

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 двух методов иерархической кластеризации:

- 1 - это агломеративный, что подразумевает объединение кластеров;
- 2 - это дивизивный, что подразумевает разбиение больших кластеров.

Выполняем работу, смотря методические указания. В одном из заданий сказано посмотреть разницу между методами отображения дендрограммы, и вот в чем их разница :

Метод средней связи (average linkage). На каждом следующем шаге объединяет два ближайших кластера, рассчитывая среднюю арифметическую дистанцию между всеми парами объектов.

Метод полной связи (complete linkage). Расстояние между кластерами равно максимальному расстоянию между точками из разных кластеров.

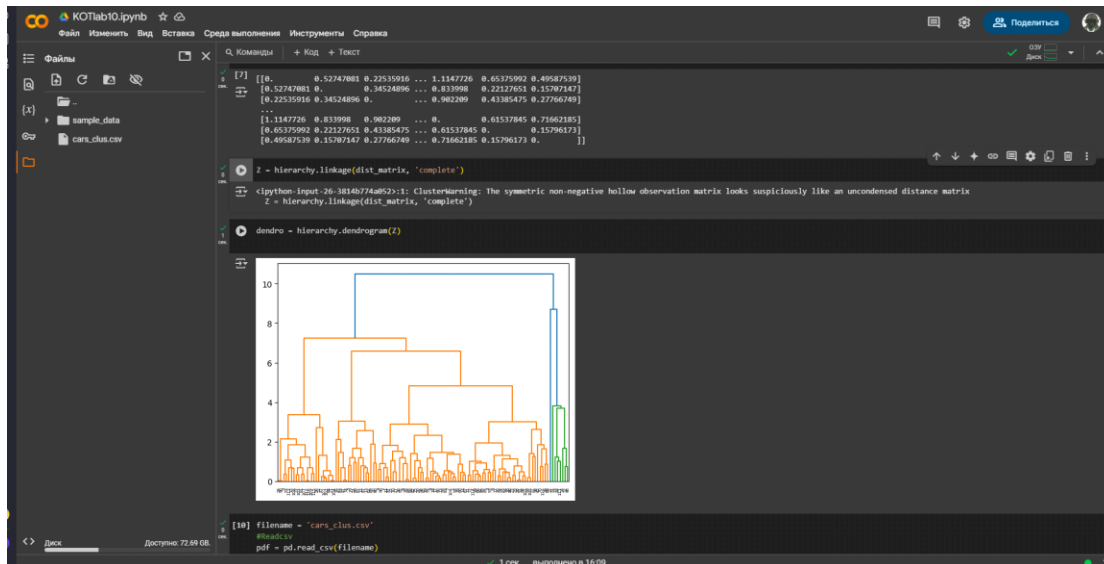
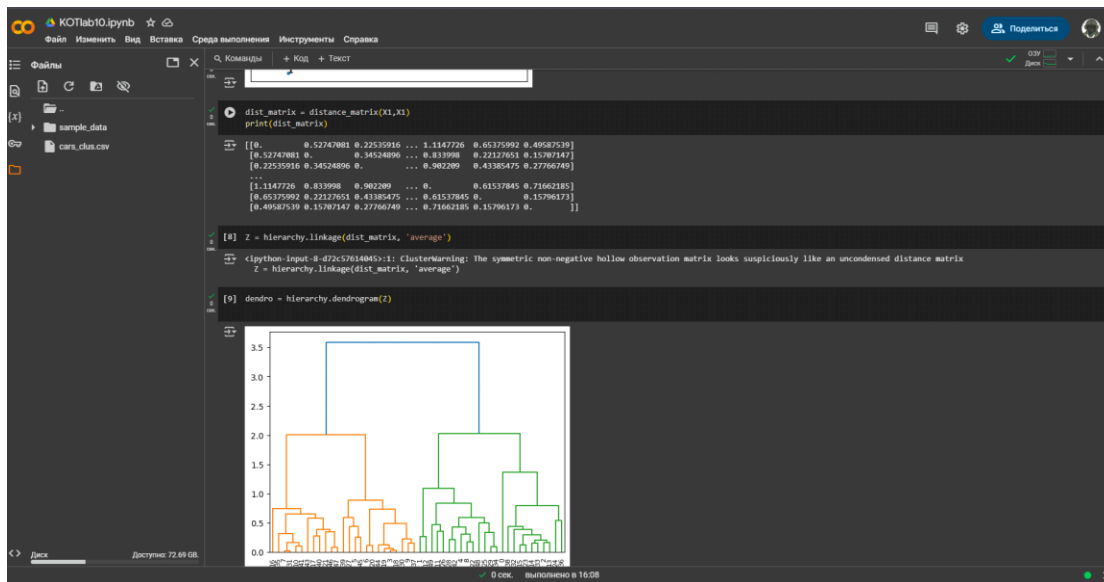


Рисунок 1, 2 - визуализация дендрограммы average и complete

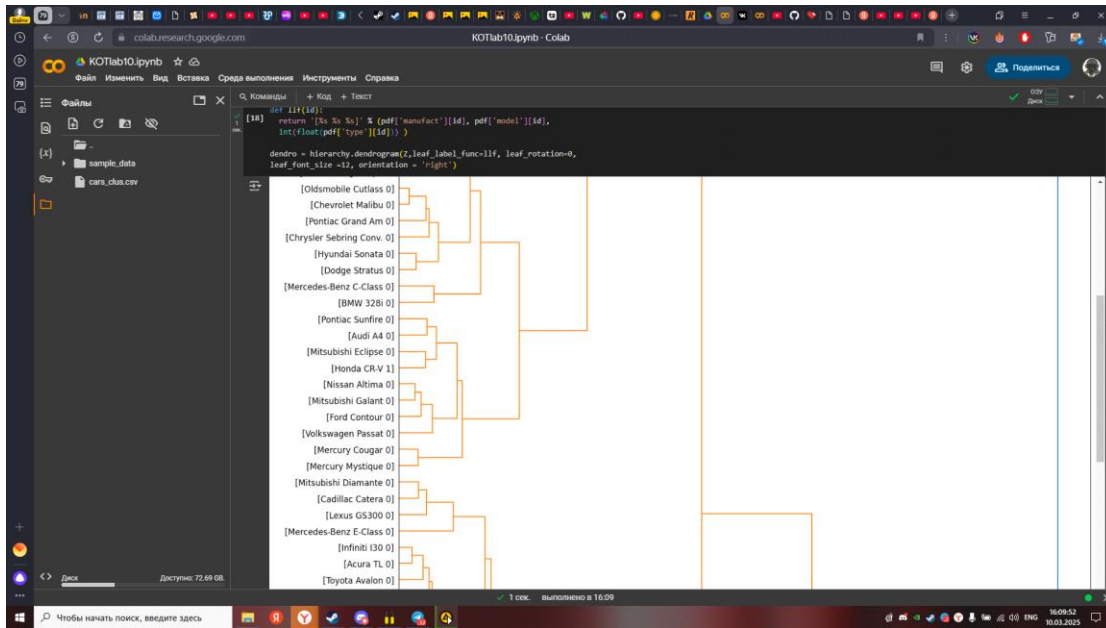


Рисунок 3 - большая, страшная дендрограмма

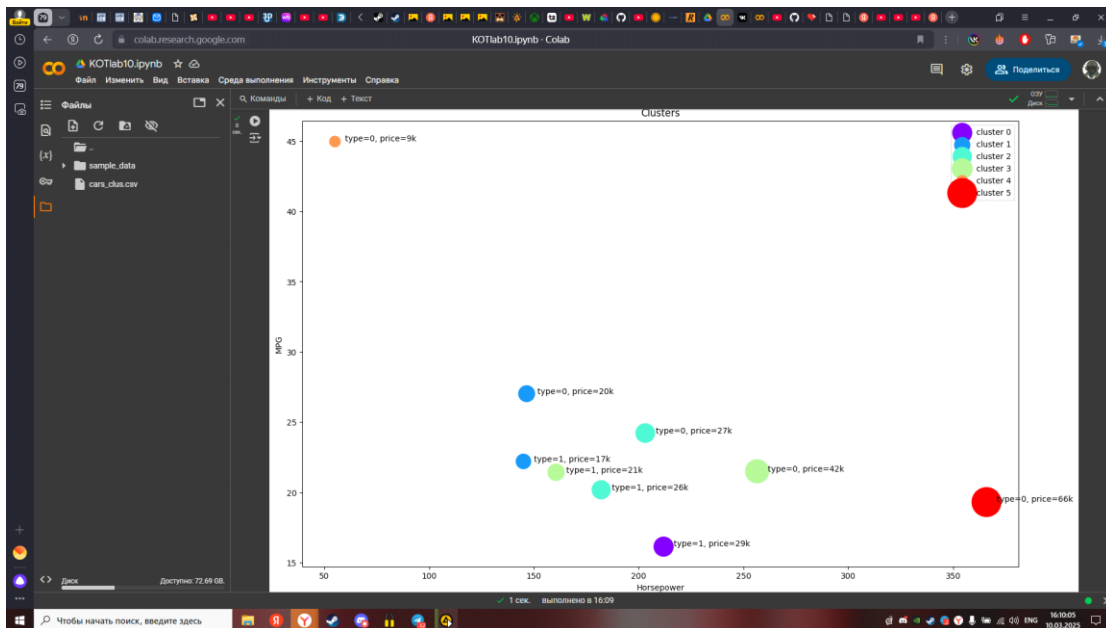


Рисунок 4 - финальная визуализация кластеров машин

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