# Исследование объявлений о продаже квартир

Поздравляем! Вы прошли курс в тренажёре. Самое время проверить полученные знания на практике и решить аналитический кейс. Выполнять работу вы будете самостоятельно.

Как закончите работу над проектом, отправьте его на проверку ревьюеру. В течение суток вы получите комментарии. Их нужно учесть: доработать проект и вернуть ревьюеру обновлённый вариант.

Скорее всего, вы снова получите комментарии по кейсу. Это нормально — доработка может проходить в несколько этапов.

Проект завершён, когда засчитаны все исправления.

## **Описание проекта**

В вашем распоряжении данные сервиса Яндекс.Недвижимость — архив объявлений о продаже квартир в Санкт-Петербурге и соседних населённых пунктах за несколько лет. Нужно научиться определять рыночную стоимость объектов недвижимости. Ваша задача — установить параметры. Это позволит построить автоматизированную систему: она отследит аномалии и мошенническую деятельность.

По каждой квартире на продажу доступны два вида данных. Первые вписаны пользователем, вторые — получены автоматически на основе картографических данных. Например, расстояние до центра, аэропорта, ближайшего парка и водоёма.

### Инструкция по выполнению проекта

**Шаг 1**. **Откройте файл с данными и изучите общую информацию**

Путь к файлу: */datasets/real\_estate\_data.csv*. [Скачать датасет](https://code.s3.yandex.net/datasets/real_estate_data.csv)

**Шаг 2. Предобработка данных**

определите и изучите пропущенные значения:

для некоторых пропущенных значений можно предположить логичную замену. Например, если человек не указал число балконов — скорее всего, их нет. Такие пропуски правильно заменить на 0. Для других типов данных нет подходящего значения на замену. В этом случае правильно оставить эти значения пустыми. Отсутствие значения — тоже важный сигнал, который не нужно прятать;

заполните пропуски, где это уместно. Опишите, почему вы решили заполнить пропуски именно в этих столбцах и как выбрали значения;

укажите причины, которые могли привести к пропускам в данных.

приведите данные к нужным типам:

поясните, в каких столбцах нужно изменить тип данных и почему.

**Шаг 3. Посчитайте и добавьте в таблицу:**

цену квадратного метра;

день недели, месяц и год публикации объявления;

этаж квартиры; варианты — первый, последний, другой;

соотношение жилой и общей площади, а также отношение площади кухни к общей.

**Шаг 4. Проведите исследовательский анализ данных и выполните инструкции:**

Изучите следующие параметры: площадь, цена, число комнат, высота потолков. Постройте гистограммы для каждого параметра.

Изучите время продажи квартиры. Постройте гистограмму. Посчитайте среднее и медиану. Опишите, сколько обычно занимает продажа. Когда можно считать, что продажи прошли очень быстро, а когда необычно долго?

Уберите редкие и выбивающиеся значения. Опишите, какие особенности обнаружили.

Какие факторы больше всего влияют на стоимость квартиры? Изучите, зависит ли цена от площади, числа комнат, удалённости от центра. Изучите зависимость цены от того, на каком этаже расположена квартира: первом, последнем или другом. Также изучите зависимость от даты размещения: дня недели, месяца и года.

Выберите 10 населённых пунктов с наибольшим числом объявлений. Посчитайте среднюю цену квадратного метра в этих населённых пунктах. Выделите среди них населённые пункты с самой высокой и низкой стоимостью жилья. Эти данные можно найти по имени в столбце '*locality\_name'*.

Изучите предложения квартир: для каждой квартиры есть информация о расстоянии до центра. Выделите квартиры в Санкт-Петербурге (*'locality\_name'*). Ваша задача — выяснить, какая область входит в центр. Создайте столбец с расстоянием до центра в километрах: округлите до целых значений. После этого посчитайте среднюю цену для каждого километра. Постройте график: он должен показывать, как цена зависит от удалённости от центра. Определите границу, где график сильно меняется — это и будет центральная зона.

Выделите сегмент квартир в центре. Проанализируйте эту территорию и изучите следующие параметры: площадь, цена, число комнат, высота потолков. Также выделите факторы, которые влияют на стоимость квартиры (число комнат, этаж, удалённость от центра, дата размещения объявления). Сделайте выводы. Отличаются ли они от общих выводов по всему городу?

**Шаг 5. Напишите общий вывод**

**Оформление:** Выполните задание в *Jupyter Notebook*. Заполните программный код в ячейках типа *code,* текстовые пояснения — в ячейках типа *markdown*. Примените форматирование и заголовки.

### Описание данных

*airports\_nearest* — расстояние до ближайшего аэропорта в метрах (м)

*balcony* — число балконов

*ceiling\_height* — высота потолков (м)

*cityCenters\_nearest* — расстояние до центра города (м)

*days\_exposition* — сколько дней было размещено объявление (от публикации до снятия)

*first\_day\_exposition* — дата публикации

*floor* — этаж

*floors\_total* — всего этажей в доме

*is\_apartment* — апартаменты (булев тип)

*kitchen\_area* — площадь кухни в квадратных метрах (м²)

*last\_price* — цена на момент снятия с публикации

*living\_area* — жилая площадь в квадратных метрах(м²)

*locality\_name* — название населённого пункта

*open\_plan* — свободная планировка (булев тип)

*parks\_around3000* — число парков в радиусе 3 км

*parks\_nearest* — расстояние до ближайшего парка (м)

*ponds\_around3000* — число водоёмов в радиусе 3 км

*ponds\_nearest* — расстояние до ближайшего водоёма (м)

*rooms* — число комнат

*studio* — квартира-студия (булев тип)

*total\_area* — площадь квартиры в квадратных метрах (м²)

*total\_images* — число фотографий квартиры в объявлении

*Пояснение*: апартаменты — это нежилые помещения, не относящиеся к жилому фонду, но имеющие необходимые условия для проживания.

## **Как будут проверять мой проект?**

Мы подготовили критерии оценки проекта. Прежде чем приступить к решению кейса, внимательно их изучите.

На что обращают внимание ревьюеры, проверяя проект:

Как вы описываете выявленные в данных проблемы?

Какие способы обработки пропусков применяете?

Как используете срезы данных?

Решают ли ваши графики определенный вид задач?

Какие методы построения графиков применяете?

Выводите финальные данные в сводных таблицах?

Автоматизируете построение графиков?

Считаете ли показатели взаимосвязи в данных и правильно их объясняете?

Соблюдаете структуру проекта и поддерживаете аккуратность кода?

Какие выводы делаете?

Оставляете комментарии к шагам?

Всё, что нужно для выполнения этого проекта, есть в шпаргалках и конспектах прошлых уроков.

Успехов!

[Вперёд](https://praktikum.yandex.ru/learn/data-scientist/courses/bcdd857f-43ed-4877-8321-44433134943f/sprints/2956/topics/46014da3-7d6a-4faf-a647-0bde32bf3fd0/lessons/20b2411c-8baf-4388-a9dd-9cf9aada0a49/)

Задание