

Информатика. Звёзды и кластеры.

1 Задача

Учёный решил провести кластеризацию некоторого множества звёзд по их расположению на карте звёздного неба. Кластер звёзд – это набор звёзд (точек) на графике, лежащий внутри прямоугольника высотой H и шириной W . Каждая звезда обязательно принадлежит только одному из кластеров.

Истинный центр кластера, или центроид, – это одна из звёзд на графике, сумма расстояний от которой до всех остальных звёзд кластера минимальна. Под расстоянием понимается расстояние Евклида между двумя точками $A(x_1, y_1)$ и $B(x_2, y_2)$ на плоскости, которое вычисляется по формуле:

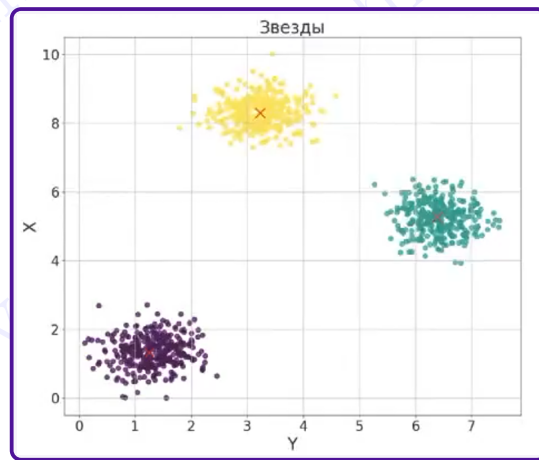
$$d(A, B) = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

В файле А хранятся данные о звёздах **двух** кластеров, где $H = 3$, $W = 3$ для каждого кластера. В каждой строке записана информация о расположении на карте одной звезды: сначала координата x , затем координата y . Значения даны в условных единицах. Известно, что количество звёзд не превышает 1000.

В файле Б хранятся данные о звёздах **трёх** кластеров, где $H = 3$, $W = 3$ для каждого кластера. Известно, что количество звёзд не превышает 10 000. Структура хранения информации о звездах в файле Б аналогична файлу А.

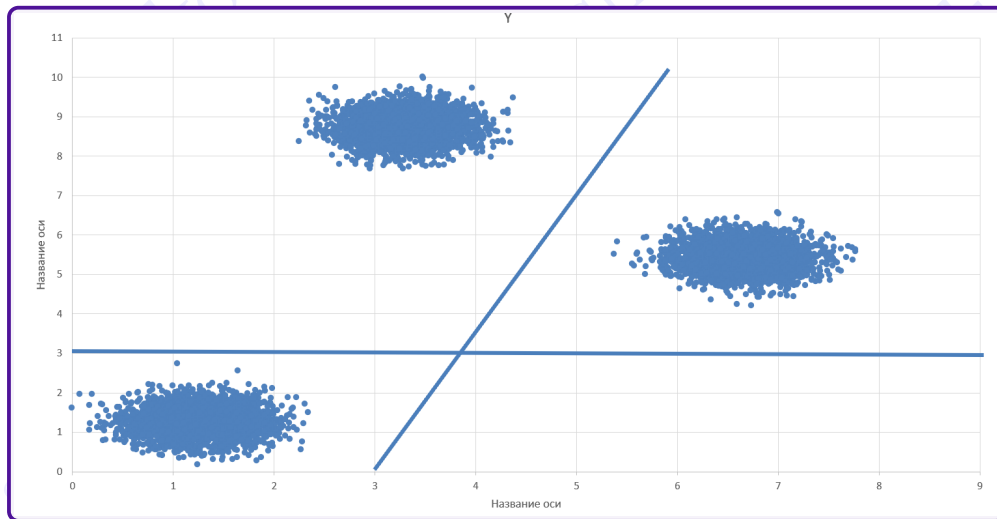
Для каждого файла определите координаты центра каждого кластера, затем вычислите два числа: P_x – среднее арифметическое абсцисс центров кластеров, и P_y – среднее арифметическое ординат центров кластеров. В ответе запишите четыре числа: в первой строке сначала целую часть произведения $P_x \times 10\,000$, затем целую часть произведения $P_y \times 10\,000$ для файла А, во второй строке – аналогичные данные для файла Б. Возможные данные одного из файлов иллюстрированы графиком.

Внимание! График приведён в иллюстративных целях для произвольных значений, не имеющих отношения к заданию. Для выполнения задания используйте данные из прилагаемого файла.



Решение:

Открываем файл Excel. Переходим в раздел *Вставка* → *Иллюстрации* → *Фигуры* и разделяем кластеры прямыми.



Чтобы найти уравнение прямой, воспользуемся формулой:

$$\frac{y - y_1}{y_2 - y_1} = \frac{x - x_1}{x_2 - x_1}$$

Подставим точки (3;0) и (6;11), через которые проходит прямая:

$$\frac{y - 0}{11 - 0} = \frac{x - 3}{6 - 3}$$

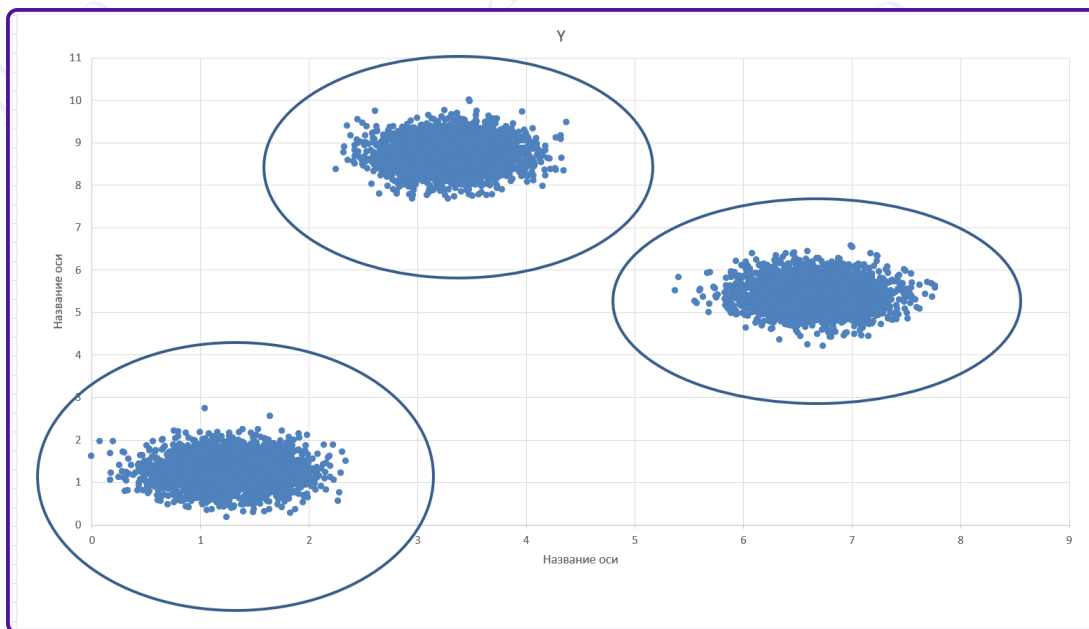
$$\frac{y}{11} = \frac{x - 3}{3}$$

```

f = open('demo_2025_27_B.txt')
s = f.readline()
a = [list(map(float, i.replace(',', '.').split())) for i in f]
c = [[], [], []] # 3 кластера
for i in a:
    # Если звезда ниже прямой 'y = 3'
    if i[1] <= 3:
        c[0].append(i) # Добавляем ее в первый кластер
    # Иначе если звезда выше прямой 'y/11=(x-3)/3'
    elif (i[1] / 11) > ((i[0]-3) / 3):
        c[1].append(i) # Добавляем ее во второй кластер
    # Если звезда ниже прямых 'y = 3' и 'y/11=(x-3)/3'
    else:
        c[2].append(i) # Добавляем ее в третий кластер
# Поиск центроидов
sx = sy = 0 # Искомые средн. арифметические
for k in range(3): # Для каждого кластера
    mn = 1000000500000000000 # Минимальное расстояние
    for j in range(len(c[k])): # Для каждой звезды кластера
        star = c[k][j] # Звезда, от которой ищем расстояние
        l = 0 # Длина расстояния
        for i in range(len(c[k])):
            x1, y1 = star # Координаты звезды, от которой ищем расстояние
            x2, y2 = c[k][i] # Координаты звезды, до которой ищем расстояние
            l += ((x2-x1)**2 + (y2-y1)**2)**0.5
        if l < mn: # Если расстояние меньше минимума
            mn = l # Обновляем mn
            x, y = x1, y1 # Запоминаем координаты предполагаемого центроида
sx += x # Добавляем координату 'x' найденного центроида
sy += y # Добавляем координату 'y' найденного центроида
print(int((sx/3)*10000), int((sy/3)*10000))

```

Представим, что мы не можем разделить кластеры с помощью прямых. Тогда разделим на кластеры с помощью уравнения окружности.



Изменим первую часть решения. Поиск центроидов остается прежним.

```
f = open('demo_2025_27_B.txt')
s = f.readline()
a = [list(map(float, i.replace(',', ' ').split())) for i in f]
c = [[], [], []] # 3 кластера
for i in a:
    # Если звезда внутри окр-ти с центром (1;2) и радиусом 2
    if (i[0]-1)**2 + (i[1]-2)**2 <= 4:
        c[0].append(i) # Добавляем ее в первый кластер
    # Если звезда внутри окр-ти с центром (3;9) и радиусом 2
    elif (i[0]-3)**2 + (i[1]-9)**2 <= 4:
        c[1].append(i) # Добавляем ее во второй кластер
    # Если звезда внутри окр-ти с центром (7;5) и радиусом 2
    elif (i[0]-7)**2 + (i[1]-5)**2 <= 4:
        c[2].append(i) # Добавляем ее в третий кластер
```

Ответ: 37522 51277