

Uppgift: Generera data och visualisera

Bakgrund

Visualisering är ett kraftfullt verktyg för att kommunicera data och information. Det kan hjälpa oss att upptäcka mönster och trender i datan och förstå vad den betyder. I denna uppgift kommer du att få träna på att generera data och visualisera den.

Exempel: Försäljning av produkter över tid

Skapa en visualisering som tydligt visar att försäljningen av produkter är högre under andra halvan av året än under första halvan.

Steg 1: Generera datan

Denna kod genererar en lista med försäljningsdata för varje dag under 2023. Du kan använda denna data för att skapa en visualisering.

```
import pandas as pd
import numpy as np

# Number of days in each month
days_in_month = [31, 28, 31, 30, 31, 30, 31, 31, 30, 31, 30, 31]

# Create an empty list to store sales data
sales_data = []

# Generate sales data for each month
for month, days in enumerate(days_in_month):
    if month < 6: # First half of the year
        sales = np.random.randint(50, 100, days).tolist()
    else: # Second half of the year
        sales = np.random.randint(100, 150, days).tolist()

    sales_data.extend(sales)

# Create a dataframe with the sales data
df = pd.DataFrame({
    'Date': pd.date_range(start="2023-01-01", end="2023-12-31"),
    'Sales': sales_data
})
```

Steg 2: Visualisera datan

Här kan du använda ett bibliotek som matplotlib, seaborn eller något annat visualiseringsverktyg:

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

# Set a style
sns.set(style="whitegrid")

# Create a figure and axes
plt.figure(figsize=(14, 6))
sns.lineplot(data=df, x="Date", y="Sales")

# Set title and labels
plt.title("Försäljning av produkter under 2023")
plt.xlabel("Datum")
plt.ylabel("Antal sålda produkter")

# Show the plot
plt.show()
```

För att visualisera ett 10-dagars genomsnitt av försäljningen kan du använda en glidande medelvärdesfunktion (i Pandas kallas detta "rolling mean").

```
# ...
# Calculate the 10-day average
df['10_Day_Avg'] = df['Sales'].rolling(window=10).mean()
sns.lineplot(data=df, x="Date", y="10_Day_Avg", label="10-dagars genomsnitt",
color='red')
# ...
```

Uppgift

1. **Generera och visualisera data som visar på** en ökning av användare på en webbplats efter en marknadsföringskampanj.
2. **Generera och visualisera data som visar på** sambandet mellan nedbördsmängd och antalet besökare i en utomhuspark.
3. **Generera och visualisera data som visar på** hur medarbetarnas prestation varierar beroende på antal timmars sömn de får per natt.
4. **Generera och visualisera data som visar på** effekten av temperaturhöjning på försäljningen av isglass.
5. **Generera och visualisera data som visar på** antalet olyckor på en viss korsning före och efter installationen av ett stoppljus.

6. **Generera och visualisera data som visar på** kundrecensioners inverkan på försäljningen av en ny bok.
7. **Generera och visualisera data som visar på** skillnader i daglig energiförbrukning mellan sommar- och vintermånader.
8. **Generera och visualisera data som visar på** relationen mellan kaffeintag och produktivitet under arbetsdagen.
9. **Generera och visualisera data som visar på** hur studenternas prestationer förändras över tid när de genomgår en ny undervisningsmetod.
10. **Generera och visualisera data som visar på** skillnaden i trafik till en e-handelswebbplats mellan veckodagar och helger.
11. **Generera och visualisera data som visar på** sambandet mellan priset på en vara och dess försäljningshastighet.
12. **Generera och visualisera data som visar på** påverkan av en influencer's rekommendation på försäljningen av en skönhetsprodukt.
13. **Generera och visualisera data som visar på** minskningen av koldioxidutsläpp efter introduktionen av en grön teknologi.
14. **Generera och visualisera data som visar på** förändringen i gymmedlemskap före och efter nyår.
15. **Generera och visualisera data som visar på** relationen mellan antalet timmar spenderade på sociala medier och studenternas betyg.