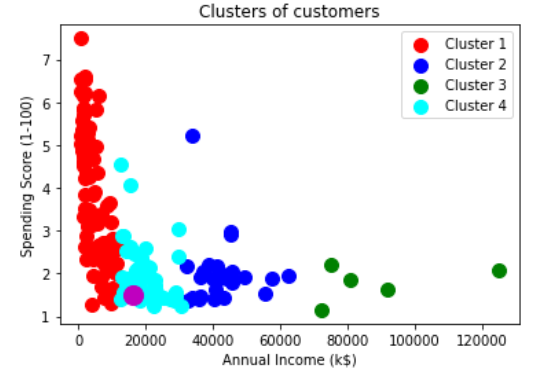
**Лабораторная работа № 5**

**Тема**: Алгоритмы кластеризации.

1. Для работы используйте данные подходящие для кластеризации на свой выбор (если ничего не найдете, то можно использовать *Country-data.csv*)
2. Из датасета выберите 2 параметра, характеризующих вашу цель исследования.
3. Разделите данные на кластеры методом K-means, обоснуйте выбор количества кластеров с помощью метода локтя.
4. Разделите данные на кластеры методом иерархической кластеризации, выберите с помощью дендрограммы оптимальное количество кластеров.
5. Визуализируйте результаты кластеризации двух методов и сравните их.
6. Оцените качество кластеризации методами K-means и иерархической кластеризации, рассчитав пару метрик качества кластеризации (модуль **sklearn.metrics**). Например, силуэт для выборки **silhouette\_score()** и др.
7. Из датасета выберите любой конкретный объект (если вы делаете модель на датасете *Country-data.csv*, то выберите любую страну) и визуализируйте этот объект в виде точки отличного цвета и размера на графике кластеров (*пример на рисунке, точка пурпурного цвета*).

**Вопросы:**

1. Что решают задачи кластеризации в машинном обучении?
2. Расскажите принцип работы метода K-means.
3. Как можно выбрать оптимальное количество кластеров в K-means?
4. Расскажите принцип работы метода иерархической кластеризации.
5. Для чего можно использовать дендрограмму в методе иерархической кластеризации?
6. Какие метрики используют для оценки качества кластеризации?