

# notebook

October 20, 2023

## 1 Tarea Semana 4

### 1.0.1 Jessy Suarez A. / LIDE / Modulo Python

#### 1.1 Instrucciones

Utilizando los datos sobre los partidos de la copa de fútbol masculina en español aquí: <https://github.com/cienciadedatos/datos-de-miercoles/tree/master/datos/2019/2019-04-10> van a crear dos gráficos que comparen dos o más países.

Acceder los datos de manera remota, es decir no bajándolos a su computadora.

Crear dos gráficos en el que comparen a al menos dos países diferentes. Si quieren incluir a más países pueden hacerlo.

Gráfico 1 deberá ser un gráfico en el que los datos de los países comparados aparezcan en el mismo cuadro

Gráfico 2 deberá ser un gráfico en el que los datos de cada país aparezca en un subgráfico diferente. Es decir, una figura tenga al menos dos recuadros (uno por cada país en la comparación).

Los gráficos deberán tener un título general

Los gráficos deberán tener etiquetas en cada eje

Deberán mostrar cómo guardar un gráfico

Deberán incluir solamente documentos relevantes en su repositorio

#### 1.2 Codigo

##### 1.2.1 Importar paquetes

```
[1]: import numpy as np
import pandas as pd
```

##### 1.2.2 Importar datos

```
[2]: partidos_cm=pd.read_csv('https://raw.githubusercontent.com/cienciadedatos/
↳datos-de-miercoles/master/datos/2019/2019-04-10/partidos.txt',
↳delimiter='\t')
partidos_cm
```

```
[2]:      anio anfitrión ... equipo_1_final equipo_2_final
0    1930   Uruguay ...           4           1
1    1930   Uruguay ...           3           0
2    1930   Uruguay ...           2           1
3    1930   Uruguay ...           3           1
4    1930   Uruguay ...           1           0
..    ...    ...    ...
895  2018    Rusia ...           0           2
896  2018    Rusia ...           1           0
897  2018    Rusia ...           2           1
898  2018    Rusia ...           2           0
899  2018    Rusia ...           4           2
```

[900 rows x 10 columns]

### 1.2.3 Líneas de código

```
[3]: partidos_cm.columns.values
```

```
[3]: array(['anio', 'anfitrión', 'estadio', 'ciudad', 'partido_orden', 'fecha',
        'equipo_1', 'equipo_2', 'equipo_1_final', 'equipo_2_final'],
        dtype=object)
```

```
[ ]: # Países que deseo analizar
países = ['Francia', 'Argentina', 'México', 'Países Bajos', 'Corea del Sur']
```

```
[4]: partidos_equipos =partidos_cm[(partidos_cm['equipo_1'].isin(['Francia',
↵ 'Argentina', 'Mexico','Países Bajos','Corea del Sur'])) &
↵ (partidos_cm['equipo_2'].isin(['Francia', 'Argentina', 'Mexico','Países
↵ Bajos','Corea del Sur'])))]

partidos_equipos
```

```
[4]:      anio      anfitrión ... equipo_1_final equipo_2_final
0    1930      Uruguay ...           4           1
4    1930      Uruguay ...           1           0
10   1930      Uruguay ...           6           3
84   1954        Suiza ...           3           2
172  1966  Inglaterra ...           1           1
257  1974  Alemania occidental ...           4           0
278  1978      Argentina ...           2           1
307  1978      Argentina ...           1           3
363  1986        Mexico ...           3           1
525  1998      Francia ...           1           3
540  1998      Francia ...           5           0
557  1998      Francia ...           2           2
575  1998      Francia ...           2           1
```

672	2006	Alemania	...	1	0
680	2006	Alemania	...	0	0
693	2006	Alemania	...	2	1
725	2010	Sudáfrica	...	0	2
726	2010	Sudáfrica	...	4	1
759	2010	Sudáfrica	...	3	1
822	2014	Brasil	...	2	1
833	2014	Brasil	...	2	4
862	2018	Rusia	...	1	2
885	2018	Rusia	...	4	3

[23 rows x 10 columns]

```
[41]: # suma de goles totales por paises seleccionados en equipo_1_final
sum_pais_1 = partidos Equipos.groupby('equipo_1')['equipo_1_final'].sum()
sum_pais_1
```

```
[41]: equipo_1
Argentina      21
Corea del Sur   2
Francia        13
Países Bajos   18
Name: equipo_1_final, dtype: int64
```

```
[42]: # suma de goles totales por paises seleccionados en equipo_2_final
sum_pais_2 = partidos Equipos.groupby('equipo_2')['equipo_2_final'].sum()
sum_pais_2
```

```
[42]: equipo_2
Argentina      11
Corea del Sur   2
Francia         1
Mexico         19
Name: equipo_2_final, dtype: int64
```

```
[43]: # suma de puntajes del mayor goleador
sum_puntajes = pd.concat([sum_pais_1, sum_pais_2])
puntajes_totales = sum_pais_1.add(sum_pais_2, fill_value=0)
puntajes_totales
```

```
[43]: Argentina      32.0
Corea del Sur      4.0
Francia            14.0
Mexico            19.0
Países Bajos       18.0
dtype: float64
```

```
[44]: # DataFrame con la columna "total"
nuevo_dataframe = pd.DataFrame({'País': puntajes_totales.index, 'Total':
    ↪puntajes_totales.values})
nuevo_dataframe=nuevo_dataframe.sort_values(by='Total',ascending=False)
nuevo_dataframe
```

```
[44]:
```

	País	Total
0	Argentina	32.0
3	Mexico	19.0
4	Países Bajos	18.0
2	Francia	14.0
1	Corea del Sur	4.0

Podemos observar que de los paises elegidos para el analisis el Pais con mayor numero de anotaciones es Argentina con 32 goles a los largo del tiempo.

```
[64]: #analisis de goles por anio para comparacion por paises
goles_por_anio_equipo_1 = partidos Equipos.groupby(['anio',
    ↪'equipo_1'])['equipo_1_final'].sum().reset_index()
goles_por_anio_equipo_2 = partidos Equipos.groupby(['anio',
    ↪'equipo_2'])['equipo_2_final'].sum().reset_index()
```

```
[65]: # Renombrar las columnas
goles_por_anio_equipo_1 = goles_por_anio_equipo_1.rename(columns={'equipo_1':
    ↪'Equipo', 'equipo_1_final': 'Goles'})
goles_por_anio_equipo_2 = goles_por_anio_equipo_2.rename(columns={'equipo_2':
    ↪'Equipo', 'equipo_2_final': 'Goles'})

# Combinar los DataFrames
goles_por_anio = pd.concat([goles_por_anio_equipo_1, goles_por_anio_equipo_2],
    ↪ignore_index=True)

# Ordenar los datos por año y equipo
goles_por_anio = goles_por_anio.sort_values(by=['anio', 'Equipo'])

print("Goles por año y equipo:")
goles_por_anio
```

Goles por año y equipo:

```
[65]:
```

	anio	Equipo	Goles
0	1930	Argentina	7
1	1930	Francia	4
18	1930	Francia	0
19	1930	Mexico	4
2	1954	Francia	3
20	1954	Mexico	2

3	1966	Francia	1
21	1966	Mexico	1
22	1974	Argentina	0
4	1974	Países Bajos	4
5	1978	Argentina	2
23	1978	Argentina	3
24	1978	Francia	1
6	1978	Países Bajos	1
7	1986	Argentina	3
25	1986	Corea del Sur	1
26	1998	Argentina	1
8	1998	Corea del Sur	1
27	1998	Corea del Sur	0
28	1998	Mexico	5
9	1998	Países Bajos	9
10	2006	Argentina	2
29	2006	Argentina	0
30	2006	Corea del Sur	0
11	2006	Francia	1
31	2006	Mexico	1
12	2006	Países Bajos	0
13	2010	Argentina	7
32	2010	Corea del Sur	1
14	2010	Francia	0
33	2010	Mexico	3
34	2014	Argentina	4
35	2014	Mexico	1
15	2014	Países Bajos	4
36	2018	Argentina	3
16	2018	Corea del Sur	1
17	2018	Francia	4
37	2018	Mexico	2

#### 1.2.4 Resumen tabla final de goles por año por cada equipo seleccionado

```
[66]: # Agrupar y sumar los goles por año y equipo
goles_por_anio_paises = goles_por_anio.groupby(['anio', 'Equipo'])['Goles'].
    ↪sum().reset_index()

# Renombrar la columna de goles
goles_por_anio_paises = goles_por_anio_paises.rename(columns={'Goles': 'Goles_
    ↪Totales'})

# Mostrar la tabla final
print("Goles por año y por países:")
goles_por_anio_paises
```

Goles por año y por países:

```
[66]:
```

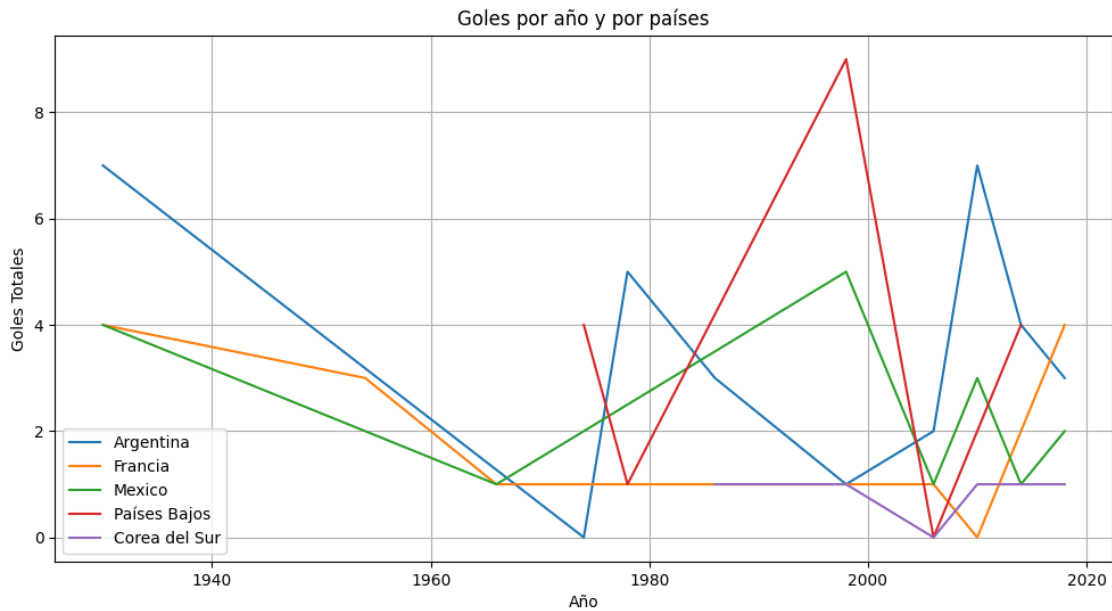
	anio	Equipo	Goles Totales
0	1930	Argentina	7
1	1930	Francia	4
2	1930	Mexico	4
3	1954	Francia	3
4	1954	Mexico	2
5	1966	Francia	1
6	1966	Mexico	1
7	1974	Argentina	0
8	1974	Países Bajos	4
9	1978	Argentina	5
10	1978	Francia	1
11	1978	Países Bajos	1
12	1986	Argentina	3
13	1986	Corea del Sur	1
14	1998	Argentina	1
15	1998	Corea del Sur	1
16	1998	Mexico	5
17	1998	Países Bajos	9
18	2006	Argentina	2
19	2006	Corea del Sur	0
20	2006	Francia	1
21	2006	Mexico	1
22	2006	Países Bajos	0
23	2010	Argentina	7
24	2010	Corea del Sur	1
25	2010	Francia	0
26	2010	Mexico	3
27	2014	Argentina	4
28	2014	Mexico	1
29	2014	Países Bajos	4
30	2018	Argentina	3
31	2018	Corea del Sur	1
32	2018	Francia	4
33	2018	Mexico	2

### 1.2.5 Crear graficos y guardar

```
[68]: import matplotlib.pyplot as plt
```

```
[69]: plt.figure(figsize=(12, 6))
      for equipo in goles_por_anio_paises['Equipo'].unique():
          data = goles_por_anio_paises[goles_por_anio_paises['Equipo'] == equipo]
          plt.plot(data['anio'], data['Goles Totales'], label=equipo)
```

```
plt.xlabel('Año')
plt.ylabel('Goles Totales')
plt.title('Goles por año y por países')
plt.legend()
plt.grid(True)
plt.show()
plt.savefig('grafico_1.png')
```



<Figure size 640x480 with 0 Axes>

```
[70]: #colores para cada pais
colores = ['steelblue', 'coral', 'green', 'red', 'purple']

# lista de países
países_2 = goles_por_anio_paises['Equipo'].unique()

# Crear subgráficos para cada país
fig, axes = plt.subplots(len(países_2), 1, figsize=(10, 5 * len(países_2)),
    ↳sharex=True)

# crear un gráfico para cada uno con colores
for i, pais in enumerate(países_2):
    data = goles_por_anio_paises[goles_por_anio_paises['Equipo'] == pais]
    ax = axes[i]
    color = colores[i % len(colores)]
    ax.plot(data['anio'], data['Goles Totales'], color=color)
    ax.set_title(f'Goles por año en {pais}')
```

```
ax.set_xlabel('Año')
ax.set_ylabel('Goles Totales')

# Espacio entre los subgráficos
plt.tight_layout()

# Guardar el gráfico como un archivo PNG
plt.savefig('grafico_2.png')

plt.show()
```



