

«Talento Tech»

Data Analytics

con Python

Clase 13



Clase 13 | Visualización con Matplotlib

Temario:

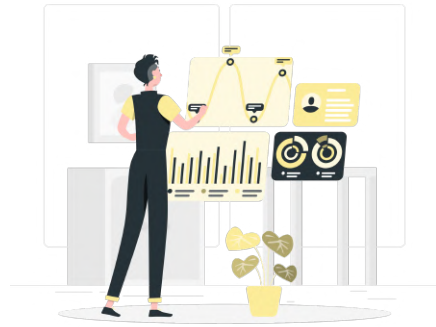
- Repaso sobre tipos de gráficos básicos.
 - Profundización de gráficos básicos: líneas, barras y dispersión utilizando Matplotlib.
 - Personalización.
-

Objetivos de la Clase:

- **Comprender los fundamentos de la visualización de datos** utilizando Matplotlib.
- **Aprender a personalizar gráficos** para mejorar la comunicación visual, garantizando que la información resaltada sea fácilmente interpretable.
- **Identificar las dificultades** que pueden surgir al no personalizar gráficos, reconociendo la importancia de una presentación profesional y atractiva de los datos.

Data-Viz con Matplotlib

En el análisis de datos, la visualización es fundamental para presentar la información de manera clara y efectiva. En esta clase, nos enfocaremos en los gráficos básicos más utilizados: gráficos de líneas, gráficos de barras y gráficos de dispersión, utilizando la biblioteca Matplotlib de Python.



Personalización de gráficos

La personalización de gráficos en Matplotlib no sólo **mejora la estética visual** de las visualizaciones, sino que también **potencia la efectividad en la comunicación de la información**. Exploremos las ventajas de personalizar los gráficos y las dificultades que pueden surgir al no hacerlo.

Ventajas de la personalización de gráficos

- **Mejora la Comprensión de los Datos**

Personalizar un gráfico ayuda a **destacar información clave**. Por ejemplo, al agregar títulos descriptivos y etiquetas en los ejes, los usuarios pueden comprender rápidamente qué datos están viendo y su significado. Esto es importante en las presentaciones, donde el tiempo es limitado y la comunicación debe ser más efectiva.

- **Aumenta la Atractividad Visual**

Los gráficos personalizados son más atractivos y pueden **captar mejor la atención de la audiencia**. Usar diferentes colores, tipos de marcadores y estilos de línea puede hacer que un gráfico sea visualmente más impactante y agradable, animando a los espectadores a enfocarse en los detalles que se desean resaltar.

- **Facilita la Comparación**

Al personalizar gráficos, como en los gráficos de barras, se pueden utilizar colores diferentes para cada barra, lo que facilita la comparación entre categorías. Esto **permite que las diferencias sean más evidentes** y ayuda a la audiencia a recordar esos puntos clave.

- **Aumenta la Profesionalidad**

La presentación de gráficos bien personalizados transmite un nivel de profesionalismo y cuidado. Esto es especialmente importante en contextos académicos y empresariales, donde **la impresión visual puede influir en la percepción de la calidad del trabajo presentado**. Un gráfico desorganizado o poco atractivo puede desmerecer incluso un análisis profundo.

- **Refuerza el Storytelling**

La personalización permite contar una historia con los datos. Se pueden usar colores y formatos para **enfatar tendencias o patrones importantes**, ayudando a guiar a la audiencia a través del análisis y hacer que el mensaje sea más convincente.

Dificultades al no personalizar gráficos

- **Confusión en la Comunicación**

Un gráfico sin personalización puede ser confuso. La falta de etiquetas claras, títulos y una paleta de colores apropiada puede hacer que la audiencia no comprenda el mensaje. Esto es problemático durante las presentaciones, donde la claridad es importante.

- **Pérdida de Información Crucial**

Sin la personalización necesaria, es fácil que la información relevante pase desapercibida. Por ejemplo, un gráfico de líneas que no cambia de color para marcar diferentes series de datos puede dificultar la comparación entre ellas.

- **Dificultades en la Toma de Decisiones**

Cuando los gráficos no son claros, la interpretación de los datos se complica. En el contexto empresarial, esto puede llevar a decisiones mal informadas basadas en una visualización que no comunica adecuadamente las tendencias o patrones de los datos.

- **Menor Impacto Visual**

Gráficos sin personalización suelen ser planos y poco atractivos, lo que puede llevar a que la audiencia pierda interés rápidamente. Cuando los gráficos no tienen un diseño atractivo, es más probable que se pasen por alto los mensajes importantes.

- **Falta de Contexto**

La personalización también ayuda a proporcionar contexto adicional que puede ser vital para la interpretación de los datos. Sin etiquetas adecuadas o anotaciones explicativas, los gráficos carecen del contexto que se necesita para que los espectadores comprendan completamente los datos.

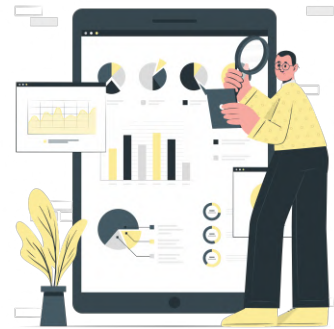


Tipos de data-viz básicos y sus personalizaciones

Gráficos de Líneas

Los gráficos de líneas son ideales para mostrar la **evolución de una variable a lo largo del tiempo**. Son especialmente útiles en series temporales, donde queremos observar tendencias.

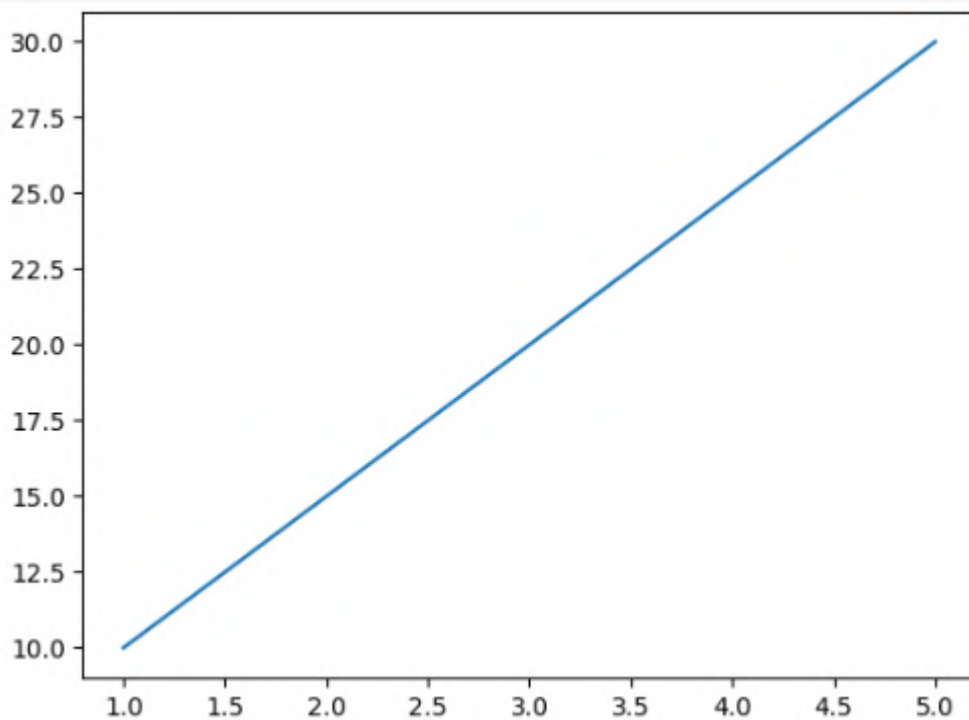
Para crear un gráfico de líneas en Matplotlib, debemos recordar importar la biblioteca. A continuación, se presenta un ejemplo básico:



```
import matplotlib.pyplot as plt

# Datos
tiempo = [1, 2, 3, 4, 5]
ventas = [10, 15, 20, 25, 30]

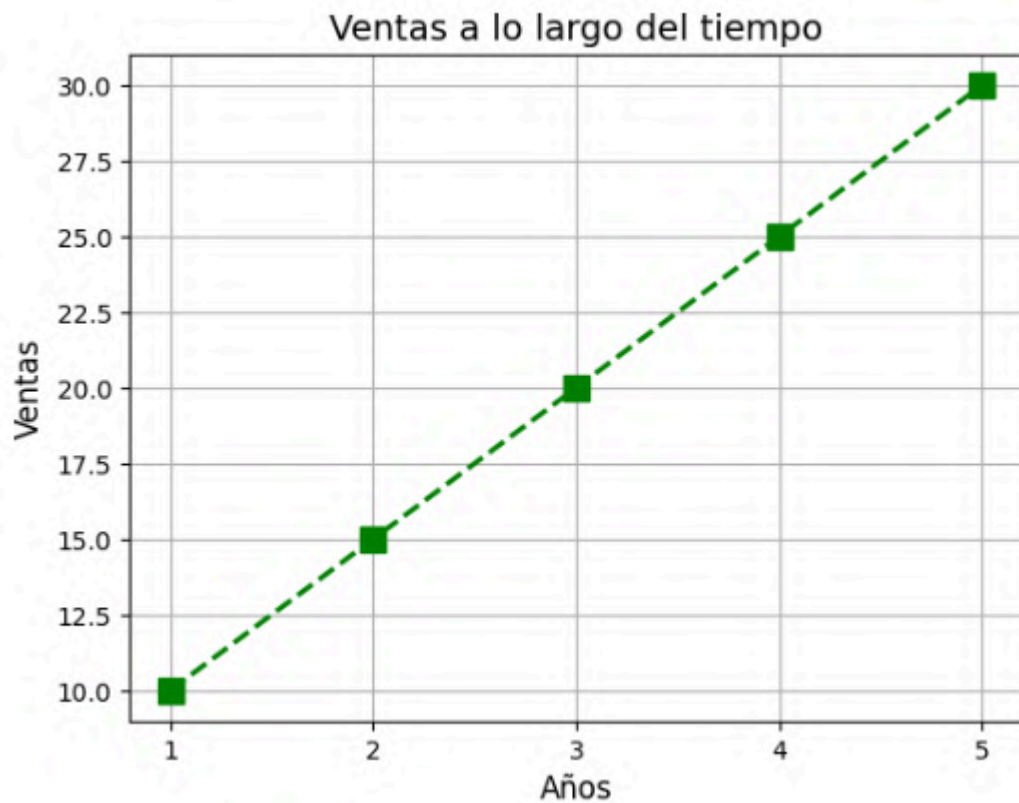
# Gráfico de líneas
plt.plot(tiempo, ventas)
plt.show()
```



Personalización del Gráfico de Líneas

Podemos personalizar nuestros gráficos para hacerlos más informativos. Algunas opciones incluyen cambiar el color, agregar etiquetas y ajustar el estilo de las líneas. Por ejemplo:

```
plt.plot(tiempo, ventas, marker='s', linestyle='--',
color='green', linewidth=2, markersize=10)
plt.title('Ventas a lo largo del tiempo', fontsize=14)
plt.xlabel('Años', fontsize=12)
plt.ylabel('Ventas', fontsize=12)
plt.grid(True)
plt.xticks(tiempo)
plt.show()
```

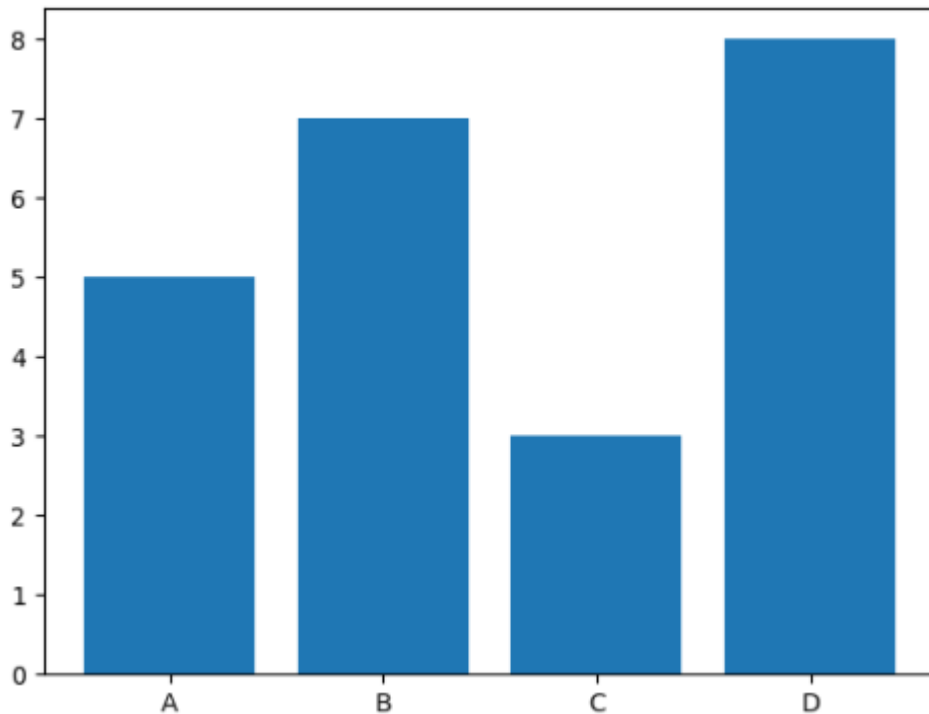


Gráficos de Barras

Los gráficos de barras son excelentes para **comparar diferentes grupos de datos**. Muestran la cantidad de cada categoría de manera visual. A continuación, les mostramos un ejemplo:

```
# Datos
categorias = ['A', 'B', 'C', 'D']
valores = [5, 7, 3, 8]

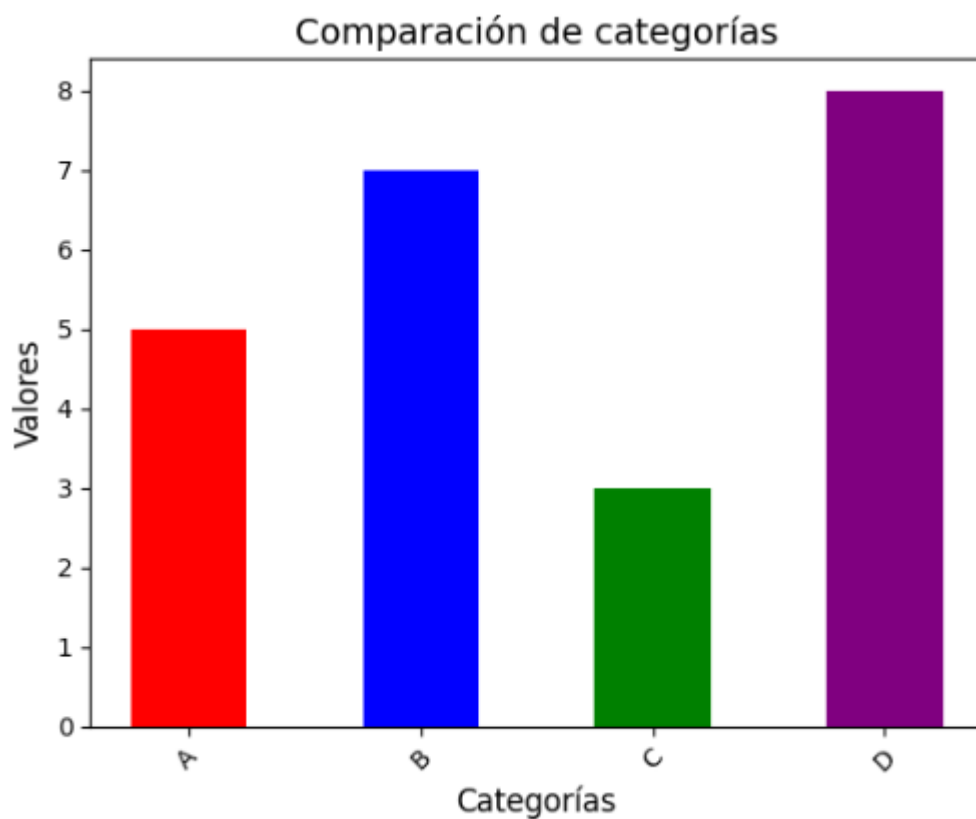
# Gráfico de barras
plt.bar(categorias, valores)
plt.show()
```



Personalización del Gráfico de Barras

Al igual que los gráficos de líneas, los gráficos de barras pueden ser personalizados. Podemos agregar colores, etiquetas y cambiar el ancho de las barras:

```
plt.bar(categorias, valores, color=['red', 'blue', 'green',  
'purple'], width=0.5)  
plt.title('Comparación de categorías', fontsize=14)  
plt.xlabel('Categorías', fontsize=12)  
plt.ylabel('Valores', fontsize=12)  
plt.xticks(rotation=45)  
plt.show()
```



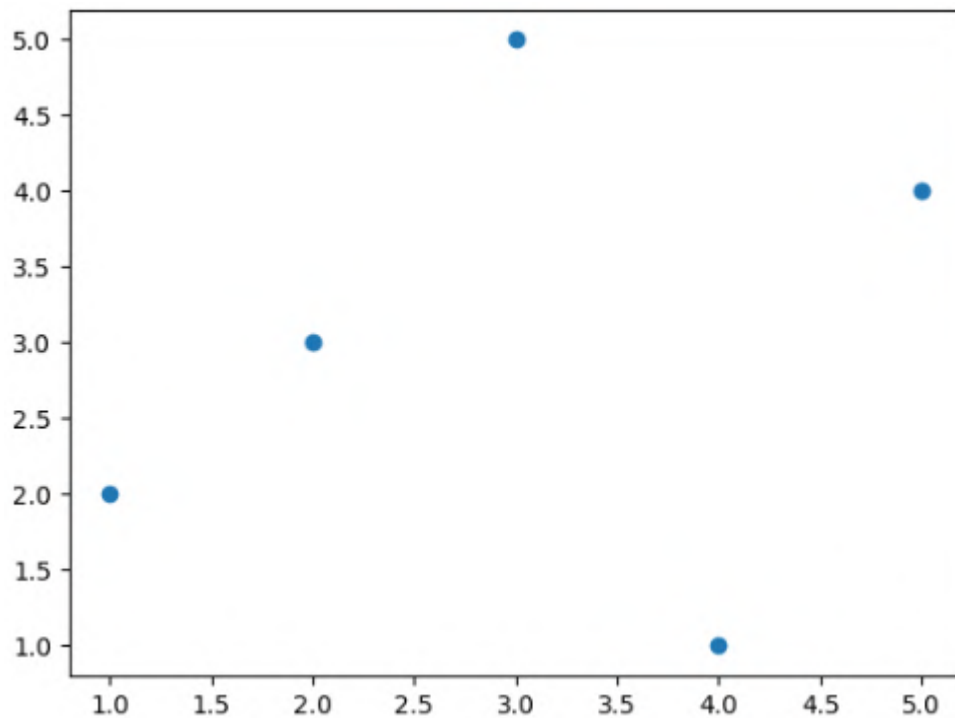
Gráficos de Dispersión

Los gráficos de dispersión permiten **visualizar la relación entre dos variables**. Cada punto representa un conjunto de valores (x, y). Este tipo de gráfico es útil para identificar correlaciones.

Así creamos un gráfico de dispersión:

```
x = [1, 2, 3, 4, 5]
y = [2, 3, 5, 1, 4]

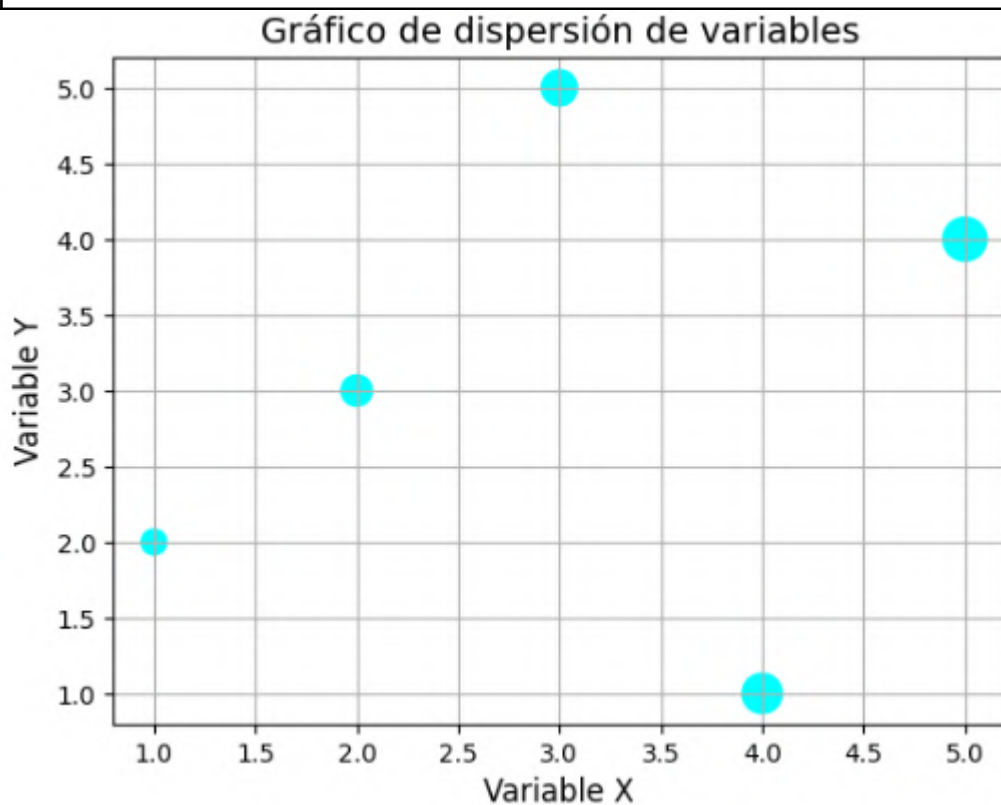
# Gráfico de dispersión
plt.scatter(x, y)
plt.show()
```



Personalización del Gráfico de Dispersión

Al igual que los otros tipos de gráficos, se pueden personalizar los gráficos de dispersión:

```
plt.scatter(x, y, color='cyan', s=[100, 150, 200, 250, 300])  
plt.title('Gráfico de dispersión de variables', fontsize=14)  
plt.xlabel('Variable X', fontsize=12)  
plt.ylabel('Variable Y', fontsize=12)  
plt.grid(True)  
plt.show()
```



Reflexión Final

La personalización de gráficos en Matplotlib mejora la **claridad y comprensión de los gráficos, aumenta su atractivo visual y profesionalismo**. Por el contrario, no personalizar los gráficos puede acarrear diversas dificultades, que van desde la confusión en la comunicación hasta la falta de impacto visual.

En el análisis de datos, donde la interpretación y presentación son clave, asegurarse de que los gráficos sean personalizados y claros es una de las mejores inversiones que se pueden hacer para que el mensaje llegue de manera efectiva.



Materiales y recursos adicionales

- [Matplotlib.org](https://matplotlib.org)

Próximos Pasos

- Crear visualizaciones estadísticas usando Seaborn.
- Uso de paletas y estilos.
- Gráficas orientadas a objetos.

Ejercicios Prácticos



Actividad 1: Análisis de Ventas a lo Largo del Tiempo

Contexto



Como pasante en SynthData, Silvia, tu mentora y Project Manager, te asigna la tarea de analizar las ventas de un producto para entender su evolución a lo largo del tiempo. Este análisis es necesario para que el cliente pueda tomar decisiones estratégicas sobre los productos y las campañas de marketing de la marca.

Objetivos

El objetivo de esta actividad es aprender a crear gráficos de líneas en Matplotlib para visualizar cómo las ventas han cambiado durante los meses del año pasado. Esta habilidad es fundamental en data analytics para presentar tendencias y patrones de datos.

Ejercicio práctico

- Crear un gráfico de líneas que muestre las ventas mensuales.
- Redactar un párrafo breve que explique la gráfica.

Dataset

```
# Datos de ventas mensuales (en unidades)
meses = ['Ene', 'Feb', 'Mar', 'Abr', 'May', 'Jun', 'Jul',
'Ago', 'Sep', 'Oct', 'Nov', 'Dic'] # Lista de los meses del
año
ventas = [100, 120, 130, 150, 180, 170, 200, 210, 250, 270,
300, 320] # Lista de las unidades vendidas correspondientes a
cada mes
```

¿Por qué importa esto en SynthData?

Visualizar las ventas mensuales de manera eficaz te permite identificar patrones estacionales y momentos de crecimiento o decrecimiento, ayudando al equipo a proponer estrategias de ventas más efectivas. Además, el análisis de tendencias es

fundamental para cumplir con las expectativas de los clientes y realizar ajustes proactivos en la oferta.

Actividad 2: Comparación de Productos en una Campaña

Contexto



En la siguiente tarea, Luis, el Analista de BI, te pide que compares el rendimiento de cuatro productos durante una reciente campaña publicitaria. Este análisis es vital para evaluar los resultados de la campaña y decidir en qué productos enfocar esfuerzos futuros.

Objetivos

El objetivo es crear un gráfico de barras que permita visualizar fácilmente cuál fue el producto más vendido y cómo se comparan las ventas entre ellos. Este tipo de visualización es muy útil en ambientes de negocios para obtener insights de desempeño.

Ejercicio práctico

- Crear un gráfico que muestre las ventas por artículo.
- Agregar el contexto al gráfico y buscar una forma de resaltar el artículo más vendido

Dataset

```
# Datos de ventas por producto
productos = ['Producto A', 'Producto B', 'Producto C',
'Producto D'] # Lista de los nombres de los productos
ventas = [300, 250, 400, 350] # Lista de las unidades vendidas
de cada producto en la campaña.
```

¿Por qué importa esto en SynthData?

Al comparar los productos, podemos identificar cuáles están rindiendo mejor y cuáles necesitan mejoras. Esta información permite al equipo de marketing reorientar sus esfuerzos y maximizar el retorno de la inversión en futuras campañas publicitarias.

⚠️ **Estos ejercicios son una simulación de cómo se podría resolver el problema en este contexto específico. Las soluciones encontradas no aplican de ninguna manera a todos los casos.**

Recuerda que las soluciones dependen de los sets de datos, el contexto y los requerimientos específicos de los stakeholders y las organizaciones.



Buenos Aires
aprende

Agencia de Habilidades para el Futuro

BA Buenos
Aires
Ciudad