КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ТАРАСА ШЕВЧЕНКА

Кафедра інтелектуальних та інформаційних систем

Лабораторна робота № 4

з дисципліни

“Попередній аналіз та підготовка даних”

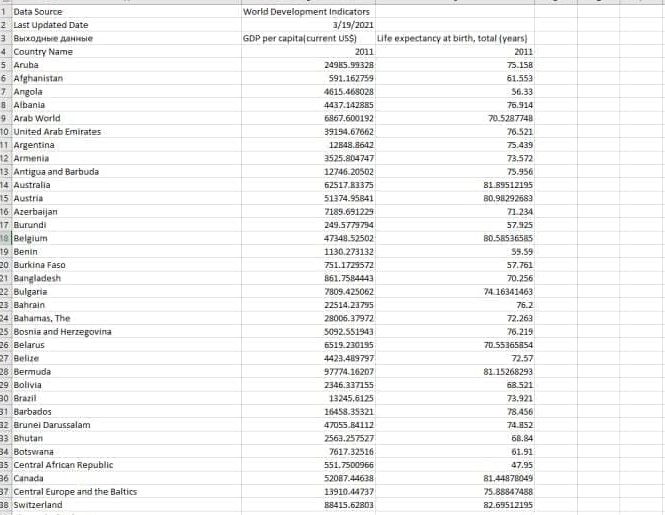
Виконав студент

групи КН-31

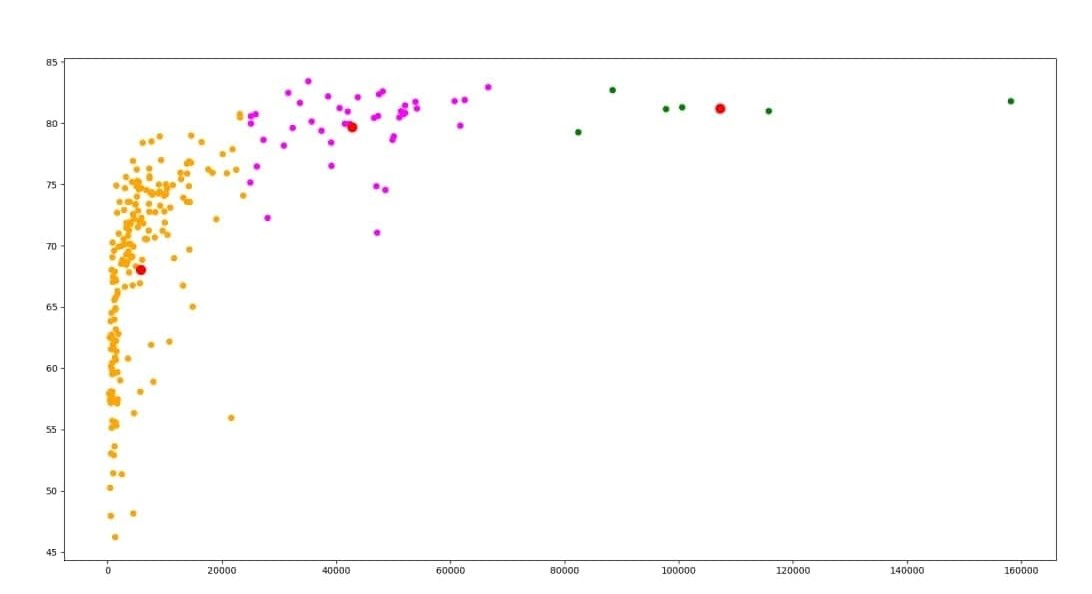
Пашковський Павло Володимирович

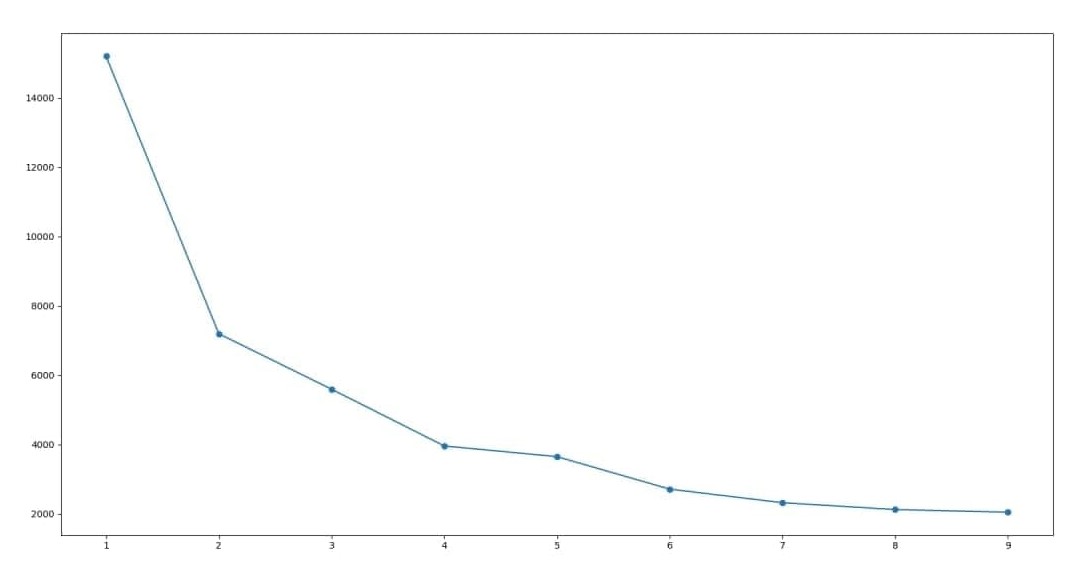
Київ-2021

Выходные данные



Результат выполнения программы





Код программы  
  
from openpyxl import load\_workbook

import matplotlib.pyplot as plt

import numpy as np

from sklearn.cluster import KMeans

from scipy.spatial.distance import cdist

#использование библиотек для кластеризации данных,а так же работы с файлам иксель(в том числе парсинг значений)

wb = load\_workbook('data.xlsx')#открываем файл,ниже открывает нужный лист книги

sheet = wb.get\_sheet\_by\_name('data')

data = []#объявляем массив,для заполнения его данными из нашего файла

for cellObj in sheet['B5':'C240']:

data.append([cell.value for cell in cellObj])

lock = []

for cl\_n in range(1, 10):

kmeans = KMeans(n\_clusters=cl\_n)

y\_kmeans = kmeans.fit\_predict(data)

#использование функций библиотеки для kmeans

data = np.array(data)

lock.append(sum(np.min(cdist(data, kmeans.cluster\_centers\_, 'euclidean'), axis=1)) / data.shape[0])

plt.subplot(111)

plt.plot(range(1, 10), lock, 'o-')

plt.show()

#демонастрация графика в окне

cl\_n = int(input("Введіть бажану кулькість кластерів"))

kmeans = KMeans(n\_clusters=cl\_n)

y\_kmeans = kmeans.fit\_predict(data)

#ввод количества кластеров

colors = np.array(['orange', 'magenta', 'green', 'blue', 'purple', 'brown', 'black', 'gray'])

plt.scatter([i[0] for i in data], [i[1] for i in data], c=colors[y\_kmeans])

plt.scatter(kmeans.cluster\_centers\_[:, 0], kmeans.cluster\_centers\_[:, 1], c='r', s=100)

plt.show()

#демонастрация графика согласно введённым данным