

Отчет по лабораторной работе №1
ФКН группа 2.1
Хоменко Полина
Вариант 20

Анализируем данные филиалов сети магазинов. Руководство сети желает разделить филиалы на группы, чтобы затем сравнить продажи. Данный анализ может помочь в выборе направления развития и оптимизации сети.

Часть переменных представлена в текстовом виде:

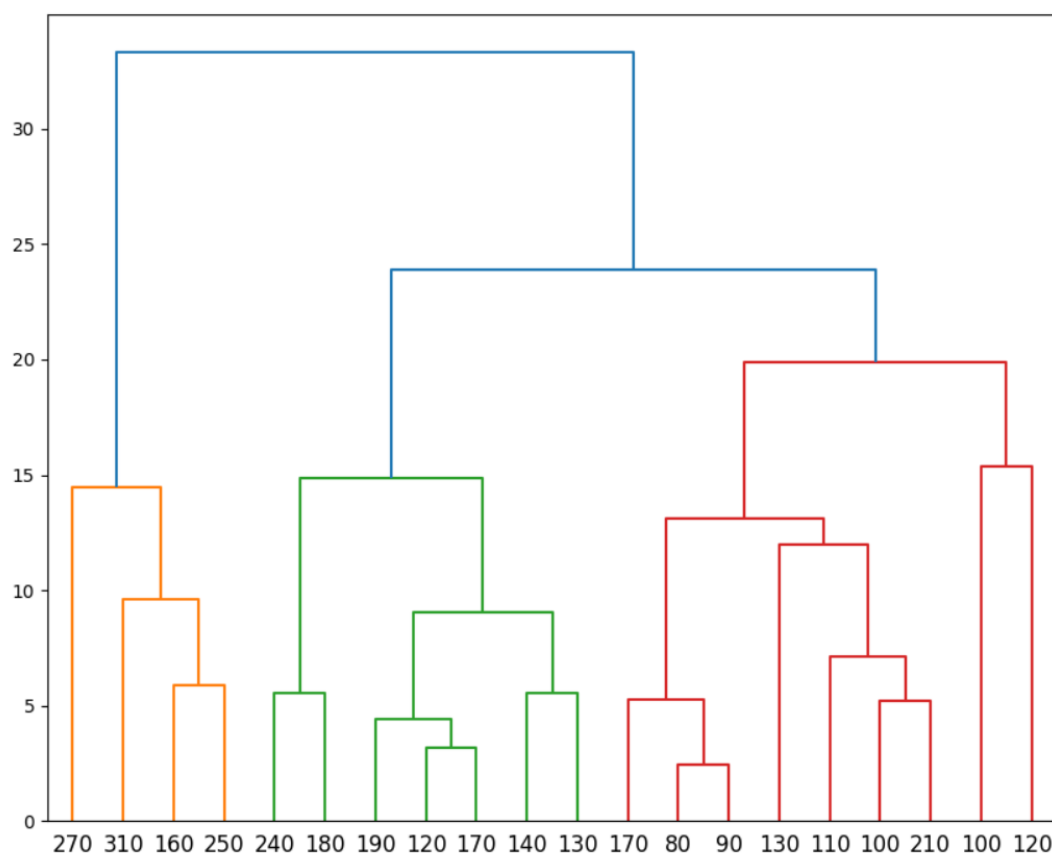
- * **название** Условный код филиала
- * **площадь** Площадь торгового зала, кв.м.
- * **проходим** Проходимость, ср. кол-во входящих в магазин за 10 минут (с 14-00 до 18-00)
- * **ассортим** Ассортимент (текстовая переменная!)
- * **конкурент** Как выглядит филиал по совокупности показателей по отношению к ближайшему конкуренту (текстовая переменная!)
- * **метро** За сколько минут можно дойти до ближайшей станции метро
- * **консульт** Наличие в торговом зале консультантов (текстовая переменная!)
- * **дизайн** Наличие или отсутствие следующих компонентов: вывеска, витрина, световая вывеска (текстовая переменная!)
- * **цены** Индекс цен по отношению к базовым.
- * **продажи** Совокупные продажи за 2 последних месяца
- * **ассорт2** Ассортимент (числовая переменная после перекодировки!)
- * **конкур2** Как выглядит филиал по совокупности показателей по отношению к ближайшему конкуренту (числовая переменная после перекодировки!)
- * **консул2** Наличие в торговом зале консультантов (числовая переменная после перекодировки!)
- * **дизайн2** Наличие или отсутствие следующих компонентов: вывеска, витрина, световая вывеска (числовая переменная после перекодировки!)

Удалим информацию о названии и введем переобозначение текстовых переменных, необходимых для кластеризации – в данном случае кластеризация будет вестись без отбора, так как каждый признак несёт важную роль в определении качества работы филиала сети магазинов. Так как необходимо сравнить полученные группы по уровню продаж так же уберем переменную продажи.

```
assortmentDict = {'миним': 0, 'средний': 1, 'широкий': 2, 'макс': 3}
df['assortment'] = df['assortment'].map(assortmentDict)
competitorDict = {'хуже': -1, 'одинак': 0, 'лучше': 1}
df['competitor'] = df['competitor'].map(competitorDict)
consultantDict = {'Нет': 0, 'Есть': 1}
df['consultant'] = df['consultant'].map(consultantDict)
designDict = {'Вывеска': 0, 'Витрина': 1, 'Св+Ви': 2, 'Световая': 3, 'Бедно': 4, 'Вы+Ви': 5}
df['design'] = df['design'].map(designDict)
```

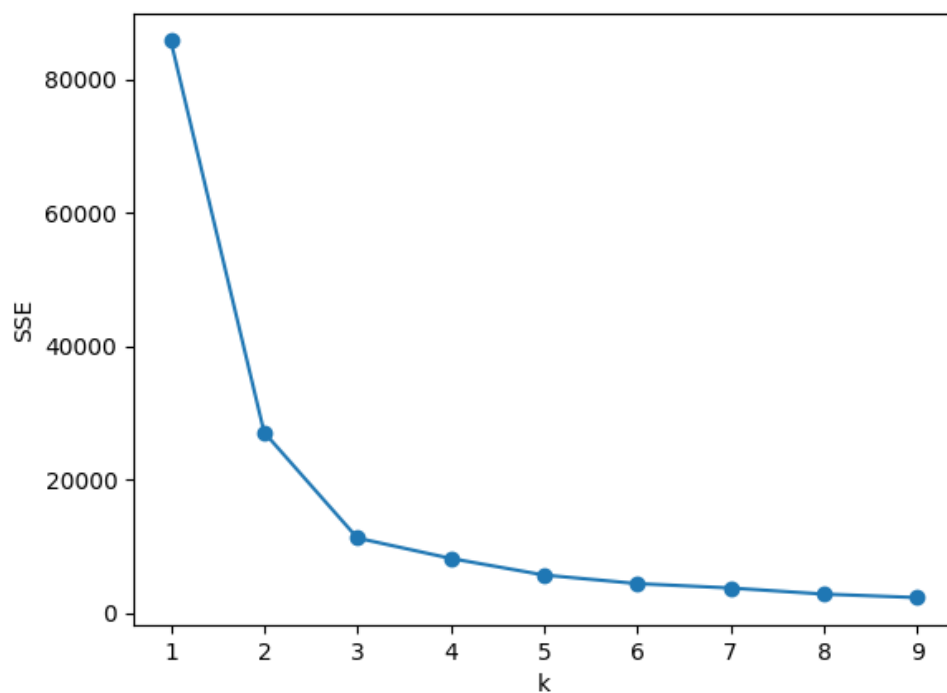
Для иерархической кластеризации была выбрана библиотека Scipy.

Приведем график дендрограммы:

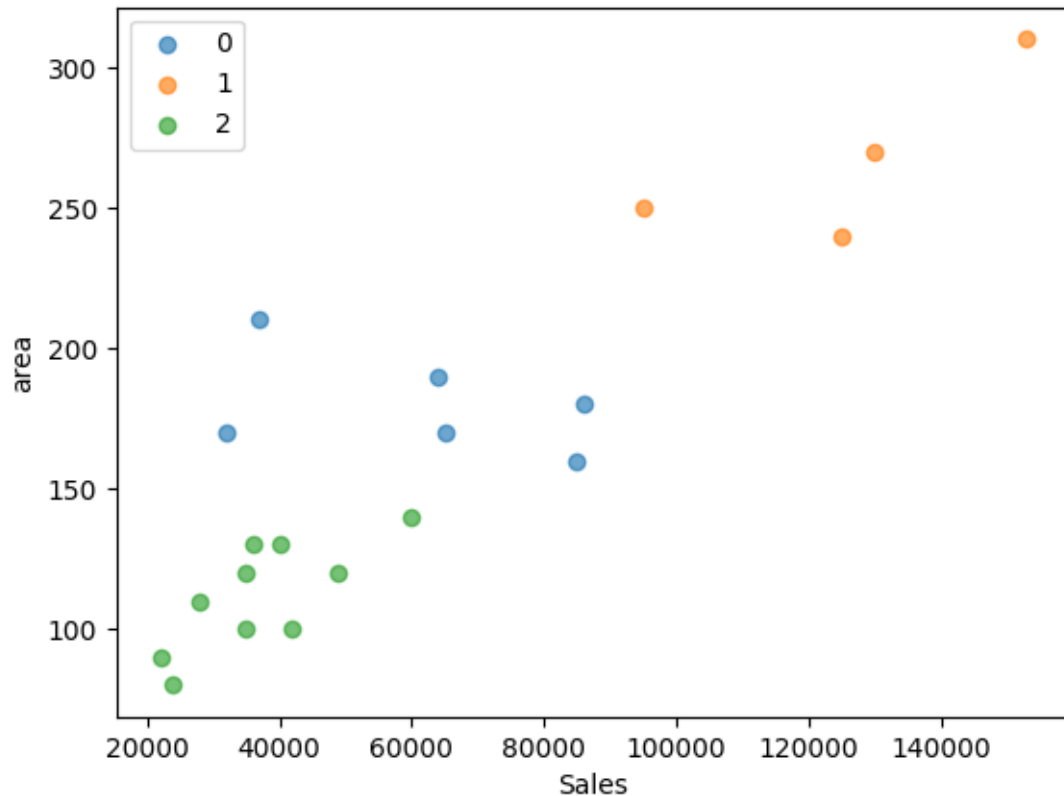


Метод k средних - использовалась библиотека sklearn:

Для определения оптимального числа кластеров воспользуемся методом Каменистая осыпь(Локоть).



Из этого можно сделать вывод, что кластеров необходимо 3, так как, один из самых наибольших изгибов. Выбираем 3, как оптимальное количество кластеров:



На графике выделяются 3 основных кластера – можно сделать вывод, что площадь является наиболее значимым параметром для повышения продаж.

Вывод:

Делая вывод основываясь на результатах кластерного анализа, для увеличения продаж - сеть должна развивать магазины большой площади, а в последующем филиалы с максимальным ассортиментом, с высокой проходимостью и тд.