

~~HENRY~~



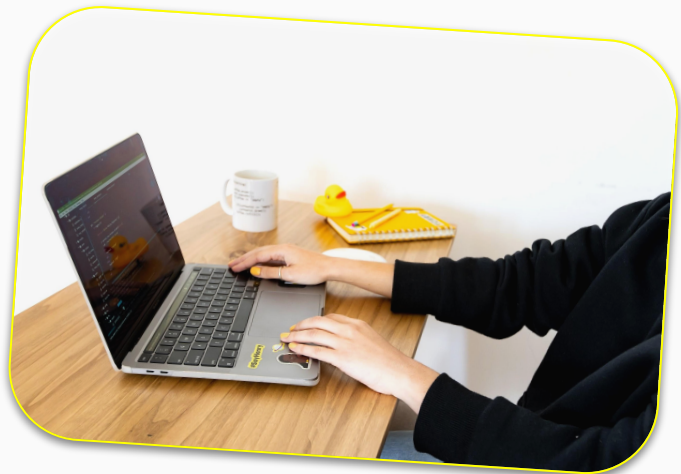
seaborn

Data Science





Agenda



- Script de Python en Power BI
- Cargar datos mediante un script
- Visualizaciones de python en Power BI
- Limitaciones conocidas
- Seaborn



OBJETIVOS DE LA CLASE

Al finalizar esta lecture estarás en la capacidad de...

- Integrar Python en Power BI.
- Habilitar secuencias de comandos de Python.
- Conocer Seaborn



Al **finalizar** cada uno de los temas,
tendremos un **espacio de consultas**.



Hay un **mentor** asignado para
responder el **Q&A**.

¡Pregunta, pregunta, pregunta! :D



Script de Python en Power BI





¿Cómo?

- En Power BI, se pueden ejecutar scripts de Python e importar los conjuntos de datos resultantes en un modelo de datos de Power BI Desktop.
- La integración de Power BI con Python requiere la instalación de dos paquetes de Python:



pandas

matplotlib



Habilitar secuencias de comandos de Python

Para habilitar las secuencias de comandos de Python:

1. En Power BI Desktop, seleccionar:
 Archivo > **Opciones y configuración** > **Opciones** > **Creación de scripts de Python**. Aparecerá la página de opciones de script de Python.
2. Si es necesario, especificar la ruta de instalación local de Python en Directorios principales de Python detectados.
3. Seleccionar **Aceptar**.

Una vez que se especifique la instalación de Python, estará listo para comenzar a ejecutar scripts de Python en Power BI Desktop.



Preparar un script de Python

- Primero, crear un script en su entorno de desarrollo local de Python y asegurarse de que se ejecute correctamente.

Por ejemplo:

```
import pandas as pd
data = [['Alex',10],['Bob',12],['Clarke',13]]
df = pd.DataFrame(data,columns=['Name','Age'],dtype=float)
print (df)
```

(script de Python simple que importa pandas y usa un DataFrame)



Preparar un script de Python

Limitaciones:

- Sólo se importan tramas de datos pandas (los datos a importar a Power BI deben estar representados en un DataFrame).
- Cualquier script de Python que se ejecute durante más de 30 minutos de tiempo de espera devolverá un error.
- Las llamadas interactivas en el script de Python, como esperar la entrada del usuario, detienen la ejecución del script.
- Al establecer el directorio de trabajo dentro del script de Python, debe definir una ruta completa al directorio de trabajo, en lugar de una ruta relativa.
- Actualmente no se admiten tablas anidadas.

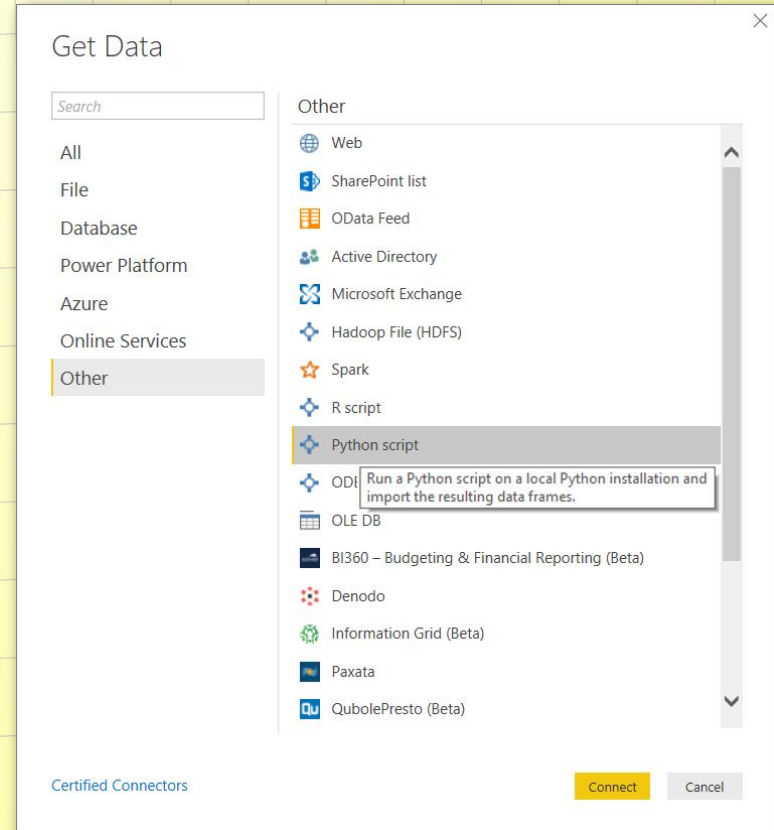


**cargar datos
mediante un script**



Pasos

1. En la cinta Inicio, seleccionar: **Obtener datos** > **Más**. También se puede escribir "python" en el buscador.
2. Seleccionar **Otros** > **Script de Python**.
3. Seleccionar **Conectar**.

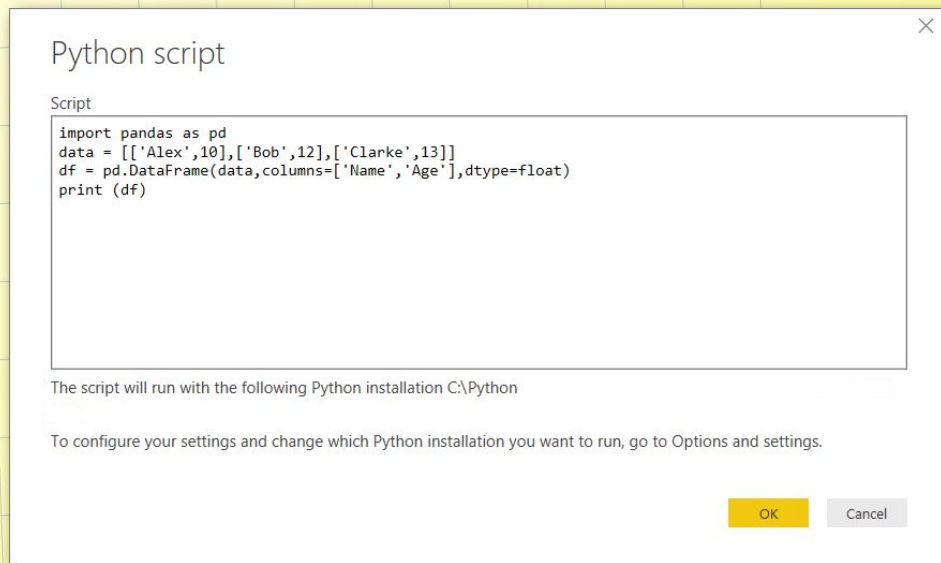




Pasos

4. Copiar el script en el cuadro de diálogo de script de Python que aparece.

5. Seleccionar **OK**.

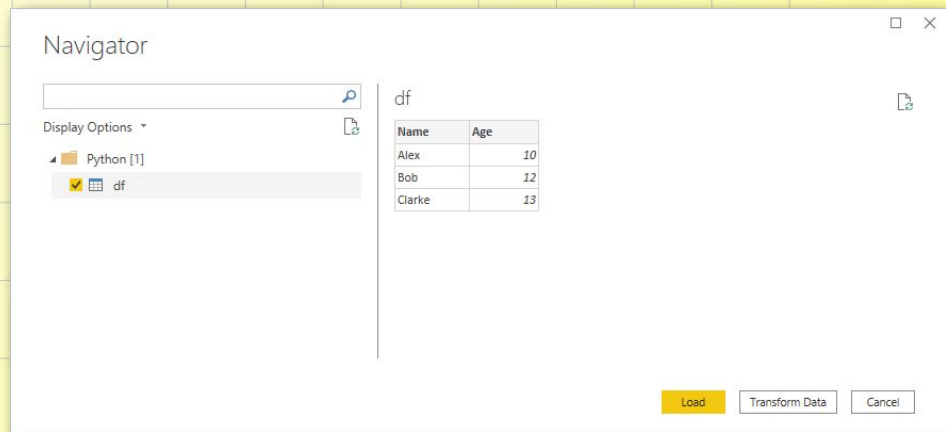




Pasos

Si el script se ejecuta correctamente, aparece el Navegador y se pueden cargar los datos y utilizarlos.

Para el ejemplo, se debe seleccionar **df**, como se muestra en la imagen y, a continuación, **Cargar**.





visualizaciones de python en Power BI





Gráfico de dispersión

```
import matplotlib.pyplot as plt
dataset.plot(kind='scatter', x='Age', y='Weight', color='red')
plt.show()
```

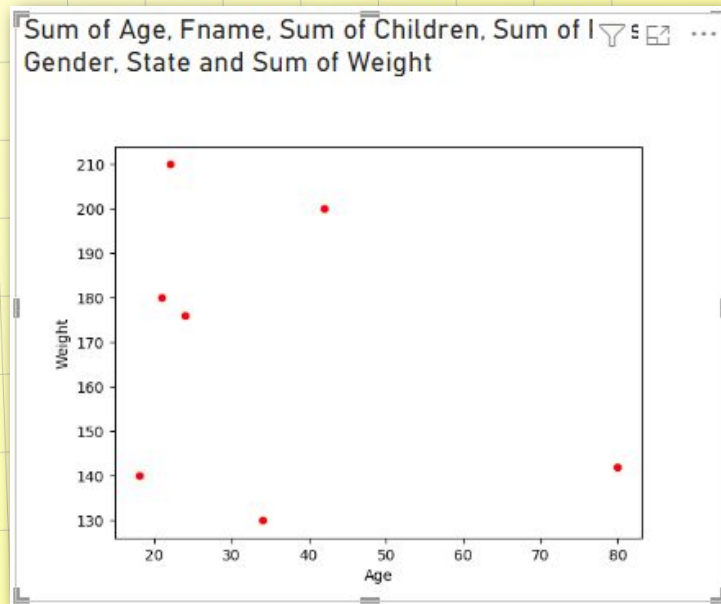




Gráfico de líneas

```
import matplotlib.pyplot as plt
ax = plt.gca()
dataset.plot(kind='line',x='Fname',y='Children',ax=ax)
dataset.plot(kind='line',x='Fname',y='Pets', color='red', ax=ax)
plt.show()
```

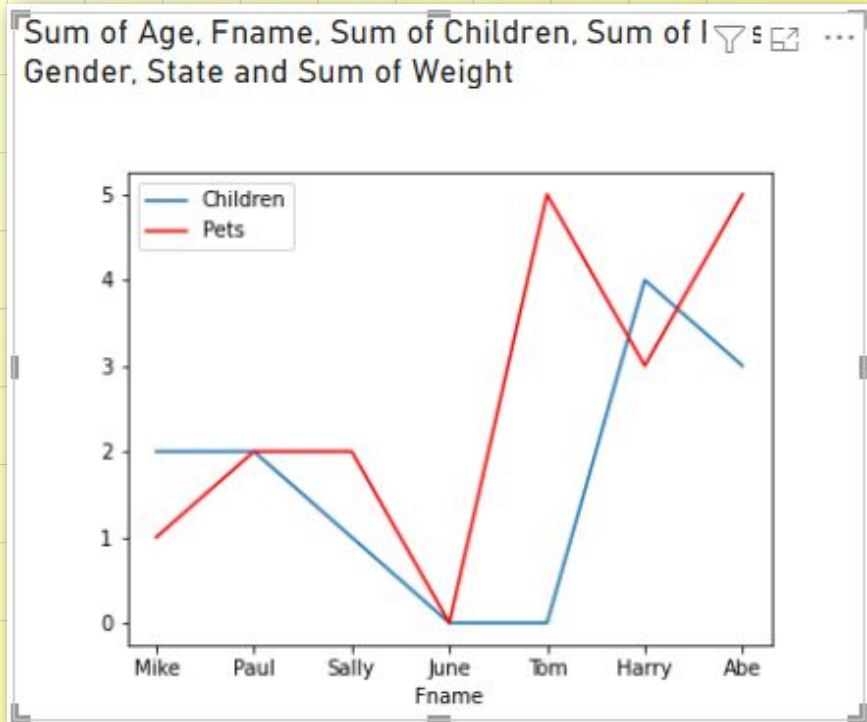
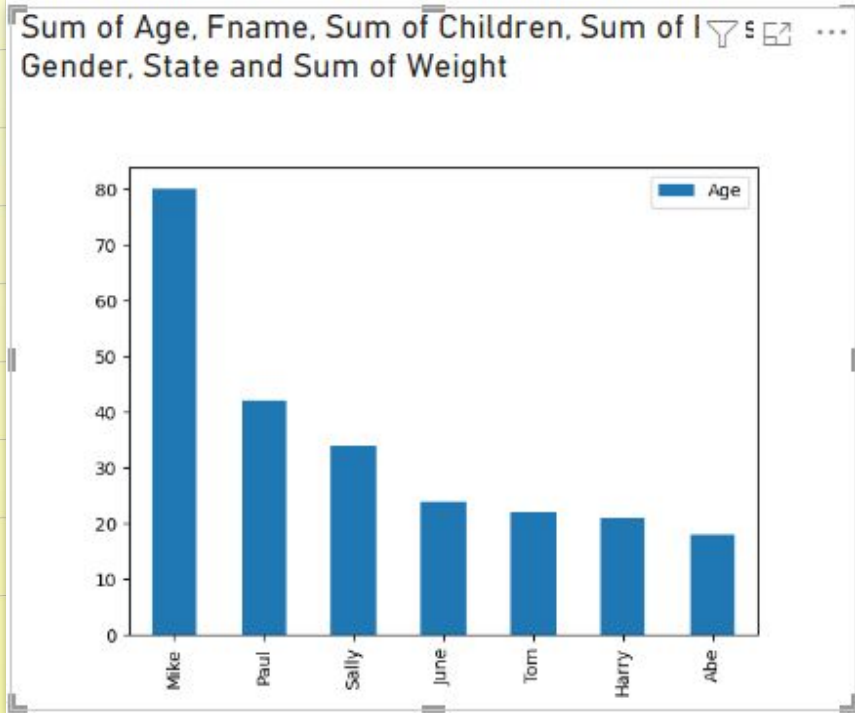




Gráfico de barras

```
import matplotlib.pyplot as plt
dataset.plot(kind='bar',x='Fname',y='Age')
plt.show()
```





Limitaciones conocidas





¿Cuáles?

- Limitaciones de tamaño de datos (están limitados a 150.000 filas. Si se seleccionan más de 150.000 filas, sólo se utilizan las 150.000 filas principales).
- Los datos de entrada tienen un límite de 250 MB.
- Si el conjunto de datos de entrada tiene una columna con un valor de cadena superior a 32766 caracteres, ese valor se trunca.
- Resolución. Todas las imágenes de Python se muestran a 72 DPI.
- Limitación de tiempo de cálculo. Si un cálculo visual de Python supera los cinco minutos, se agota el tiempo de espera de ejecución, lo que resulta en un error.



¿Cuáles?

- Relaciones. Si se seleccionan campos de datos de diferentes tablas sin una relación definida entre ellos, se produce un error.
- Los objetos visuales de Python se actualizan tras las actualizaciones de datos, el filtrado y el resaltado. La imagen en sí no es interactiva y no puede ser la fuente de filtrado cruzado.
- Los elementos visuales de Python responden a resaltar otros elementos visuales, pero no puede hacer clic en elementos en el objeto visual de Python para filtrar otros elementos.



¿Cuáles?

- Sólo los gráficos que se trazan en el dispositivo de visualización predeterminado de Python se muestran correctamente en el lienzo. Se debe evitar usar explícitamente un dispositivo de visualización Python diferente.
- Los objetos visuales de Python no admiten el cambio de nombre de las columnas de entrada. Las columnas se denominarán por su nombre original durante la ejecución del script.



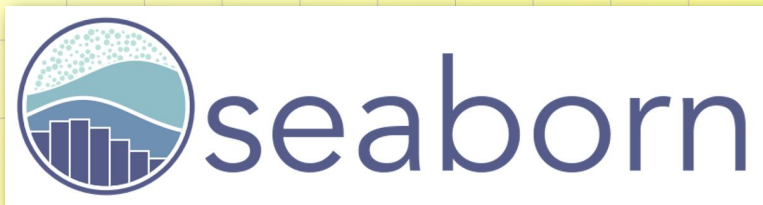
seaborn





¿Qué es?

- Seaborn es una biblioteca utilizada principalmente para el trazado estadístico en Python.
- Está construido sobre Matplotlib y proporciona estilos predeterminados y paletas de colores para hacer que las gráficas estadísticas sean más atractivas.
- Esta librería nos permite generar gráficos a partir de DataFrames de Pandas de manera muy cómoda y accesible.





scatterplot()





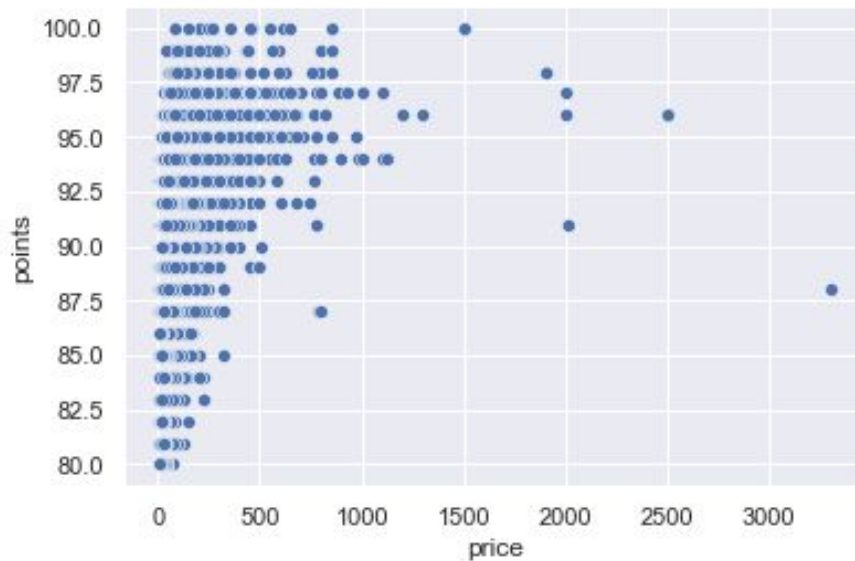
¿Qué es?

- Scatterplot se puede utilizar con varias agrupaciones semánticas que pueden ayudar a comprender bien en un gráfico.
- Pueden trazar gráficos bidimensionales que se pueden mejorar asignando hasta tres variables adicionales mientras se utiliza la semántica de los parámetros de tono, tamaño y estilo.
- Todos los parámetros controlan la semántica visual que se utilizan para identificar los diferentes subconjuntos.



Ejemplo

```
ax = sns.scatterplot(x="price", y="points", data=wine_reviews)
```





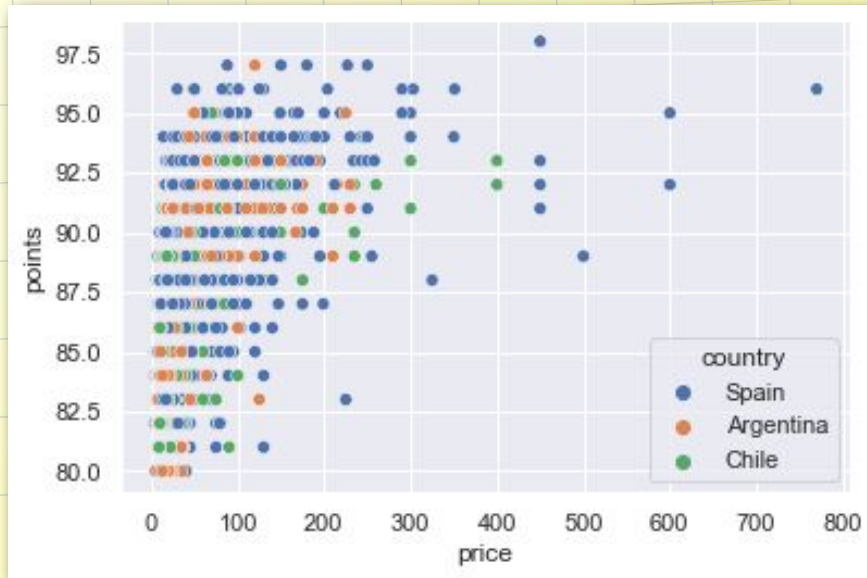
Scatter Plot con labels





Ejemplo

```
ax = sns.scatterplot(x="price", y="points", hue="country", data=wine_reviews_filtradas)
```





categorical Plots





¿Qué es?

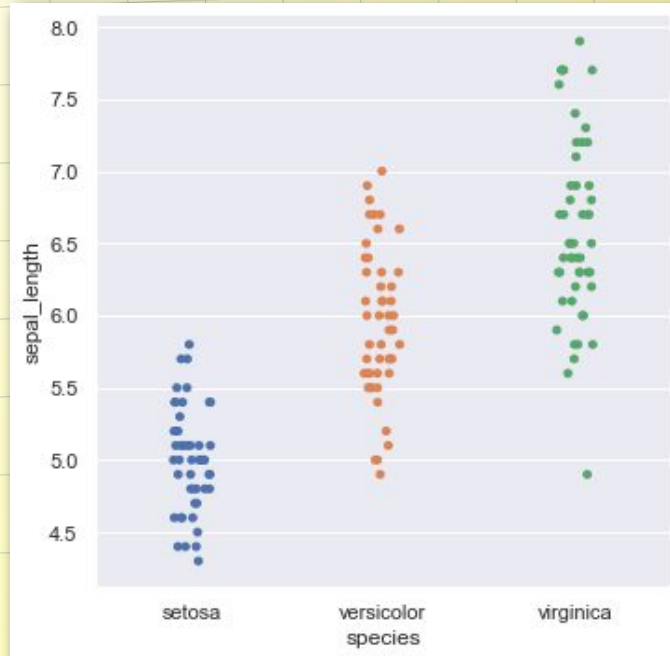
Como su nombre lo indica, los categorical plots son gráficos donde una de las variables a graficar es de tipo categórica.

Este tipo de gráficos son muy usados en Data Science y Seaborn tiene una función especial dedicada a ellos, [catplot](#).

Ejemplo



```
sns.catplot(data = iris_data, x = "species", y = "sepal_length", ci = "sd", estimator=np.median)
```





boxplot()





¿Qué es?

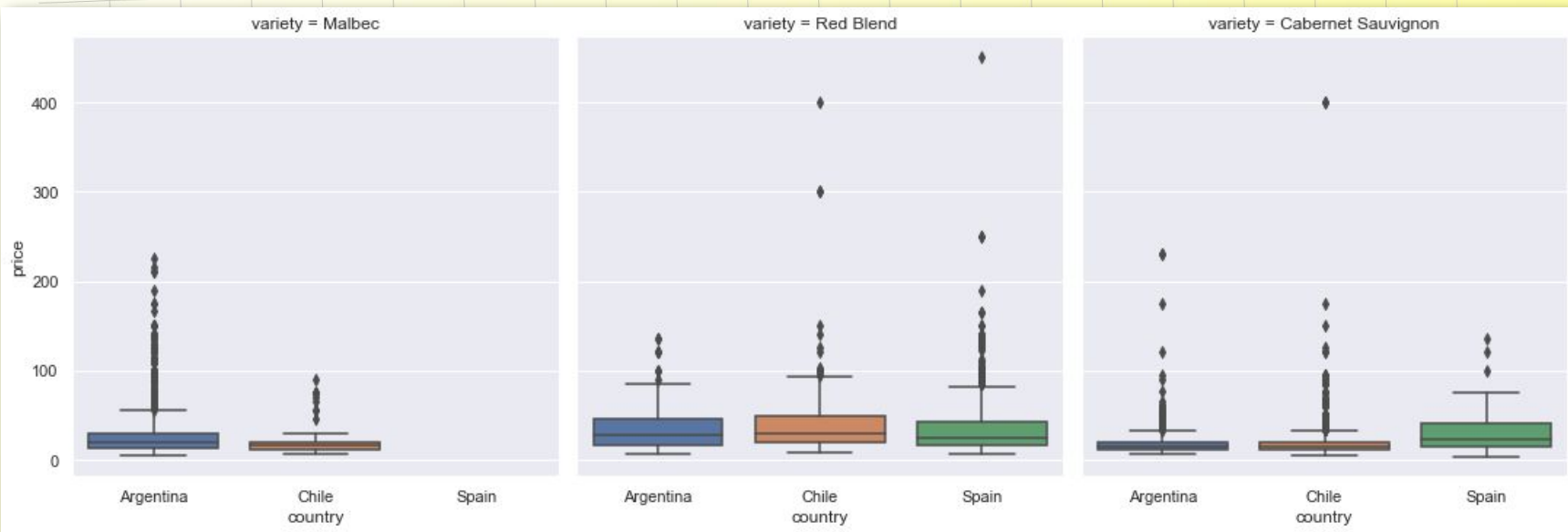
Un diagrama de caja ayuda a mantener la distribución de los datos cuantitativos de tal manera que facilita las comparaciones entre variables o entre niveles de una variable categórica.

- **Cuerpo principal**: muestra los cuartiles y los intervalos de confianza de la mediana, si está habilitado.
- **Medianas**: tienen líneas horizontales en la mediana de cada caja.
- **Bigotes**: tienen las líneas verticales que se extienden hasta los puntos de datos más extremos.

Ejemplo



```
sns.catplot(x="country", y="price", col="variety", kind='box' , data=vinos)
```





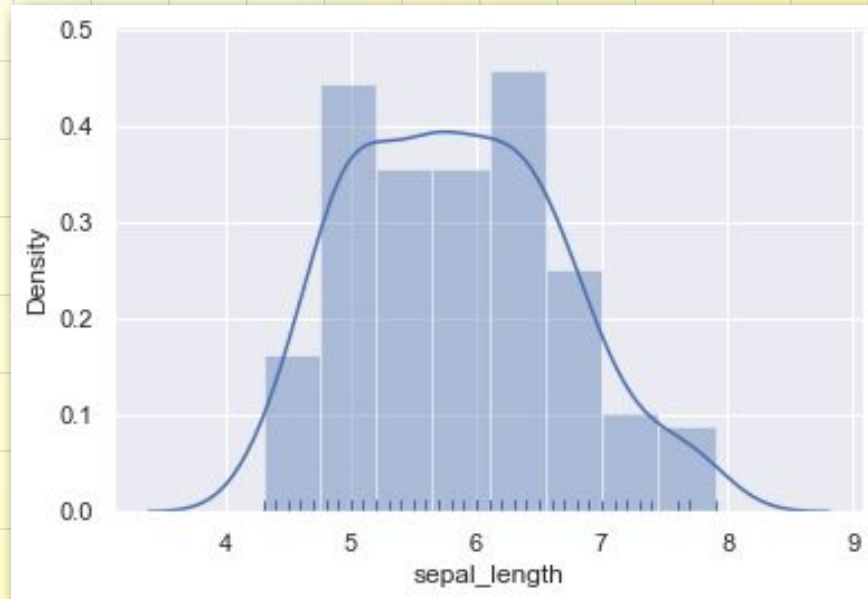
`distplot()`



Ejemplo



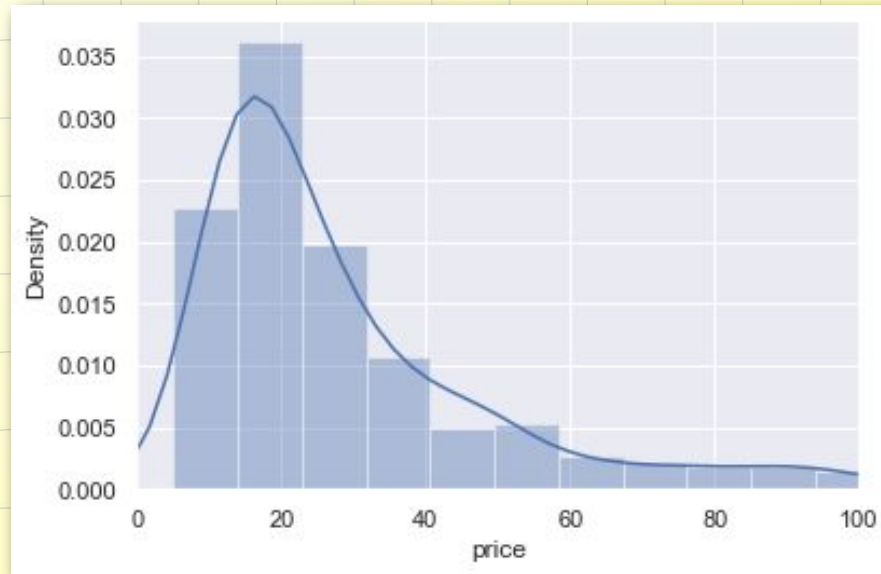
```
sns.distplot(iris_data['sepal_length'], hist=True, kde=True, rug=True)
```



Ejemplo



```
sns.distplot(vinos[(vinos['variety'] == 'Malbec') | (vinos['variety'] == 'Red Blend')].price)  
plt.xlim(0,100)
```





joinplot()

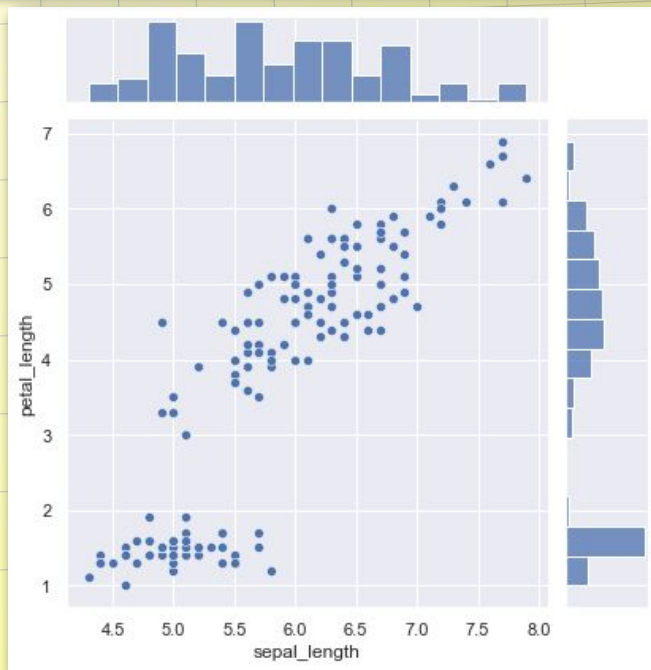


Ejemplo



```
sns.jointplot(x="sepal_length", y="petal_length", marginal_kws=dict(bins=15, rug=True), data=iris_data)
```

Dibuja una gráfica de dos variables con gráficos bivariados y univariados.





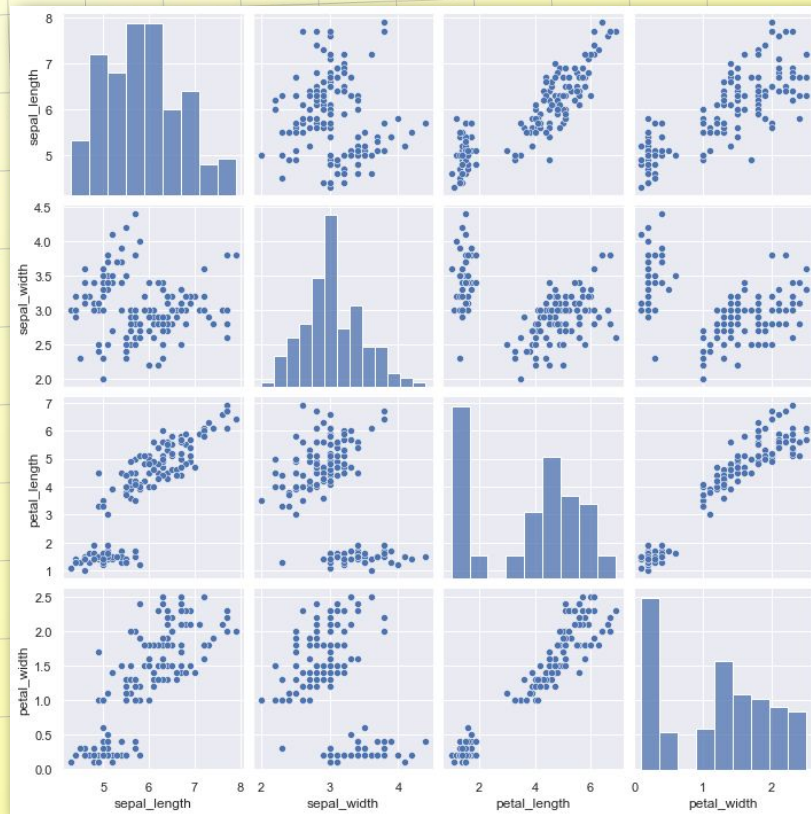
pairplot()



Ejemplo

```
sns.pairplot(data=iris_data)
```

Para trazar varias distribuciones bivariadas por pares en un conjunto de datos, puede utilizar la función `pairplot()`.





heatmap()





¿Qué es?

El mapa de calor se define como una representación gráfica de datos que utiliza colores para visualizar el valor de la matriz.

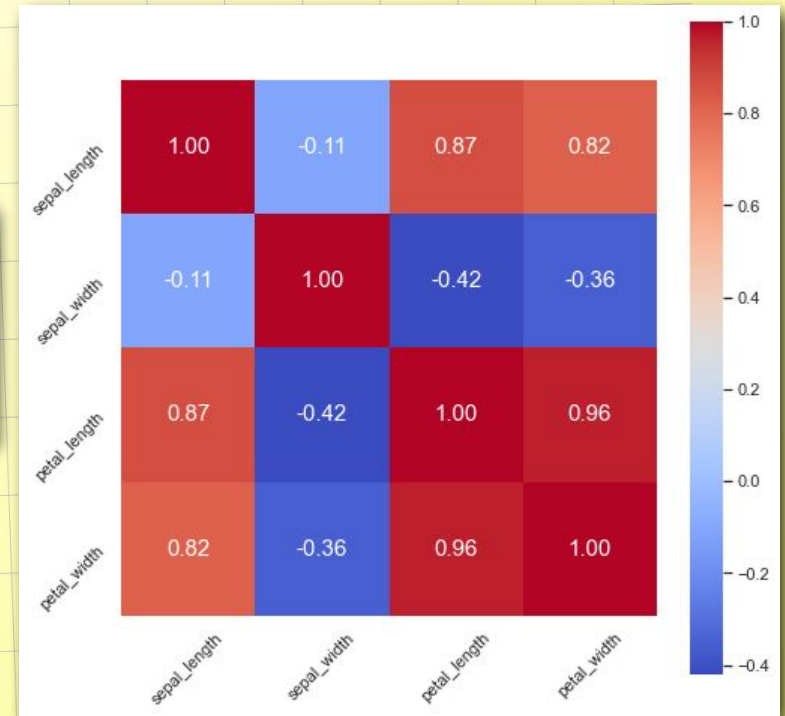
- Para representar valores más comunes o actividades más altas se utilizan **colores más brillantes**, básicamente colores rojizos.
- Para representar valores menos comunes o de poca actividad, se prefieren **colores más oscuros**.

El mapa de calor también se define por el nombre de la matriz de sombreado.

Los mapas de calor en Seaborn se pueden trazar utilizando la función **seaborn.heatmap()**.

Ejemplo

```
corr = iris_data.drop(columns = 'species').corr() #is used for find corelation
plt.figure(figsize=(8,8))
sns.heatmap(corr, cbar = True, square = True, annot=True, fmt= '.2f',annot_kws={'size': 15},
            xticklabels= iris_data.drop(columns = 'species').columns,
            yticklabels= iris_data.drop(columns = 'species').columns,
            cmap= 'coolwarm')
plt.xticks(rotation = 45)
plt.yticks(rotation = 45)
plt.show()
```



¿PREGUNTAS?



¿Alguien dijo Homework?



~~HENRY~~



Próxima lecture **Tableau**





¡Feedback!

Click on me



Dispones de un **formulario** en:



Homeworks



Guías de clase



Slack

HENRY

