

# Ejercicio Visualización de Datos

## Leer archivo fifa\_eda.csv

```
In [ ]: import pandas as pd
import numpy as np
import os
import plotly.express as px

os.chdir('E:\WORK IN PROGRESS\Data Analytics course\parte 2 python\week 18')

df = pd.read_csv('fifa_eda.csv')
```

```
In [ ]: df.sample(5)
```

```
Out[ ]:
```

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Preferred
<b>11060</b>	181521	S. Saunders	34	England	64	64	Wycombe Wanderers	220.0	2.0	Right
<b>7943</b>	229446	D. Vicencio	25	Chile	67	71	San Luis de Quillota	825.0	2.0	Right
<b>14942</b>	237426	H. Nakabayashi	32	Japan	60	60	Sanfrecce Hiroshima	120.0	1.0	Right
<b>9652</b>	235971	Pablo Larrea	24	Spain	66	73	CD Numancia	775.0	3.0	Right
<b>13844</b>	243429	Alberto Redondo	21	Spain	62	72	Getafe CF	425.0	2.0	Right

## Elemento 1: Obtener matriz de correlación y dejarla como heat map

```
In [ ]: # Primero verifico cuales son las variables numéricas que estarán comprendidas en
# la matriz de correlación.
df.dtypes
```

```
Out[ ]: ID                int64
        Name                object
        Age                int64
        Nationality        object
        Overall            int64
        Potential          int64
        Club                object
        Value              float64
        Wage                float64
        Preferred Foot      object
        International Reputation float64
        Skill Moves        float64
        Position            object
        Joined              int64
        Contract Valid Until object
        Height              float64
        Weight              float64
        Release Clause      float64
        dtype: object
```

```
In [ ]: # Genero la matriz de correlaciòn
        df.corr()
```

```
Out[ ]:
```

	ID	Age	Overall	Potential	Value	Wage	International Reputation	Skill Moves	Joined	Height	Weight	Release Clause
ID	1.000000	-0.739208	-0.417025	0.047074	-0.139837	-0.204610	-0.356191	-0.056914	0.206749	-0.090090	-0.191193	-0.121297
Age	-0.739208	1.000000	0.452350	-0.253312	0.078315	0.141145	0.253765	0.027649	-0.202658	0.082506	0.229940	0.058672
Overall	-0.417025	0.452350	1.000000	0.660939	0.631848	0.571926	0.499491	0.414463	-0.169281	0.038527	0.154557	0.597821
Potential	0.047074	-0.253312	0.660939	1.000000	0.579608	0.486413	0.372993	0.354290	-0.047661	-0.009791	-0.006935	0.562346
Value	-0.139837	0.078315	0.631848	0.579608	1.000000	0.858086	0.656158	0.317246	-0.115991	0.002827	0.046702	0.973310
Wage	-0.204610	0.141145	0.571926	0.486413	0.858086	1.000000	0.668635	0.263205	-0.142337	0.019638	0.064764	0.828161
International Reputation	-0.356191	0.253765	0.499491	0.372993	0.656158	0.668635	1.000000	0.208153	-0.133009	0.034881	0.088340	0.620863
Skill Moves	-0.056914	0.027649	0.414463	0.354290	0.317246	0.263205	0.208153	1.000000	0.020674	-0.422090	-0.351193	0.297821
Joined	0.206749	-0.202658	-0.169281	-0.047661	-0.115991	-0.142337	-0.133009	0.020674	1.000000	-0.090090	-0.191193	-0.121297
Height	-0.090090	0.082506	0.038527	-0.009791	0.002827	0.019638	0.034881	-0.422090	-0.090090	1.000000	0.229940	0.058672
Weight	-0.191193	0.229940	0.154557	-0.006935	0.046702	0.064764	0.088340	-0.351193	-0.191193	0.229940	1.000000	0.058672
Release Clause	-0.121297	0.058672	0.597821	0.562346	0.973310	0.828161	0.620863	0.297821	-0.121297	0.058672	0.058672	1.000000

```
In [ ]: # Finalmente, genero el Heatmap de correlacion, el cual muestra de una mejor
        # manera los resultados
        # obtenidos en la matriz de correlaciòn.
        fig = px.imshow(df.corr(), text_auto=True)
        fig.show()
```

## Elemento 2: Gráfico de Correlación que responda

– “Cuál es la relación entre la edad y el overall?” –

## explicar la salida

```
In [ ]: # Utilizo un diagrama de dispersión para responder la pregunta del enunciado.

fig = px.scatter(df, x='Age', y='Overall')
fig.show()
```

```
In [ ]: # El diagrama de dispersion muestra la relación entre las habilidades en general
# de los futbolistas y su edad. # Como se puede apreciar, desde los 16 hasta los
# 33 años existe una relación positiva entre las dos variables.
# Esto quiere decir que a mayor edad, mayor van a ser las habilidades en general
# del futbolista. Sin embargo, después de los 33 años la relación entre las dos
# variables es negativa, ya que a mayor edad disminuyen las habilidades en
# general del futbolista.
```

## Elemento 3: Generar un gráfico de barras por club que indique el número de jugadores

```
In [ ]: # Utilizo la función "groupby" para agrupar los datos (números de jugadores)
# por club.
df_club=df.groupby('Club').size().rename('Number_of_players').reset_index()
df_club
```

```
Out[ ]:      Club  Number_of_players
0  SSV Jahn Regensburg          29
1  1. FC Heidenheim 1846          28
2    1. FC Kaiserslautern          26
3      1. FC Köln              28
4    1. FC Magdeburg          26
...      ...              ...
646  Zagłębie Sosnowiec          25
647   Çaykur Rizespor          30
648      Örebro SK            27
649   Östersunds FK           22
650   Śląsk Wrocław          25
```

651 rows × 2 columns

```
In [ ]: # Ahora, genero el gráfico de barras.
fig = px.bar(df_club,x='Club',y='Number_of_players')
fig.show()
```

```
In [ ]: # Como se puede apreciar en el gráfico, los clubes de fútbol tienen entre
# 20 y 30 jugadores aproximadamente.
```

Elemento 4: Hacer un gráfico multipanel que indique la relación altura (height) vs skill moves, siendo la variable del panel si es zurdo o derecho

```
In [ ]: # utilizo el "Facet charts" para generar los paneles.
fig = px.scatter(df, x='Height', y='Skill Moves', facet_col='Preferred Foot')
fig.show()
```

```
In [ ]: # Es interesante notar que existe una leve diferencia entre la destreza que tiene
# un jugador con respecto a su altura si usara más la pierna izquierda con
# respecto a la derecha. En el primer panel se evidencia que entre más alto sea
# el futbolista menor va a ser su destreza. Sin embargo, dicho resultado no es
# tan evidente en los futbolistas que usan más la pierna derecha, ya que aunque
# si la destreza de un futbolista tiende a disminuir mientras más aumenta su
# altura, dicha disminución no es tan evidente como la que se expresa en el
# primer panel.
# tal vez esto se debe a que hay una mayor cantidad de futbolistas que usan más
# la pierna derecha con respecto a la pierna izquierda.
```

```
In [ ]: df.groupby('Preferred Foot').size().rename('Number_of_players').reset_index()
```

```
Out[ ]: Preferred Foot  Number_of_players
0           Left           4211
1           Right          13996
```

## Elemento 5: Un gráfico que crea importante para mostrar a los potenciales cracks

```
In [ ]: # Para el gráfico que quiero hacer lo primero que hice fue filtrar el dataset
# inicial por la columna "value" de modo descendente tomando solamente los
# primeros 50 datos. De esta manera obtengo los 50 futbolistas con el mayor
# "valor".
new_df=df.sort_values(by=['Value'], ascending=False).head(50)
new_df.head(10)
```

Out[ ]:

	ID	Name	Age	Nationality	Overall	Potential	Club	Value	Wage	Prefer
2	190871	Neymar Jr	26	Brazil	92	93	Paris Saint-Germain	118500.0	290.0	F
0	158023	L. Messi	31	Argentina	94	94	FC Barcelona	110500.0	565.0	
4	192985	K. De Bruyne	27	Belgium	91	92	Manchester City	102000.0	355.0	F
5	183277	E. Hazard	27	Belgium	91	91	Chelsea	93000.0	340.0	F
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	F
25	231747	K. Mbappé	19	France	88	95	Paris Saint-Germain	81000.0	100.0	F
7	176580	L. Suárez	31	Uruguay	91	91	FC Barcelona	80000.0	455.0	F
17	194765	A. Griezmann	27	France	89	90	Atlético Madrid	78000.0	145.0	
10	188545	R. Lewandowski	29	Poland	90	90	FC Bayern München	77000.0	205.0	F

◀  ▶

```
In [ ]: # Ahora, partiendo del dataset anterior, genero un diagrama de dispersión entre
# las habilidades generales de dichos deportistas y su nacionalidad.
fig = px.scatter(new_df, x='Nationality', y='Overall',
title= "Relación entre las habilidades y el país natal del top 50 de los futbolistas",
fig.show()
```

```
In [ ]: # El gráfico automáticamente los clasifica de mayor a menor por la variable "Overall"
# Siendo en este caso Brasil, Argentina y Bélgica los países que, en promedio,
# tienen futbolistas con un mayor "valor".
```

```
In [ ]: # Sin embargo, me parecía importante determinar el número de jugadores
# que tiene cada país en esta top 50 lista.
```

```
In [ ]: new_table=new_df.groupby('Nationality').size().rename('Number_of_players').reset_index()
new_table
```

Out[ ]:

	Nationality	Number_of_players
8	France	7
17	Spain	7
0	Argentina	5
2	Brazil	4
1	Belgium	4
11	Italy	3
10	Germany	3
7	England	2
13	Portugal	2
14	Senegal	2
18	Uruguay	2
6	Egypt	1
9	Gabon	1
5	Denmark	1
4	Croatia	1
12	Poland	1
15	Serbia	1
16	Slovenia	1
3	Colombia	1
19	Wales	1

```
In [ ]: fig=px.bar(new_table, x='Nationality', y='Number_of_players', title="Cantidad de jugadores por país")
fig.show()
```

```
In [ ]: # De esta manera, se ve que Francia y España
# son los países con mayor cantidad de jugadores en el top 50. Aunque si en promedio
# filtrandolos por la variable "overall" están ubicados en la 5 y 11 posición.
# En cambio Argentina es el único país que está en el top 3 de los 2 gráficos.
```