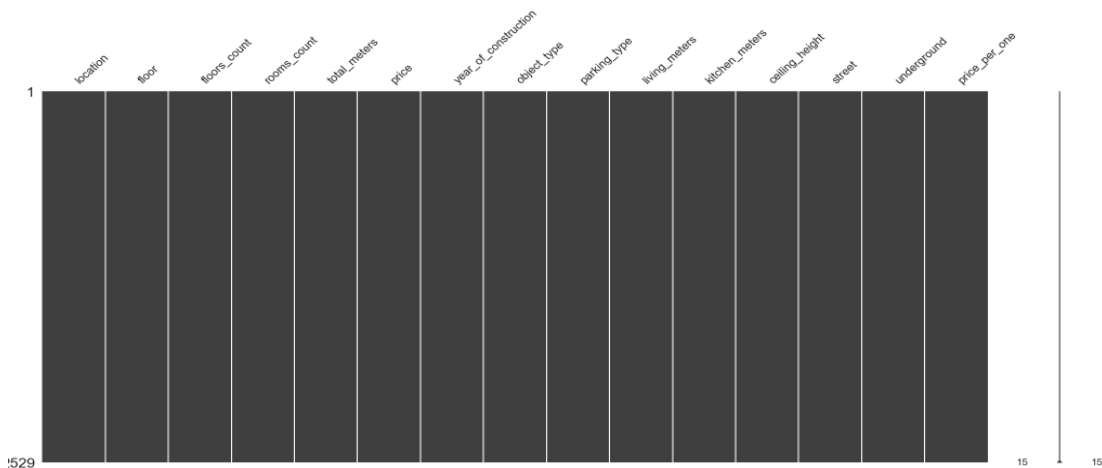


Цель проекта

Целью проекта является построение модели, способной оценивать стоимость квадратного метра жилой недвижимости в Москве и Московской области. Основой для модели служат данные по квартирам на продажу, которые были собраны и подготовлены к исследовательскому анализу.

Основные этапы проекта



Изображение 1. Готовые отсортированные данные для ML

1. Сбор данных

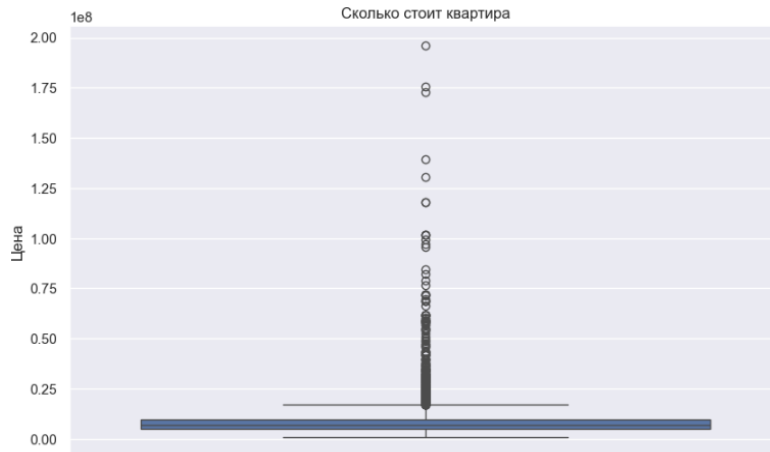
Были собраны данные о квартирах из источников объявлений, а именно Циан. Использовал библиотеку **cianparser**. Список характеристик, значимо влияющих на цену, был составлен с учетом открытых источников и личного опыта. К основным факторам относятся:

- Местоположение объекта (район),
- Общая площадь квартиры,
- Количество комнат,
- Год постройки и этажность здания,
- Материал дома и тип отделки,
- Наличие парковки.

#	Column	Non-Null	Count	Dtype
0	location	2529	non-null	int64
1	floor	2529	non-null	int64
2	floors_count	2529	non-null	int64
3	rooms_count	2529	non-null	int64
4	total_meters	2529	non-null	float64
5	price	2529	non-null	int64
6	year_of_construction	2529	non-null	int64
7	object_type	2529	non-null	int64
8	parking_type	2529	non-null	int64
9	living_meters	2529	non-null	int64
10	kitchen_meters	2529	non-null	int64
11	ceiling_height	2529	non-null	int64
12	street	2529	non-null	int64
13	underground	2529	non-null	int64
14	price_per_one	2529	non-null	float64

Изображение 2. Данные которые я отсортировал.

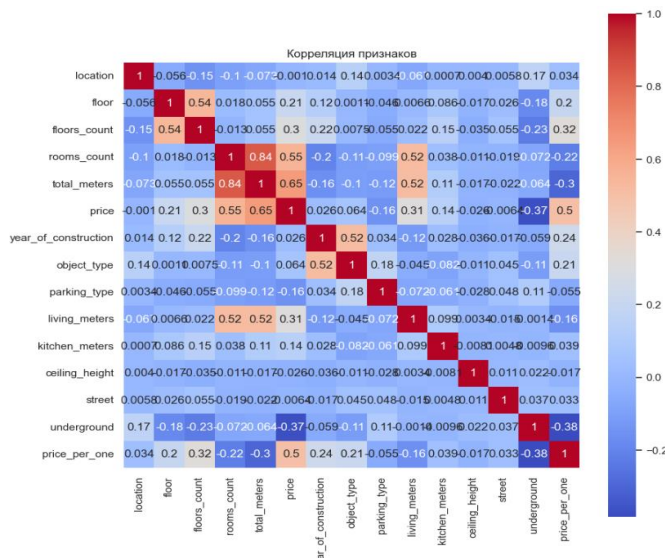
## 2. Очистка и подготовка данных



Изображение 3. Выявление выбросов из таблицы "price"

- **Обработка пропусков и ошибок:** Пропущенные значения были заполнены или удалены. Для столбцов с небольшим количеством пропусков использовалось заполнение средними или медианными значениями, а в случаях значительных пропусков данные удалялись.
- **Удаление выбросов:** были проведены проверки на наличие аномально высоких или низких значений, которые могут исказить результаты анализа. Такие выбросы были удалены.
- **Приведение данных к нужному формату:** Некоторые текстовые значения были преобразованы в числовые, лишние символы, такие как м² и д.р, были удалены.

## 3. Исследовательский анализ данных



Изображение 4. Корреляция отсортированных данных

- **Корреляционный анализ:** Проведен анализ корреляций между основными признаками. Было выявлено, что на цену квадратного метра значительно влияют такие параметры, как общая площадь, количество комнат и стоимость за квадратный метр. Меньшее влияние оказали такие факторы, как этаж и год постройки.

#### 4. Выявление ключевых факторов для модели

В результате анализа были выделены ключевые признаки, влияющие на стоимость недвижимости. Это позволяет использовать их в качестве предикторов в будущей модели оценки стоимости квадратного метра. Основные признаки, которые показали наибольшее влияние:

- Площадь квартиры,
- Количество комнат,
- Цена за квадратный метр.

#### Выводы и рекомендации

- **Основные влияющие факторы:** На цену квартиры больше всего влияют её площадь и количество комнат, что подтвердил корреляционный анализ. Также стоит учитывать цену за квадратный метр, которая сама по себе сильно коррелирует с общей ценой.
- **Для построения модели оценки:** Следует уделить внимание именно этим факторам и дополнить модель дополнительными признаками, такими как наличие отделки и д.р.
- **Пропуски:** На этапе очистки данных удалось избавиться от большинства выбросов, что повысит точность и стабильность модели при прогнозировании цен.

#### Заключение

В результате проделанной работы был проведен анализ факторов, влияющих на цену квадратного метра, собраны и подготовлены данные для построения модели, а также определены ключевые параметры для использования в оценочной модели. Эти данные и выводы могут быть полезны как для разработки модели машинного обучения.