

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

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

Работу выполнил:
студент группы № ИСП-22
Богомяков Максим
Проверил:
Преподаватель информатики
Базяк Г.В. Коновалов И. В.

Долгопрудный, 2024

СОДЕРЖАНИЕ

СОДЕРЖАНИЕ.....	2
ВВЕДЕНИЕ	3
Основная часть.....	5
1.1 Методология.....	5
1.2 Анализ данных	7
Рефлексия	10
Заключение.....	10

ВВЕДЕНИЕ

Актуальность задачи

Оценка недвижимости - важная составляющая девелоперского бизнеса. Информация, о реальной цене квартиры исходя из рынка, интересна для покупателей продавцов, застройщиков, агентов и др.

Эту информацию можно использовать по-разному, в частности, у бизнеса есть потребность использовать такую модель для лидогенерации клиентов. Потенциальный клиент заходит на сайт, хочет оценить свою квартиру, вводит параметры: квартира, на 3 этаже, 3-х комнатная, дизайнерский ремонт. Затем нажимает кнопку «узнать цену», и, прежде чем дать ответ, мы хотим запросить у клиента номер телефона или адрес почты.

Для этого нужна модель оценки цены квартиры по Московскому региону: Москва, Новая Москва, Московская область.

Постановка задачи

Цель: Собрать данные и провести разведочный исследовательский анализ данных (EDA) для построения модели, которая будет оценивать цену квадратного метра недвижимости в Московском регионе (Москва, Новая Москва, Московская область).

Задачи:

Задание базового (обязательного) уровня)

- Используя открытые источники и личный опыт, составить список параметров, значительно влияющих на цену квадратного метра жилой площади.
- С учётом выявленных выше факторов произвести парсинг данных по квартирам на продажу, используя различные парсеры. Данные получаем, используя различные сайты с объявлениями о продаже недвижимости: Циан, Авито, ДомКлик и др.
- Произвести подготовку данных для анализа: проверка на пропуски, выбросы и ошибки. Обработать выявленные аномалии (удалить / заполнить)
- Проведите Исследовательский Анализ Данных (EDA). Постройте распределение основных параметров; визуализируйте взаимосвязи между ними; определите признаки, оказывающие наиболее сильное влияние на целевую переменную.

Основная часть

1.1 Методология

Для сбора данных использовался `cianparser`, с его помощью удалось собрать 8393 квартиры. Данные были собраны с платформы Циан, поскольку она предоставляет более обширные и полезные для анализа показатели по сравнению с другими платформами, такими как ДомКлик.

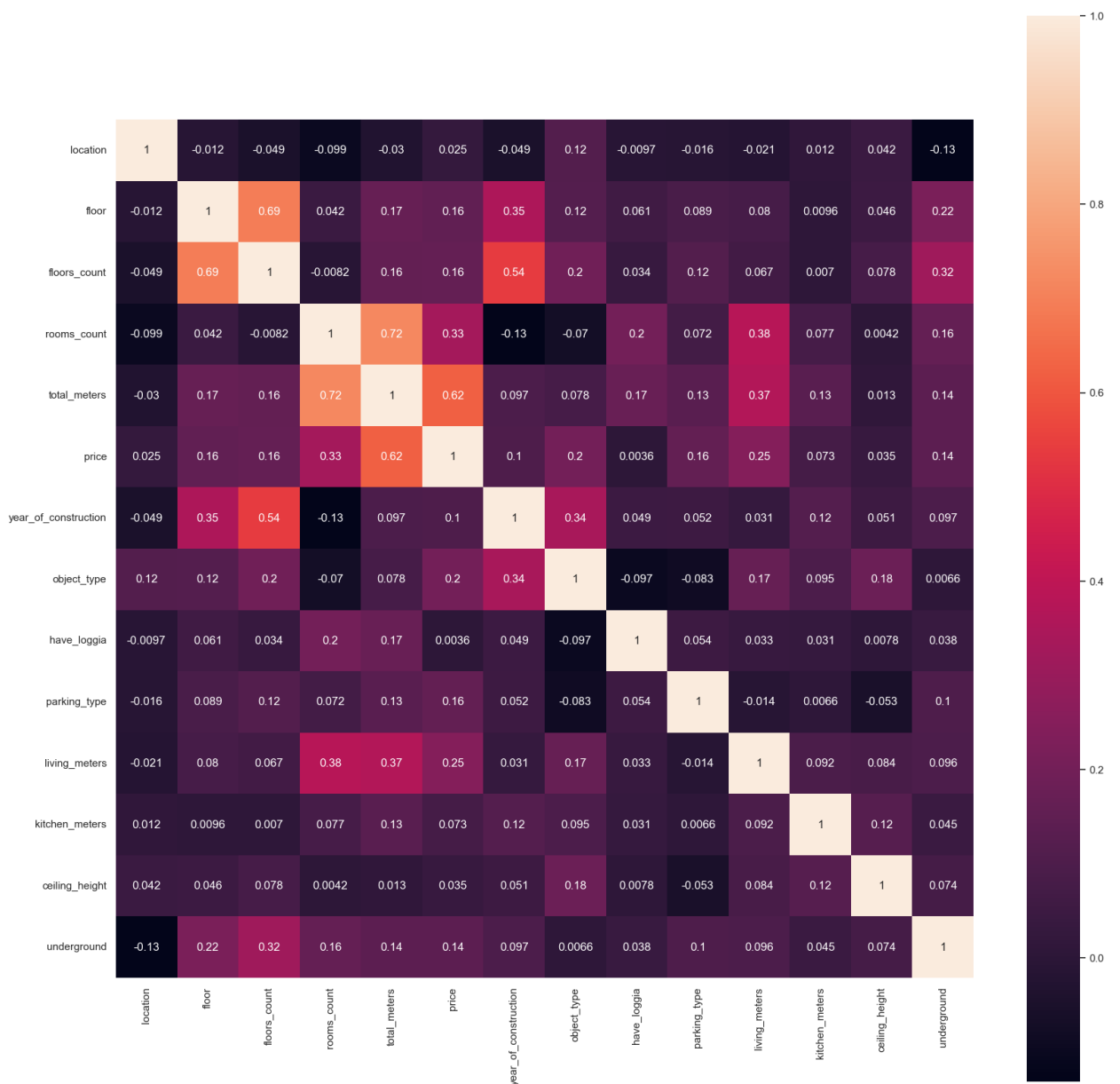
Основные колонки данных:

Колонка	Значение
url	ссылка
location	город
deal_type	тип объявления (продажа/аренда)
accommodation_type	тип жилья
floor	этаж
floors_count	количество этажей
total_meters	площадь квартиры
price	стоимость
seller.agent.is_agent	является ли продавец агентом
year_of_construction	год постройки здания
object_type	тип жилья
have_loggia	наличие лоджий
house_material_type	тип дома (монолитный, панельный и т.д.)
heating_type	тип отопления
finish_type	отделка
living_meters	жилая площадь
kitchen_meters	площадь кухни
phone	телефон
ceiling_height	высота потолка
district	район
street	улица
house_number	номер дома
underground	метро
residential_complex	жилой комплекс

1.2 Анализ данных

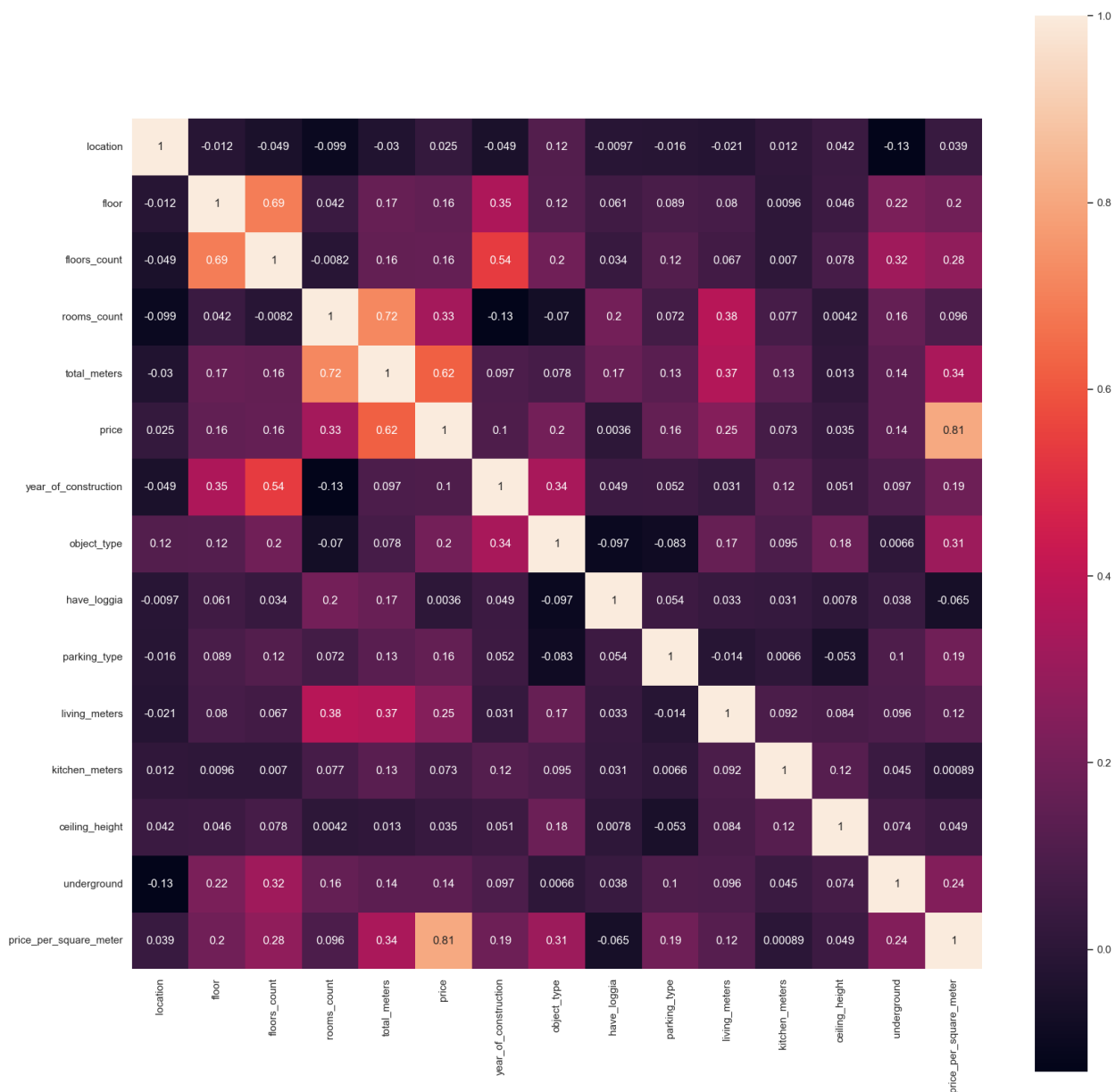
После получения данных мы их очищаем, удаляем лишние столбцы, проверяем на аномальные значения и удаляем строки с аномалиями. Затем превращаем все наши данные в числовые.

Для выявления зависимостей между характеристиками квартир и их ценой был построен график корреляций. По графику видно, что на цену больше всего влияет общее число метров квартиры (0,62), затем уже идет количество комнат (0,33)



Далее мы можем получить нашу целевую - цену за квадратный метр.

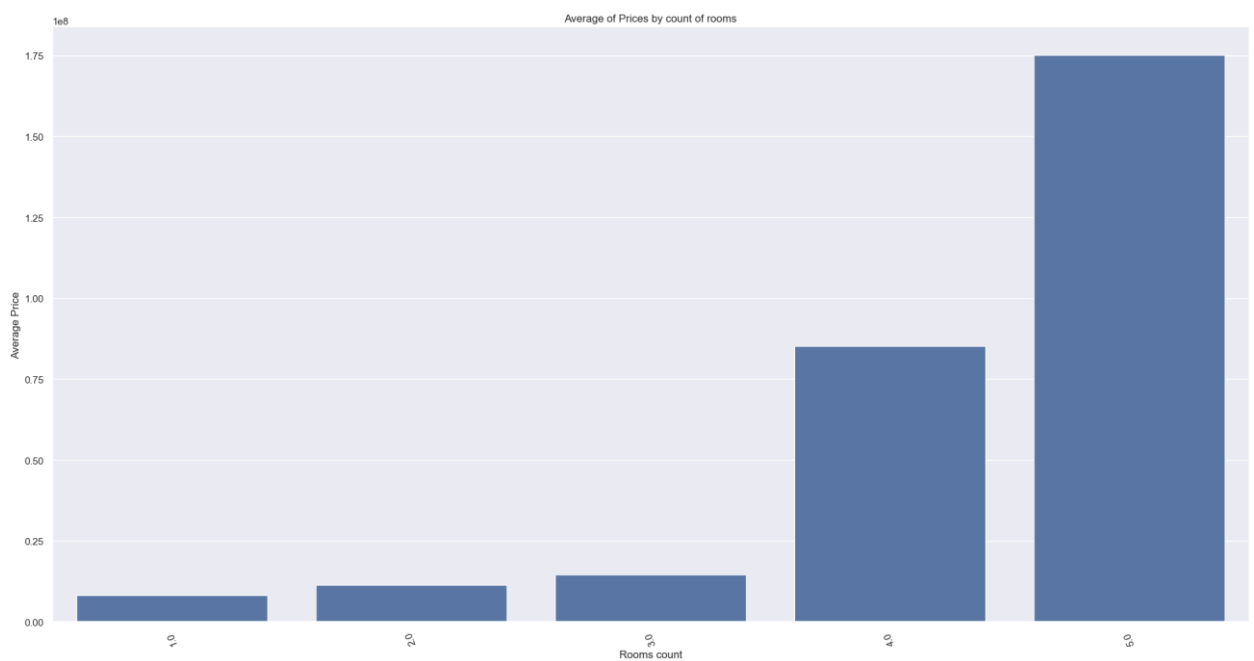
Матрица корреляций с ценой за квадратный метр:



На этом графике видно, что на нашу целевую влияют: всего метров, цена, наличие метро, число этажей, этаж.



По графику мы видим, что локация тоже значительно влияет на цену



На данном графике видно, что цена так же зависит от количества комнат, что довольно логично

Рефлексия

Что получилось:

- Выявить основные параметры наиболее сильно влияющие на нашу целевую.

Что можно улучшить:

- Избавиться от остальных выбросов
- Колонку "have_loggia" заполнить не нулями и единицами, а другим способом

Заключение

В ходе проведенного анализа были выявлены ключевые факторы, влияющие на целевой показатель — цену за квадратный метр.

Наибольшее влияние оказывают следующие параметры:

- Площадь квартиры — основной критерий, прямо пропорционально влияющий на стоимость.
- Локация — важный фактор, определяющий уровень цен в зависимости от расположения объекта недвижимости.

И уже чуть менее влияет: количество этажей, наличие метро и этаж.

Данные результаты создают основу для разработки модели, способной прогнозировать цену квартиры на основе этих ключевых характеристик.