

# Лабораторна робота №1

Вернигора Артем. КП-72. 4 курс

## Аналіз та візуалізація даних у Python

Мета: продемонструвати свої знання про життєвий цикл аналізу даних, використовуючи заданий набір даних та вказані інструменти

### Передумови / сценарій

У цій лабораторній роботі ви імпортуєте деякі пакети Python, необхідні для аналізу набору даних, що містить інформацію про злочини в Сан-Франциско. Потрібно використати засоби Python та Jupyter, щоб підготувати ці дані до аналізу, проаналізувати їх, побудувати графіки та повідомити про свої результати. Завдання до лабораторної роботи знаходиться за посиланням: <https://static-course-assets.s3.amazonaws.com/loTFBDA201/en/course/files/2.2.4.5%20Lab%20-%20San%20Francisco%20Crime.html>

### Відповіді на питання

Скільки змінних міститься у фреймі даних SF?

**Відповідь: 12 змінних**

Якого виду злочину було скоєно найбільше?

**Відповідь : LARCENY/THEFT**

В якому PdDistrict було найбільше випадків зареєстрованих злочинів? Надайте команди Python, які використовуються для підтримки вашої відповіді.

**Відповідь: SOUTHERN PdDistrict**

Скільки випадків злочинів було за серпень?

**Відповідь: 9720**

Скільки квартирних крадіжок було зареєстровано у серпні?

**Відповідь: 1257**

Зверніть увагу на діагональ зліва направо у таблиці кореляцій, сформованій вище. Чому діагональ заповнена 1s? Це випадковість? Поясніть.

**Відповідь: Кореляційна матриця являє собою квадратну матрицю розмірності  $M \times M$ , де  $M$  - кількість завдань, симетричну відносно головної діагоналі.**

Не дивлячись на таблицю кореляцій вище, зауважте, що значення відображаються дзеркально; значення нижче 1 діагоналі мають дзеркальний аналог вище 1 діагоналі. Це випадковість? Поясніть.

Відповідь: Кореляційна матриця завжди симетрична (коефіцієнт кореляції між змінними  $X$  і  $Y$  дорівнює коефіцієнту кореляції між змінними  $Y$  і  $X$ ), і на головній діагоналі такої матриці стоять 1 (кореляція змінної самої з собою дорівнює 1).

Багато пар змінних мають кореляцію, близьку до нуля. Що це означає?

Відповідь : Коефіцієнт кореляції, близький до нульового значення, свідчить про те, що між  $x$  і  $y$  зв'язку не існує.

Навіщо розділяти статі?

Відповідь : Щоб результати не спотворювались через різницю в чоловічому та жіночому тілах.

Які змінні мають сильнішу кореляцію з розміром мозку (MRI\_Count)? Це очікується? Поясніть.

Відповідь : для жінок - це Weight, PIQ, для чоловіків - PIQ, FSIQ, VIQ

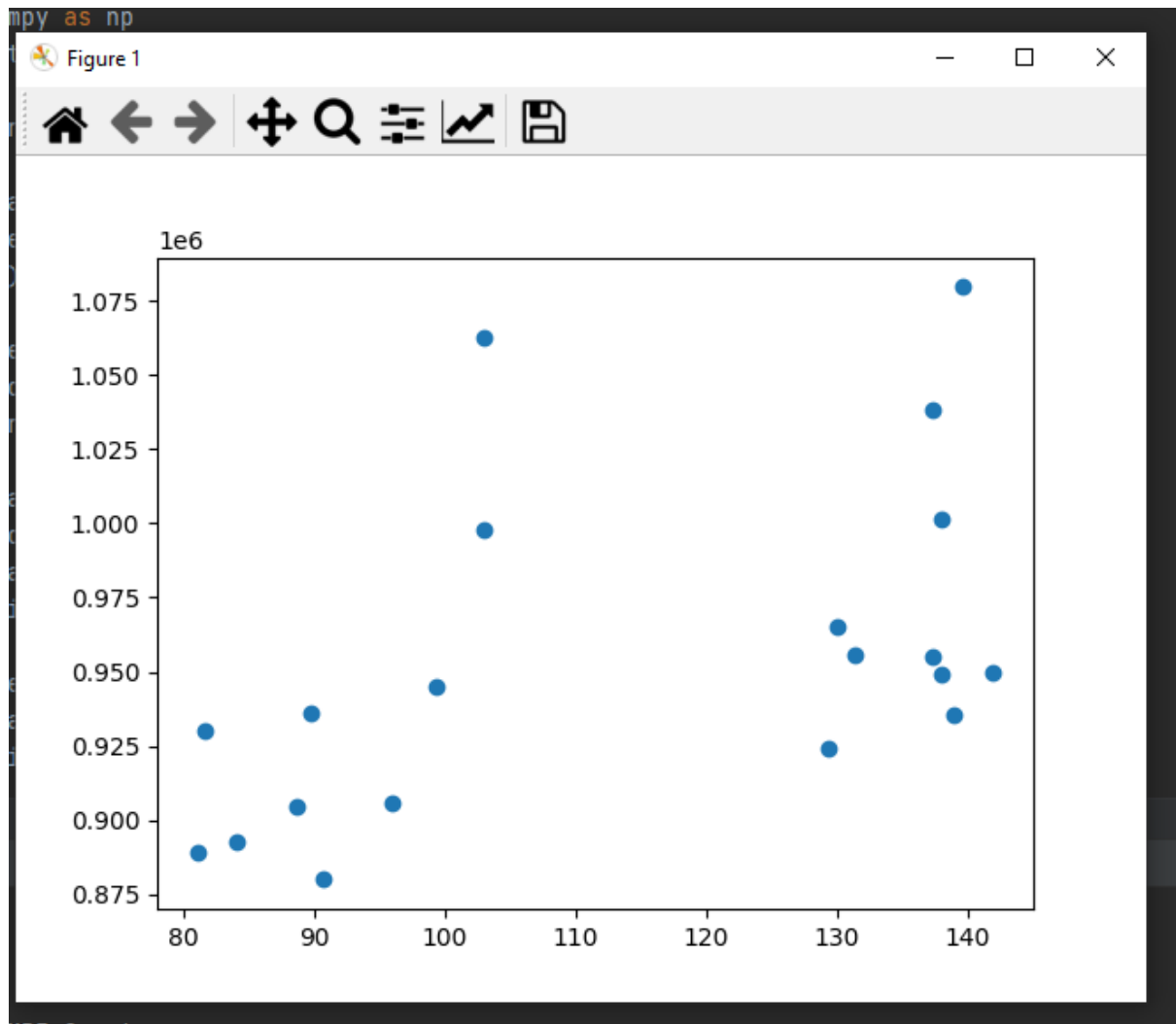
За даними коефіцієнта кореляції, який тип кореляції ви спостерігали між річними чистими продажами та кількістю магазинів у районі?

Відповідь: Спостерігаємо негативну кореляцію. Тобто збільшення (або зменшення) значень однієї змінної (stores) веде до закономірного зменшення (або збільшення) іншої змінної (sales).

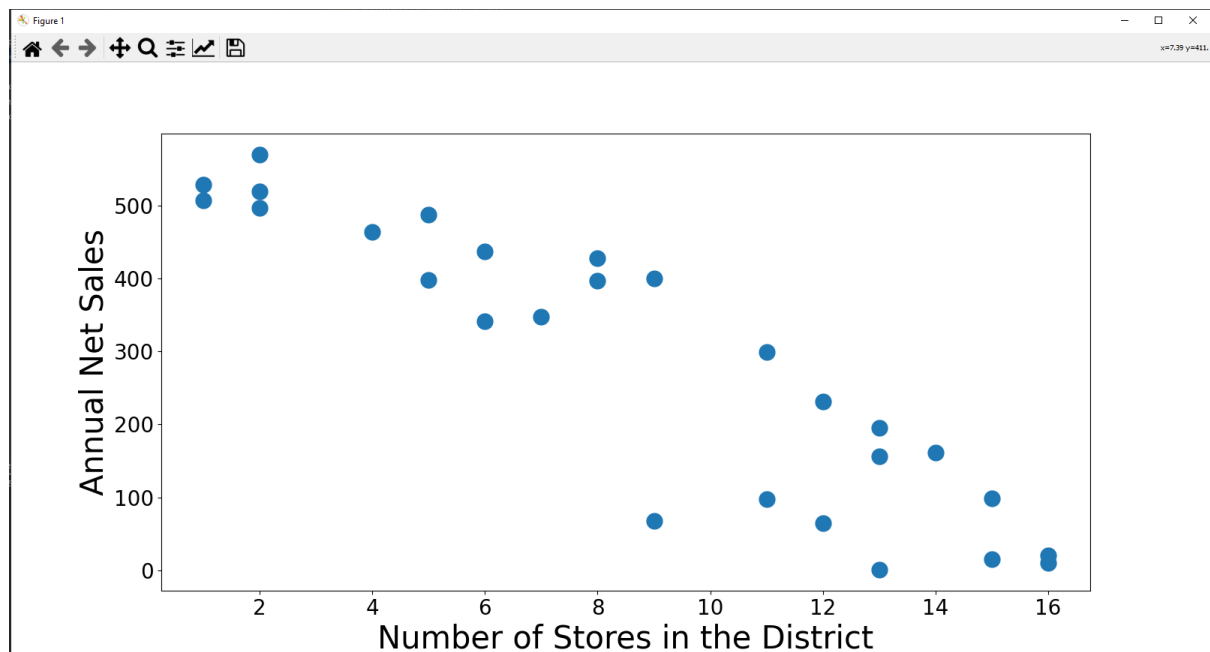
Який прогнозований чистий продаж, якщо в районі є 4 магазини?

Відповідь : 456.2313681207654

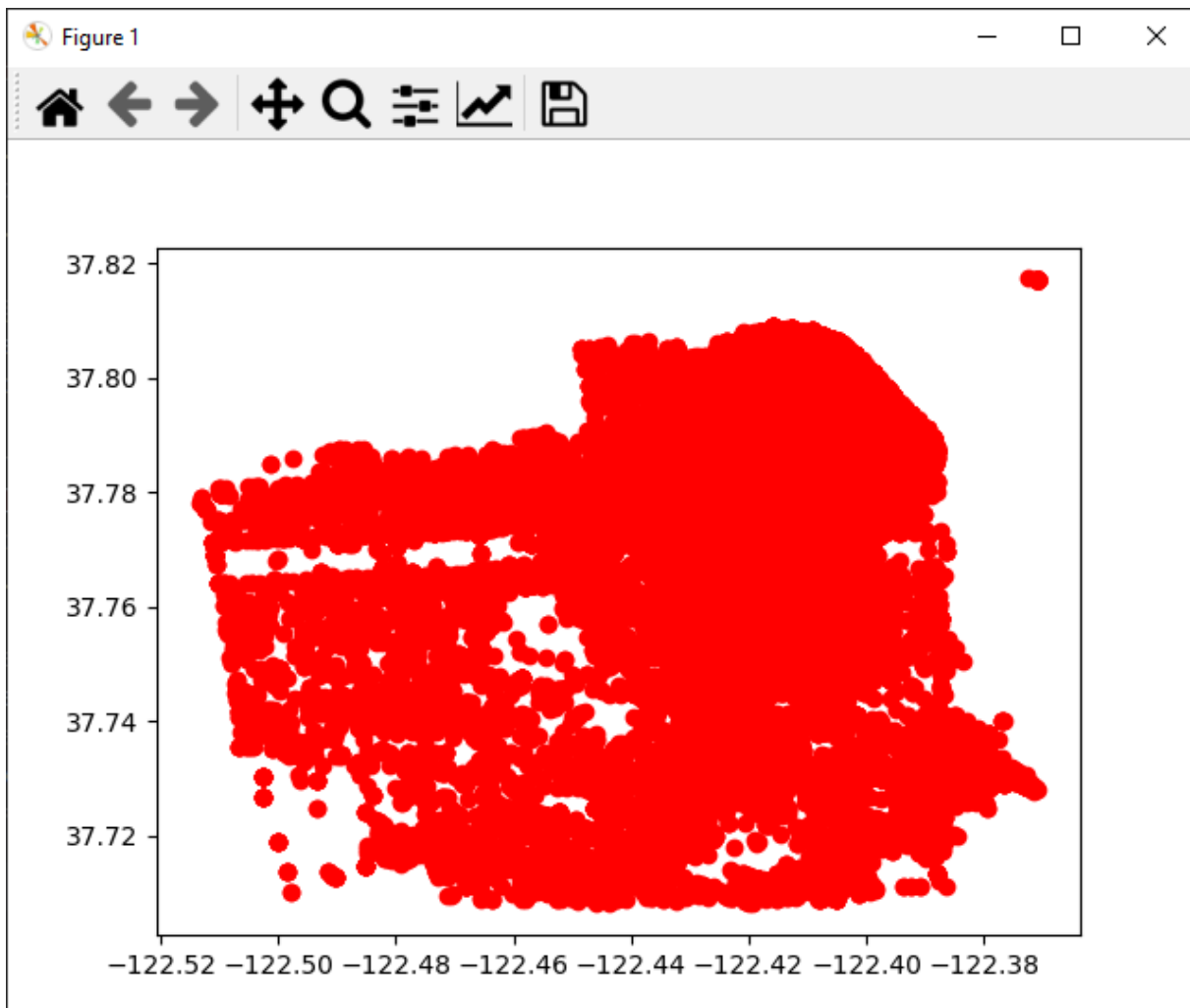
**Скріншоти з графіками**



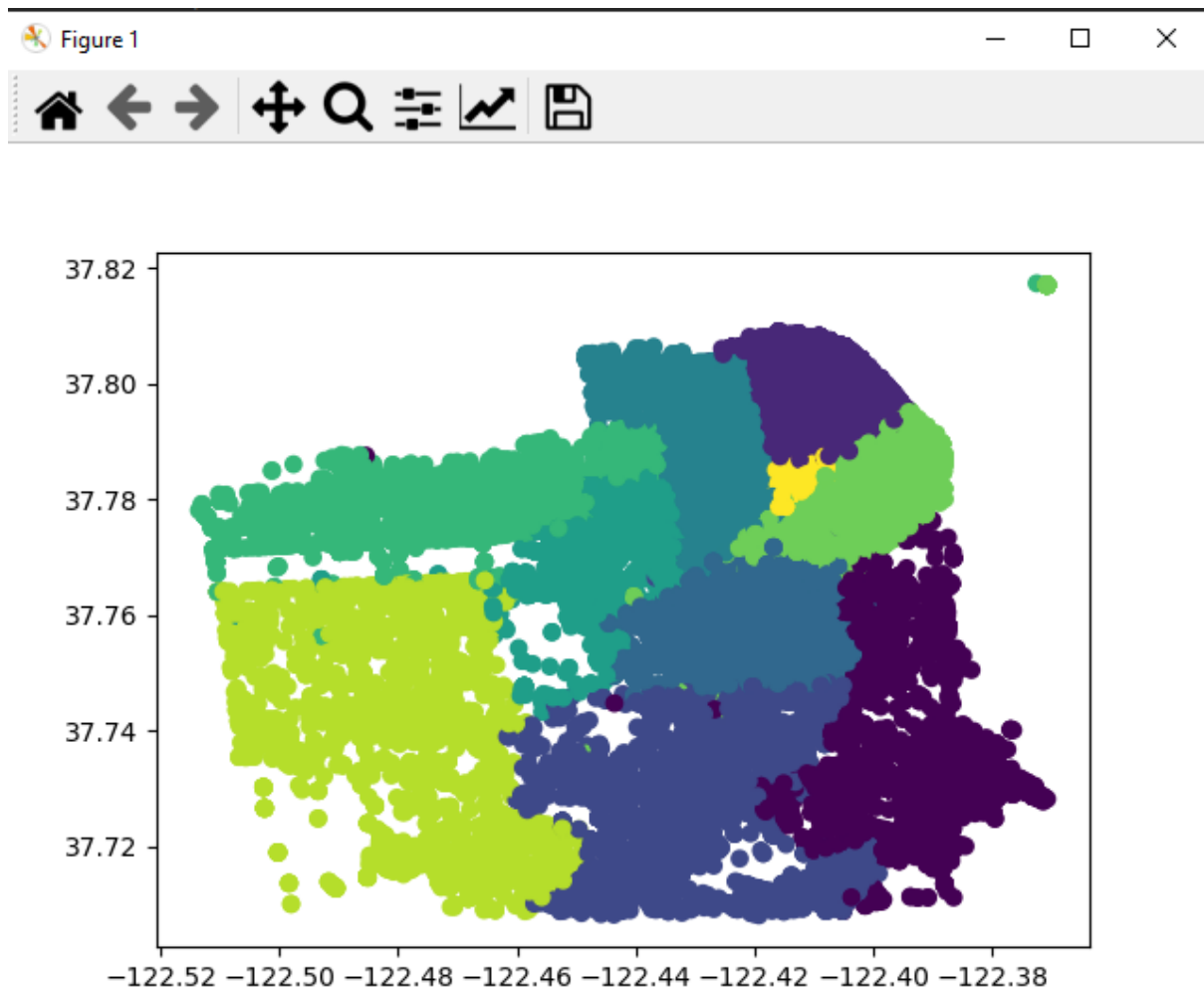
brainsize



# stores



data before clustering



clustered data

### Висновки

Під час виконання цієї лабораторної роботи я покращив навички програмування на Пайтоні, застосував різні алгоритми для оброблення статистичних даних, та використовувати багатопоточність для прискорення швидкості роботи програм та алгоритмів