

Actividad Módulo 16: Data Massaging 1

Para el archivo fifa_eda.csv:

1. Mostrar las primeras 20 filas del archivo, las últimas 5 y un sample de 10
2. Generar data estadística con `.describe()` y además los tipos de datos del dataset
3. Si es necesario, pasar a numéricas por lo menos 2 columnas que contengan números para incluirlas en el `.describe()`
4. Añadir una columna 'Years Playing' que calcule el año actual menos la columna 'Joined'
5. Buscar y mostrar a todos los jugadores de México
6. Ordenar y mostrar los datos por la columna Release Clause (sueldo)
7. Generar un nuevo dataset que contenga el año y el número de jugadores
8. Opcional: Generar un gráfico que contenga, por año, el número de jugadores

```
In [ ]: # Importar Librerías y archivo
import pandas as pd

df = pd.read_csv('d:/Documentos/DataAnalysis/EBAC/Python/Modulo14/fifa_eda.csv')
# 1.1 - Primeras 20 filas del df
df.head(20)
```

Out[]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Preferred
0	158023	L. Messi	31	Argentina	94	94	FC Barcelona	110500.0	565.0	
1	20801	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000.0	405.0	R
2	190871	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500.0	290.0	R
3	193080	De Gea	27	Spain	91	93	Manchester United	72000.0	260.0	R
4	192985	K. De Bruyne	27	Belgium	91	92	Manchester City	102000.0	355.0	R
5	183277	E. Hazard	27	Belgium	91	91	Chelsea	93000.0	340.0	R
6	177003	L. Modrić	32	Croatia	91	91	Real Madrid	67000.0	420.0	R
7	176580	L. Suárez	31	Uruguay	91	91	FC Barcelona	80000.0	455.0	R
8	155862	Sergio Ramos	32	Spain	91	91	Real Madrid	51000.0	380.0	R
9	200389	J. Oblak	25	Slovenia	90	93	Atlético Madrid	68000.0	94.0	R
10	188545	R. Lewandowski	29	Poland	90	90	FC Bayern München	77000.0	205.0	R
11	182521	T. Kroos	28	Germany	90	90	Real Madrid	76500.0	355.0	R
12	182493	D. Godín	32	Uruguay	90	90	Atlético Madrid	44000.0	125.0	R
13	168542	David Silva	32	Spain	90	90	Manchester City	60000.0	285.0	
14	215914	N. Kanté	27	France	89	90	Chelsea	63000.0	225.0	R
15	211110	P. Dybala	24	Argentina	89	94	Juventus	89000.0	205.0	
16	202126	H. Kane	24	England	89	91	Tottenham Hotspur	83500.0	205.0	R
17	194765	A. Griezmann	27	France	89	90	Atlético Madrid	78000.0	145.0	
18	192448	M. ter Stegen	26	Germany	89	92	FC Barcelona	58000.0	240.0	R
19	192119	T. Courtois	26	Belgium	89	90	Real Madrid	53500.0	240.0	

```
In [ ]: # 1.2 - Últimas 5 filas del df
df.tail(5)
```

Out[]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Pref
18202	238813	J. Lundstram	19	England	47	65	Crewe Alexandra	60.0	1.0	
18203	243165	N. Christoffersson	19	Sweden	47	63	Trelleborgs FF	60.0	1.0	
18204	241638	B. Worman	16	England	47	67	Cambridge United	60.0	1.0	
18205	246268	D. Walker-Rice	17	England	47	66	Tranmere Rovers	60.0	1.0	
18206	246269	G. Nugent	16	England	46	66	Tranmere Rovers	60.0	1.0	

```
In [ ]: # 1.3 - Sample de 10 filas del df
df.sample(10)
```

Out[]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Prefer F
5453	214217	B. Vidal	27	Chile	70	72	CD Palestino	1700.0	6.0	Ri
15891	242721	E. Dick	23	United States	58	65	Sporting Kansas City	130.0	1.0	Ri
5057	239441	Jordi Mboula	19	Spain	70	85	AS Monaco	3800.0	15.0	Ri
3126	190942	K. Hansen	29	Denmark	73	73	FC Midtjylland	3100.0	18.0	Ri
10737	178120	M. Bruccini	32	Italy	65	65	Cosenza	400.0	1.0	Ri
11081	225557	R. Poole	20	Wales	64	75	Manchester United	675.0	13.0	Ri
12346	215112	R. Contreras	22	Argentina	63	71	CD Antofagasta	550.0	1.0	Ri
7817	205094	Jo Young Cheol	29	Korea Republic	67	67	Gyeongnam FC	750.0	3.0	Ri
9145	244305	R. Ramírez	22	Venezuela	66	71	Fortuna Sittard	725.0	2.0	Ri
1558	212729	Tozé	25	Portugal	76	80	Vitória Guimarães	10000.0	11.0	Ri

```
In [ ]: # 2.1 Estadísticas de describe()
df.describe()
```

Out[]:

	ID	Age	Overall	Potential	Value	Wage	Inte Rt
count	18207.000000	18207.000000	18207.000000	18207.000000	17955.000000	18207.000000	18159.000000
mean	214298.338606	25.122206	66.238699	71.307299	2444.530214	9.731312	18.159000
std	29965.244204	4.669943	6.908930	6.136496	5626.715434	21.999290	1.815900
min	16.000000	16.000000	46.000000	48.000000	10.000000	0.000000	16.000000
25%	200315.500000	21.000000	62.000000	67.000000	325.000000	1.000000	16.000000
50%	221759.000000	25.000000	66.000000	71.000000	700.000000	3.000000	17.000000
75%	236529.500000	28.000000	71.000000	75.000000	2100.000000	9.000000	18.000000
max	246620.000000	45.000000	94.000000	95.000000	118500.000000	565.000000	19.000000

In []:

```
# 2.2 Tipo de datos del df
df.info()

# A todos los campos numéricos del df les fue asignado el tipo correcto de valor
```

```
<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 []:

```
# 4 - Añadir columna Years Playing
df['Years Playing'] = 2022 - df['Joined']
df
```

Out[]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage
0	158023	L. Messi	31	Argentina	94	94	FC Barcelona	110500.0	565.0
1	20801	Cristiano Ronaldo	33	Portugal	94	94	Juventus	77000.0	405.0
2	190871	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500.0	290.0
3	193080	De Gea	27	Spain	91	93	Manchester United	72000.0	260.0
4	192985	K. De Bruyne	27	Belgium	91	92	Manchester City	102000.0	355.0
...
18202	238813	J. Lundstram	19	England	47	65	Crewe Alexandra	60.0	1.0
18203	243165	N. Christoffersson	19	Sweden	47	63	Trelleborgs FF	60.0	1.0
18204	241638	B. Worman	16	England	47	67	Cambridge United	60.0	1.0
18205	246268	D. Walker-Rice	17	England	47	66	Tranmere Rovers	60.0	1.0
18206	246269	G. Nugent	16	England	46	66	Tranmere Rovers	60.0	1.0

18207 rows × 19 columns

```
In [ ]: # 5 - Mostrar a todos los jugadores de México
df.loc[df['Nationality'] == 'Mexico']
```

Out[]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Preferre Foot
306	171897	A. Guardado	31	Mexico	82	82	Real Betis	19000.0	35.0	Le
329	221992	H. Lozano	22	Mexico	81	86	PSV	24000.0	22.0	Rigl
371	193165	J. Corona	25	Mexico	81	83	FC Porto	21500.0	18.0	Rigl
397	169416	C. Vela	29	Mexico	81	81	Los Angeles FC	17500.0	15.0	Le
406	156519	H. Herrera	28	Mexico	81	81	FC Porto	17500.0	20.0	Rigl
...	
17373	244828	R. Gutiérrez	22	Mexico	54	64	Monarcas Morelia	80.0	1.0	Rigl
17614	239739	A. Ocejo	20	Mexico	53	62	Santos Laguna	90.0	2.0	Le
18037	246089	C. Landa	19	Mexico	50	60	Tiburones Rojos de Veracruz	50.0	1.0	Le
18068	240286	J. García	20	Mexico	50	62	Santos Laguna	40.0	1.0	Rigl
18113	237045	R. Pasquel	22	Mexico	50	60	Deportivo Toluca	40.0	2.0	Rigl

366 rows × 19 columns

```
In [ ]: # 6 - Ordenar datos por la columna 'Release Clause'
df.sort_values(by='Release Clause')
```

Out[]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage
16204	176860	C. Månsson	37	Sweden	58	58	Kristiansund BK	10.0	1.0
17209	237751	Y. Nishibe	37	Japan	55	55	Shimizu S-Pulse	10.0	1.0
16249	176631	B. Lekström	37	Sweden	58	58	Hammarby IF	10.0	1.0
16565	232893	K. Kitamoto	36	Japan	57	57	Vissel Kobe	10.0	1.0
16628	102881	K. Stamatopoulos	38	Canada	57	57	AIK	10.0	1.0
...
25	231747	K. Mbappé	19	France	88	95	Paris Saint-Germain	81000.0	100.0
5	183277	E. Hazard	27	Belgium	91	91	Chelsea	93000.0	340.0
4	192985	K. De Bruyne	27	Belgium	91	92	Manchester City	102000.0	355.0
0	158023	L. Messi	31	Argentina	94	94	FC Barcelona	110500.0	565.0
2	190871	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500.0	290.0

18207 rows × 19 columns

```
In [ ]: # 7 - Generar un dataset que contenga el año (joined) y número de jugadores
df_new = df.groupby(by='Joined').size().rename('Players').reset_index()
df_new
```

```
Out[ ]:
```

	Joined	Players
0	1991	1
1	1998	3
2	1999	1
3	2000	2
4	2001	2
5	2002	10
6	2003	13
7	2004	12
8	2005	17
9	2006	18
10	2007	38
11	2008	53
12	2009	78
13	2010	131
14	2011	201
15	2012	340
16	2013	458
17	2014	818
18	2015	1336
19	2016	3799
20	2017	4307
21	2018	6569

```
In [ ]: # 8 - Generar un gráfico que contenga el año y número de jugadores
df_new.plot.bar(x='Joined', y='Players')
```

```
Out[ ]: <AxesSubplot: xlabel='Joined'>
```


