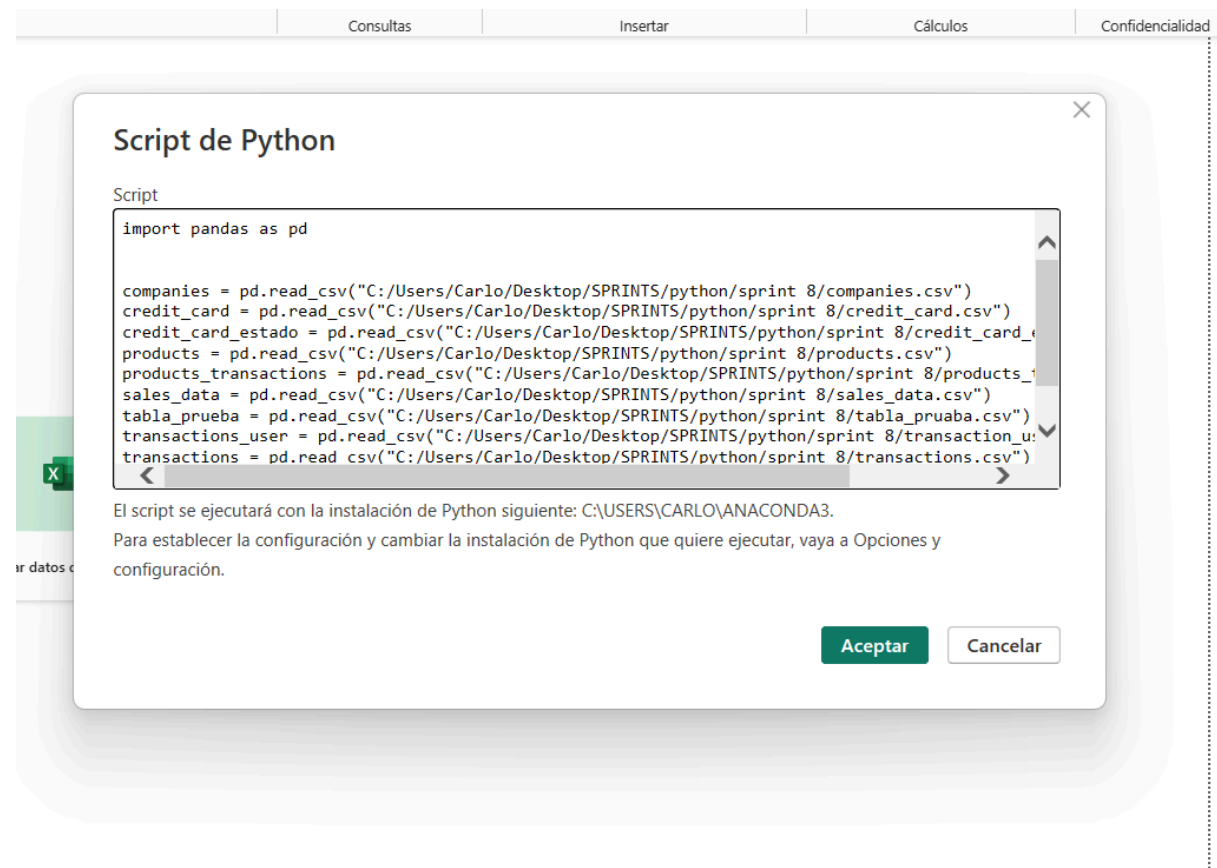
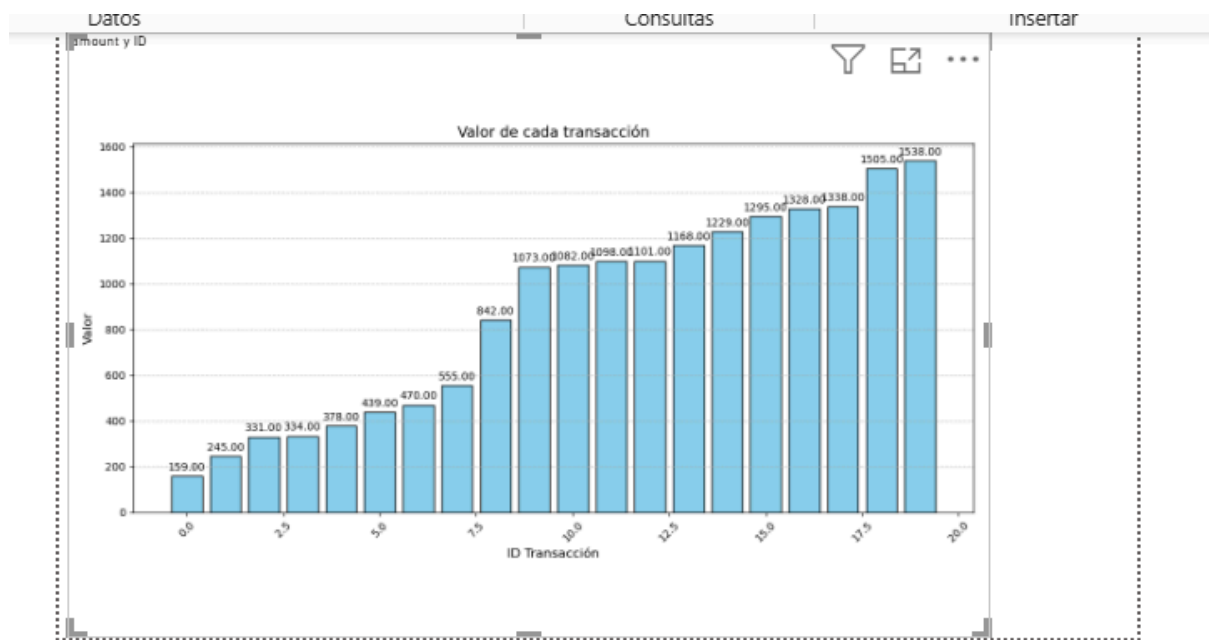


El primer paso es crear un script para cargar los archivos csv creados al final del ejercicio 8.1



Nivel I Ej 1



```
import pandas as pd
import seaborn as sns
import sqlalchemy
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np
import scipy.stats as stats

import matplotlib.pyplot as plt

grafico1 = dataset["amount"].head(20)

plt.figure(figsize=(12, 6))

bars = plt.bar(grafico1.index, grafico1, color="skyblue", edgecolor="black")

plt.xlabel("ID Transacción", fontsize=12)
plt.ylabel("Valor", fontsize=12)
plt.title("Valor de cada transacción", fontsize=14)

plt.xticks(rotation=45)

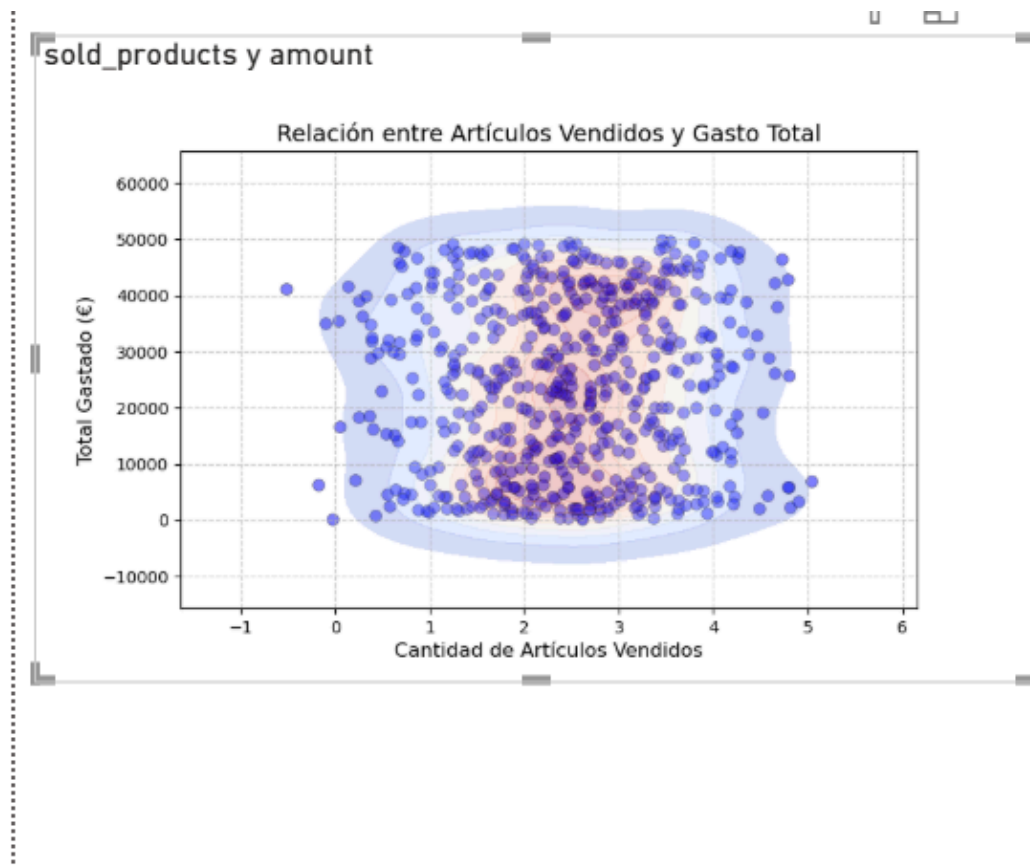
plt.grid(axis="y", linestyle="--", alpha=0.7)

for i, valor in enumerate(grafico1):
    plt.text(i, valor + max(grafico1)*0.01, f"{valor:.2f}", ha="center",
va="bottom", fontsize=10)

plt.tight_layout()

plt.show()
```

Nivel I Ej 2



Editor de scripts de Python

```
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

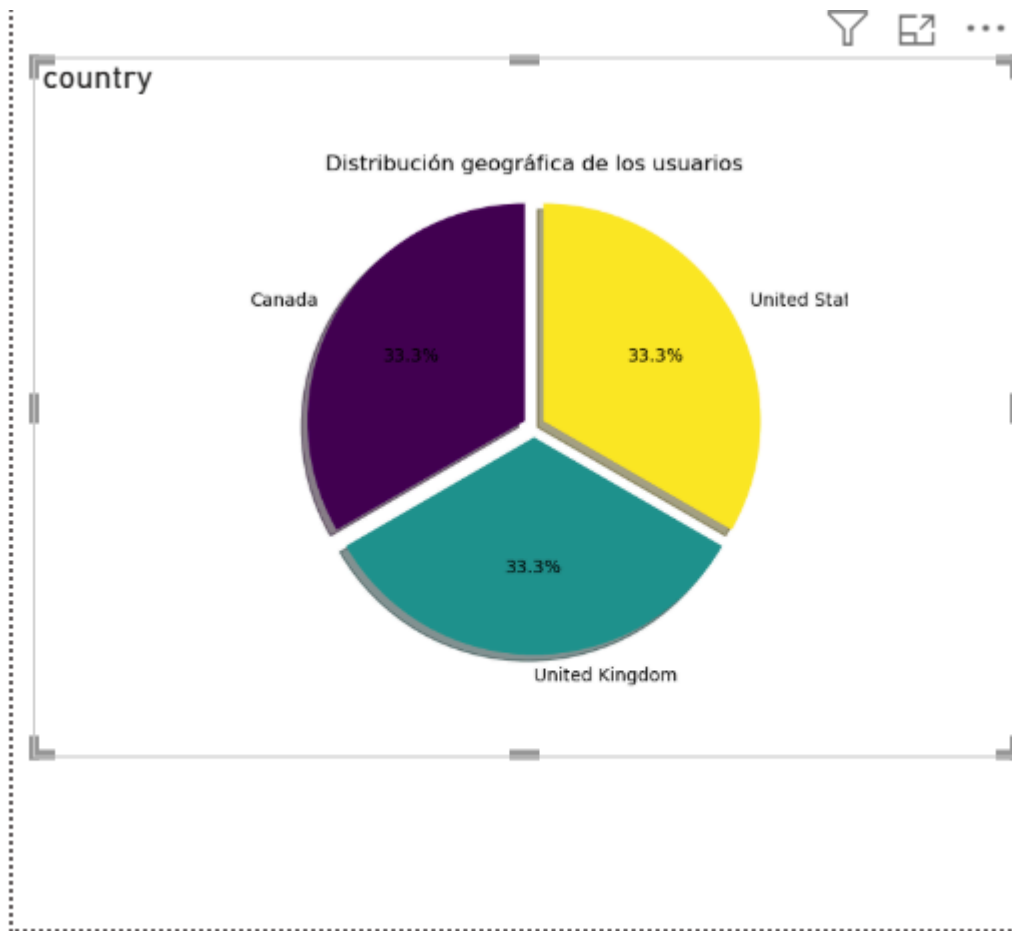
df_sales = dataset

plt.figure(figsize=(8, 5))

jitter_sold = df_sales["sold_products"] + np.random.normal(scale=0.5,
size=len(df_sales))
jitter_amount = df_sales["amount"] + np.random.normal(scale=5,
size=len(df_sales))

plt.scatter(jitter_sold, jitter_amount, color="blue", alpha=0.5, s=50,
edgecolors="black", linewidth=0.5)
sns.kdeplot(x=jitter_sold, y=jitter_amount, cmap="coolwarm", fill=True,
alpha=0.25)
plt.xlabel("Cantidad de Artículos Vendidos", fontsize=12)
plt.ylabel("Total Gastado (€)", fontsize=12)
plt.grid(True, linestyle="--", alpha=0.6)
plt.show()
```

Nivel I Ej 3



tor de scripts de Python

```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

datos = dataset
datos = dataset["country"].value_counts()
colores = plt.cm.viridis(np.linspace(0, 1, len(datos)))
explode = [0.05] * len(datos)
plt.figure(figsize=(5, 5))
plt.pie(datos,
        labels=datos.index,
        autopct="%1.1f%%",
        startangle=90,
        shadow=True,
        explode=explode,
        colors=colores)
plt.axis('equal')
plt.show()
```

Nivel I Ej 4



Editor de scripts de Python

```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

grafico4 = dataset
grafico4 = dataset.groupby("country")["ID"].count()

plt.figure(figsize=(10, 6))

# crear el gráfico de barras con colores personalizados
bars = grafico4.plot(kind="bar", color="skyblue", edgecolor="black")

plt.xlabel("País", fontsize=12)
plt.ylabel("Cantidad de IDs", fontsize=12)

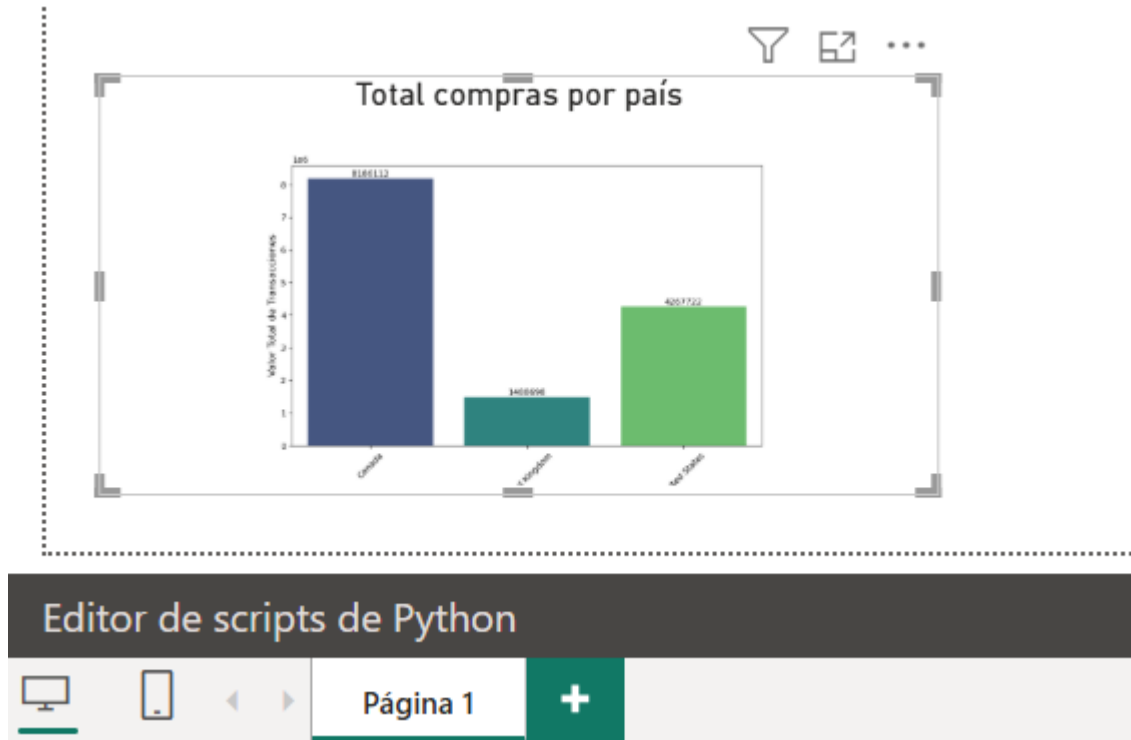
plt.xticks(rotation=45)

# cuadrícula
plt.grid(axis="y", linestyle="--", alpha=0.7)

# valores encima de las barras
for i, valor in enumerate(grafico4):
    plt.text(i, valor + max(grafico4)*0.01, f"{valor}", ha="center",
             fontsize=10)

# ajustar el layout para evitar solapamientos
plt.tight_layout()

plt.show()
```



```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

conteo =
dataset.groupby(["country"])["amount"].sum().rename(columns={"amount":
"Frecuency"}).reset_index()

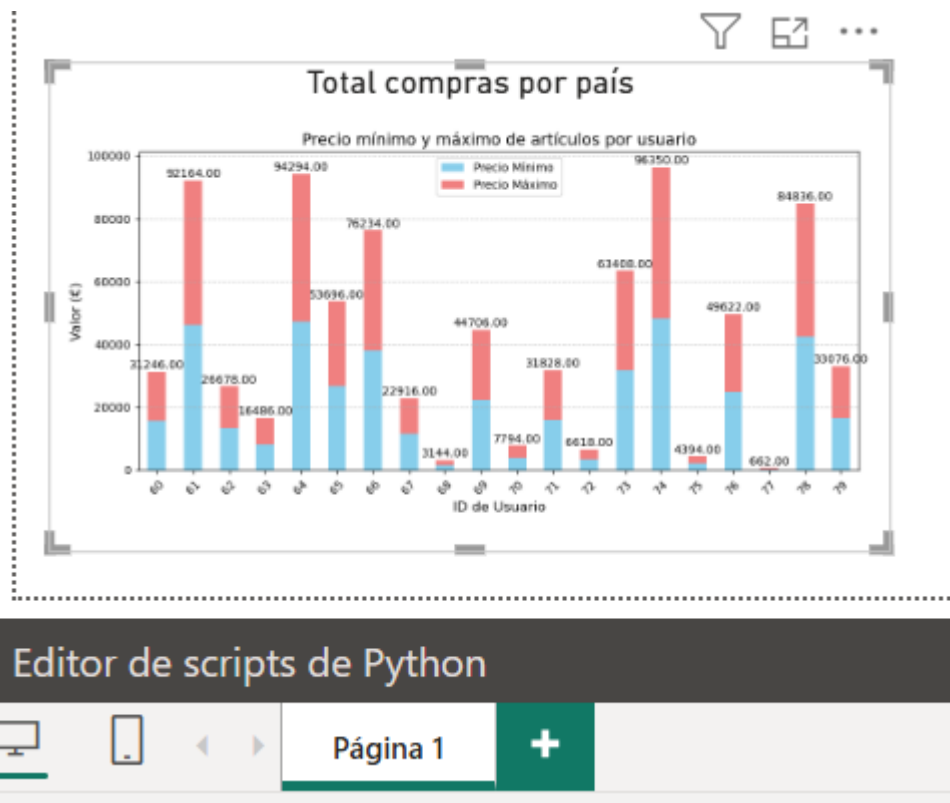
plt.figure(figsize=(10, 6))
ax = sns.barplot(x="country", y="Frecuency", hue="country", data=conteo,
palette="viridis")

plt.xlabel("País", fontsize=12)
plt.ylabel("Valor Total de Transacciones", fontsize=12)

plt.xticks(rotation=45, fontsize=10)

for container in ax.containers:
    ax.bar_label(container, fmt="%d", fontsize=10)

plt.show()
```



```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

grafico6 = dataset

grafico6 = dataset.groupby("user_id")["amount"].agg(['min', 'max']).head(20)

grafico6.plot(kind="bar", stacked=True, figsize=(10, 5), color=["skyblue",
"lightcoral"])

plt.xlabel("ID de Usuario", fontsize=12)
plt.ylabel("Valor (€)", fontsize=12)
plt.title("Precio mínimo y máximo de artículos por usuario", fontsize=14)

plt.xticks(rotation=45, fontsize=10)

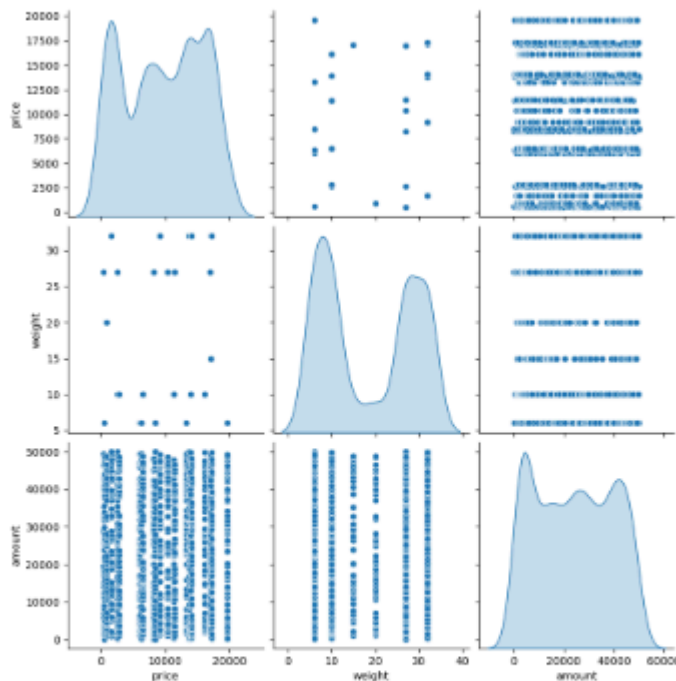
plt.grid(axis="y", linestyle="--", alpha=0.7)

plt.legend(["Precio Mínimo", "Precio Máximo"], fontsize=10)

for i, valor in enumerate(grafico6.sum(axis=1)):
plt.text(i, valor + max(grafico6.sum(axis=1)) * 0.01, f"{valor:.2f}",
ha="center", fontsize=10)
```

```
plt.tight_layout()  
plt.subplots_adjust(bottom=0.15)  
  
plt.show()
```


Relación peso, precio y valor

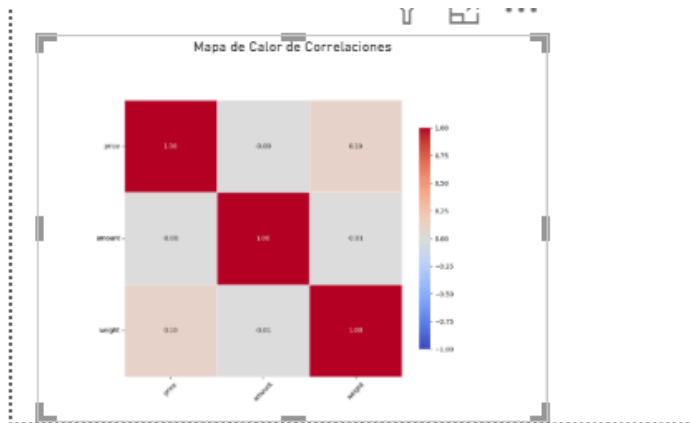


Editor de scripts de Python

```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

grafico4 = dataset
grafico4 = dataset.groupby("country")["ID"].count()
plt.figure(figsize=(10, 6))
bars = grafico4.plot(kind="bar", color="skyblue", edgecolor="black")
plt.xlabel("País", fontsize=12)
plt.ylabel("Cantidad de IDs", fontsize=12)
plt.xticks(rotation=45)
plt.grid(axis="y", linestyle="--", alpha=0.7)
for i, valor in enumerate(grafico4):
    plt.text(i, valor + max(grafico4)*0.01, f"{valor}", ha="center",
    fontsize=10)
plt.tight_layout()

plt.show()
```



Editor de scripts de Python

```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

df_productos = dataset

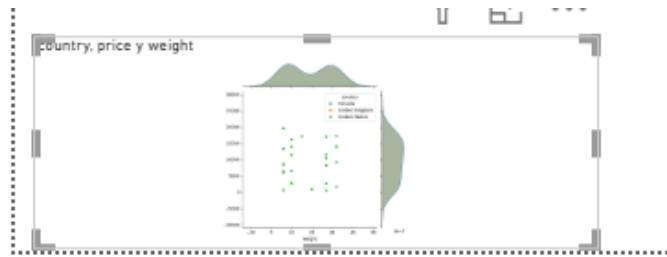
correlacion = df_productos.corr()

plt.figure(figsize=(10, 8))

sns.heatmap(correlacion, annot=True, cmap="coolwarm", fmt=".2f",
            vmin=-1, vmax=1, linewidths=0.5, cbar_kws={"shrink":
0.8})

plt.xticks(rotation=45, fontsize=10)
plt.yticks(rotation=0, fontsize=10)

plt.show()
```



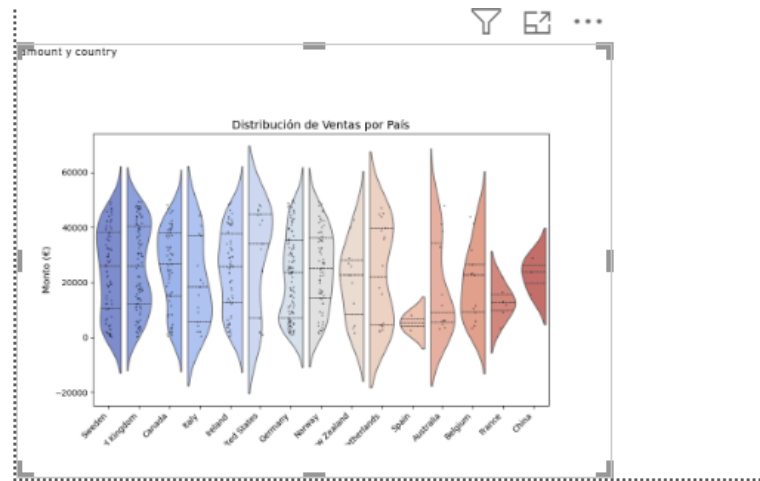
Editor de scripts de Python

⚠ Las filas duplicadas se quitarán de los datos.

```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

sns.jointplot(x="weight", y="price", data=dataset, kind="scatter",
hue="country")

plt.show()
```



Editor de scripts de Python

```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

df_company = dataset

plt.figure(figsize=(10, 6))

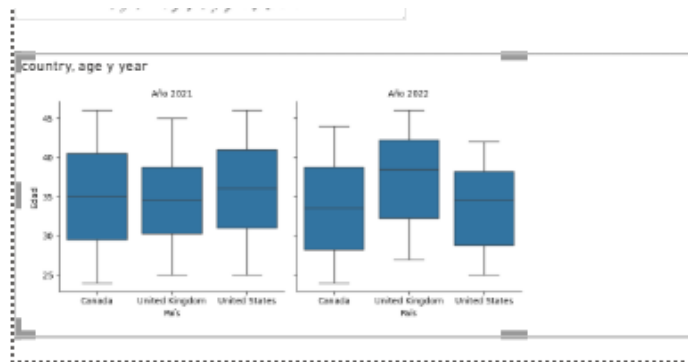
sns.violinplot(x="country", y="amount", hue="country",
data=df_company,
              palette="coolwarm", split=True,
density_norm="width", inner="quartile",
              legend=False, alpha=0.8)

sns.stripplot(x="country", y="amount", data=df_company,
color="black", alpha=0.5, jitter=True, size=2)

plt.xlabel("País", fontsize=12)
plt.ylabel("Monto (€)", fontsize=12)
plt.title("Distribución de Ventas por País", fontsize=14)

plt.xticks(rotation=45, ha="right")

plt.show()
```



Editor de scripts de Python

```
import pandas as pd
import seaborn as sns
import matplotlib as mpl
import matplotlib.pyplot as plt
import numpy as np

df_sales2 = dataset

g = sns.FacetGrid(df_sales2, col="year", height=10, aspect=1,
col_wrap=3)

g.map(sns.boxplot, "country", "age")

g.fig.suptitle("Distribución de Edad por País y Año", fontsize=16,
fontweight="bold")
g.fig.subplots_adjust(top=0.9, bottom=0.15, left=0.3, right=0.9)

g.set_axis_labels("País", "Edad")
g.set_titles("Año {col_name}")

plt.show()
```