

# Описание проекта

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 наставнику. В течение суток он пришлёт комментарии. Их нужно учесть; доработать проект и вернуть наставнику обновлённый вариант.

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

Проект завершён, когда наставник засчитает все доработки.

## Описание проекта

В вашем распоряжении данные сервиса Яндекс.Недвижимость — архив объявлений о продаже квартир в Санкт-Петербурге за несколько лет. Научимся определять рыночную стоимость объектов недвижимости. Вы установите параметры для построения автоматизированной системы, отслеживающей аномалии и мошеннические действия.

По каждой продаваемой квартире у вас есть два вида данных. Первые — вносили сами пользователи при публикации объявлений, вторые — получены на основе картографических данных: расстояния до центра, аэропорта, ближайшего парка и водоёма.

## Инструкция по выполнению проекта

### Задание 1. Начните исследование

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

Путь к файлу: `/datasets/real_estate_data.csv`

## Шаг 2. Подготовьте данные

- определите и заполните пропущенные значения;
- приведите данные к нужным типам.

Посчитайте и добавьте в таблицу:

- цену квадратного метра;
- день недели, месяц и год публикации объявления;
- этаж, на котором расположена квартира; варианты — первый, последний, другой;
- соотношение жилой и общей площади, а также отношение площади кухни к общей.

Поясните:

- какие пропущенные значения обнаружены;
- в чём возможные причины возникновения пропусков;
- в каких столбцах потребовалось изменить тип данных и почему;
- по какому принципу заполнены пропуски.

Оформление:

Выполните **первое задание** в *Jupyter Notebook*. Свои действия описывайте в ячейках типа *markdown*. Чтобы открыть рабочую тетрадь, нажмите кнопку «Далее».

## Задание 2. Опишите выводы

**Шаг 1. Проведите исследовательский анализ данных и ответьте на вопросы**

- Каковы типичные параметры продаваемых квартир (например, площадь, цена)? Укажите наиболее важные характеристики и постройте графики. Сколько обычно длится процесс продажи?
- Уберите квартиры с очень высокой ценой и другими необычными параметрами. Опишите, какие особенности обнаружили.
- Какие факторы больше всего влияют на стоимость квартиры? Отдельно изучите, зависит ли цена квадратного метра от числа комнат, этажа (первого или последнего), удалённости от центра и даты размещения: дня недели, месяца и года.
- Что можно сказать о цене квадратного метра в пригородах? Выберите 10 населённых пунктов с наибольшим числом объявлений. Эти данные можно отличить по имени в столбце `'locality_name'`.
- Выделите сегменты типичных квартир в центре (по удалённости, числу комнат и площади) и вне центра. Границы определите самостоятельно. Выберите достаточно крупные сегменты с не менее 1000 объявлений. Сравните корреляцию основных факторов с ценой по всем предложениям и объявлениям в вашей выборке.
- Подберите наиболее характерные параметры для апартаментов, студий и квартир свободной планировки. Влияют ли они на цену? Сравните со схожими параметрами квартир типовой планировки.

## Шаг 2. Напишите общий вывод

Оформление:

**Второе задание** представьте как текст в *Jupyter Notebook*, в ячейке 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 (м<sup>2</sup>)
- *last\_price* — цена на момент снятия с публикации
- *living\_area* — жилая площадь (м<sup>2</sup>)
- *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 (м<sup>2</sup>)
- *total\_images* — число фотографий квартиры в объявлении

## Как будут проверять мой проект?

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

На что обращают внимание наставники, 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 как вы 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!