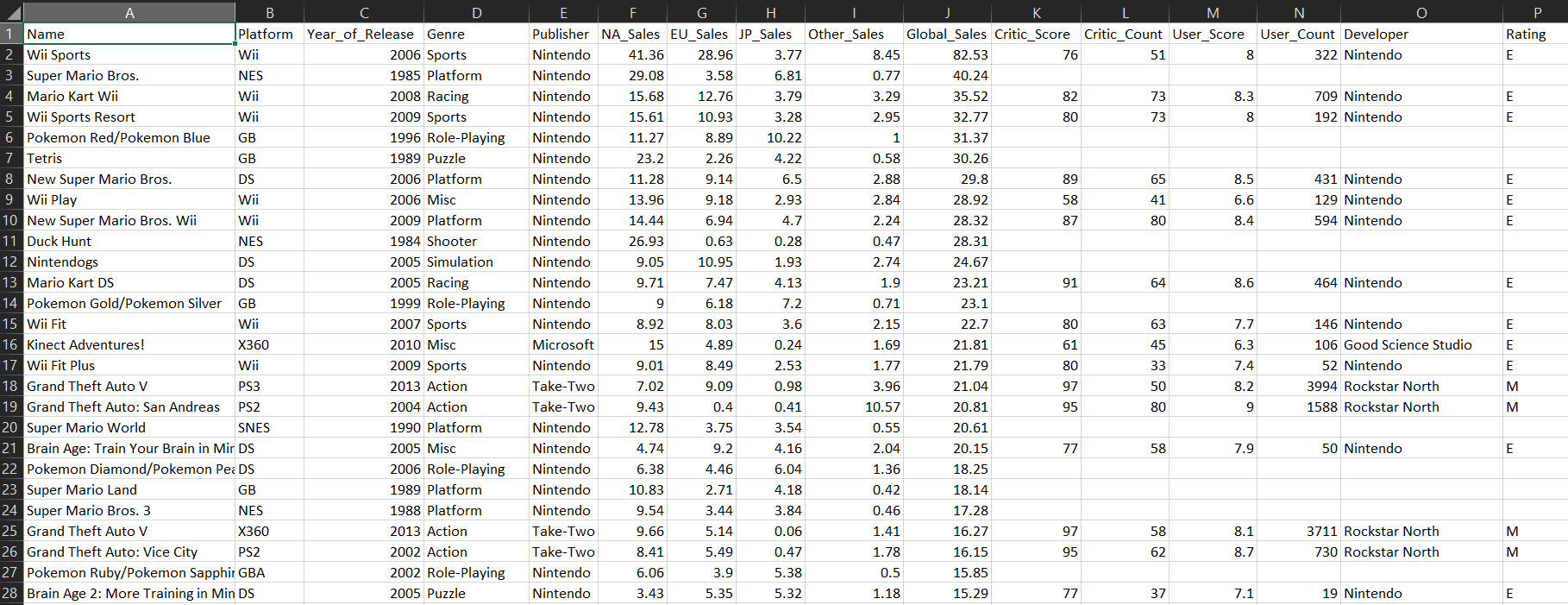
Застосування нечіткої кластеризації C-середніх (FCM algorithm) для аналізу даних на мові Python.

Файл вхідного набору даних про продажі відеоігор:



**Хід роботи**

Для використання алгоритму нечіткої кластеризації потрібна множина точок, що утворюють між собою скупчення в умовному просторі. Для того, щоб була можливість зобразити результати графічно та інтерпретувати їх, будемо виводити утворені кластери у двовимірному просторі.

1. Імпорт бібліотек

import numpy as np  
from fcmeans import FCM  
from matplotlib import pyplot as plt  
import csv

2. Створюємо порожній масив для попарних ознак

X = []

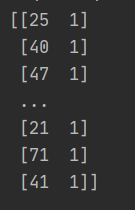
3. Зчитуємо csv-файл і виділяємо дві ознаки, що нас цікавлять. В даному випадку ми з’ясовуємо, як пов’язані вік людини та наявність у неї ліцензії водія. Обрані ознаки приєднуються до списку X у форматі [a, b]

with open('test.csv') as File:  
 reader = csv.DictReader(File)  
 for row in reader:  
 X.append([int(row['Age']), int(row['Driving\_License'])])

4. Приводимо отриманий масив до формату numpy.array

X=np.array(X)  
print(X)

Його вигляд:



5. З бібліотекою fcmeans навчаємо модель та виводимо результат:

fcm = FCM(n\_clusters=2)  
fcm.fit(X)  
  
# outputs  
fcm\_centers = fcm.centers  
fcm\_labels = fcm.predict(X)  
  
# plot result  
f, axes = plt.subplots(1, 2, figsize=(11, 5))  
axes[0].scatter(X[:, 0], X[:, 1], alpha=.1)  
axes[1].scatter(X[:, 0], X[:, 1], c=fcm\_labels, alpha=.1)  
axes[1].scatter(fcm\_centers[:, 0], fcm\_centers[:, 1], marker="+", s=500, c='w')  
  
axes[0].set\_xlabel("Вік")  
axes[0].set\_ylabel("Ліцензія")

**Висновок**: в ході даної роботи було використано реалізацію методу кластеризації C-середніх Python-бібліотеки *fcmeans*, для аналізу даних з продажу відеоігор та представлення кластерів аналізу цих даних у графічному вигляді.