**Міністерство освіти і науки України**

**Національний університет харчових технологій**

Кафедра інформаційних технологій, штучного

інтелекту та кібербезпеки

**Звіт**

із лабораторної роботи № **8**

з дисципліни «Прикладний аналіз даних мовами VBA та Python»

на тему: «Кластеризація, класифікація та дерева в машинному навчанні»

Виконав:

Студент групи КН-1-3М

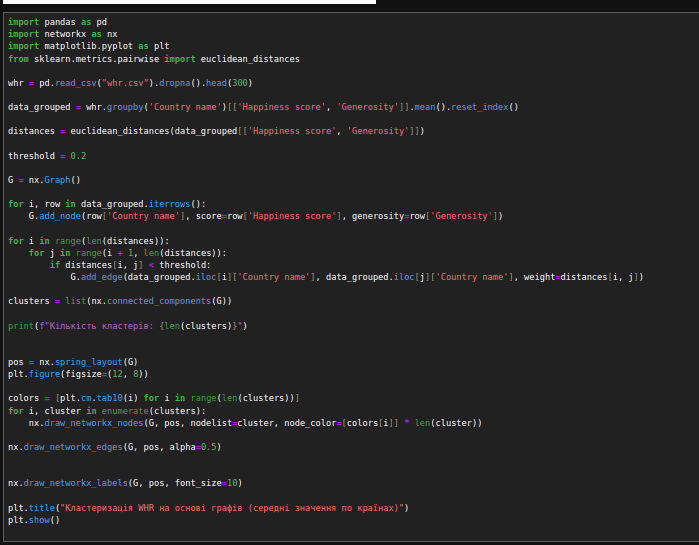
Кучерявий М. В.

**Київ — 2024**

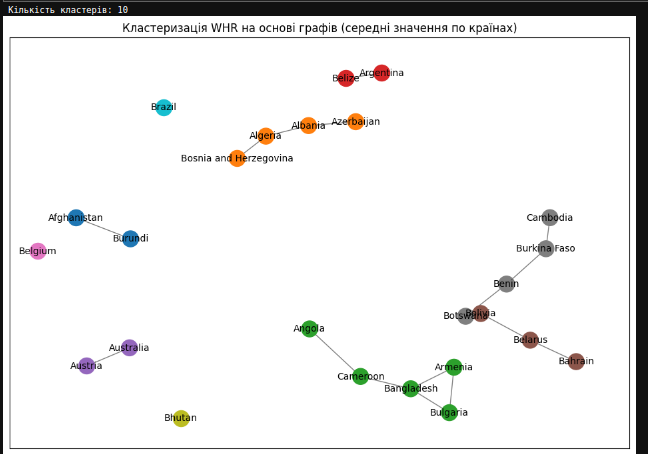
**Хід роботи:**

**Кластеризація за допомогою графів:**

Щоб провести кластеризацію даних за допомогою використанням графів, перетворимо таблицю World Happines Report використовуючи значення **Happiness Score** і **Generosity. Зв’язки між вузлами будуються на основі близькості значень цих показників. Кожен запис стане вузлом графа. Ребра між вузлами будуються на основі схожості.**



Результат дії

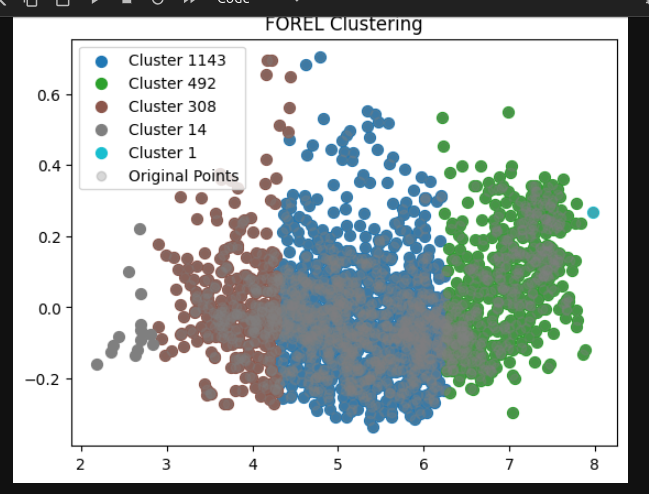
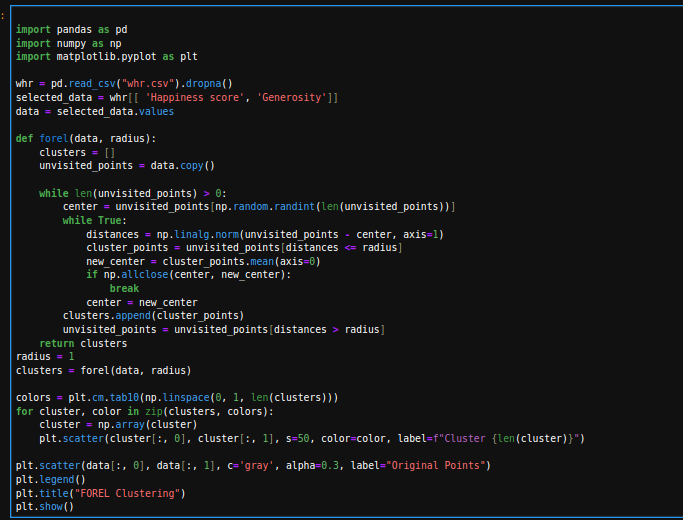


****Алгоритм FOREL****

Ф**ормує кластери, виокремлюючи щільні групи точок у просторі, які знаходяться на близьких відстанях одна від одної.**

**Задається радіус сфери (r), який визначає "зону притягання" для кластерів. Точки, що знаходяться в межах цього радіусу одна від одної, можуть бути об'єднані в один кластер.**

**Вибирається точка, з якої розпочнеться процес кластеризації. Ця точка стає центром першого кластера, і алгоритм шукає всі інші точки, що знаходяться в межах заданого радіусу (r) від неї.**



**K-середніх (K-means)**

Він використовується для поділу даних на **k** кластерів на основі мінімізації внутрішньокластерного розкиду (мінімізації середньоквадратичної помилки між точками та центроїдами кластерів). Це алгоритм навчання без вчителя, який працює з числовими даними.

Процес кластеризації включає такі етапи:

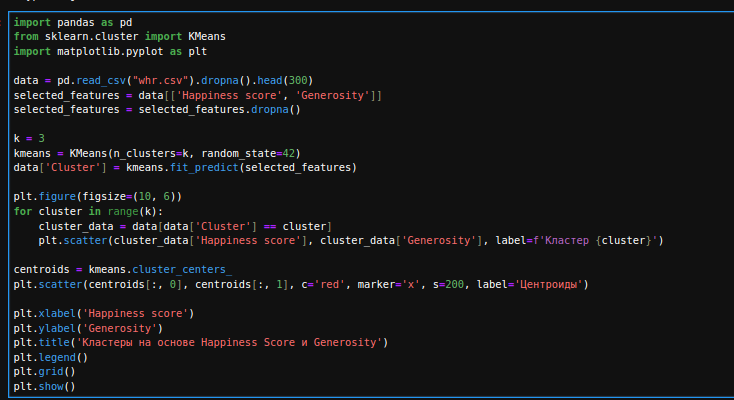
Вибір **k** початкових центроїдів (зазвичай випадково).

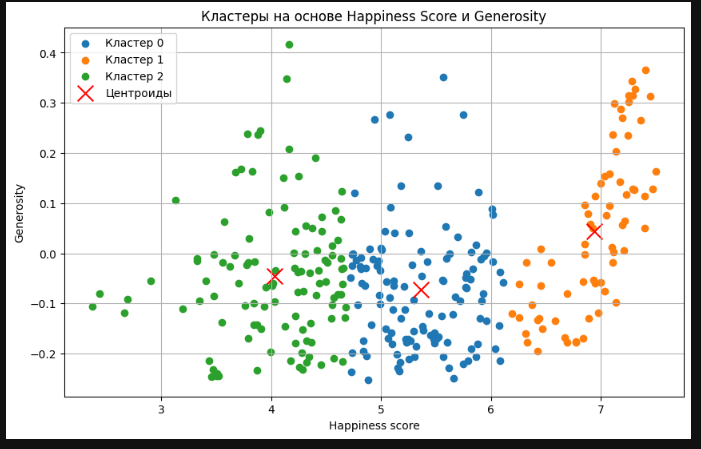
Призначення кожної точки до найближчого центроїда.

Оновлення центроїдів на основі середнього значення точок, що належать кластеру.

Повторення кроків 2 і 3 до досягнення стійкого стану, коли центроїди не змінюються значно.

Алгоритм K-середніх широко використовується через свою простоту та ефективність, особливо для великих наборів даних.





****Вибір оптимальної кількості кластерів**:**

### **Метод "лікоть" (Elbow Method)**

Цей метод допомагає знайти оптимальну кількість кластерів шляхом оцінки зміни інерції (відстані від точок до центроїдів кластерів) із ростом кількості кластерів.

Обчисліть модель кластеризації для різних значень k (кількість кластерів).

Для кожного k обчисліть **інерцію** (значення атрибута inertia\_ у моделі KMeans).

Побудуйте графік інерції залежно від k.

Виберіть точку, де зміна інерції починає сповільнюватися ("лікоть").

### **Силуетний аналіз (Silhouette Score)**

Силуетний коефіцієнт оцінює, наскільки добре кожна точка належить своєму кластеру. Значення варіюються від -1 до 1:

