**Отчет**

Тема: примеры кластеризации с использованием библиотеки pandas.

Источником данных для датафрейма является excel-файл с данной таблицей. Brinnel – шкала твердости , density – плотность(кг/м3).

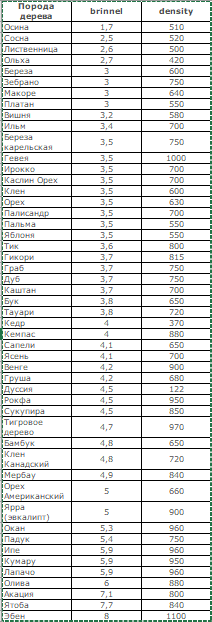


Таблица характеристик древесины.

**Метод кластеризации k-means.**

Код программы:

import pandas as pd #для обработки и анализа данных

import matplotlib.pyplot as plt #для визуализации

from sklearn.cluster import KMeans #машинное обучение

data1 = pd.read\_excel(r'C:\Users\Wotan\Desktop\pandas\excel.xlsx') #путь к файлу

df = pd.DataFrame(data1, columns = ['brinnel', 'density'])  #наименование столбцов

# print (df)

#df = pd.DataFrame(cars,columns=['brinnel','density'])

kmeans = KMeans(n\_clusters=3).fit(df)

centroids = kmeans.cluster\_centers\_

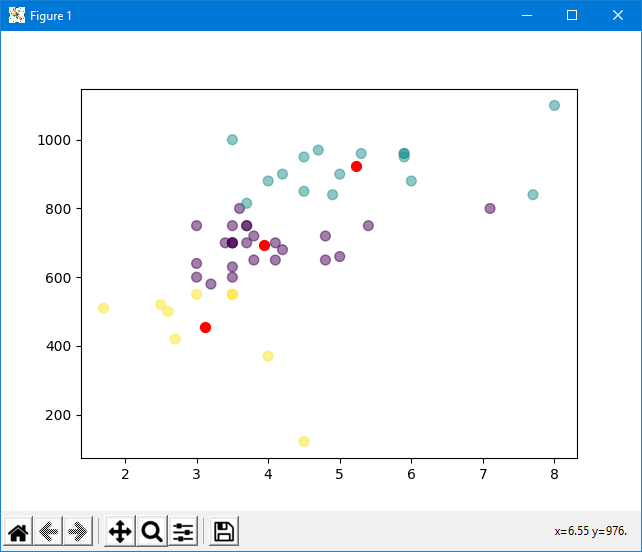
print(centroids)

plt.scatter(df['brinnel'], df['density'], c= kmeans.labels\_.astype(float), s=50, alpha=0.5)

plt.scatter(centroids[:, 0], centroids[:, 1], c='red', s=50)

plt.show()

Где n\_clusters – количество кластеров , сentroids – количество центройдов. Используемые библиотеки: pandas, mathplotlib, sklearn



Результат кластеризации.

**Иерархической кластеризация**

Код программы:

import pandas as pd #для обработки и анализа данных

import matplotlib.pyplot as plt #для визуализации

from sklearn.manifold import TSNE #машинное обучение

from sklearn import datasets

from scipy.cluster.hierarchy import linkage, dendrogram

data1 = pd.read\_excel(r'C:\Users\Wotan\Desktop\pandas\excel.xlsx') #путь к файлу

df = pd.DataFrame(data1, columns = ['brinnel', 'density'])  #наименование столбцов

# Реализация иерархической кластеризации при помощи функции linkage

mergings = linkage(df, method='complete')

# Строим дендрограмму, указав параметры удобные для отображения

dendrogram(mergings,

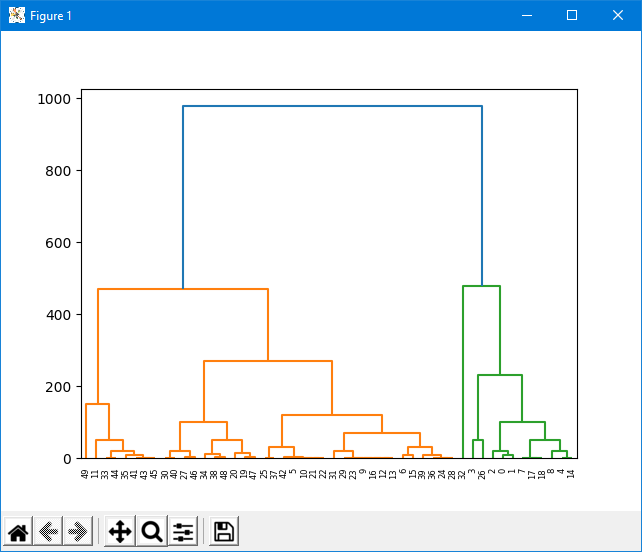
        #    labels=varieties,

           leaf\_rotation=90,

           leaf\_font\_size=6,)

plt.show()

Используемые библиотеки: pandas, mathplotlib, sklearn



Результат кластеризации.

**Пример DBSCAN**

Код программы:

import pandas as pd #для обработки и анализа данных

import matplotlib.pyplot as plt #для визуализации

from sklearn.cluster import DBSCAN #машинное обучение

data1 = pd.read\_excel(r'C:\Users\Wotan\Desktop\pandas\excel.xlsx') #путь к файлу

df = pd.DataFrame(data1, columns = ['brinnel', 'density'])  #наименование столбцов

# print (df)

#df = pd.DataFrame(cars,columns=['brinnel','density'])

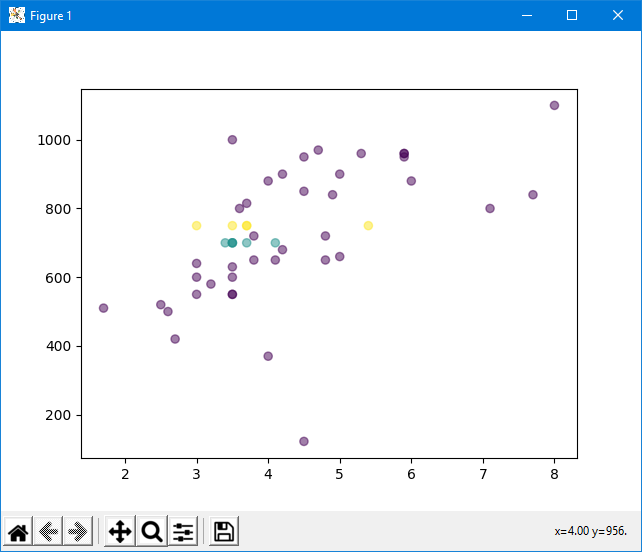
dbscan = DBSCAN(eps=2, min\_samples=5).fit(df)

plt.scatter(df['brinnel'], df['density'], c= dbscan.labels\_.astype(float), s=35, alpha=0.5)

plt.show()

Где eps – максимальное расстояние между двумя образцами, чтобы один считался соседним с другим, min\_samples – количество базовых точек.

Используемые библиотеки: pandas, mathplotlib, sklearn

****

Результат кластеризации