

## 7. Лабораторна робота. Застосування бібліотеки Pandas до групування та агрегування даних

**Мета:** засвоїти можливості роботи бібліотеки Pandas до групування даних, агрегуванню даних.

### Теоретичні відомості та методичні рекомендації

DataFrameGroupBy – це об'єкт у Pandas, який створюється після виклику методу `groupby()`. Він дозволяє групувати дані за одним або кількома стовпцями і виконувати агрегатні операції, такі як підрахунок, сума, середнє значення тощо.

Синтаксис:

```
df.groupby("стовпець")
```

Приклад використання `groupby()`:

```
import pandas as pd

# Створення DataFrame
data = {
    "Група": ["А", "В", "А", "В", "А", "С"],
    "Студент": ["Анна", "Богдан", "Віктор", "Ганна",
               "Денис", "Єва"],
    "Оцінка": [90, 85, 88, 92, 75, 80]
}

df = pd.DataFrame(data)

# Групування за "Група" та знаходження середньої оцінки
grouped = df.groupby("Група")["Оцінка"].mean()
print(grouped)
```

Результат:

```
Група
А      84.333333
В      88.500000
С      80.000000
Name: Оцінка, dtype: float64
```

Агрегуючі методи `groupby()`:

- `mean()` – середнє значення;
- `sum()` – сума значень;
- `count()` – кількість елементів у групі;
- `min()` – мінімальне значення;
- `max()` – максимальне значення;
- `median()` – медіана;
- `std()` – стандартне відхилення;

– `describe()` – повна статистика групи.

Наприклад, сумування всіх числових даних в DataFrame (див. рис. 7.1).

```
In [14]: df.groupby('Day_name').sum()
```

Out[14]:

	ProductID	Quantity	Price	Total
Day_name				
Friday	906881	6	33973.80	33973.80
Monday	765576	6	23435.39	38971.89
Saturday	143185	1	4334.33	4334.33
Sunday	767407	5	37180.35	43145.85
Thursday	3925651	10	41439.05	59075.05
Tuesday	641762	11	25618.08	64991.43
Wednesday	717171	6	32873.79	37488.07

Рисунок 7.1 – Агрегація по дням тижня

Очевидно, що `groupby()` – це потужний інструмент для аналізу даних.

### 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. Виконайте групування та агрегування даних у файлі. Надайте пояснення.

### Питання для самоконтролю

1. Охарактеризуйте об'єкт `DataFrameGroupBy` з бібліотеки `Pandas`.
2. Наведіть агрегуючі методи.
3. Чи можливе групування за кількома стовпцями?
4. Який метод дозволяє залишити тільки групи, які відповідають певній умові?
5. Чи можливо застосовувати кілька агрегуючих функцій?