

8. Лабораторна робота. Застосування бібліотеки Pandas до візуалізації даних

Мета: засвоїти можливості роботи бібліотеки Pandas до групуванню даних, агрегуванню даних, візуалізації даних.

Теоретичні відомості та методичні рекомендації

У бібліотеці Pandas є вбудовані засоби для візуалізації даних, які базуються на бібліотеці Matplotlib. Вони дозволяють швидко будувати графіки без додаткових налаштувань.

Основні методи візуалізації у Pandas:

1. Лінійний графік line plot. За замовчуванням `df.plot()` будує лінійний графік (див. рис. 8.1). Деякі властивості:

- `kind="line"` – вказує, що будуємо лінійний графік;
- `marker="o"` – додає точки на лінії.

```
import pandas as pd
import matplotlib.pyplot as plt

# Створюємо DataFrame
data = {"Місяць": ["Січ", "Лют", "Бер", "Кві", "Тра"],
        "Продажі": [100, 150, 130, 170, 200]}

df = pd.DataFrame(data)

# Будуємо графік
df.plot(x="Місяць", y="Продажі", kind="line", marker="o",
figsize=(8, 5))

plt.title("Динаміка продажів")
plt.xlabel("Місяць")
plt.ylabel("Продажі")
plt.grid(True)
plt.show()
```

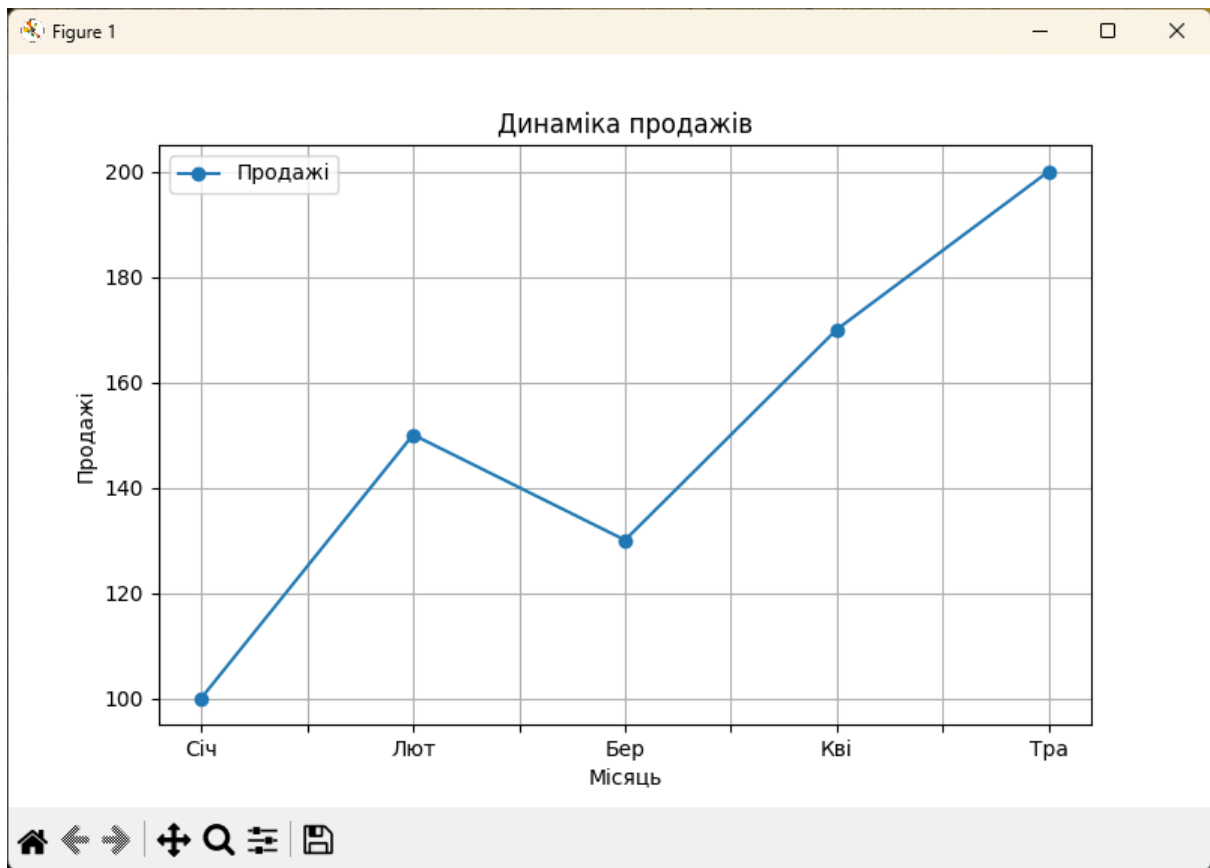


Рисунок 8.1 – Лінійний графік

2. Стовпчикова діаграма bar plot (див. рис. 8.2).

```
import pandas as pd
import matplotlib.pyplot as plt

data = {"Місяць": ["Січ", "Лют", "Бер", "Кві", "Тра"],
        "Продажі": [100, 150, 130, 170, 200]}

df = pd.DataFrame(data)

df.plot(x="Місяць", y="Продажі", kind="bar",
        color="skyblue", figsize=(8, 5))

plt.title("Продажі за місяцями")
plt.xlabel("Місяць")
plt.ylabel("Продажі")
plt.show()
```

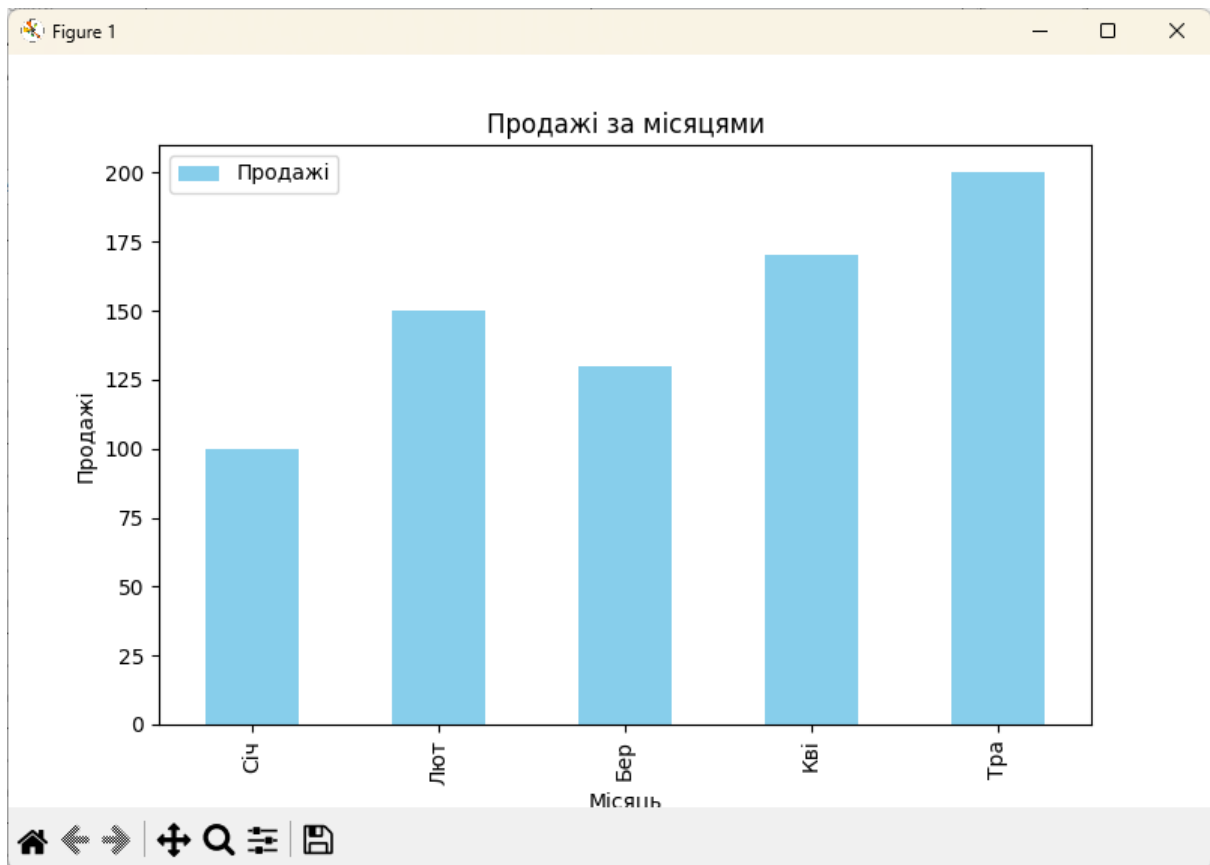


Рисунок 8.2 – Стовпчикова діаграма

3. Кругова діаграма pie chart (див. рис. 8.3).

```
import pandas as pd
import matplotlib.pyplot as plt

data = {"Місяць": ["Січ", "Лют", "Бер", "Кві", "Тра"],
        "Продажі": [100, 150, 130, 170, 200]}

df = pd.DataFrame(data)

df.plot(kind="pie", y="Продажі", labels=df["Місяць"],
autopct="%1.1f%%", legend=False)

plt.title("Розподіл продажів")
plt.ylabel("") # Прибираємо підпис осі Y
plt.show()
```

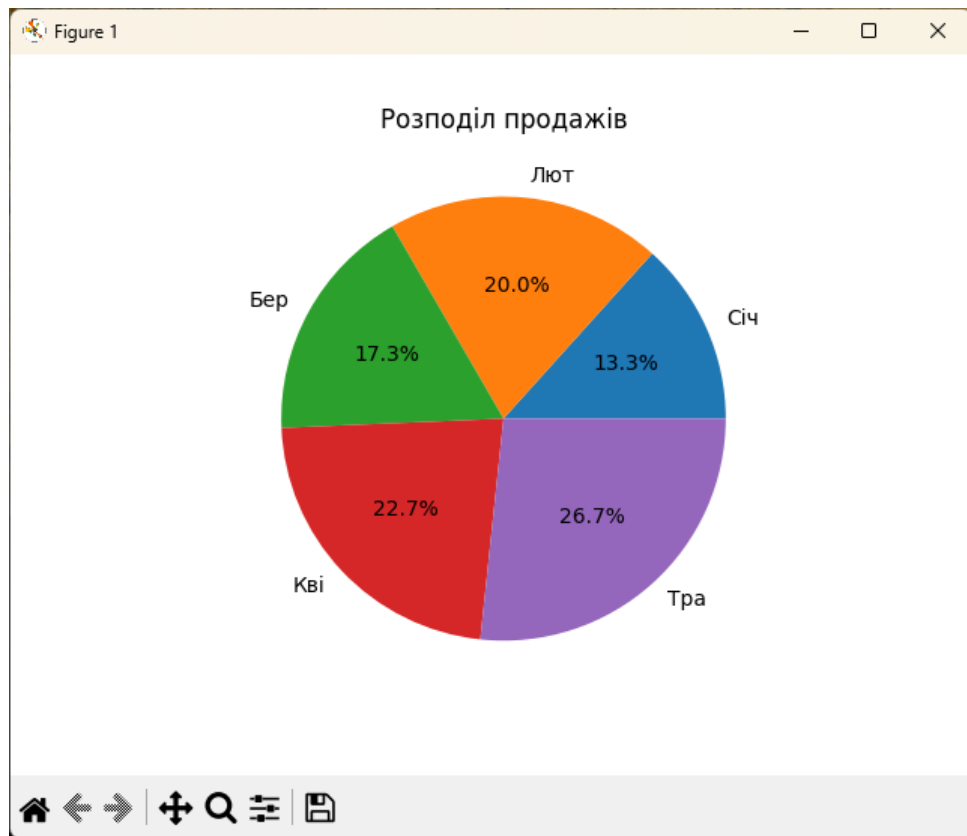


Рисунок 8.3 – Кругова діаграма

4. Графік розсіювання scatter plot (див. рис. 8.4).

```
import pandas as pd
import matplotlib.pyplot as plt

data = {"Місяць": ["Січ", "Лют", "Бер", "Кві", "Тра"],
        "Продажі": [100, 150, 130, 170, 200]}

df = pd.DataFrame(data)

df.plot(kind="scatter",          x="Місяць",          y="Продажі",
color="red")

plt.title("Зв'язок між місяцем і продажами")
plt.xlabel("Місяць")
plt.ylabel("Продажі")
plt.show()
```

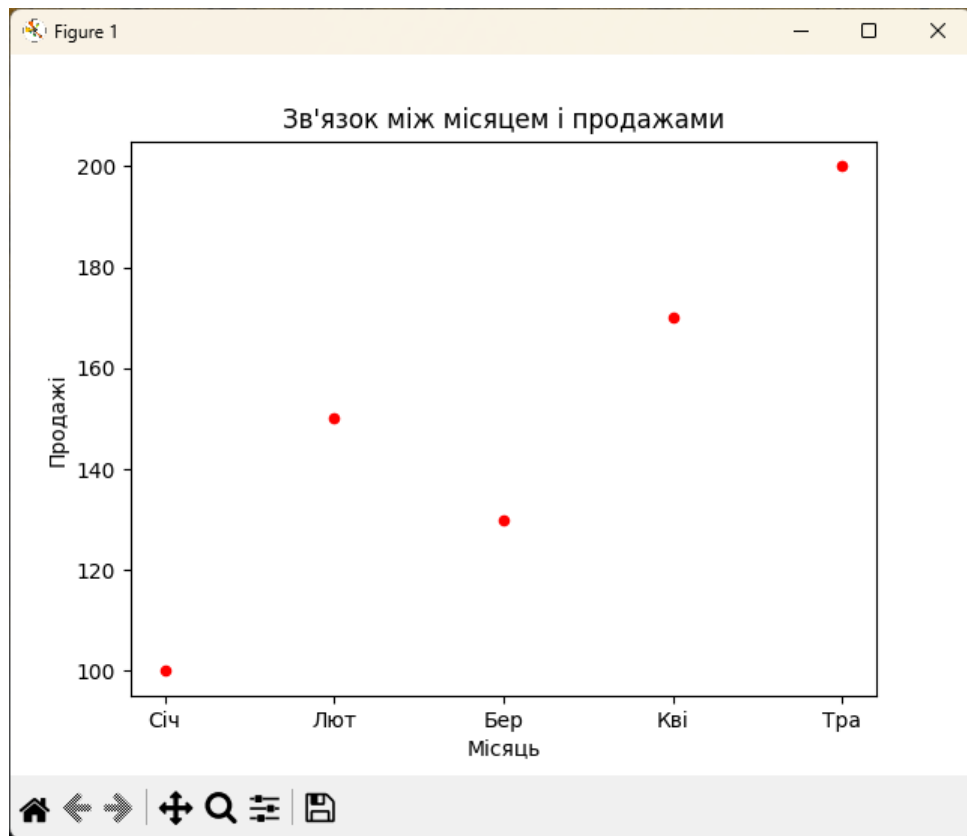


Рисунок 8.4 – Графік розсіювання

Крім того, у бібліотеці Pandas можна візуалізувати дані штатними можливостями, не використовуючи бібліотеку Matplotlib. Для побудови гістограми існує метод `plot.bar()`, а для побудови кругової діаграми існує метод `plot.pie()` (див. рис. 8.5). Отже, Pandas має вбудовані методи візуалізації, які достатні для базового аналізу.

```
In [22]: plot = res.plot.pie(y='sum', figsize=(7, 7))
```

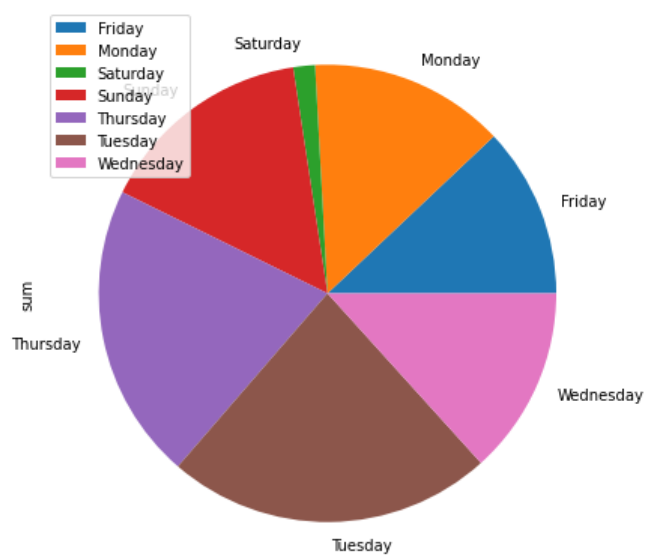


Рисунок 8.5 – Візуалізація

Завдання до лабораторної роботи

1. Завантажити з <https://www.kaggle.com/datasets> Public Datasets (csv-файл з даними). Провести аналіз даних.
2. Виберіть колонки з числовими даними. Дайте характеристику вибраним даним.
3. Побудуйте стовпчикову діаграму, використовуючи метод `bar()`. Зробіть висновки.
4. Побудуйте кругову діаграму, використовуючи метод `plot.pie()`. Зробіть висновки.

Питання для самоконтролю

1. Який метод використовується для створення графіків у Pandas?
2. Як побудувати лінійний графік у Pandas?
3. Який параметр `kind` потрібно вказати, щоб створити стовпчикову діаграму?
4. Як змінити розмір графіка в Pandas?
5. Як додати заголовок до графіка?
6. Як зробити точки на лінійному графіку більш помітними?
7. Чим відрізняється `bar` від `barh` у Pandas?
8. Як побудувати гістограму (графік розподілу)?
9. Як змінити колір графіка у Pandas?
10. Як зробити підписи у відсотках на круговій діаграмі?
11. Як побудувати графік розсіювання у Pandas?
12. Що таке `bins` у гістограмі?
13. Як змінити підписи осей на графіку?
14. Як можна покращити оформлення графіків у Pandas за допомогою Matplotlib?
15. Чи можна будувати кілька графіків одночасно? Якщо так, як?
16. Як вибрати певні стовпці для візуалізації?