

Государственное бюджетное профессиональное  
образовательное учреждение Московской области  
«Физико-технический колледж»

## **Исследовательский аналитический анализ:**

***«Модель оценки цены квартиры на рынке по  
Московскому региону: Москва, Новая Москва,  
Московская область»***

Работу выполнил:  
Студент группы № ИСП-21  
Коряковский А. Р.  
Проверил:  
преподаватель  
Базяк Г. В.

Долгопрудный, 2024

## Аннотация

В наше время особенно важно следить за динамикой цен на жильё и выбирать подходящий момент для покупки недвижимости в Москве и Московской области. Регулярный анализ стоимости и учёт рыночных тенденций помогают не только определить оптимальное время для приобретения, но и оценить факторы, которые могут повлиять на цену в будущем: развитие инфраструктуры, появление новых жилых комплексов, улучшение транспортной доступности и так далее.

**Цель отчета:** провести сбор данных и выполнить разведочный исследовательский анализ (EDA) для построения модели, оценивающей цену квадратного метра недвижимости в Московском регионе.

Задачи:

1. Составить список ключевых факторов, влияющих на цену квадратного метра жилой площади, на основе открытых источников и личного опыта.
2. Провести парсинг данных по квартирам на продажу с платформ, таких как Циан, Авито и ДомКлик.
3. Подготовить данные для анализа, проверяя их на наличие пропусков и аномалий.
4. Выполнить исследовательский анализ данных, визуализируя взаимосвязи между признаками и выявляя наиболее значимые из них.

## Введение

Рынок жилой недвижимости в Москве постоянно меняется, и его анализ помогает определить оптимальное время для покупки жилья. Это может помочь потенциальным покупателям существенно сэкономить средства.

Москва — столица России, поэтому она привлекает много людей, которые хотят приобрести недвижимость. Из-за этого спрос на жильё в городе остаётся стабильно высоким.

Рынок столичной недвижимости чутко реагирует на различные экономические и политические факторы: изменения ставок по ипотеке, экономическую и политическую ситуацию, государственные меры регулирования. Всё это может серьёзно влиять на цены и создавать как благоприятные, так и неблагоприятные условия для покупателей.

Изучение этих факторов и их влияния на стоимость жилья позволяет лучше понять, какие изменения могут произойти на рынке в ближайшее время и в долгосрочной перспективе.

## Основная часть

Изначально был получен датасет размерностью 10582 строки, состоящий из 19 столбцов, из них 13 – категориальных и 6 с числовыми данными.

Ключевой переменной в данной работе будет являться переменная «price\_of\_meter» (цена за метр квадратный), который вычисляется из столбцов «price» (цена квартиры) и «total\_meters» (общая площадь). Так как эти два столбца являются необходимыми для существования ключевой переменной, то записи, в которых отсутствует значения в столбцах – будут удалены из датасета.

Просмотр выбросов данных по ключевой переменной «price\_of\_meter» представлен на рисунке 1.

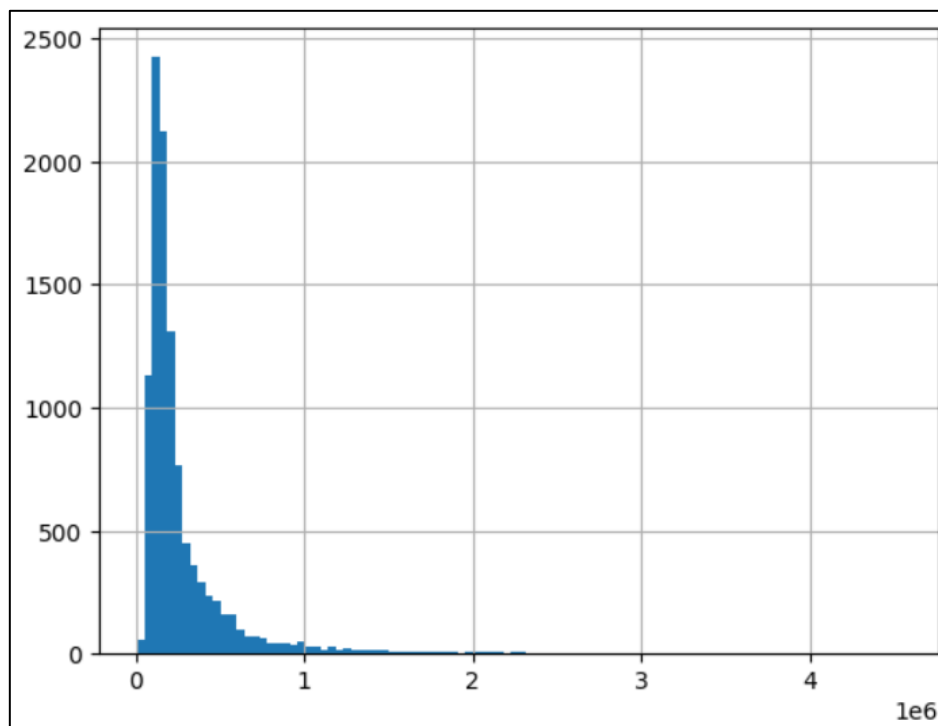


Рисунок 1 – диаграмма выбросов.

Посмотрев на график заметных выбросов не видно, поэтому этот столбец остается без изменений.

Далее необходимо рассмотреть пропущенные значения в записях датасета. Для этого отображена тепловая карта на рисунке 2.

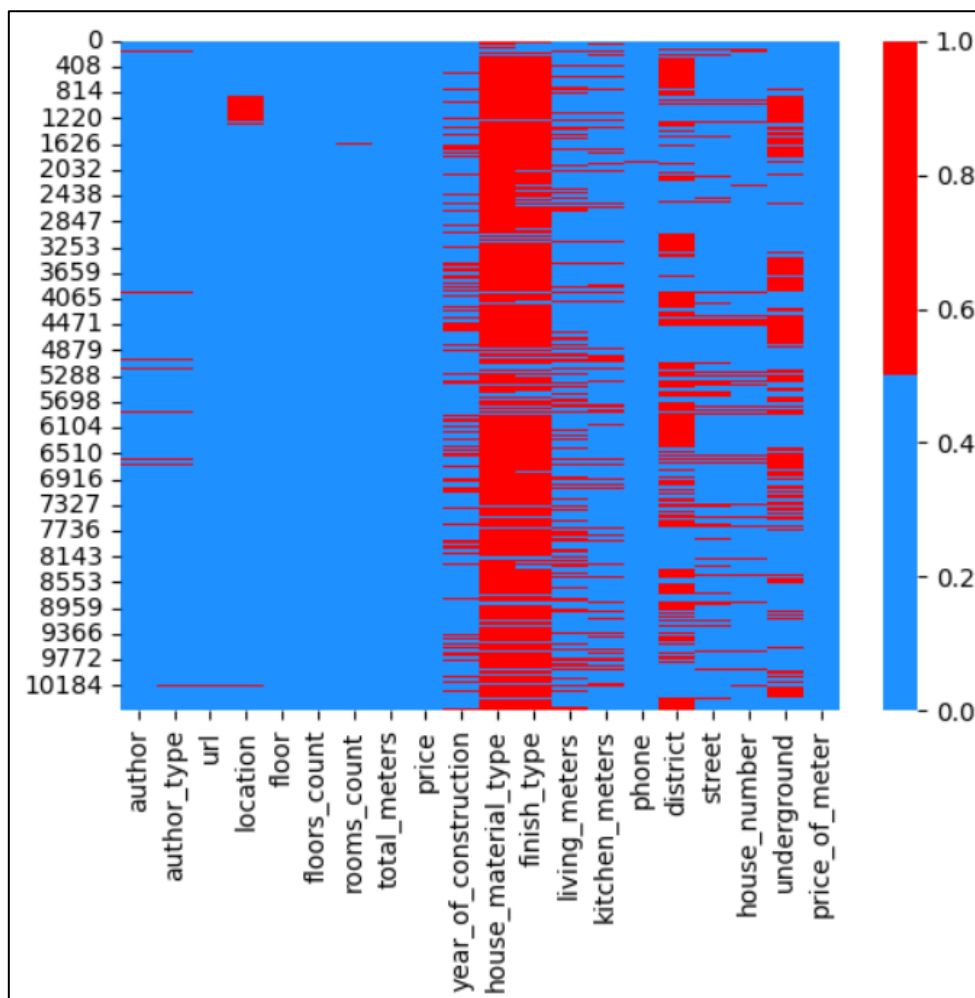


Рисунок 2 – тепловая карта пропущенных значений.

Глянув на тепловую карту, видно, что больше всего пропусков в столбцах «house\_material\_type» и «finish\_type». Поэтому данные два столбца полностью удаляются из датасета.

Также для очистки датасета от записей с пропусками удаляем записи, где отсутствуют значения в столбцах «location» и «street», так как объявления не несут смысла без наличия минимального адреса квартиры (город, улица).

Очистка столбца «year\_of\_construction» (года постройки). Убираются все записи, где год постройки меньше 1924 или более 2024, так как либо это очень старые здания, которые не будут рассматриваться для покупки, либо это еще не построенные дома.

Удаляются повторяющиеся записи (дубликаты) командой методом: `drop_duplicates()`.

Пропуски в таких столбцах как «underground», «district» и «house\_number» заполняются значением «\_MISSING\_», а не удаляются, так как это не особо важные столбцы и записи при пустом значении остаются ценными.

Заменяются пропуски в столбцах «living\_meters» и «kitchen\_meters» средними значениями.

После проделанных манипуляций тепловая карта (рисунок 3) стала выглядеть красивее – практически не осталось пропущенных.

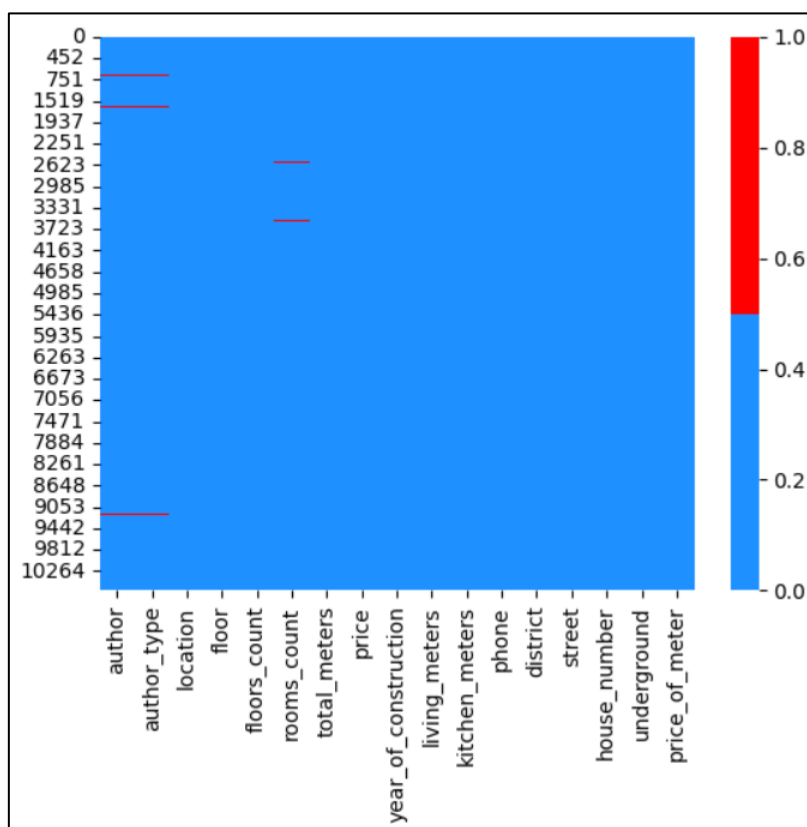


Рисунок 3 – обновлённая тепловая карта пропущенных значений.

В результате проведённой обработки датасет был приведён в более структурированный вид. Пропуски данных были либо заполнены, либо удалены, что позволило избежать искажений и неточностей в анализе. Удаление дубликатов обеспечило уникальность записей. Количество записей сократилось до 6375, а количество столбцов до 17.

После очистки и заполнения данных мы преобразовали всю информацию в числовой формат. Это позволит нам определить факторы, которые влияют на ключевую переменную — «price\_of\_meter» (цена за метр квадратный).

Отобразим матрицу корреляции (рисунок 4), на которой будут видны влияющие факторы.

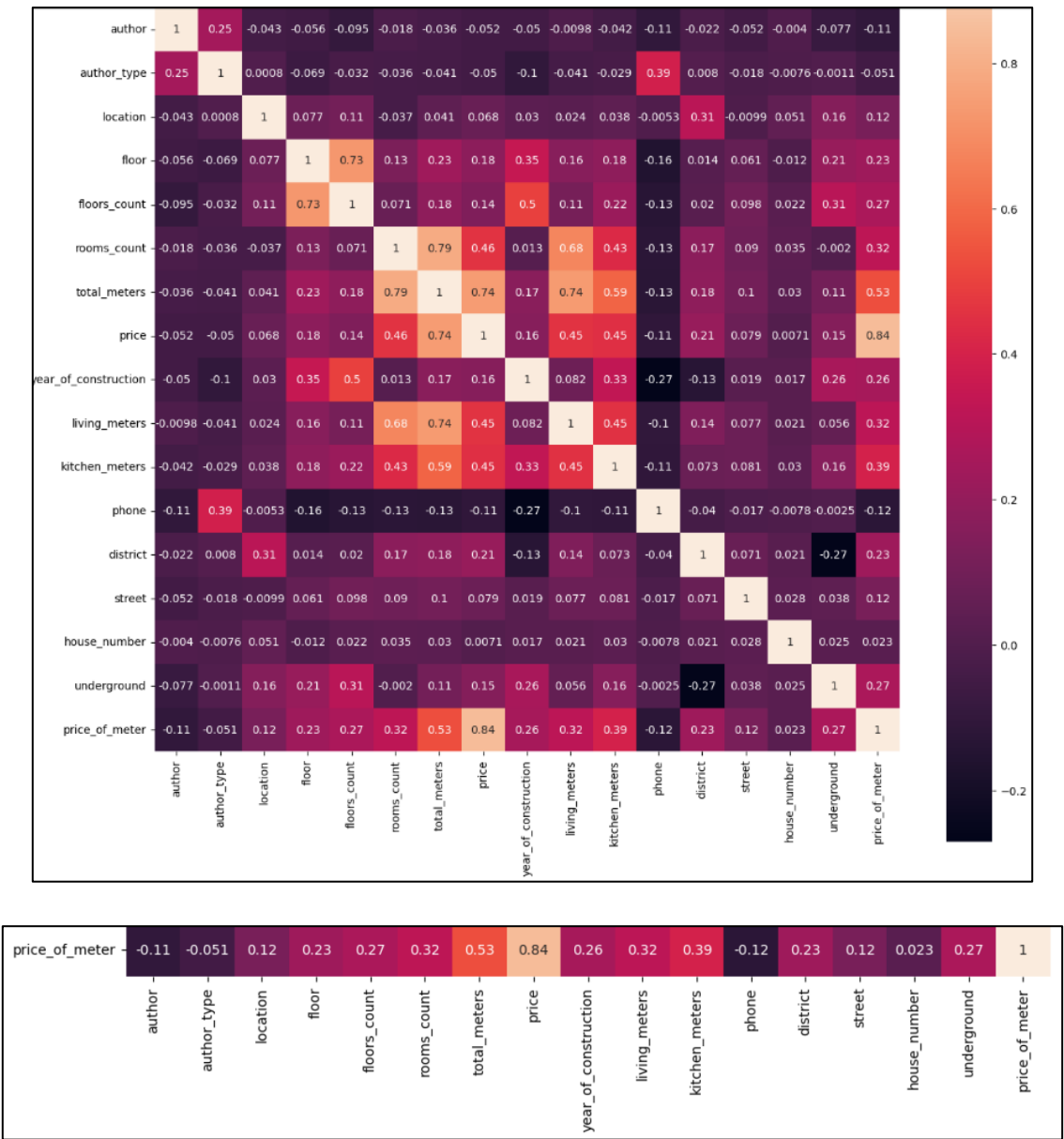


Рисунок 4 – матрица корреляции.

Если посмотреть на матрицу, то можно выделить 5 основных факторов, влияющих на ключевой столбец: total\_meters, rooms\_count, floors\_count, living\_meters и kitchen\_meters. Для каждого фактора построена диаграмма (рисунок 5 - 9).

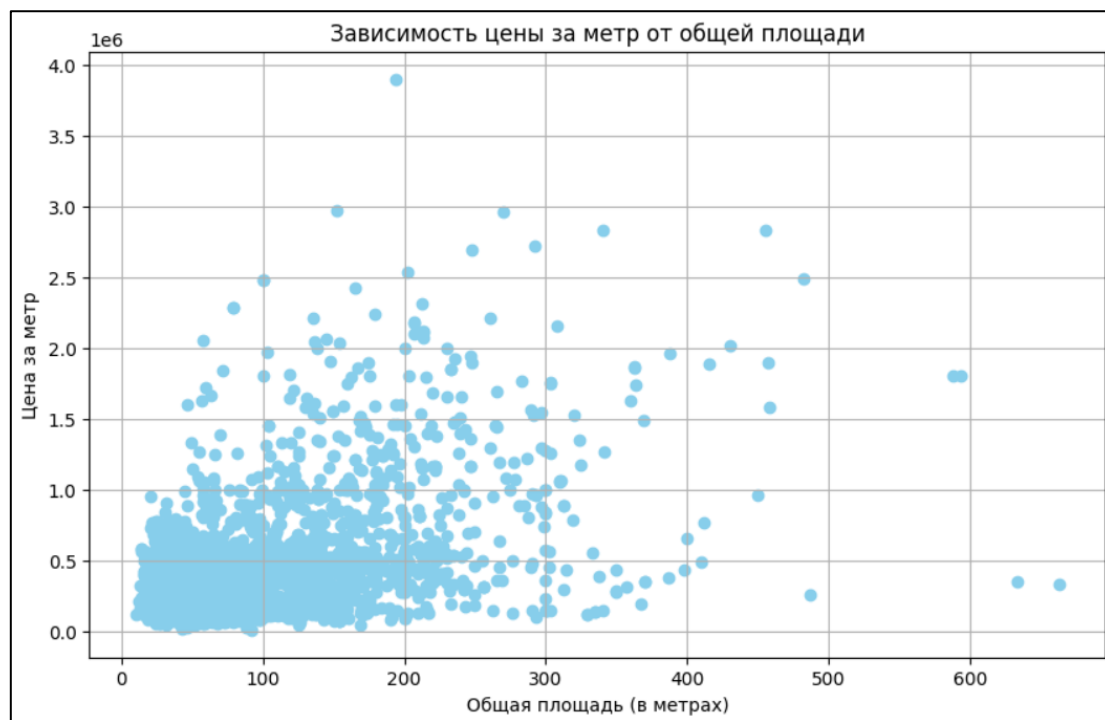


Рисунок 5 – график зависимости цены за метр от общей площади.

По графику (рисунок 5) можно сделать вывод, что чем больше общая площадь квартиры, тем выше цена за метр, однако присутствуют критические значения, которые либо имеют очень большую стоимость, либо наоборот очень дешёвые для своей площади.

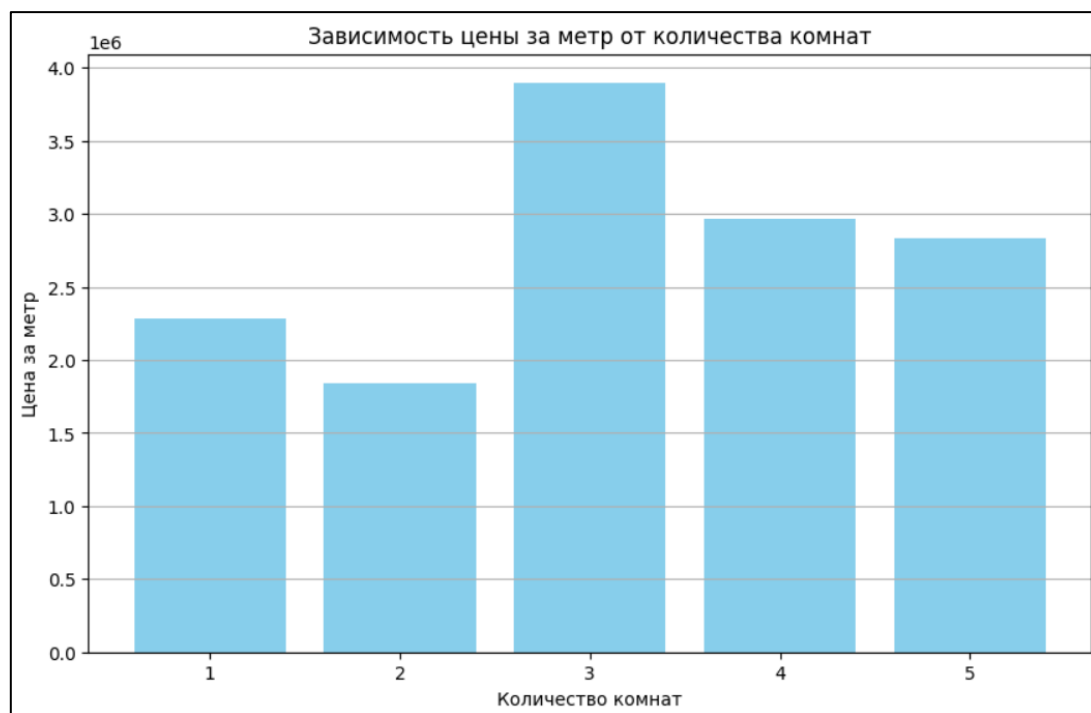


Рисунок 6 – график зависимости цены за метр от количества комнат.



На рисунке 6 не особо можно заметить прямой зависимости, так, как например у двухкомнатной квартиры цена за метр ниже, чем у однокомнатной, та же ситуация с 5 и 4 комнатными квартирами.



Рисунок 7 – график зависимости цены за метр от количества этажей.

По графику (рисунок 7) видно, что чем больше количество этажей в доме, тем более цена за метр, но также присутствуют критические значения, которые либо имеют очень большую стоимость, либо наоборот низкую.

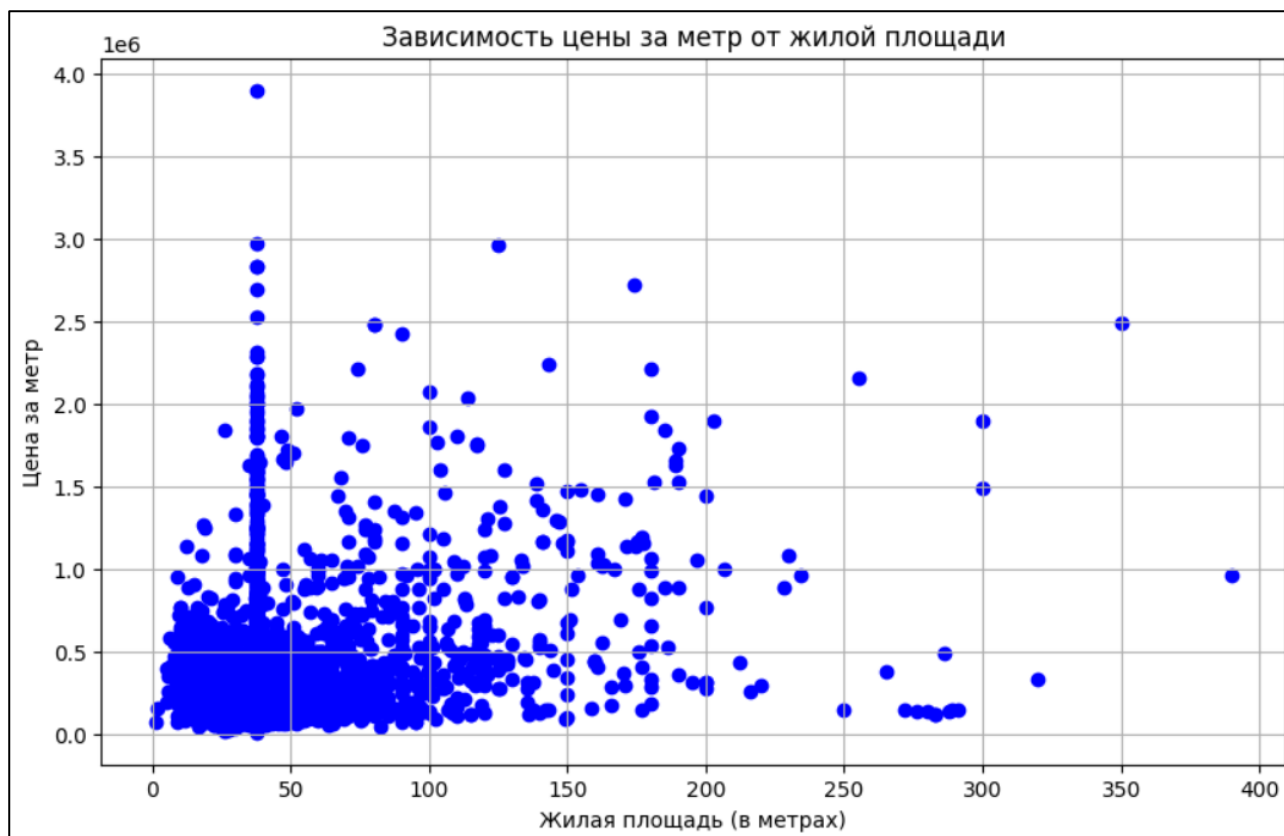


Рисунок 8 – график зависимости цены за метр от жилой площади.

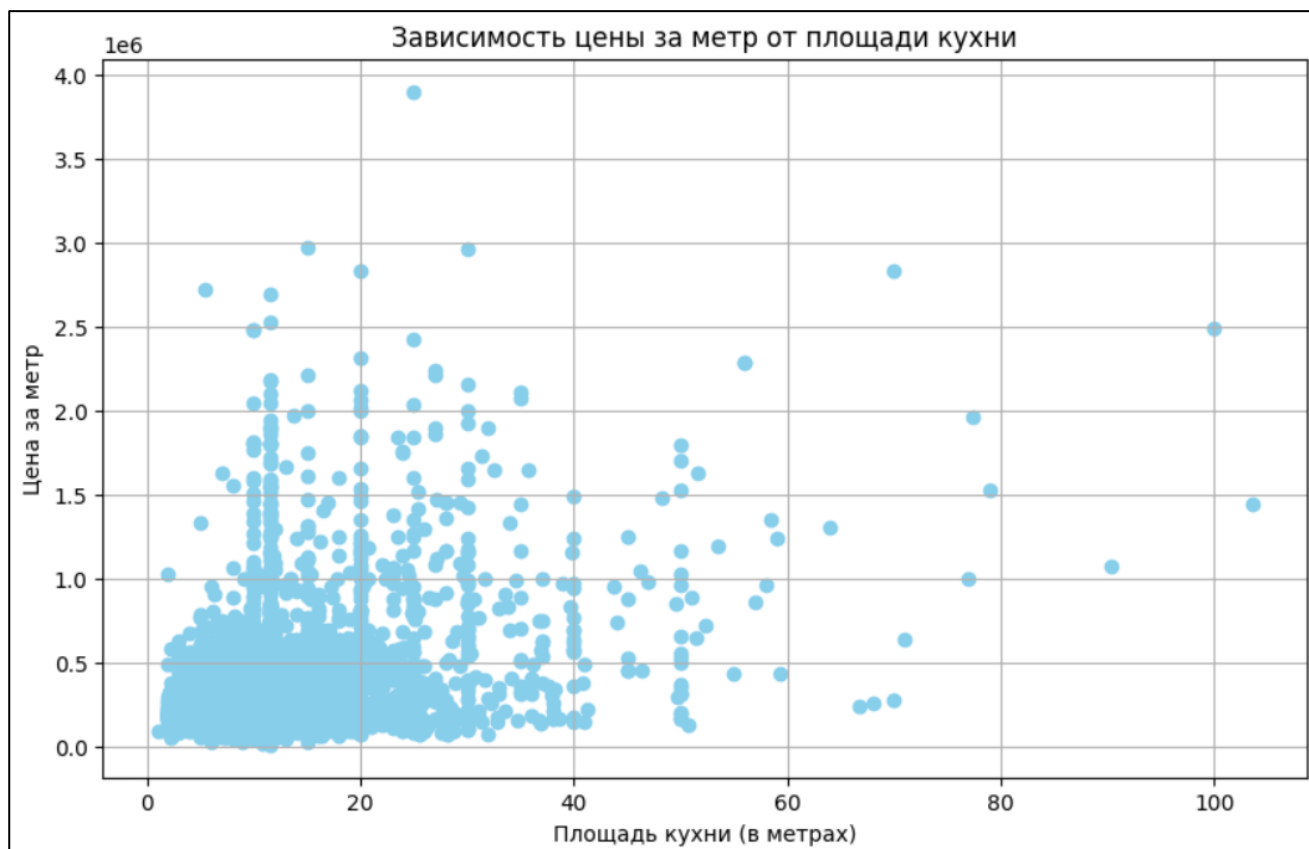


Рисунок 9 – график зависимости цены за метр от площади кухни.

## Результаты работы

В результате проведённой работы с предоставленным датасетом были проверены данные на выбросы, удалены записи с пропусками и дубликаты, что позволило обеспечить высокое качество данных для дальнейшего анализа. Были проведены исследования, направленные на выявление факторов, которые оказывают влияние на ключевую переменную, такие как общая площадь (total\_meters), количество комнат (rooms\_count), количество этажей (floors\_count), жилая площадь (living\_meters) и площадь кухни (kitchen\_meters). Это позволило лучше понять взаимосвязи между различными характеристиками объектов и их влияние на стоимость квадратного метра.