

## Задание 1

Самостоятельно ознакомьтесь с библиотекой Pandas языка Python. Желающие могут воспользоваться альтернативными средствами языка R.

В ходе знакомства выполните предварительную подготовку данных по продажам квартир для анализа. Познакомьтесь с объектами Series и Dataframe и способами индексирования данных в них. Для преобразования данных скорее всего, потребуется использовать такие методы Pandas как apply, to\_numeric, replace, sort, drop\_duplicates, value\_counts.

1. Прочитайте данные по продающимся квартирам из файла .csv (используя соответствующую функцию Pandas, не руками).
2. Проверьте, что данные по цене имеют числовой тип. Если нет, сделайте его числовым.
3. В данных имеются повторения – одна и та же квартира может помещаться несколько раз чтобы показать, что она находится рядом с несколькими станциями метро. Нужно удалить дубли. Дублями считаются те записи, у которых совпадают значения полей «Улица», «Номер дома», «Дом», «Площадь». Поле цена может отличаться, в этом случае при удалении дублей нужно оставить запись с минимальной ценой.
4. Теперь займёмся полем «Дом». В файле записи вида «1/5 Б» в этом столбце означает, что квартира находится на 1 этаже 5 этажного блочного дома (Б – блочный). На основе этих данных надо сделать несколько новых полей:
  - a. количество этажей в доме;
  - b. этаж, на котором находится квартира;
  - c. первый этаж (булевское поле);
  - d. крайний (первый или последний) этаж (булевское поле);
  - e. отношение этажа квартиры к количеству этажей дома;
  - f. тип дома.

Учитывайте, что часть информации может быть не известна. Для обозначения неизвестных значений используйте nan (это распространённый способ обозначения неизвестных значений в Pandas).

5. Посмотрим на частотную информацию по типам домов:
  - a. распечатайте число записей каждого типа в штуках и процентах;
  - b. два типа дома представлены только одной записью – эти типы надо пометить как пропущенные значения, заменив на nan;
  - c. для оставшихся типов сформируйте отдельные булевские поля, выполнив при этом два объединения:
    - i. объедините типы Б (блочный) и П (панельный) в один тип;
    - ii. объедините типы М (монолит) и М-К (монолит кирпич) в один тип.