

Практическая работа №10

Hierarchical Clustering

Цель работы: изучение и применение методов иерархической кластеризации для анализа данных, а также создание иерархии кластеров с использованием агломеративного метода.

Ход работы

В данной практической нам необходимо классифицировать данные, объединяя или разделяя их в дендограмму с использованием агломеративного метода:

Иерархическая кластеризация обычно визуализируется в виде дендрограммы, как показано в следующей ячейке. Каждое слияние представлено горизонтальной линией. Координата у горизонтальной линии отражает сходство двух кластеров, которые были объединены, где города рассматриваются как одноэлементные кластеры. Двигаясь вверх от нижнего слоя к верхнему узлу, дендрограмма позволяет нам восстановить историю слияний, которые привели к изображенной кластеризации.

Изменим привязку с полной на метод Уорда (минимизация увеличения дисперсии):

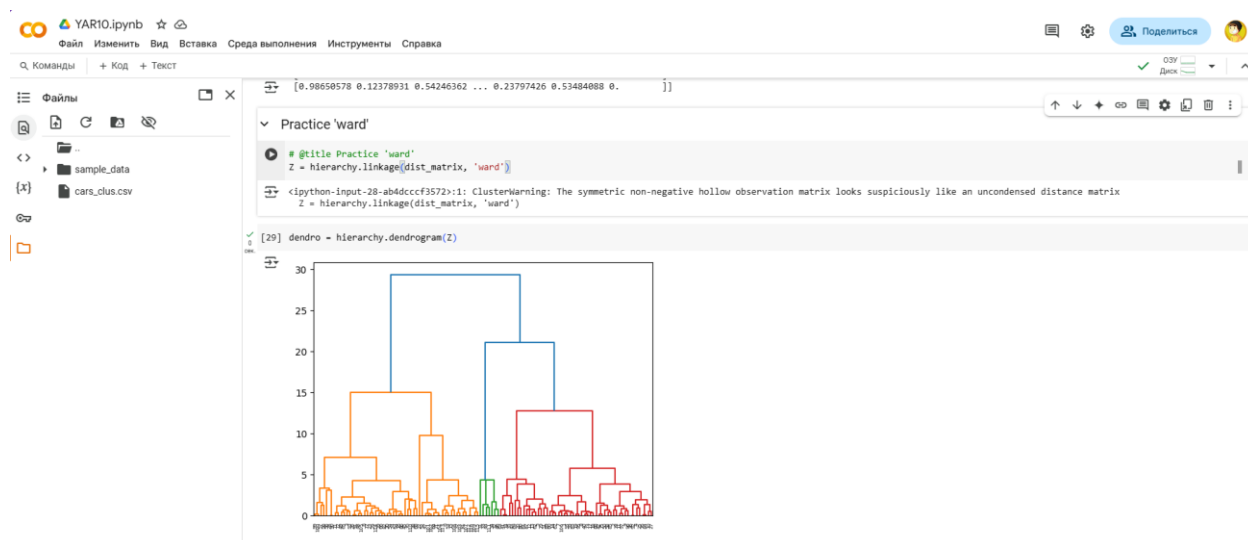


Рисунок 1 – Метод Уорда

Используем функцию "**AgglomerativeClustering**" из библиотеки `scikit-learn` для кластеризации набора данных. `AgglomerativeClustering` выполняет иерархическую кластеризацию, используя подход "снизу вверх".

Критерии связывания определяют показатель, используемый для стратегии объединения Ward минимизирует сумму квадратов различий во всех кластерах. Это подход, направленный на минимизацию дисперсии, и в этом смысле он аналогичен целевой функции k-средних, но применяется с использованием агломеративного иерархического подхода. Максимальная или полная привязка минимизирует максимальное расстояние между наблюдениями пар кластеров. Средняя привязка минимизирует среднее значение расстояний между всеми наблюдениями пар кластеров.

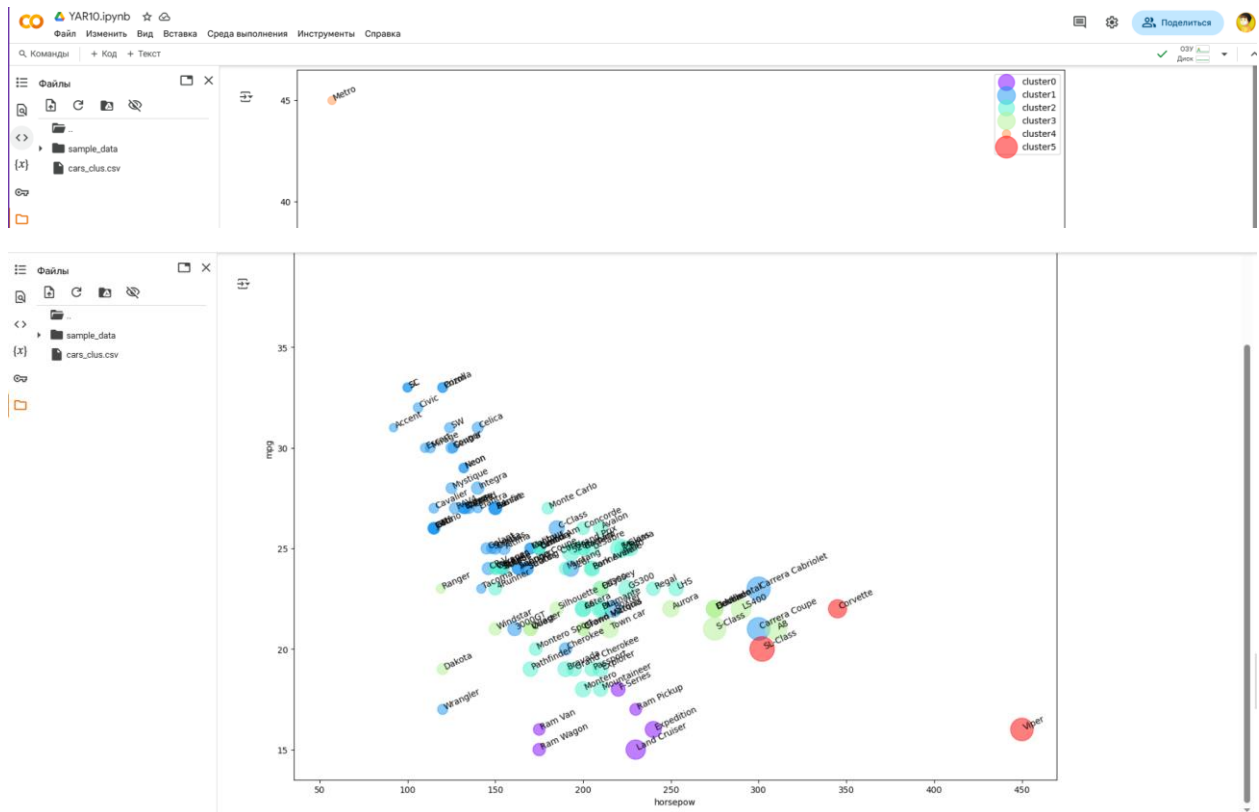


Рисунок 1.1, 1.2 – точечная дендограмма распределения каждого кластера

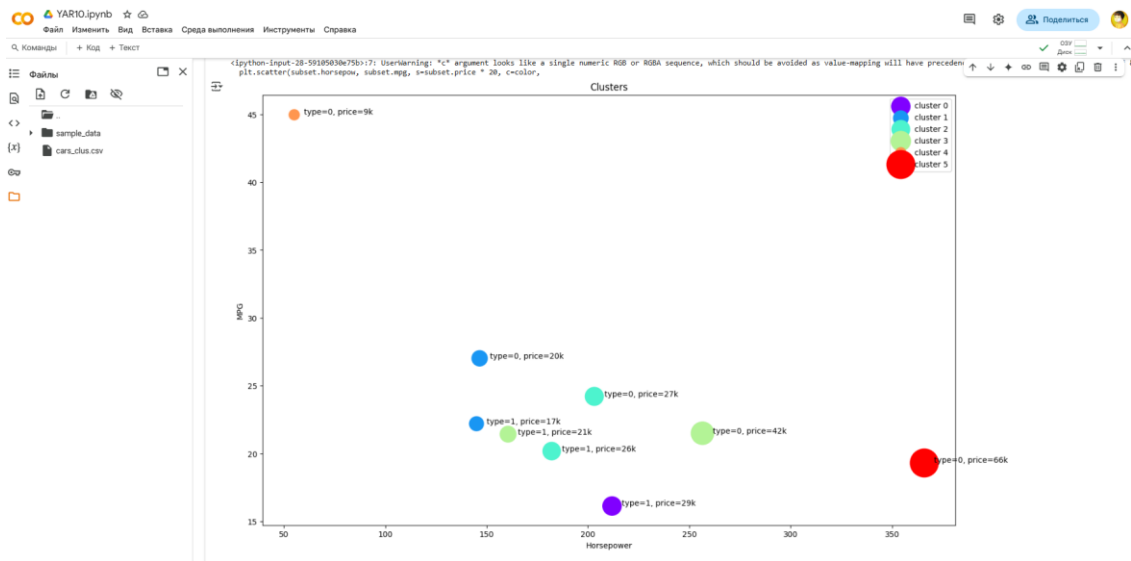


Рисунок 1.3 – визуализация кластеров автомобилей

Вывод: в результате практической работы мы научились применять методы иерархической кластеризации для анализа данных, и в итоге построили наглядную визуализацию кластеров автомобилей по цене и расходу топлива.