МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ РАДІОЕЛЕКТРОНІКИ

Звіт

до лабораторної

роботи з дисципліни

**"інтелектуальний аналіз даних"**

для студентів освітньо-кваліфікаційного рівня «бакалавр»

за спеціальністю 122 «Комп’ютерні науки та інформаційні технології», освітня програма «Інформатика»

Виконав студент группи ІТІНФ-18-2

Туркін Микита

ХАРКІВ 2020

**Лабораторна робота №5. Застосування кластерного аналізу**

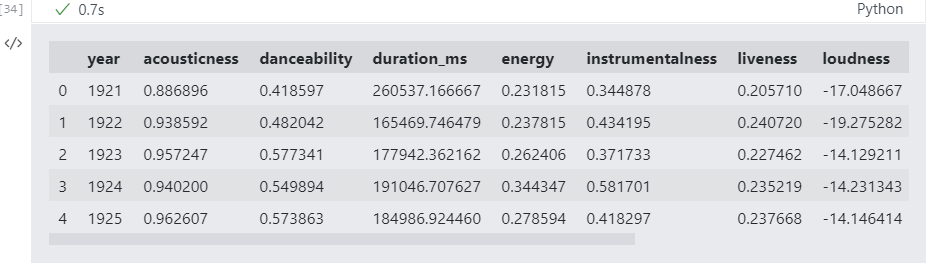
**Мета:** Освоїти практичні засоби застосування кластерного аналізу для даних, що описуються множиною векторів з числовими компонентами при вирішенні задач ІАД.

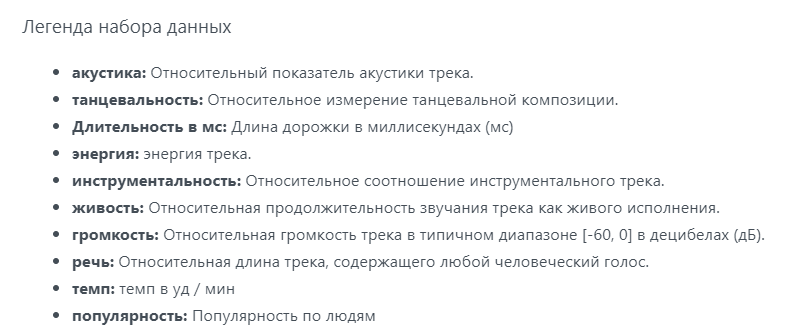
**Xід роботи**

Візьмемо набір данних, на якому будемо робити кластерний аналіз

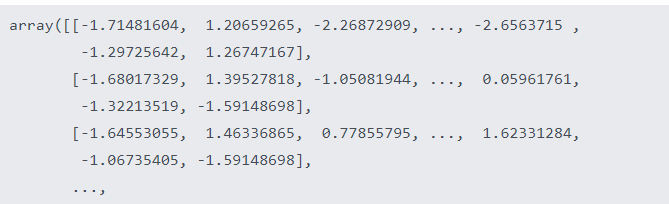
Джерело: <https://www.kaggle.com/yamaerenay/spotify-dataset-19212020-160k-tracks>

«Audio features of 160k+ songs released in between 1921 and 2020»





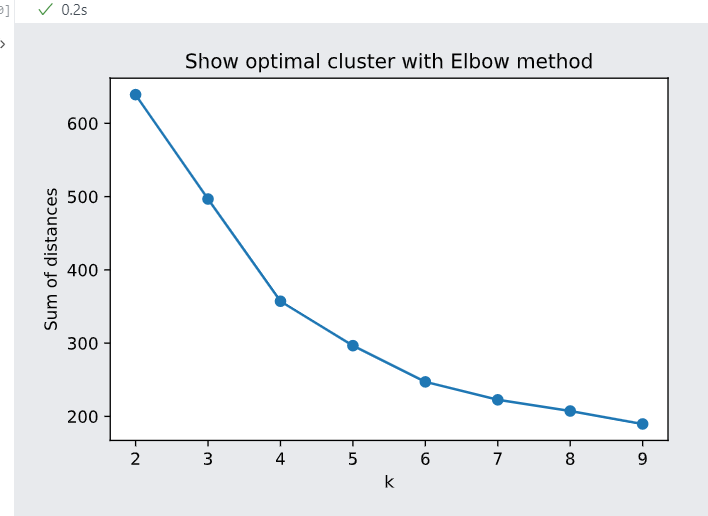
Перед початком кластеризації проведемо операцію нормалізації данних:



Кластеризацію будемо проводити задомогою алгоритму k-середніх

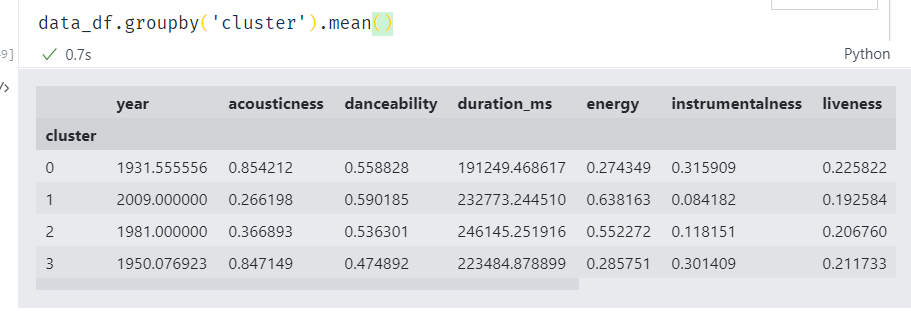
Для цього алгоритму нам потрібно завчасно знати кількіть кластерів. Оскільки ми розглядаемо зміни характеристик у музичних треках у період з 1920 по 2020 роки, можно зробити припущення, що кластерів може бути від 2 до 10

Спробуємо побувати різні кластеризації для цього діапозону кластерів та побудуємо графік на якому зображено як змінюєтся сумма відстаней від центрів кластеризації:



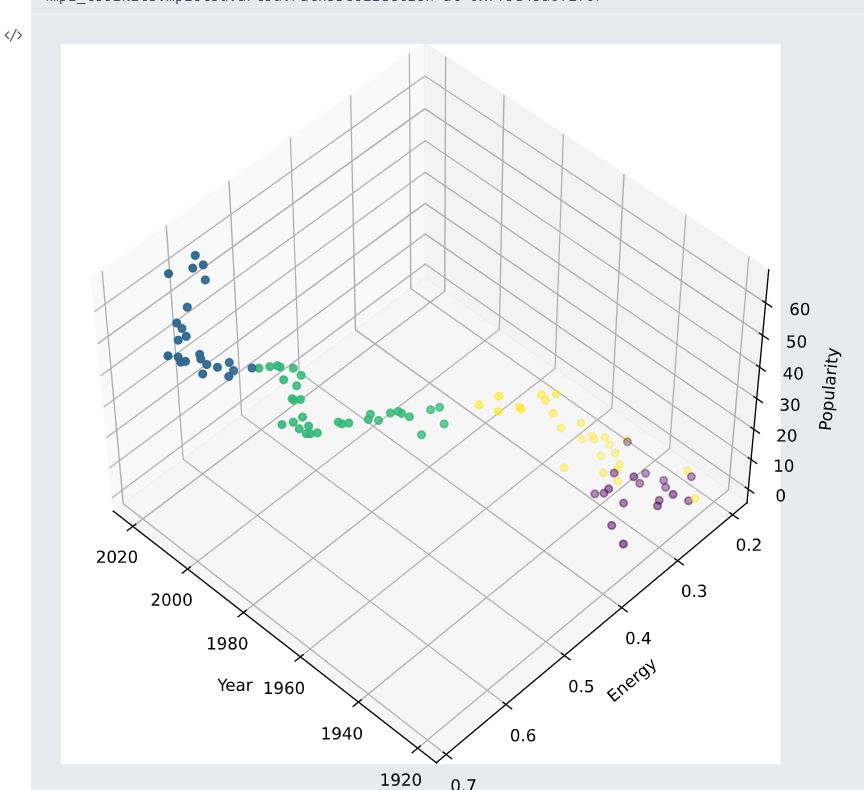
На 4х кластерах можно відбувается перелом графіку, тому можно зробити припущення, що кількість різних кластерів наших даних дорівнює 4

Зробивши кластерізацію на 4 класси можно порахувати середні значення, для кожного кластеру:

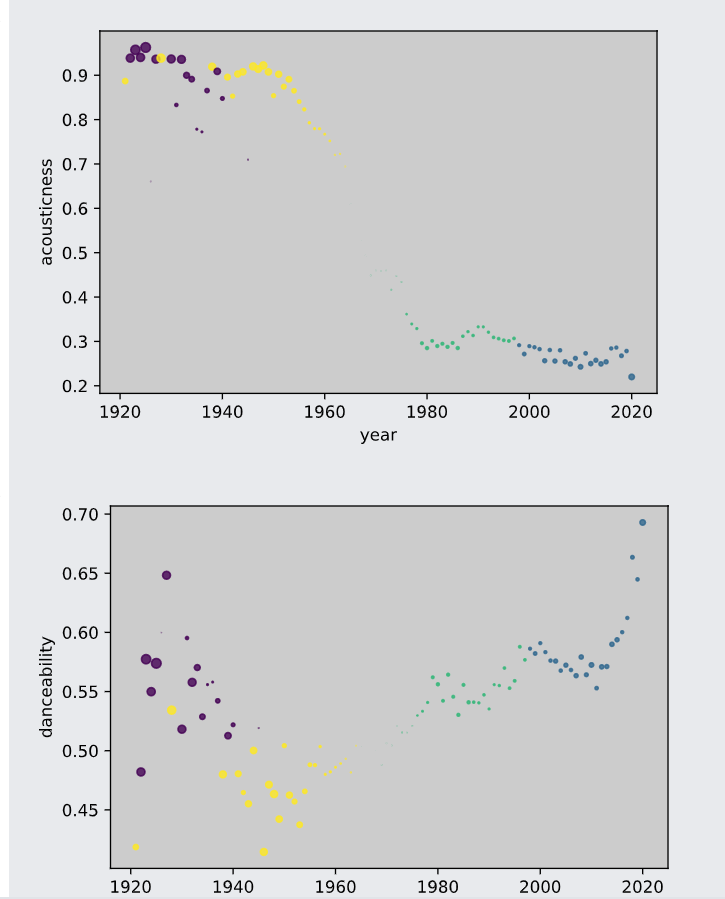


Бачимо що кластери по рокам розділились приблизно на: 1930ті, 1950ті, 1980ті та 2010ті

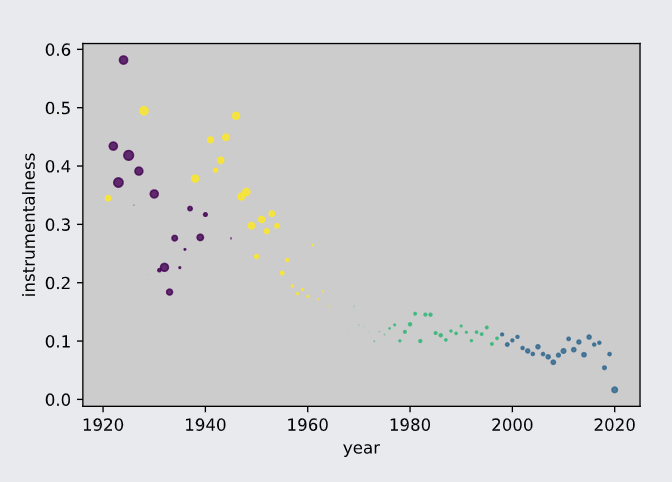
Тепер побудуємо графіки щоб побачити як змінюются характеристики для різних параметрів:



На цьому графіку можна побачити як змінюєтся енергійність та полярність музики від 1920 до 2020



Графіки акустичності та танцювального характеру пісень у періоді з 1920 до 2020



Тут можна побачити, що інструментальність музики з часом йде до появи пісень з синтезованою музикою.



Гучність концертів і музики зростає, при чому почало зростати тільки для кластерів 1980 і 2010х

**Висновок:** Освоїв практичні засоби застосування кластерного аналізу для даних, що описуються множиною векторів з числовими компонентами при вирішенні задач ІАД.