050000
Algeria	5.947222	167.250000	8490.020720	0.000000
Andorra	6.000000	174.000000	384.000000	0.000000
Angola	5.877778	166.533333	3623.741441	0.000000
...	...	...	...	...
Uzbekistan	5.958333	164.000000	6725.000000	0.000000
Venezuela	5.894728	164.671330	4203.207643	0.000000
Wales	5.960594	166.403101	3714.370469	0.015504
Zambia	5.907407	162.000000	2174.000000	0.000000
Zimbabwe	5.852564	164.000000	5319.384615	0.000000

[164 rows x 13 columns]

MAXIMO

Nationality	ID	Age	Overall	Potential	Value	Wage
Afghanistan	244175	26	63	71	450.0	2.0
Albania	245792	34	81	86	20000.0	58.0
Algeria	245565	35	85	85	40500.0	205.0
Andorra	192937	28	62	64	290.0	1.0
Angola	242188	34	78	82	10000.0	44.0
...	...	...	...	...	...	...

```
Uzbekistan   197597   30      75        75    6000.0   17.0
Venezuela    246388   34      81        84   22500.0   42.0
Wales        246227   37      88        88   60000.0  355.0
Zambia       246144   29      73        83    3500.0    7.0
Zimbabwe     239438   34      78        83   12000.0   27.0
```

```
               International Reputation  Skill Moves  Joined    Height  Weight  \
Nationality
Afghanistan                         1.0          3.0    2018  5.833333   152.0
Albania                             2.0          3.0    2018  6.416667   196.0
Algeria                             3.0          5.0    2018  6.416667   207.0
Andorra                             1.0          2.0    2018  6.000000   174.0
Angola                              2.0          4.0    2018  6.166667   192.0
...                                 ...          ...     ...       ...     ...
Uzbekistan                          1.0          3.0    2017  6.000000   165.0
Venezuela                           2.0          4.0    2018  6.500000   216.0
Wales                               4.0          4.0    2018  6.500000   207.0
Zambia                              1.0          3.0    2018  6.250000   201.0
Zimbabwe                            1.0          4.0    2018  6.166667   185.0
```

```
             Release Clause  Candidato
Nationality
Afghanistan           878.0      False
Albania             35500.0       True
Algeria             78000.0      False
Andorra               384.0      False
Angola              17800.0      False
...                     ...        ...
Uzbekistan          13200.0      False
Venezuela           35400.0      False
Wales              123000.0       True
Zambia               6600.0      False
Zimbabwe            18900.0      False
```

```
[164 rows x 13 columns]
```

```
---
```

```
MÍNIMO
                 ID  Age  Overall  Potential  Value  Wage  \
Nationality
Afghanistan  223989   20       59         64  220.0   1.0
Albania      170612   18       52         62   80.0   1.0
Algeria       49659   19       58         62  160.0   1.0
Andorra      192937   28       62         64  290.0   1.0
Angola       144037   19       60         64  325.0   1.0
...             ...  ...      ...        ...    ...   ...
Uzbekistan   181573   29       60         60  200.0   1.0
Venezuela    162123   18       53         55   80.0   0.0
Wales           657   17       47         56   30.0   0.0
Zambia       190276   18       56         66  190.0   1.0
Zimbabwe     189171   22       64         66  475.0   1.0
```

```
               International Reputation  Skill Moves  Joined    Height  Weight  \
Nationality
Afghanistan                         1.0          2.0    2016  5.666667   143.0
```

```
Albania                             1.0      1.0   2008  5.583333  137.0
Algeria                             1.0      1.0   2011  5.500000  132.0
Andorra                             1.0      2.0   2018  6.000000  174.0
Angola                              1.0      2.0   2014  5.583333  143.0
...                                 ...      ...    ...       ...    ...
Uzbekistan                          1.0      2.0   2017  5.916667  163.0
Venezuela                           1.0      1.0   2013  5.250000  121.0
Wales                               1.0      1.0   2006  5.500000  132.0
Zambia                              1.0      2.0   2017  5.583333  141.0
Zimbabwe                            1.0      2.0   2011  5.500000  132.0

              Release Clause   Candidato
Nationality
Afghanistan             374.0      False
Albania                 116.0      False
Algeria                 259.0      False
Andorra                 384.0      False
Angola                  463.0      False
...                       ...        ...
Uzbekistan              250.0      False
Venezuela               108.0      False
Wales                    53.0      False
Zambia                  356.0      False
Zimbabwe                594.0      False

[164 rows x 13 columns]
```

In [ ]:
```python
# Generar dos jugadores ficticios
jugadores = [
    (246621, 'C. Del Valle', 23, 'Mexico', 10, 15, 'Mezcaleros F.C.', 200, 0, 'Righ
    (246621, 'S. Name', 19, 'Mexico', 15, 30, 'Mezcaleros F.C.', 500, 0, 'Right', 0
]

df_nuevos_jugadores = pd.DataFrame(jugadores, columns=list(df.columns))
df = pd.concat([df,df_nuevos_jugadores]).reset_index(drop=True)
df.tail()
```

Out[ ]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Preferred Foot
18204	241638	B. Worman	16	England	47	67	Cambridge United	60.0	1.0	Right
18205	246268	D. Walker-Rice	17	England	47	66	Tranmere Rovers	60.0	1.0	Right
18206	246269	G. Nugent	16	England	46	66	Tranmere Rovers	60.0	1.0	Right
18207	246621	C. Del Valle	23	Mexico	10	15	Mezcaleros F.C.	200.0	0.0	Right
18208	246621	S. Name	19	Mexico	15	30	Mezcaleros F.C.	500.0	0.0	Right

```
In [ ]:  print(df.mean(numeric_only=True), '\n\n---\n')
         print(df.min(numeric_only=True), '\n\n---\n')
         print(df.max(numeric_only=True))
```

```
ID                        214301.888791
Age                           25.121753
Overall                       66.232797
Potential                     71.301939
Value                       2444.296932
Wage                           9.730243
International Reputation        1.113099
Skill Moves                    2.361048
Joined                      2016.421220
Height                         5.946777
Weight                       165.981438
Release Clause              4584.612285
dtype: float64

---

ID                            16.000000
Age                           16.000000
Overall                       10.000000
Potential                     15.000000
Value                         10.000000
Wage                           0.000000
International Reputation        0.000000
Skill Moves                    0.000000
Joined                      1991.000000
Height                         5.083333
Weight                       110.000000
Release Clause                13.000000
dtype: float64

---

ID                           246621.00
Age                              45.00
Overall                          94.00
Potential                        95.00
Value                        118500.00
Wage                            565.00
International Reputation           5.00
Skill Moves                       5.00
Joined                         2022.00
Height                            6.75
Weight                          243.00
Release Clause               228100.00
dtype: float64
```

Debido a que los dos jugadores agregados son provenientes de un equipo amateur y tienen estadísticas muy bajas, varios de los valores mínimos del dataframe fueron modificados tales como Overall, Potential, Value, International Reputation y Skill Moves. Los valores máximos del df no se vieron afectados al igual que el promedio debido al gran tamaño de la base de datos, dos personas no modificaron significativamente los valores promedio

```python
# Try except para ver qué nombres de países no se soporta la librería CountryInfo p
country_list = list(df['Nationality'].sort_values().unique())
missing_countries = []
for country in country_list:
    try:
        CountryInfo(country).iso(2)
    except KeyError:
        missing_countries.append(country)

missing_countries
```

```
Out[ ]: ['Andorra',
 'Antigua & Barbuda',
 'Bosnia Herzegovina',
 'Central African Rep.',
 'China PR',
 'Congo',
 'Curacao',
 'England',
 'FYR Macedonia',
 'Gambia',
 'Guinea Bissau',
 'Korea DPR',
 'Korea Republic',
 'Kosovo',
 'Montenegro',
 'Northern Ireland',
 'Palestine',
 'St Kitts Nevis',
 'St Lucia',
 'São Tomé & Príncipe',
 'Trinidad & Tobago']
```

In [ ]:
```python
# Diccionario country_helper para cambiar el nombre de algunos países que no soport
country_helper = {
    'Andorra': 'France',
    'Antigua & Barbuda': 'Antigua and Barbuda',
    'Bosnia Herzegovina': 'Bosnia and Herzegovina',
    'Central African Rep.': 'Central African Republic',
    'China PR': 'China',
    'Congo': 'Democratic Republic of the Congo',
    'Curacao': 'Venezuela',
    'England': 'Great Britain',
    'FYR Macedonia': 'Albania',
    'Gambia': 'Senegal',
    'Guinea Bissau': 'Guinea-Bissau',
    'Korea DPR': 'North Korea',
    'Korea Republic': 'South Korea',
    'Kosovo': 'Albania',
    'Montenegro': 'Albania',
    'Northern Ireland': 'Great Britain',
    'Palestine': 'Lebanon',
    'St Kitts Nevis': 'Saint Kitts and Nevis',
    'St Lucia': 'Saint Lucia',
    'São Tomé & Príncipe': 'Nigeria',
    'Trinidad & Tobago': 'Trinidad and Tobago'
    }
```

In [ ]:
```python
# 4. Lambda que agregue zona horaria y capital del país de nacionalidad

# Agregar una columna Helper Country para cambiar el país de Nationality a los sopo
df['HelperCountry'] = df['Nationality'].apply(lambda x : country_helper[x] if x in
df
```

Out[ ]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Prefer Fo
0	158023	L. Messi	31	Argentina	94	94	FC Barcelona	110500.0	565.0	l
1	20801	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000.0	405.0	Ri
2	190871	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500.0	290.0	Ri
3	193080	De Gea	27	Spain	91	93	Manchester United	72000.0	260.0	Ri
4	192985	K. De Bruyne	27	Belgium	91	92	Manchester City	102000.0	355.0	Ri
...	...	...	...	...	...	...	...	...	...	
18204	241638	B. Worman	16	England	47	67	Cambridge United	60.0	1.0	Ri
18205	246268	D. Walker-Rice	17	England	47	66	Tranmere Rovers	60.0	1.0	Ri
18206	246269	G. Nugent	16	England	46	66	Tranmere Rovers	60.0	1.0	Ri
18207	246621	C. Del Valle	23	Mexico	10	15	Mezcaleros F.C.	200.0	0.0	Ri
18208	246621	S. Name	19	Mexico	15	30	Mezcaleros F.C.	500.0	0.0	Ri

18209 rows × 19 columns

In [ ]:
```python
# Crear un diccionario de ayuda para cambiar el nombre del país por su abreviación
country_iso = {}
for country in list(df['HelperCountry'].unique()):
    country_iso[country] = CountryInfo(country).iso(2)

# Escocia tiene un error en la librería donde su ISO aparece vacío
country_iso['Scotland'] = 'GB'
```

In [ ]:
```python
# Cambiar el nombre de helper country a su abreviación ISO de dos letras para poder
# df['HelperCountry'] = df['HelperCountry'].apply(lambda x : country_iso[x])

# Cambiar el nombre de helper country a su abreviación ISO de dos letras para poder
df['Timezone'] = df['HelperCountry'].apply(lambda x : pytz.country_timezones[countr
```

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

Out[ ]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Prefer Fc
0	158023	L. Messi	31	Argentina	94	94	FC Barcelona	110500.0	565.0	I
1	20801	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000.0	405.0	Ri
2	190871	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500.0	290.0	Ri
3	193080	De Gea	27	Spain	91	93	Manchester United	72000.0	260.0	Ri
4	192985	K. De Bruyne	27	Belgium	91	92	Manchester City	102000.0	355.0	Ri
...	...	...	...	...	...	...	...	...	...	
18204	241638	B. Worman	16	England	47	67	Cambridge United	60.0	1.0	Ri
18205	246268	D. Walker-Rice	17	England	47	66	Tranmere Rovers	60.0	1.0	Ri
18206	246269	G. Nugent	16	England	46	66	Tranmere Rovers	60.0	1.0	Ri
18207	246621	C. Del Valle	23	Mexico	10	15	Mezcaleros F.C.	200.0	0.0	Ri
18208	246621	S. Name	19	Mexico	15	30	Mezcaleros F.C.	500.0	0.0	Ri

18209 rows × 20 columns

In [ ]:
```python
# Generar lista de todos los países que no no soporte countryinfo().capital()
not_supported = []
for country in list(df['Nationality'].unique()):
    try:
        CountryInfo(country).capital()
    except KeyError:
        not_supported.append(country)

# Estos serán países que no soporta la función .capital
not_supported
```

```
Out[ ]: ['England',
         'Wales',
         'Bosnia Herzegovina',
         'Korea Republic',
         'Montenegro',
         'Scotland',
         'Central African Rep.',
         'FYR Macedonia',
         'China PR',
         'Guinea Bissau',
         'Kosovo',
         'Gambia',
         'Congo',
         'Northern Ireland',
         'Curacao',
         'Trinidad & Tobago',
         'São Tomé & Príncipe',
         'Korea DPR',
         'St Kitts Nevis',
         'Antigua & Barbuda',
         'Palestine',
         'St Lucia',
         'Andorra']
```

```python
In [ ]:  # Casi todos los países en la lista los corregimos antes, pero algunos los cambiamo

         # Diccionario con los países que se corrigieron

         country_helper = {
             'Antigua & Barbuda': 'Antigua and Barbuda',
             'Bosnia Herzegovina': 'Bosnia and Herzegovina',
             'Central African Rep.': 'Central African Republic',
             'China PR': 'China',
             'Congo': 'Democratic Republic of the Congo',
             'England': 'Great Britain',
             'Gambia': 'The Gambia',
             'Guinea Bissau': 'Guinea-Bissau',
             'Korea DPR': 'North Korea',
             'Korea Republic': 'South Korea',
             'St Kitts Nevis': 'Saint Kitts and Nevis',
             'St Lucia': 'Saint Lucia',
             'Trinidad & Tobago': 'Trinidad and Tobago'
             }


         # Lista con los países que se tuvieron que cambiar y sus capitales de manera manual
         wrong_countries  = {
             'Andorra': 'Andorra la Vella',
             'Curacao': 'Willemstad',
             'FYR Macedonia': 'Skopje',
             'Kosovo': 'Pristina',
             'Montenegro': 'Podgorica',
             'Northern Ireland': 'Belfast',
             'Palestine': 'Jerusalem, Ramallah',
             'São Tomé & Príncipe': 'São Tomé',
             'Scotland': 'Edinburgh',
             'Wales': 'Cardiff'
             }

         # Generar un diccionario de países y sus capitales apoyado por el diccionario y la
         capital_helper = {}

         for country in list(df['Nationality'].unique()):
             if country in country_helper.keys():
                 capital_helper[country] = CountryInfo(country_helper[country]).capital()
             elif country in wrong_countries.keys():
                 capital_helper[country] = wrong_countries[country]
             else:
                 capital_helper[country] = CountryInfo(country).capital()

         capital_helper
```

```
Out[ ]: {'Argentina': 'Buenos Aires',
         'Portugal': 'Lisbon',
         'Brazil': 'Brasília',
         'Spain': 'Madrid',
         'Belgium': 'Brussels',
         'Croatia': 'Zagreb',
         'Uruguay': 'Montevideo',
         'Slovenia': 'Ljubljana',
         'Poland': 'Warsaw',
         'Germany': 'Berlin',
         'France': 'Paris',
         'England': 'London',
         'Italy': 'Rome',
         'Egypt': 'Cairo',
         'Colombia': 'Bogotá',
         'Denmark': 'Copenhagen',
         'Gabon': 'Libreville',
         'Wales': 'Cardiff',
         'Senegal': 'Dakar',
         'Costa Rica': 'San José',
         'Slovakia': 'Bratislava',
         'Netherlands': 'Amsterdam',
         'Bosnia Herzegovina': 'Sarajevo',
         'Morocco': 'Rabat',
         'Serbia': 'Belgrade',
         'Algeria': 'Algiers',
         'Austria': 'Vienna',
         'Greece': 'Athens',
         'Chile': 'Santiago',
         'Sweden': 'Stockholm',
         'Korea Republic': 'Seoul',
         'Finland': 'Helsinki',
         'Guinea': 'Conakry',
         'Montenegro': 'Podgorica',
         'Armenia': 'Yerevan',
         'Switzerland': 'Bern',
         'Norway': 'Oslo',
         'Czech Republic': 'Prague',
         'Scotland': 'Edinburgh',
         'Ghana': 'Accra',
         'Central African Rep.': 'Bangui',
         'DR Congo': 'Kinshasa',
         'Ivory Coast': 'Yamoussoukro',
         'Russia': 'Moscow',
         'Ukraine': 'Kiev',
         'Iceland': 'Reykjavik',
         'Mexico': 'Mexico City',
         'Jamaica': 'Kingston',
         'Albania': 'Tirana',
         'Venezuela': 'Caracas',
         'Japan': 'Tokyo',
         'Turkey': 'Ankara',
         'Ecuador': 'Quito',
         'Paraguay': 'Asunción',
         'Mali': 'Bamako',
         'Nigeria': 'Abuja',
```

```
                'Cameroon': 'Yaoundé',
                'Dominican Republic': 'Santo Domingo',
                'Israel': 'Jerusalem',
                'Kenya': 'Nairobi',
                'Hungary': 'Budapest',
                'Republic of Ireland': 'Dublin',
                'Romania': 'Bucharest',
                'United States': 'Washington D.C.',
                'Cape Verde': 'Praia',
                'Australia': 'Canberra',
                'Peru': 'Lima',
                'Togo': 'Lomé',
                'Syria': 'Damascus',
                'Zimbabwe': 'Harare',
                'Angola': 'Luanda',
                'Burkina Faso': 'Ouagadougou',
                'Iran': 'Tehran',
                'Estonia': 'Tallinn',
                'Tunisia': 'Tunis',
                'Equatorial Guinea': 'Malabo',
                'New Zealand': 'Wellington',
                'FYR Macedonia': 'Skopje',
                'United Arab Emirates': 'Abu Dhabi',
                'China PR': 'Beijing',
                'Guinea Bissau': 'Bissau',
                'Bulgaria': 'Sofia',
                'Kosovo': 'Pristina',
                'South Africa': 'Pretoria',
                'Madagascar': 'Antananarivo',
                'Georgia': 'Tbilisi',
                'Tanzania': 'Dodoma',
                'Gambia': 'Banjul',
                'Cuba': 'Havana',
                'Belarus': 'Minsk',
                'Uzbekistan': 'Tashkent',
                'Benin': 'Porto-Novo',
                'Congo': 'Kinshasa',
                'Mozambique': 'Maputo',
                'Honduras': 'Tegucigalpa',
                'Canada': 'Ottawa',
                'Northern Ireland': 'Belfast',
                'Cyprus': 'Nicosia',
                'Saudi Arabia': 'Riyadh',
                'Curacao': 'Willemstad',
                'Moldova': 'Chişinău',
                'Bolivia': 'Sucre',
                'Trinidad & Tobago': 'Port of Spain',
                'Sierra Leone': 'Freetown',
                'Zambia': 'Lusaka',
                'Chad': "N'Djamena",
                'Philippines': 'Manila',
                'Haiti': 'Port-au-Prince',
                'Comoros': 'Moroni',
                'Libya': 'Tripoli',
                'Panama': 'Panama City',
                'São Tomé & Príncipe': 'São Tomé',
```

```
 'Eritrea': 'Asmara',
 'Oman': 'Muscat',
 'Iraq': 'Baghdad',
 'Burundi': 'Bujumbura',
 'Fiji': 'Suva',
 'New Caledonia': 'Nouméa',
 'Lithuania': 'Vilnius',
 'Luxembourg': 'Luxembourg',
 'Korea DPR': 'Pyongyang',
 'Liechtenstein': 'Vaduz',
 'St Kitts Nevis': 'Basseterre',
 'Latvia': 'Riga',
 'Suriname': 'Paramaribo',
 'Uganda': 'Kampala',
 'El Salvador': 'San Salvador',
 'Bermuda': 'Hamilton',
 'Kuwait': 'Kuwait City',
 'Antigua & Barbuda': "Saint John's",
 'Thailand': 'Bangkok',
 'Mauritius': 'Port Louis',
 'Guatemala': 'Guatemala City',
 'Liberia': 'Monrovia',
 'Kazakhstan': 'Astana',
 'Niger': 'Niamey',
 'Mauritania': 'Nouakchott',
 'Montserrat': 'Plymouth',
 'Namibia': 'Windhoek',
 'Azerbaijan': 'Baku',
 'Guam': 'Hagåtña',
 'Faroe Islands': 'Tórshavn',
 'India': 'New Delhi',
 'Nicaragua': 'Managua',
 'Barbados': 'Bridgetown',
 'Lebanon': 'Beirut',
 'Palestine': 'Jerusalem, Ramallah',
 'Guyana': 'Georgetown',
 'Sudan': 'Khartoum',
 'St Lucia': 'Castries',
 'Ethiopia': 'Addis Ababa',
 'Puerto Rico': 'San Juan',
 'Grenada': "St. George's",
 'Jordan': 'Amman',
 'Rwanda': 'Kigali',
 'Qatar': 'Doha',
 'Afghanistan': 'Kabul',
 'Hong Kong': 'City of Victoria',
 'Andorra': 'Andorra la Vella',
 'Malta': 'Valletta',
 'Belize': 'Belmopan',
 'South Sudan': 'Juba',
 'Indonesia': 'Jakarta',
 'Botswana': 'Gaborone'}
```

```python
In [ ]:  # Generar la columna capital y buscar la capital en el diccionario antes creado
         df['Capital'] = df['Nationality'].apply(lambda x : capital_helper[x])
         df.sample(10)
```

Out[ ]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Prefe
12159	241172	Diego Barri	22	Spain	63	74	Albacete BP	550.0	2.0	F
4222	217167	G. Sandoval	22	Mexico	71	80	Guadalajara	3600.0	20.0	F
14210	224097	T. Denton	22	England	61	70	Peterborough United	325.0	5.0	
5749	239645	M. de los Santos	25	Uruguay	69	72	Millonarios FC	1100.0	2.0	F
16836	235139	Long Wei	23	China PR	56	63	Henan Jianye FC	110.0	2.0	F
5979	216416	J. Gyamerah	23	Germany	69	77	VfL Bochum 1848	1300.0	7.0	F
18062	243091	Gao Dalun	21	China PR	50	58	Jiangsu Suning FC	40.0	2.0	F
11192	224302	M. Wieteska	21	Poland	64	72	Legia Warszawa	575.0	2.0	F
5600	224756	A. Lusamba	21	France	70	79	KSV Cercle Brugge	2800.0	12.0	F
9285	211059	R. Godínez	25	Mexico	66	71	Lobos BUAP	725.0	2.0	F

10 rows × 21 columns

In [ ]:
```python
# Generar una columna 'Candidato' para los candidatos a Balón de Oro el siguiente a
df['Candidato'] = df['Potential'].apply(lambda x : x > 85)
df.sample(10)
```

Out[ ]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Pre
18010	242421	Zhu Zhengyu	23	China PR	51	57	Guizhou Hengfeng FC	50.0	2.0	
7205	211857	Alex Muñoz	23	Spain	68	74	Real Zaragoza	1000.0	4.0	
7846	238636	R. Preißinger	21	Germany	67	75	1. FC Magdeburg	975.0	4.0	
12782	241588	Toni Moya	20	Spain	63	79	Atlético Madrid	800.0	6.0	
1744	195668	Joel Robles	28	Spain	75	76	Real Betis	5500.0	17.0	
3839	214483	C. Arboleda	32	Colombia	72	72	Independiente Santa Fe	1700.0	4.0	
2173	177974	A. Aguilar	33	Colombia	74	74	FC Dallas	2300.0	7.0	
5292	188824	E. Imorou	29	Benin	70	70	Stade Malherbe Caen	1300.0	12.0	
120	207566	William Carvalho	26	Portugal	84	86	Real Betis	31500.0	38.0	
1045	143745	A. Turan	31	Turkey	77	77	Medipol Başakşehir FK	7500.0	120.0	

10 rows × 22 columns