import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

from sklearn.cluster import KMeans

from sklearn.metrics import silhouette\_score

# 读取数据，指定编码

file\_path = 保存路径'

data = pd.read\_excel(file\_path)

# 查看数据前几行

print(data.head())

# 打印列名

print(data.columns)

# 确定目标列名

target\_column = '肥胖等级'

# 确定最佳K值

# 使用肘部法则

loss = []

for i in range(2, 10):

    kmeans = KMeans(n\_clusters=i, random\_state=42)

    kmeans.fit(data.drop(target\_column, axis=1))

    loss.append(kmeans.inertia\_)

plt.plot(range(2, 10), loss)

plt.xlabel('k')

plt.ylabel('loss')

plt.title('Elbow Method for Optimal k')

plt.show()

# 使用轮廓系数

score = []

for i in range(2, 10):

    kmeans = KMeans(n\_clusters=i, random\_state=42)

    labels = kmeans.fit\_predict(data.drop(target\_column, axis=1))

    score.append(silhouette\_score(data.drop(target\_column, axis=1), labels, metric='euclidean'))

plt.plot(range(2, 10), score)

plt.xlabel('k')

plt.ylabel('silhouette\_score')

plt.title('Silhouette Score for Optimal k')

plt.show()

# 选择最佳K值

best\_k = 3  # 例如，我们选择K=3作为最佳聚类数

# 进行聚类

kmeans = KMeans(n\_clusters=best\_k, random\_state=42)

labels = kmeans.fit\_predict(data.drop(target\_column, axis=1))

data['cluster'] = labels

# 可视化聚类结果

fig, axs = plt.subplots(2, 2, figsize=(12, 10))

axs[0, 0].scatter(data['年纪'], data['体重'], c=labels, cmap='viridis', label='Cluster')

axs[0, 0].set\_title('Age vs Weight')

axs[0, 1].scatter(data['身高'], data['体重'], c=labels, cmap='viridis', label='Cluster')

axs[0, 1].set\_title('Height vs Weight')

axs[1, 0].scatter(data['卡路里消耗频率'], data['身体活动频率'], c=labels, cmap='viridis', label='Cluster')

axs[1, 0].set\_title('Calorie Consumption vs Physical Activity')

axs[1, 1].scatter(data[target\_column], data['cluster'], c='red', label='Obesity Level')

axs[1, 1].set\_title('Obesity Level vs Cluster')

plt.show()

# 聚类效果评估

# 计算每个聚类的平均年纪、身高、体重、卡路里消耗频率、主餐次数、蔬菜食用、身体活动频率

avg\_df = data.groupby('cluster', as\_index=False).mean()

# 查看平均值数据

print(avg\_df)

# 可视化聚类效果

fig, axs = plt.subplots(2, 2, figsize=(12, 10))

axs[0, 0].bar(avg\_df['cluster'], avg\_df['年纪'], color='blue')

axs[0, 0].set\_title('Average Age by Cluster')

axs[0, 1].bar(avg\_df['cluster'], avg\_df['体重'], color='blue')

axs[0, 1].set\_title('Average Weight by Cluster')

axs[1, 0].bar(avg\_df['cluster'], avg\_df['身高'], color='blue')

axs[1, 0].set\_title('Average Height by Cluster')

axs[1, 1].bar(avg\_df['cluster'], avg\_df['卡路里消耗频率'], color='blue')

axs[1, 1].set\_title('Average Calorie Consumption by Cluster')

plt.show()







