

# Ценовая сегментация брендов

Большое домашнее задание по курсу «Введение в анализ данных» № 1

27 апреля 2019 г.

## Содержание

1. Общая постановка задания	1
2. Продуктовые категории	2
3. Сбор данных	2
4. Методология сегментации	2
5. Присваивание сегмента брендам	3
6. Инструменты и темы курса	3
7. Организационные вопросы	3

## 1. Общая постановка задания

Полка воды в супермаркетах довольно часто вызывает удивление. Бутылка 0.5 л негазированной воды может стоить от 15 рублей до 250 рублей. Понятно, что большая часть цены имеет маленькое отношение к составу и качеству воды, скорее, цена связана с другими аспектами, например, позиционированием бренда или логистическими издержками.

В этом задании вам предстоит присвоить брендам одной продуктовой категории один из 4-х ценовых сегментов:

- Low priced
- Middle priced
- Upper middle priced
- High priced

Для того чтобы выполнить это задание, вам нужно:

1. разбиться в команды по 2-3 человека;
2. выбрать продуктовую категорию;
3. собрать данные;
4. продумать методологию сегментации;
5. присвоить брендам ценовой сегмент.

На выходе должна получиться таблица:

Бренд	Ценовой сегмент

Выбор продуктовой категории за вами. Подбор брендов и данных - одна из ваших задач, т.к. сбор данных - один из ключевых навыков аналитиков. Детальнее обо всем в следующих секциях.

## 2. Продуктовые категории

Под продуктовой категорией понимаем тип продуктов<sup>1</sup> (по иерархии из товарного маркетинга): чай, кофе, йогурт, хлеб и т.п.

**Требования:** выбрать категорию, в которой не менее 5 брендов и не менее 15 товаров.

## 3. Сбор данных

Данные собираются самостоятельно и вручную. Команды должны посетить 3 магазина сети X5 Retail Group (Перекресток, Пятерочка, Карусель) и собрать информацию о товарах выбранной категории: наименование, бренд, цена, вес.

**Совет:** пойти в один Перекресток и в две Пятерочки.

Результат сбора данных нужно представить в виде таблицы (на каждой странице — данные из каждого магазина), где на первой строке в левой ячейке указывается наименование супермаркета, а в соседней — адрес. На второй строке — название колонок, и начиная с третьей — сами данные.

Перекрёсток	г. Москва, ул. Дубравная, д. 34/29		
Наименование	Бренд	Цена	Вес
GREENF.Чай ГОЛДЕН ЦЕЙЛОН чер.лист.200г	GREENFIELD	139.90	0.200

## 4. Методология сегментации

Это творческий этап - именно вы должны определить ценовые границы каждого из сегментов.

---

<sup>1</sup><https://studfiles.net/preview/3560003/>

**Совет:** переведите все цены на фиксированную единицу веса. Например, выше цена 200-граммового чая — 139.90 рублей, тогда цена 100 г этого чая: 69.95.

Если у вас есть две позиции одного бренда, то найдите среднюю цену на бренд. Красиво получится операция группировкой с суммированием, а затем делением одного столбца на другой (БЕЗ ЦИКЛОВ!)

Как можно выстроить методологию:

1. Можно ориентироваться на процентилях. Выберите 3 уровня процентилей, так что  $k_1 < k_2 < k_3$ , тогда low-priced бренды — те, цена которых от 0 рублей  $k_1$ .
2. Можно посчитать среднюю цену продуктов по категории и посчитать отношение цены бренда к средней цене (будем называть отношение — индексом), и дробить продукты на 4 группы по такому индексу.
3. Любые другие варианты, которые придумаете.

Эту часть надо подробно расписать внутри питоновского ноутбука. Распишите максимально подробно, чтобы мы всё поняли. Прямо как для тупых пишите.

## 5. Присваивание сегмента брендам

Когда вы переведете цены по всей категории продуктов в цены на конкретную единицу товара, остается пройти по колонке цены и в соответствии с ней присвоить бренду один из 4 сегментов.

Здесь вам пригодится функция `numpy.select`: <https://docs.scipy.org/doc/numpy/reference/generated/numpy.select.html>

## 6. Инструменты и темы курса

Содержательное нутро связано с описательными статистиками. Работаем в Jupyter Notebook.

## 7. Организационные вопросы

### Команды

Задание выполняется в командах по 2-3 человека. Индивидуальные работы не будут засчитаны.

### Дедлайн

- Задание выдаем 27.04.2019
- Задание сдаем 12.05.2019 в 23.30 на [anytask.org](https://anytask.org)

## В каком виде сдаём?

1. Ждем вашу собранную эксельку с данными
2. Ждём ірупb, в котором:
  - (a) Указаны все члены команды и номер группы.
  - (b) Прописан весь код, который читает эксельку с собранными данными и выдает эксельку с таблицей Бренд — Ценовой сегмент.
  - (c) Помимо того, что вы пишете код, вы еще в текстовых ячейках оставляете комментарии - что делали, зачем, что интересно увидели. <https://www.youtube.com/watch?v=LLF0Z2vN7ss>

Постарайтесь делать осмысленные выводы! Не будьте мудрыми, как король:



- (d) Не забывайте о красивых моментах, выводите info(), head() по входному и выходному датафрейму. Красивые закономерности, найденные в данных, изображайте на картинках.

## Оценивание

Работа оценивается по 10-бальной шкале. Она ставится на команду и нормируется на число человек в ней. Так, если в команде три человека и она получает оценку 8, она превращается в 24 балла. Далее участники команды делят эти баллы между собой в любой пропорции в зависимости от вклада каждого участника в итоговую работу. Если в команде два участника, то 8 соответственно превращается в 16. В работе оцениваются адекватность собранных данных, чистота кода, адекватность и продуманность методологии разбиения на сегменты.