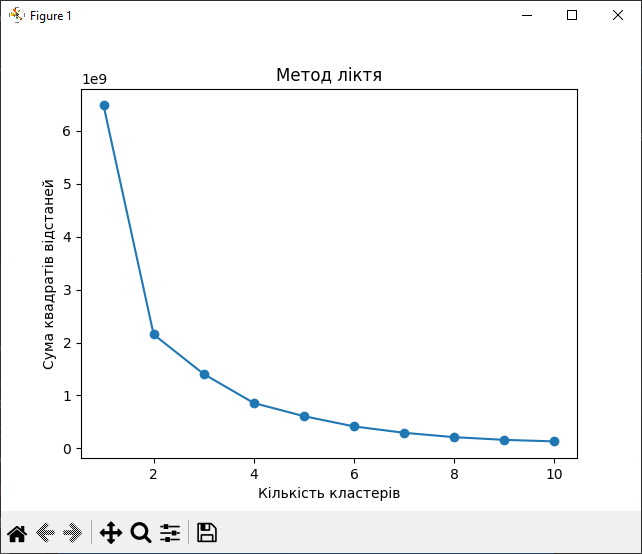
**Звіт до лабораторної роботи 7  
варіант 5**

**Завдання:** Для даних, що використовувалися у попередніх роботах, виконати поділ регіонів на кластери. Як значення ознак взяти дані у стовпчиках I, J, K

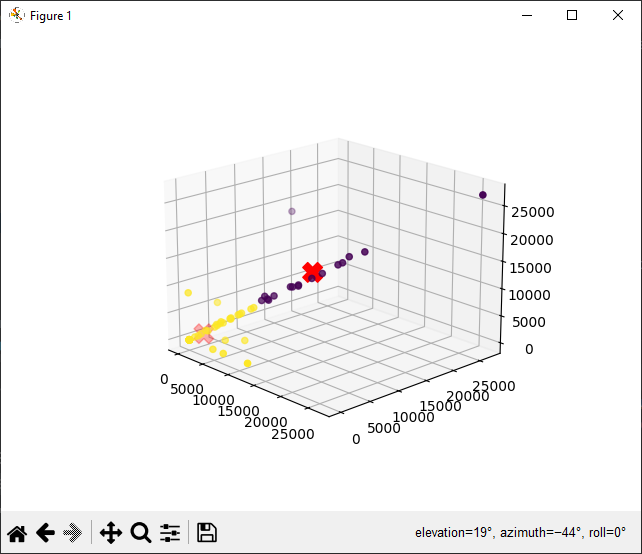
Розглянемо кластерний аналіз тривимірного вектору об’єднаного з цих даних, для цього треба визначити кількість кластерів:

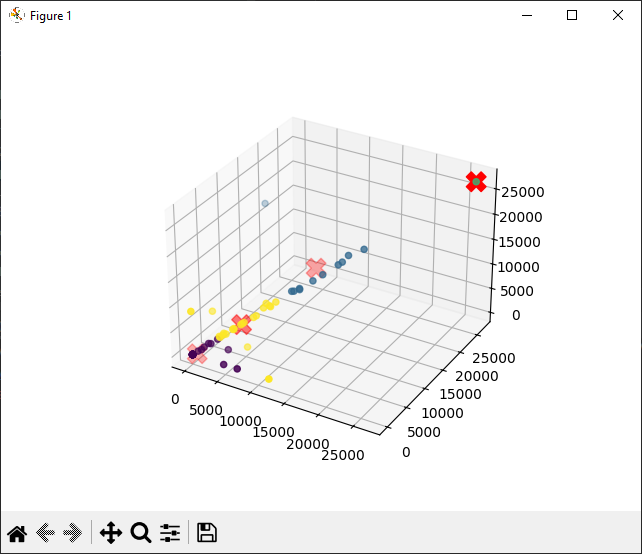
import matplotlib.pyplot as plt  
import numpy as np  
import pandas as pd  
from sklearn.cluster import KMeans  
  
df = pd.read\_csv('output.csv')  
vector\_i = df['2018'].values  
vector\_j = df['2019'].values  
vector\_k = df['2020'].values  
data = np.column\_stack((vector\_i, vector\_j, vector\_k))  
  
possible\_k = range(1, 11)  
inertia = []  
for k in possible\_k:  
 kmeans = KMeans(n\_clusters=k)  
 kmeans.fit(data)  
 inertia.append(kmeans.inertia\_)  
  
plt.plot(possible\_k, inertia, marker='o')  
plt.xlabel('Кількість кластерів')  
plt.ylabel('Сума квадратів відстаней')  
plt.title('Метод ліктя')  
plt.show()



Судячи з графіку методу ліктя можемо сказати що оптимальною кількістю кластерів буде приблизно 2 або 4. Побудуємо графік для обох варіантів.

from sklearn.cluster import KMeans  
import numpy as np  
import matplotlib.pyplot as plt  
import pandas as pd  
  
df = pd.read\_csv('output.csv')  
vector\_i = df['2018'].values  
vector\_j = df['2019'].values  
vector\_k = df['2020'].values  
  
X = np.column\_stack((vector\_i, vector\_j, vector\_k))  
  
# кількість кластерів  
k = 2  
  
# Створення та навчання моделі K-Means  
kmeans = KMeans(n\_clusters=k)  
kmeans.fit(X)  
  
# Отримання міток кластерів та центрів  
labels = kmeans.labels\_  
centers = kmeans.cluster\_centers\_  
  
# Візуалізація результатів в тривимірному просторі  
fig = plt.figure()  
ax = fig.add\_subplot(111, projection='3d')  
ax.scatter(X[:, 0], X[:, 1], X[:, 2], c=labels, cmap='viridis')  
ax.scatter(centers[:, 0], centers[:, 1], centers[:, 2], marker='X', s=200, color='red')  
plt.show()

  
Для двох



Для чотирьох.