

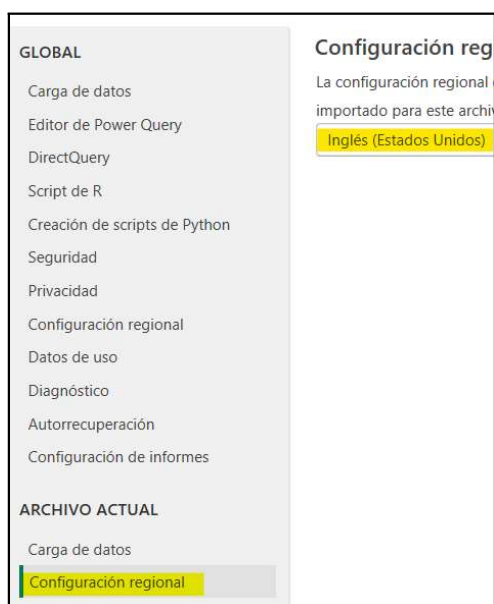
SPRINT 8

NIVEL 1

CARGA DE DATAFRAMES

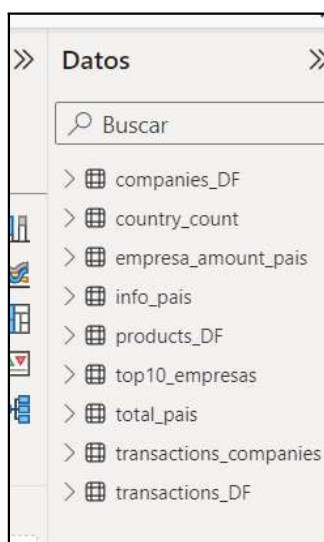
Conecto con Python y copio el script para importar todas las tablas de python de la BBDD “sprint_4”.

Algunas columnas de importes, como por ejemplo “Importe” de la tabla “transactions” han perdido su configuración original al cargar los datos por lo que cambio la configuración regional del archivo a formato inglés de US y copio de nuevo el script. Antes de cargar, transformo el formato de datos de las columnas tipo float en Python a decimal fijo en Power BI:



	1.2 Importe	AB _C declined	AB _C p
1	466,92	0	71, 1
2	49,53	0	47, 9
3	92,61	0	47, 6
4	394,18	0	89, 8
5	279,93	0	43, 3
6	340,87	1	47, 2
7	303,05	1	67, 7
8	430,49	0	29, 4
9	288,81	1	19, 4
0	103,44	1	89, 3
1	428,69	1	83, 4
2	252,47	1	7, 47
3	447,11	0	37, 1
4	465,31	0	89, 1
5	293,57	0	59
6	27,85	1	43, 8
7	157,2	0	29
8	172,01	0	1, 67
9	348,88	1	29

Y compruebo que se hayan cargado las tablas originales de MySQL así como las auxiliares creadas en Python:



- Visualización 1

Cambio el nombre de la dataframe “transactions_DF” a “dataset” ya que el preámbulo del script interpreta que el dataframe principal se llama dataset. Este preámbulo crea y procesa un dataframe a partir de una tabla inicial.

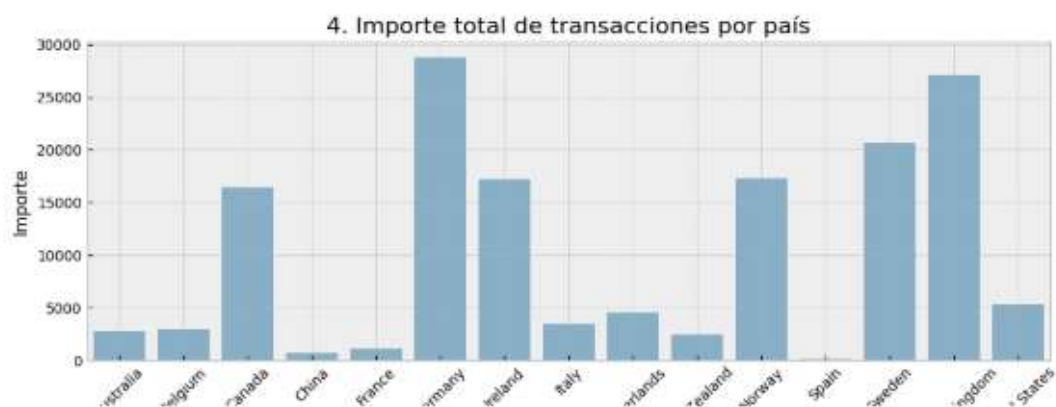
Este cambio hay que repetirlo en todos los scripts de las visualizaciones.

```
15 # dataset = pandas.DataFrame(transactions_DF)
16
17 media = round(np.mean(transactions_DF['Importe']),2)
18
19 sns.histplot(transactions_DF['Importe'])
20
21 plt.title('1. Distribución transacciones', fontsize= 20)
22 plt.xlabel('Importe', size=10)
23 plt.ylabel('Frecuencia', size=10)
24 plt.axvline(media, color='darkblue', linestyle='dashed', linewidth=1.5, label='Media')
25
26 plt.show()
```

```
15 # dataset = pandas.DataFrame(transactions_DF)
16
17 media = round(np.mean(dataset['Importe']),2)
18
19 sns.histplot(dataset['Importe'])
20
21 plt.title('1. Distribución transacciones', fontsize= 20)
22 plt.xlabel('Importe', size=10)
23 plt.ylabel('Frecuencia', size=10)
24 plt.axvline(media, color='darkblue', linestyle='dashed', linewidth=1.5, label='Media')
25
26 plt.show()
```

- Visualizaciones 4 y 6

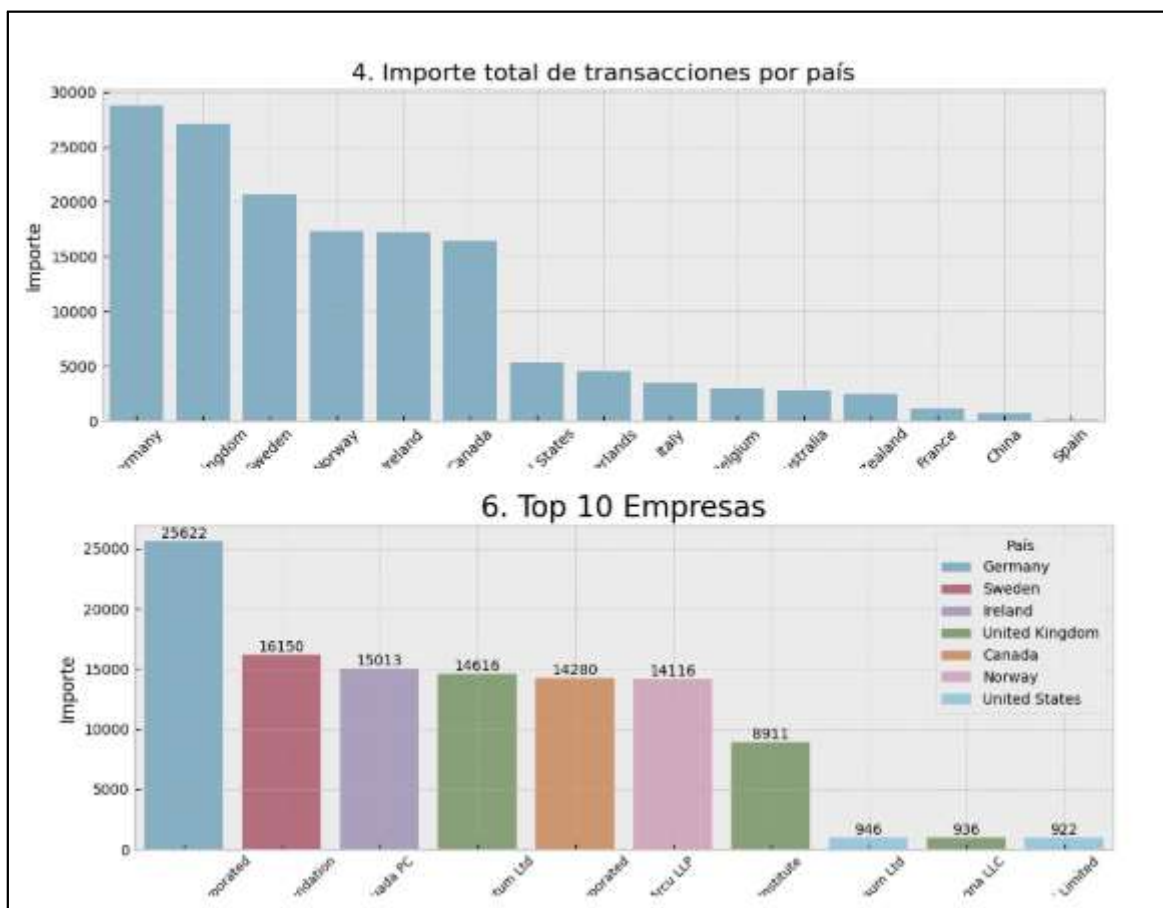
Todo y que en el código de la creación de esta dataframe en Python los valores estaban ordenados de mayor a menor, al copiar el script de la visualización en Power BI, se desordenaban las columnas:



Para solucionarlo, creo nueva variable en el escritor de scripts de python y aplico la función "sort_values":

(Repito el mismo proceso en la visualización del ejercicio 6):

```
10
11 plt.style.use('bmh')
12 plt.subplots(figsize=(12,4))
13
14 dataset_sorted = dataset.sort_values(by='Importe', ascending=False)
15
16 sns.barplot(data=dataset_sorted,
17             x='País',
18             y='Importe',
19             alpha=.6)
```



Los títulos de los ejes X salían cortados en el gráfico (visualizaciones 4 y 6):



La solución ha pasado por escribir esta función en el editor de scripts de power BI que ajusta automáticamente los márgenes de la figura para que todo el contenido, incluidas las etiquetas, se ajuste dentro de la visualización. Esto puede evitar que la etiqueta del eje X quede cortada:

```

28 plt.grid(True)
29
30 for container in ax.containers:
31     ax.bar_label(container, fmt='%d', label_type='edge')
32
33 plt.tight_layout()
34
35 plt.show()

```

