

Matplotlib y Seaborn

Previo a entrar en el detalle de lo que es Matplotlib, es fundamental tener claro que la visualización de datos es una etapa clave del análisis de datos. Después de haber recopilado, almacenado y analizado datos, es esencial transformar los resultados de esos análisis en informes y visualizaciones gráficas.

Como es sabido, el cerebro humano comprende más rápido un diagrama que una sucesión de estadísticas en forma de tabla. La “Dataviz” permite, por tanto, compartir los resultados de un análisis con los equipos no técnicos de una empresa, y en particular con sus dirigentes.

Matplotlib es una librería Python open source, desarrollada inicialmente por el neurobiólogo John Hunter en 2002. El objetivo era visualizar las señales eléctricas del cerebro de personas epilépticas. Para conseguirlo, quería replicar las funcionalidades de creación gráfica de MATLAB con Python.

Tras el fallecimiento de John Hunter en 2012, Matplotlib ha sido mejorado a lo largo del tiempo por numerosos contribuidores de la comunidad open source. Se ha utilizado para crear gráficas y diagramas de gran calidad. Es una alternativa open source a MATLAB.

En este sentido, Matplotlib es una biblioteca integral para crear visualizaciones estáticas, animadas e interactivas en Python. Por esto, Matplotlib se caracteriza por facilitar las cosas simples y hacer posibles las cosas difíciles como las siguientes:

- Crea gráficos de calidad de publicación.
- Crea figuras interactivas que se pueden hacer zoom, desplazar y actualizar.
- Personaliza el estilo visual y el diseño.
- Exporta a muchos formatos de archivo.
- Incrusta en JupyterLab y en interfaces gráficas de usuario.
- Utiliza una amplia variedad de paquetes de terceros contruidos sobre Matplotlib.

Tomando en consideración que Matplotlib es una herramienta muy completa, que permite generar visualizaciones de datos muy detalladas, a continuación, se mencionan y explican algunas de las visualizaciones que se pueden elaborar con esta herramienta:

- **Gráficos de Líneas:** Ideales para mostrar tendencias y relaciones entre datos a lo largo del tiempo o cualquier otro eje.
- **Gráficos de Barras:** Perfectos para comparar diferentes categorías o grupos a través de sus valores numéricos.
- **Gráficos de Dispersión:** Útiles para visualizar la relación entre dos variables y detectar patrones de dispersión o agrupación.
- **Histogramas:** Utilizados para representar la distribución de datos en intervalos.
- **Gráficos de Pastel:** Ideales para mostrar proporciones o porcentajes dentro de un conjunto de datos.

- **Gráficos de Contorno:** Usados para representar visualmente funciones de dos variables en tres dimensiones.
- **Gráficos de Superficie:** Permiten visualizar funciones de tres variables en un espacio tridimensional.
- **Gráficos de Mapa de Calor:** Útiles para representar matrices de datos en forma de colores, donde la intensidad del color muestra la magnitud de los valores.
- **Gráficos de Caja (Box Plots):** Sirven para representar la distribución estadística de un conjunto de datos, mostrando medianas, cuartiles y valores atípicos.
- **Gráficos 3D:** Matplotlib permite crear visualizaciones tridimensionales para representar datos en un espacio 3D.
- **Gráficos de Área:** Se utilizan para mostrar la evolución de varias series de datos superpuestas, resaltando las contribuciones relativas de cada serie.

Por otra parte, se tienen los siguientes 5 beneficios de la visualización de datos al utilizar Matplotlib:

1. **Claridad y Simplificación:** Visualizar datos nos permite simplificar información compleja y presentarla en un formato más comprensible. Mediante representaciones gráficas, podemos comprender rápidamente patrones, valores atípicos y otros conocimientos importantes que podrían no ser inmediatamente evidentes en los datos sin procesar.
2. **Identificación de Patrones y Tendencias:** Con la visualización de datos, podemos identificar patrones y tendencias con mayor facilidad. Gráficos de líneas, gráficos de barras, gráficos de dispersión y otras visualizaciones nos ayudan a identificar correlaciones, variaciones y relaciones entre diferentes variables, lo que nos permite tomar decisiones basadas en datos.
3. **Exploración y Análisis de Datos:** Visualizar datos proporciona un entorno interactivo y exploratorio para el análisis de datos. Las funciones interactivas de Matplotlib permiten a los usuarios hacer zoom, desplazarse e interactuar con los gráficos para explorar puntos de datos específicos o regiones de interés, facilitando una comprensión más profunda de los datos.
4. **Comunicación y Presentación:** La visualización de datos facilita la comunicación de hallazgos e ideas a otros. Mediante el uso de representaciones visuales convincentes, podemos transmitir de manera efectiva información compleja a partes interesadas, clientes o colegas que pueden no tener formación técnica. Las visualizaciones ayudan a contar una historia y facilitan discusiones basadas en datos.
5. **Toma de Decisiones y Planificación:** La visualización de datos ayuda a tomar decisiones informadas y a la planificación estratégica. Al visualizar datos, podemos identificar tendencias, detectar valores atípicos y descubrir patrones ocultos que influyen en los procesos de toma de decisiones en diversos ámbitos, como negocios, finanzas, atención médica y más.

Por último, otro de los elementos clave a tener en cuenta de Matplotlib es su relación con el Data Analytics, Matplotlib desempeña un papel fundamental en el ámbito de la analítica de datos. Como una biblioteca de visualización en Python, Matplotlib permite a los profesionales de análisis

de datos y científicos de datos representar y comunicar de manera efectiva los resultados de su análisis. Esta relación se hace evidente a través de los siguientes puntos:

- **Exploración de Datos:** Antes de realizar un análisis detallado, es importante comprender los datos mediante la exploración visual. Matplotlib permite crear gráficos rápidos y sencillos para visualizar patrones, distribuciones y relaciones en los datos, lo que facilita la identificación de tendencias y anomalías.
- **Comunicación de Resultados:** Los análisis de datos generan información valiosa, pero comunicar esos resultados de manera efectiva es esencial. Matplotlib permite crear gráficos de alta calidad y personalizables, lo que facilita la presentación de los hallazgos a colegas, equipos o partes interesadas.
- **Creación de Dashboards:** En el contexto de analítica de datos empresariales, los paneles de control (dashboards) son comunes para resumir y visualizar datos clave. Matplotlib puede integrarse con herramientas de creación de paneles de control para proporcionar gráficos interactivos que actualizan automáticamente con los nuevos datos.
- **Análisis Exploratorio de Datos (EDA):** En la fase inicial del proceso de análisis de datos, el EDA implica la visualización y comprensión de los datos antes de aplicar técnicas más avanzadas. Matplotlib facilita la creación de gráficos que ayudan a descubrir patrones, relaciones y posibles áreas de interés.
- **Visualización de Modelos y Resultados:** En el análisis de datos predictivos y modelado, es esencial visualizar los resultados de los modelos. Matplotlib puede utilizarse para crear gráficos que muestren la precisión del modelo, las predicciones frente a los valores reales y otros indicadores de rendimiento.
- **Presentaciones y Reportes:** Al comunicar los resultados de un análisis de datos a un público no técnico, es importante proporcionar visualizaciones claras y atractivas. Matplotlib permite crear gráficos que se pueden incorporar en presentaciones y reportes para hacer que la información sea más accesible y comprensible.

En resumen, Matplotlib desempeña un papel esencial en la analítica de datos al proporcionar las herramientas necesarias para visualizar y comunicar información de manera efectiva, lo que ayuda a los profesionales a tomar decisiones informadas basadas en los resultados de sus análisis.

En cuanto a Seaborn, esta es una librería de visualización de datos en Python que se basa en Matplotlib y proporciona una interfaz de alto nivel para crear gráficos atractivos e informativos. Seaborn se destaca por su facilidad de uso, estilo visual atractivo y funciones especializadas para la visualización de datos estadísticos y categóricos. Asimismo, Seaborn proporciona una capa más orientada a la estadística y la exploración de datos que puede ahorrar tiempo y esfuerzo al crear visualizaciones informativas y atractivas.

Además, se puede argumentar que la principal diferencia entre los gráficos de Seaborn y los de otras bibliotecas de visualización de datos, como Matplotlib o Plotly, radica en la facilidad de uso, la apariencia predeterminada y las capacidades específicas que ofrece Seaborn. Seguidamente se presentan las diferencias clave para comprender mejor esta biblioteca:

1. **Estilos Visuales Mejorados:** Ofrece estilos visuales predefinidos que mejoran la apariencia de los gráficos de manera automática, lo que puede hacer que los gráficos sean más atractivos y profesionales sin necesidad de personalización adicional. Otras bibliotecas pueden requerir ajustes manuales para lograr una apariencia similar.
2. **Sintaxis Simplificada:** Proporciona funciones específicas y simplificadas para crear gráficos complejos, lo que ahorra tiempo y reduce la cantidad de código necesario en comparación con otras bibliotecas.
3. **Enfoque Estadístico y Análisis de Datos:** Está diseñada para trabajar bien con análisis estadísticos y visualización de datos, proporcionando gráficos especializados para distribuciones, regresiones y relaciones estadísticas. Otras bibliotecas pueden no tener esta orientación específica.
4. **Integración con Pandas:** Se integra de manera fluida con DataFrames de Pandas, lo que facilita la creación de gráficos directamente desde los datos. Esto puede reducir la necesidad de preparación adicional de los datos antes de la visualización.
5. **Visualización de Datos Categóricos:** Ofrece una variedad de gráficos específicos para visualizar datos categóricos, como gráficos de barras, gráficos de conteo y gráficos de punto, lo que puede hacer que la exploración de datos categóricos sea más sencilla y efectiva.
6. **Gráficos Avanzados de Distribución:** Proporciona gráficos de distribución avanzados, como los gráficos de violín y los mapas de calor de densidad, que pueden brindar una visualización más detallada de la distribución de datos en comparación con las opciones estándar de otras bibliotecas.
7. **Personalización Simplificada:** Aunque otras bibliotecas también permiten la personalización, Seaborn ofrece una forma más sencilla de ajustar la apariencia de los gráficos mediante la selección de estilos predefinidos y paletas de colores.

En resumen, Seaborn sobresale como una biblioteca de visualización de datos altamente efectiva y especializada en Python. Su capacidad para simplificar la creación de gráficos complejos, su enfoque en la exploración y análisis de datos estadísticos, y su integración fluida con DataFrames de Pandas la convierten en una herramienta esencial para los analistas de datos.

Referencias

Akintola, S. (s.f.). Matplotlib Graphical Functions. Guide to optimise your Visualization skills using matplotlib functions.

DataScientest. (2023). Matplotlib: todo lo que tienes que saber sobre la librería Python de Dataviz. <https://datascientest.com/es/todo-sobre-matplotlib>

KeepCoding. (2023). ¿Qué es Matplotlib y cómo funciona? <https://keepcoding.io/blog/que-es-matplotlib-y-como-funciona/>

The Matplotlib Development Team. (2023). Matplotlib: Visualization with Python. <https://matplotlib.org/>

Waskom, M. (2022). Seaborn: statistical data visualization. <https://seaborn.pydata.org/>