

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

НАЦІОНАЛЬНИЙ ТЕХНІЧНИЙ УНІВЕРСИТЕТ УКРАЇНИ

“КИЇВСЬКИЙ ПОЛІТЕХНІЧНИЙ ІНСТИТУТ

імені ІГОРЯ СІКОРСЬКОГО”

Факультет прикладної математики

Кафедра програмного забезпечення комп’ютерних систем



**Лабораторна робота №2**

з дисципліни: «Технології оброблення великих даних»

на тему: «Аналіз та візуалізація даних у Python»

Виконав

студент IІІ курсу каф. ПЗКС ФПМ

групи КП-82

Мельничук Олексій Геннадійович

Перевірила

доц. каф. ПЗКС ФПМ

Олещенко Л.М.

Київ 2021

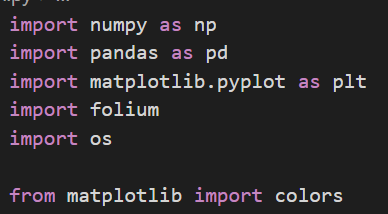
1. **Індивідуальне завдання**

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

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.

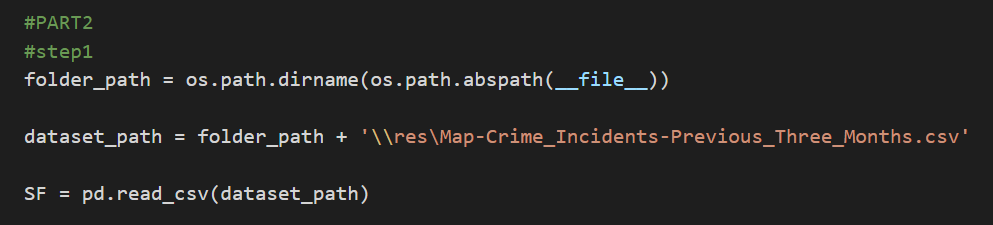
1. **Хід роботи**

**Частина 1: Імпорт пакетів Python**

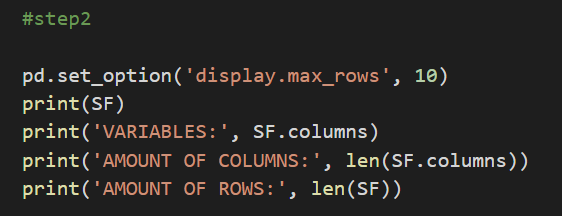


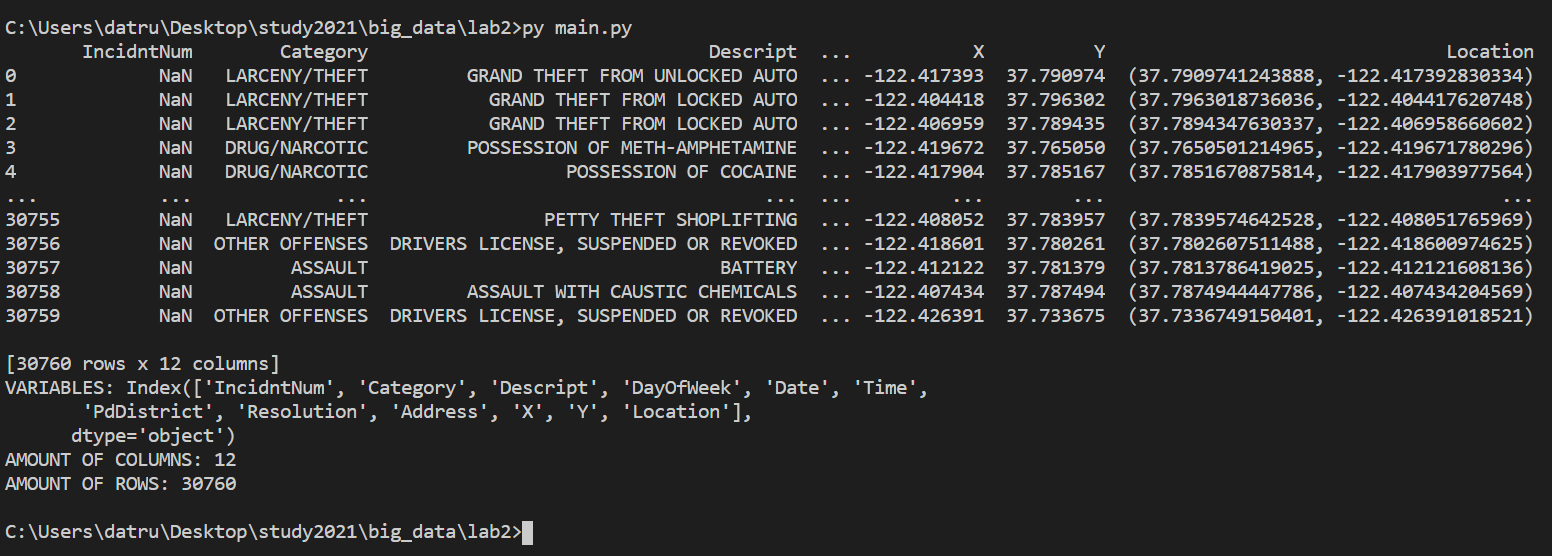
**Частина 2: Завантажте дані**

**Крок 1: Завантажте дані про злочини Сан-Франциско у дата фрейм.**



**Крок 2: Перегляньте імпортовані дані.**

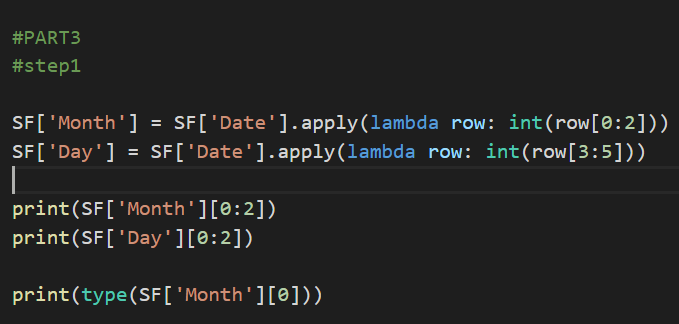


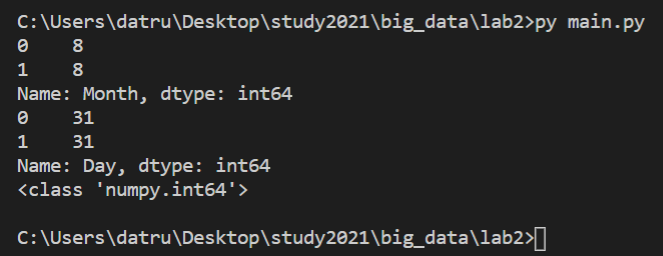


У фреймі знаходиться 12 змінних та 30760 рядків.

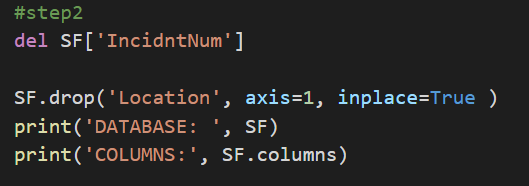
**Частина 3: Підготовка даних**

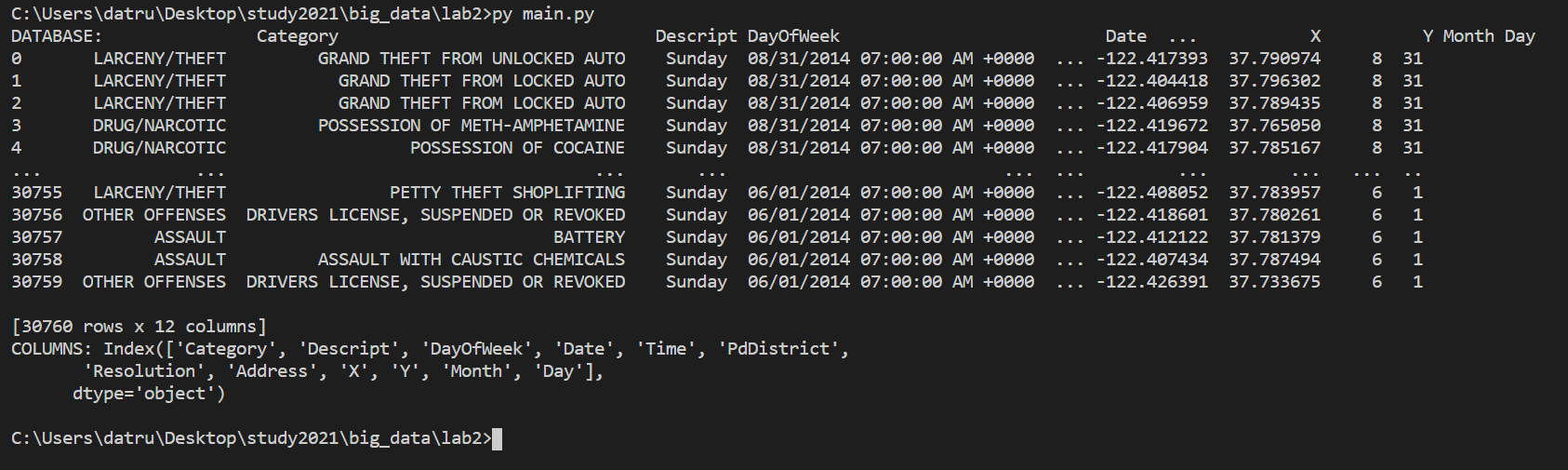
**Крок 1: Витягніть місяць і день із поля Дата.**

****

****

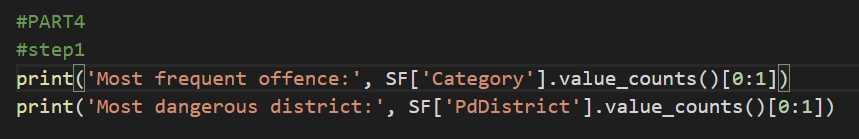
**Крок 2: Видаліть змінні з дата фрейму SF.**

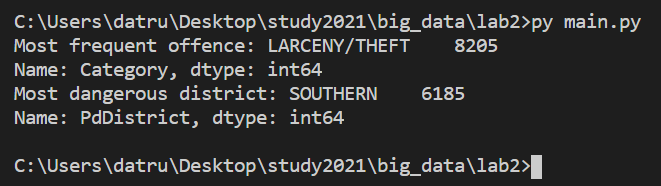
****

****

**Частина 4: Аналіз даних**

**Крок 1: Узагальнення змінних для отримання статистичної інформації.**

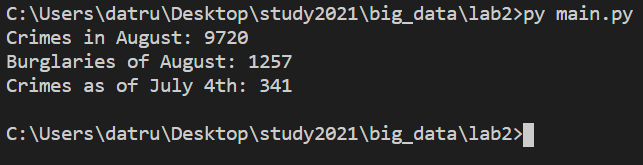
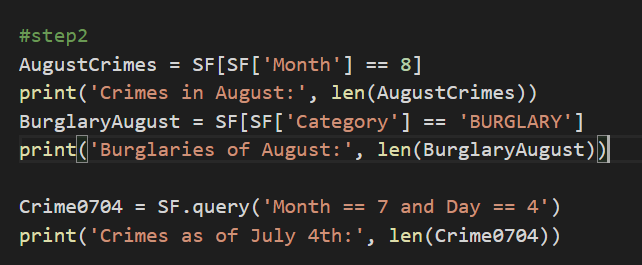
****

****

Найбільш скоєний злочин: Крадіжка, 8205 разів

Найнебезпечніший район: Південний, 6185 злочинів

**Крок 2. Підгрупуйте дані у менші дата фрейми (кадри даних).**



Кількість злочинів скоєних в Серпні: 9720

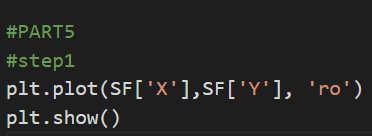
Кількість зламів скоєних в Серпні: 1257

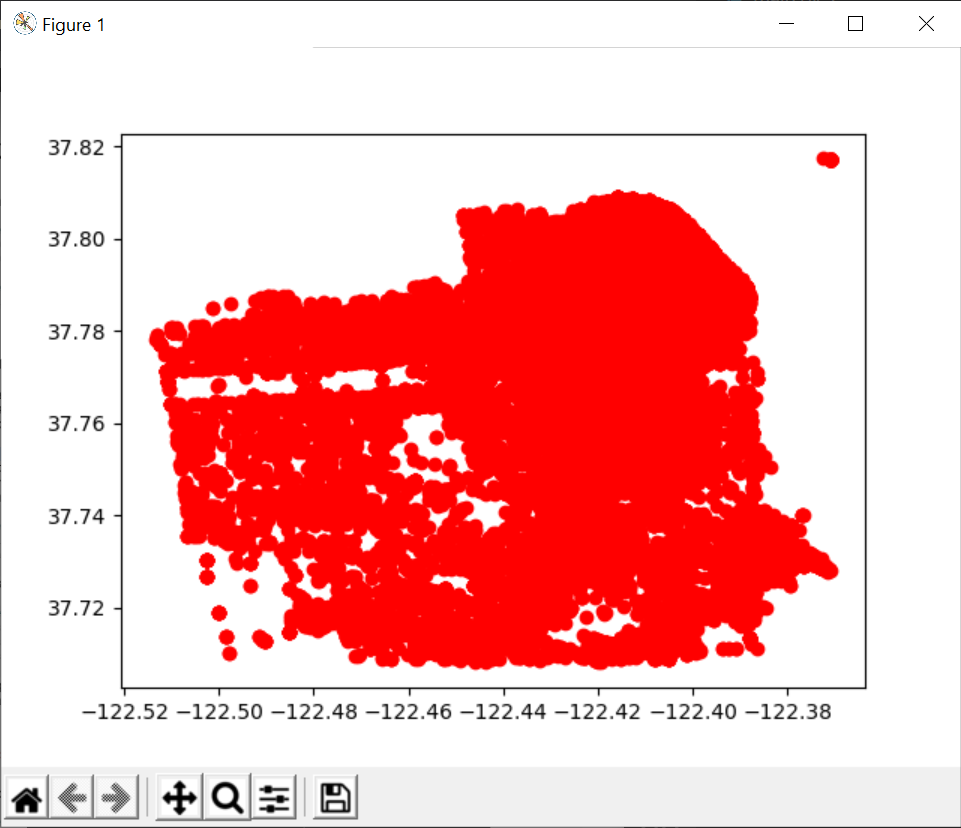
Кількість злочинів скоєних 4 Липня: 341

**Частина 5: Представлення даних**

**Крок 1: Побудуйте графік дата фрейму SF, використовуючи змінні X та Y.**

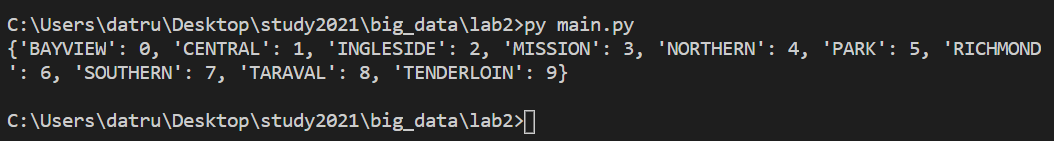
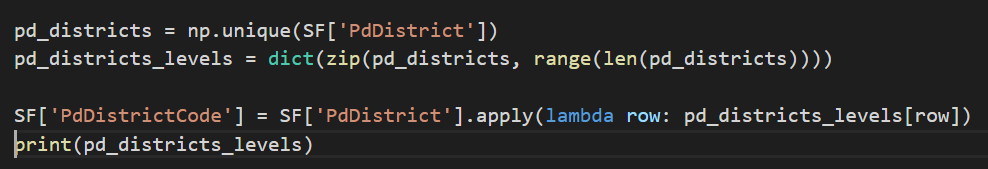
а) Використовуйте функцію plot()для побудови кадру даних SF.

****

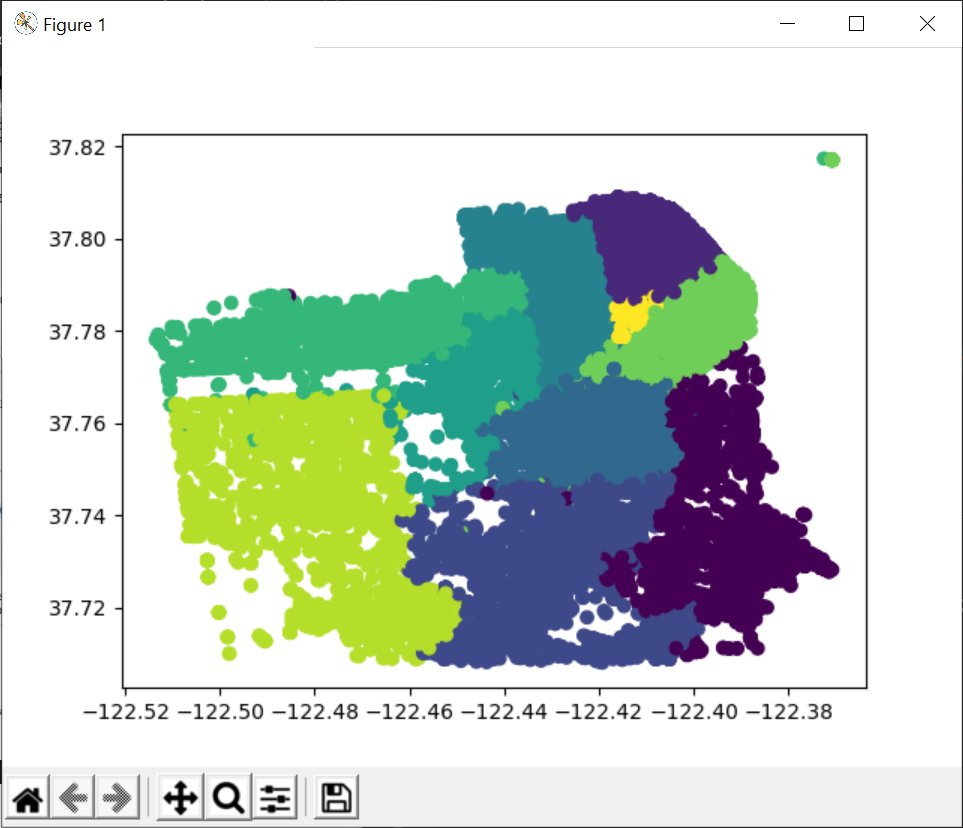
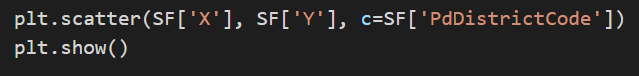


б) Визначте номери відділів поліції, складіть словник pd\_districts, щоб зв’язати їх рядок із цілим числом.

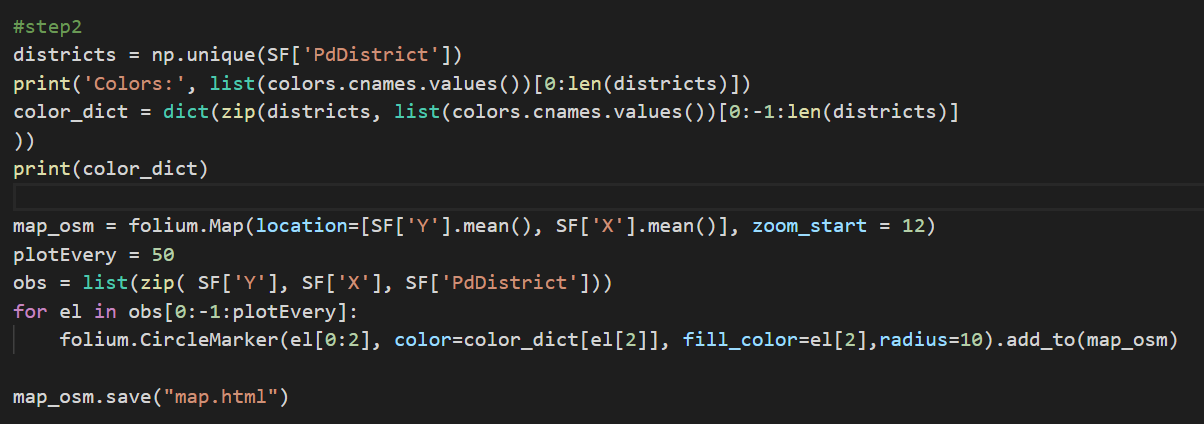
в) Використовуйте apply та, lambda, щоб додати ціле число поліцейського відділу до нового стовпця DataFrame

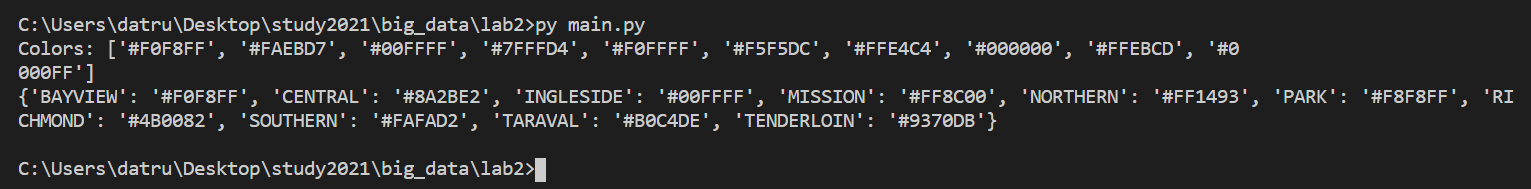


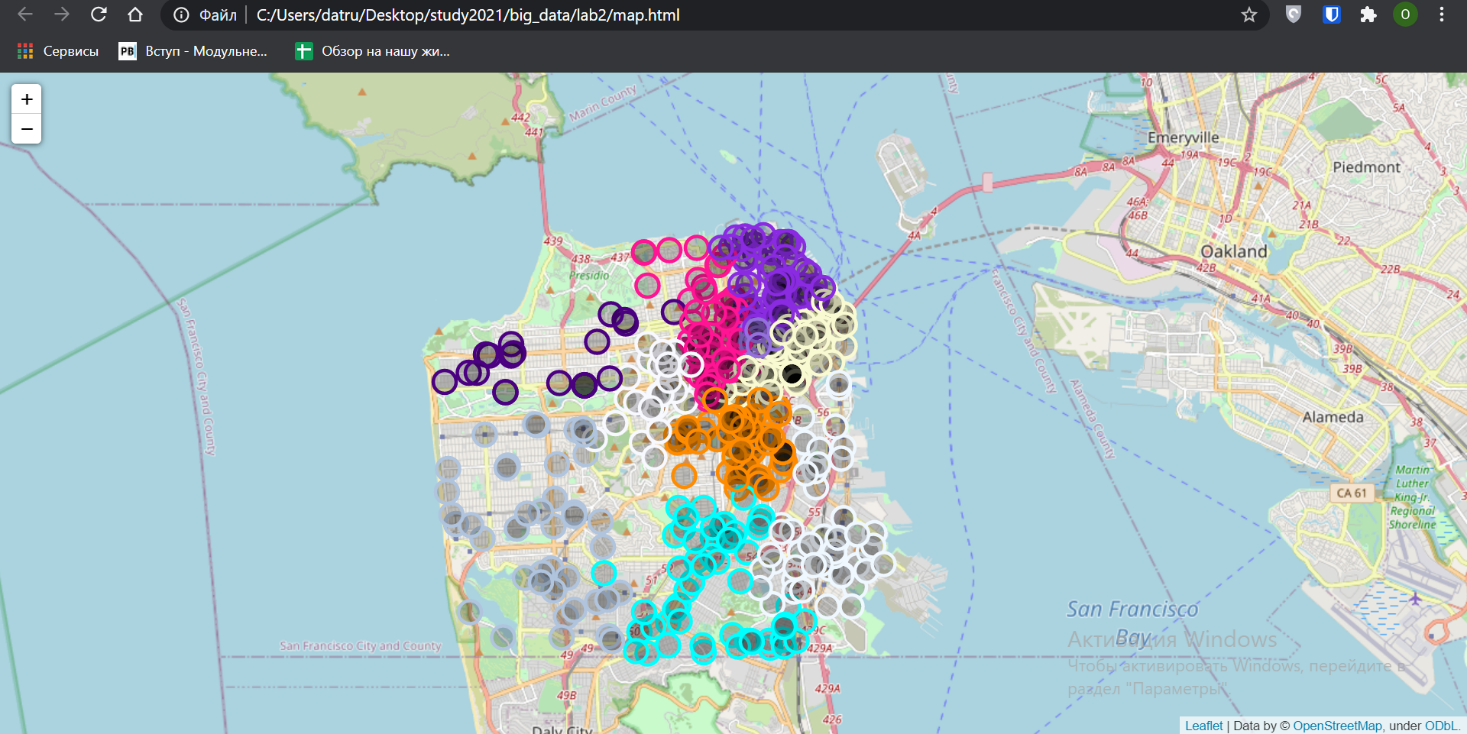
г) Використовуйте щойно створений PdDistrictCode для автоматичної зміни кольору



**Крок 2: Додайте пакети для побудови карт, щоб покращити сюжет.**

****

****

****

**Висновки**

В ході виконанні лабораторної роботи були використані навички обробки набору даних про злочини в Сан-Франциско з допомогою бібліотек для мови Python, а саме Numpy, Pandas, Matplotlib та Folium. Ці бібліотеки були використані для зручної побудови графіків для візуалізації даних з набору.

Найбільше злочинів було скоєно в Південному районі, також більше всього було скоєно крадіжок.