

Matplotlib

Introducción a la visualización de datos con Matplotlib

CertiDevs

Índice de contenidos

1. Introducción	1
2. Usos de Matplotlib	1
3. Tipos de gráficos	1
4. Limitaciones	2
5. Alternativas	2
6. Importar Matplotlib	3
7. Jerarquía de objetos en Matplotlib	3
8. EDA	3

1. Introducción

Matplotlib es una librería de trazado y gráficos **2D** y **3D** en Python.

Es una de las librerías más antiguas y ampliamente utilizadas para la **creación de gráficos** en el ámbito científico y de análisis de datos. Con Matplotlib, se pueden crear gráficos de línea, barras, áreas, histogramas, diagramas de dispersión, gráficos de contorno, entre otros.

[Sitio web oficial de Matplotlib](#)

La importancia de Matplotlib radica en su capacidad para visualizar de manera clara y efectiva grandes cantidades de datos.

La **representación gráfica de los datos** puede ayudar a identificar **patrones**, **tendencias** y **relaciones** que pueden ser difíciles de detectar mediante un análisis solo numérico.

Matplotlib es una herramienta muy flexible y personalizable, permitiendo a los usuarios controlar aspectos como el tamaño, el color, la leyenda, la etiqueta de los ejes y otros elementos de la representación gráfica.

Otro factor importante es que Matplotlib es compatible con muchos otros paquetes de Python, como **NumPy**, **Pandas** y **Seaborn**, y es fácil de integrar con otros entornos de trabajo, como **Jupyter Notebooks**.

2. Usos de Matplotlib

Se utiliza para:

1. **Visualización de datos:** Matplotlib permite representar de manera clara y efectiva grandes cantidades de datos en forma de gráficos, lo que ayuda a identificar patrones, tendencias y relaciones en los datos.
2. **Análisis exploratorio de datos:** Matplotlib es una herramienta valiosa para explorar y comprender los datos, permitiendo a los usuarios visualizar y comparar diferentes variables y su relación con el tiempo.
3. **Presentaciones y informes:** Matplotlib es una herramienta versátil para la creación de presentaciones y informes que incluyen gráficos y visualizaciones.
4. **Machine learning y modelos estadísticos:** Matplotlib es utilizado para visualizar los resultados de los modelos de machine learning y estadísticos, lo que ayuda a comprender y validar los modelos.
5. **Investigación científica:** Matplotlib es ampliamente utilizado en la investigación científica para representar los resultados de las investigaciones y experimentos en forma de gráficos y visualizaciones.

3. Tipos de gráficos

1. **Gráficos de línea:** Matplotlib permite crear gráficos de línea simples o múltiples, con o sin

marcadores y personalización de colores y estilos.

2. **Gráficos de barras:** Matplotlib permite crear gráficos de barras simples o apiladas, con o sin etiquetas y personalización de colores y estilos.
3. **Gráficos de dispersión:** Matplotlib permite crear gráficos de dispersión para representar la relación entre dos variables, con o sin marcadores y personalización de colores y tamaños.
4. **Gráficos de histograma:** Matplotlib permite crear histogramas para representar la distribución de una variable, con o sin personalización de colores y estilos.
5. **Gráficos de áreas:** Matplotlib permite crear gráficos de áreas para representar la evolución temporal de una variable, con o sin personalización de colores y estilos.
6. **Gráficos de tarta:** Matplotlib permite crear gráficos de tarta para representar la proporción de diferentes valores en un conjunto de datos.
7. **Gráficos de box plot:** Matplotlib permite crear gráficos de box plot para representar la distribución de una variable y destacar valores atípicos.
8. **Gráficos de densidad:** Matplotlib permite crear gráficos de densidad para representar la distribución de una variable.
9. **Gráficos 3D:** Matplotlib permite crear gráficos tridimensionales para representar la relación entre tres variables.
10. **Personalización de ejes:** Matplotlib permite personalizar la escala, los límites y las etiquetas de los ejes en los gráficos, así como añadir leyendas y títulos.
11. **Importación y exportación a diferentes formatos:** Matplotlib permite importar imágenes, pero también exportar los gráficos a diferentes formatos, incluyendo PNG, PDF, SVG y JPG.

4. Limitaciones

Algunas **limitaciones** de matplotlib:

1. **Complejidad:** Matplotlib puede ser una herramienta compleja de aprender y utilizar, especialmente si se desea crear gráficos complejos. Requiere un conocimiento sólido de programación y una comprensión profunda de la librería.
2. **Personalización:** Aunque Matplotlib ofrece muchas opciones de personalización, algunas veces puede ser difícil hacer ciertos cambios en los gráficos, especialmente si se desea crear un diseño único.
3. **Rendimiento:** Matplotlib puede ser lento al renderizar gráficos muy grandes o complejos. Esto puede hacer que sea difícil trabajar con grandes conjuntos de datos.
4. **Funcionalidades limitadas:** Matplotlib es una librería de visualización de 2D y aunque se pueden crear gráficos tridimensionales, su capacidad en esta área es limitada en comparación con otras librerías de visualización 3D.

5. Alternativas

Alternativas a matplotlib:

- **Seaborn:** <https://seaborn.pydata.org/>
- **Plotly:** <https://plotly.com/python/>
- **Bokeh:** <https://docs.bokeh.org/en/latest/index.html>
- **Vega-Altair:** <https://altair-viz.github.io/>

6. Importar Matplotlib

Para empezar a utilizar Matplotlib, primero se debe instalar la biblioteca.

```
pip install matplotlib
```

La siguiente línea de código importa el módulo `pyplot`, que es la interfaz más utilizada para crear gráficos en Matplotlib:

```
import matplotlib.pyplot as plt
```

Es común utilizar el alias `plt` para facilitar el acceso a las funciones de `pyplot`.

7. Jerarquía de objetos en Matplotlib

Matplotlib organiza sus elementos gráficos en una **estructura jerárquica**.

Los principales objetos en esta jerarquía son:

- **Figure:** Es el contenedor principal que contiene todos los demás elementos del gráfico. Puede contener múltiples Axes.
- **Axes:** Es el área de dibujo donde se trazan los gráficos y contiene Axis, Tick, Line2D, Text, Polygon, y otros objetos.
- **Axis:** Representa los ejes X e Y del gráfico, incluyendo las marcas de graduación (ticks) y las etiquetas de los ejes.
- **Artist:** Todos los elementos que se pueden dibujar en un gráfico, como líneas, marcadores, texto y polígonos, son objetos Artist.

Para crear gráficos en Matplotlib, generalmente se inicia creando un objeto Figure, y luego se añaden uno o más objetos Axes a la figura.

8. EDA

El **análisis exploratorio de datos** (EDA, por sus siglas en inglés) es un enfoque utilizado en estadística y ciencia de datos para analizar y resumir conjuntos de datos, principalmente mediante técnicas visuales.

El objetivo del EDA es **explorar los datos** para identificar patrones, tendencias, relaciones,

anomalías y características interesantes que puedan no ser evidentes a simple vista.

El **EDA** es un paso crucial en el proceso de **análisis de datos**, ya que ayuda a obtener una comprensión inicial de los datos antes de realizar un análisis más profundo y modelado.

Hay varias visualizaciones que se pueden utilizar en el análisis exploratorio de datos, algunas de las cuales incluyen:

- **Histograma:** Un histograma es una representación gráfica de la distribución de un conjunto de datos numéricos. Muestra la frecuencia de los datos en intervalos, lo que permite analizar la forma de la distribución, la centralidad y la dispersión.
- **Diagrama de caja (box plot):** Un diagrama de caja es una representación gráfica que muestra la mediana, los cuartiles y los valores atípicos de un conjunto de datos numéricos. Es útil para visualizar la dispersión y simetría de los datos, así como identificar posibles valores atípicos.
- **Diagrama de dispersión (scatter plot):** Un diagrama de dispersión es una representación gráfica de dos variables numéricas, en la que cada punto representa una observación y sus coordenadas corresponden a los valores de ambas variables. Es útil para visualizar la relación entre dos variables y detectar posibles correlaciones, patrones o tendencias.
- **Gráfico de barras:** Un gráfico de barras es una representación gráfica de datos categóricos en la que cada categoría se representa mediante una barra, y la altura o longitud de la barra corresponde a la frecuencia, proporción o medida de interés de esa categoría. Es útil para comparar diferentes categorías y analizar datos categóricos.
- **Gráfico de líneas:** Un gráfico de líneas es una representación gráfica de datos numéricos a lo largo del tiempo, en el que los puntos se conectan mediante líneas. Es útil para analizar tendencias, patrones y fluctuaciones a lo largo del tiempo, así como comparar series temporales de diferentes variables.
- **Gráfico de pastel (pie chart):** Un gráfico de pastel es una representación gráfica de datos categóricos en la que cada categoría se representa mediante un sector circular, y el tamaño del sector es proporcional a la frecuencia, proporción o medida de interés de esa categoría. Es útil para visualizar la composición de un conjunto de datos y comparar proporciones de diferentes categorías.
- **Gráfico de densidad (density plot):** Un gráfico de densidad es una representación suavizada de la distribución de un conjunto de datos numéricos, que estima la función de densidad de probabilidad subyacente. Es útil para analizar la forma de la distribución y comparar distribuciones de diferentes conjuntos de datos.

La **categorización** de gráficos en **univariante**, **bivariante** y **multivariante** no es exclusiva del análisis exploratorio de datos (EDA); es una clasificación más general que se refiere al número de variables representadas en un gráfico.

Esta clasificación es relevante y se utiliza comúnmente en el contexto del EDA para organizar y seleccionar las visualizaciones adecuadas según la cantidad de variables que se están analizando.

- **Gráficos univariantes:** Estos gráficos representan una sola variable. El objetivo principal de estos gráficos es analizar la distribución, la centralidad y la dispersión de la variable en cuestión. Ejemplos de gráficos univariantes incluyen histogramas, gráficos de barras, gráficos de pastel y gráficos de densidad.

- **Gráficos bivariantes:** Estos gráficos representan dos variables y se utilizan para analizar la relación entre ellas. El objetivo principal de estos gráficos es identificar posibles correlaciones, patrones y tendencias entre las dos variables. Ejemplos de gráficos bivariantes incluyen diagramas de dispersión, gráficos de líneas y gráficos de barras apiladas.
- **Gráficos multivariantes:** Estos gráficos representan tres o más variables y se utilizan para analizar las relaciones y las interacciones entre ellas. El objetivo principal de estos gráficos es identificar patrones y tendencias complejas que involucren múltiples variables. Ejemplos de gráficos multivariantes incluyen gráficos de burbujas, gráficos de barras apiladas y gráficos de coordenadas paralelas.