

Аналитический отчет по анализу данных недвижимости

Выполнил: Николаев Ярослав Владимирович

ИСП-23В

1. Введение

1.1 Цели исследования

Целью данной работы является анализ данных о недвижимости, полученных с Циан, для выявления факторов, влияющих на стоимость квадратного метра квартиры, и подготовки данных для дальнейшего использования в построении моделей машинного обучения.

1.2 Задачи:

1. Получить и очистить данные о недвижимости.
2. Провести анализ числовых и категориальных переменных.
3. Заполнить пропущенные данные и подготовить датасет для визуализации и корреляционного анализа.
4. Построить визуализации для выявления ключевых закономерностей.
5. Сформировать выводы и рекомендации для использования данных в дальнейшем.

2. Инструменты

Для выполнения поставленных задач использовались следующие инструменты и библиотеки:

- Python для обработки данных и автоматизации запросов.
- Библиотеки pandas, numpy для анализа и подготовки данных.
- Визуализационные библиотеки: seaborn и matplotlib для построения графиков и тепловой карты корреляции.

Источником данных является Циан.

3. Этапы работы

3.1 Парсинг через Python

Для загрузки данных был создан парсер на основе cianparser.

3.2 Предварительная обработка данных

После парсинга был выполнен следующий процесс:

- Создан DataFrame с нужными колонками:
- Обработка строк с пустыми значениями

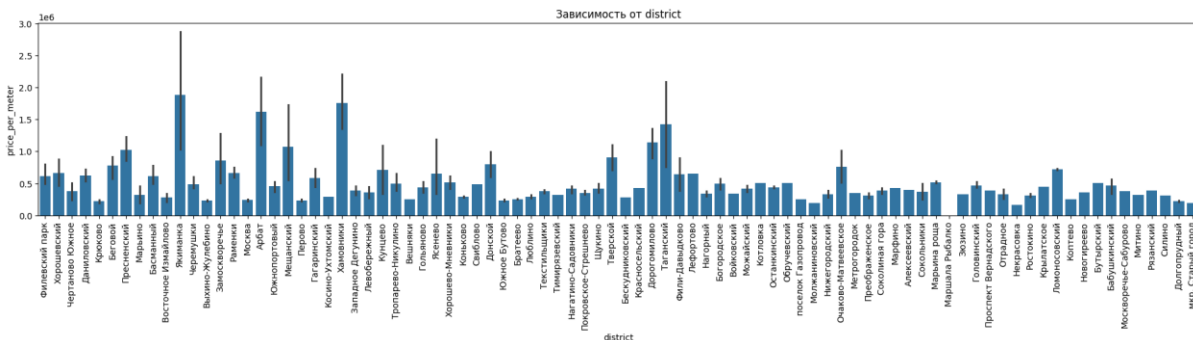
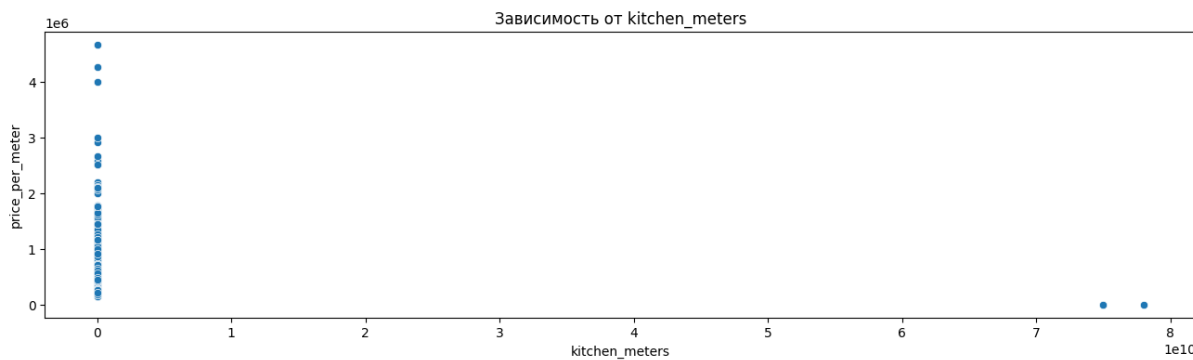
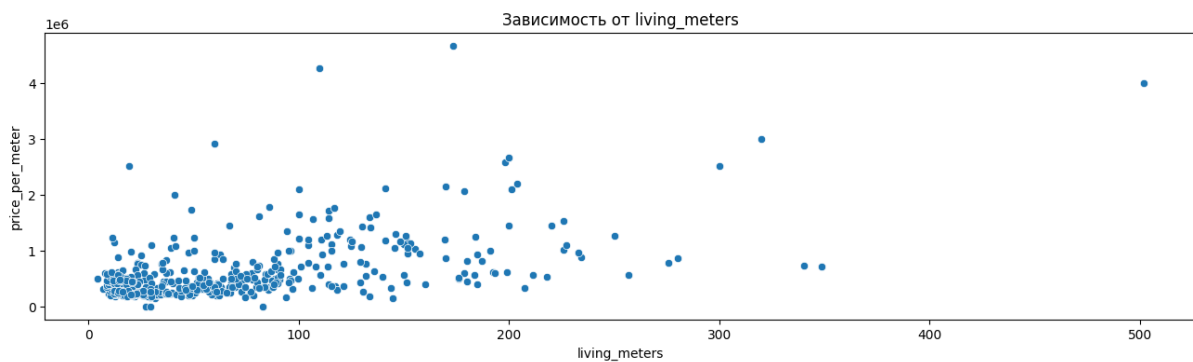
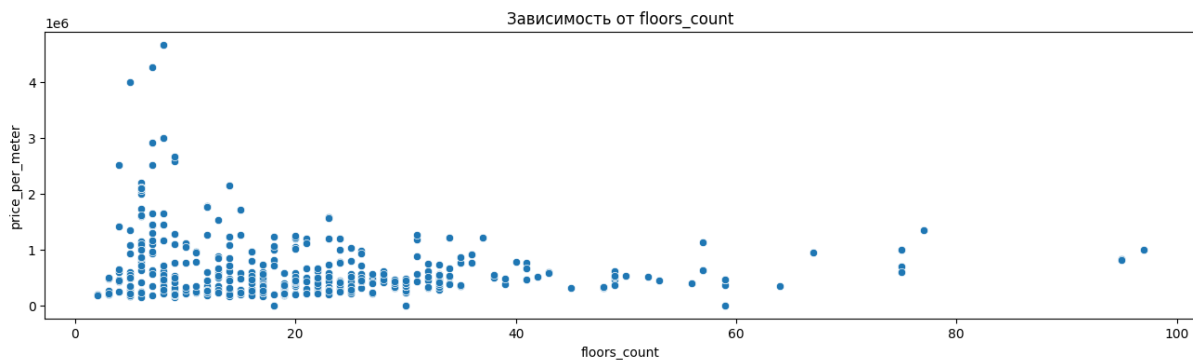
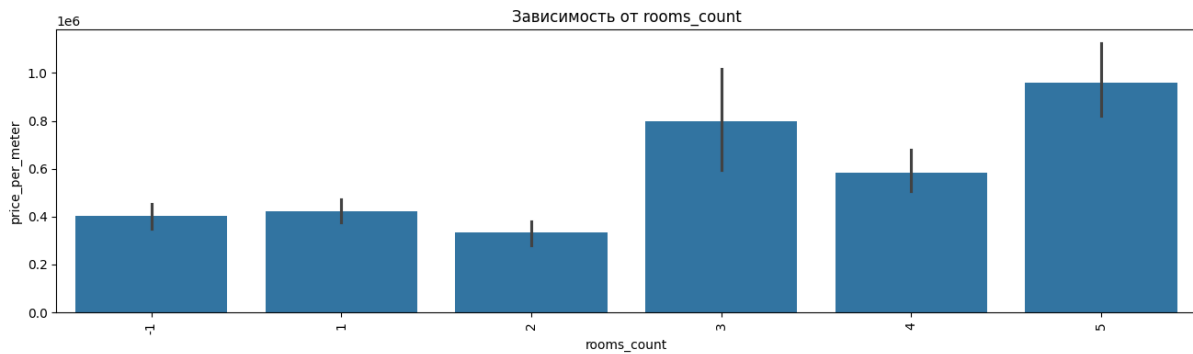
3.3 Выявление столбцов с пропущенными значениями

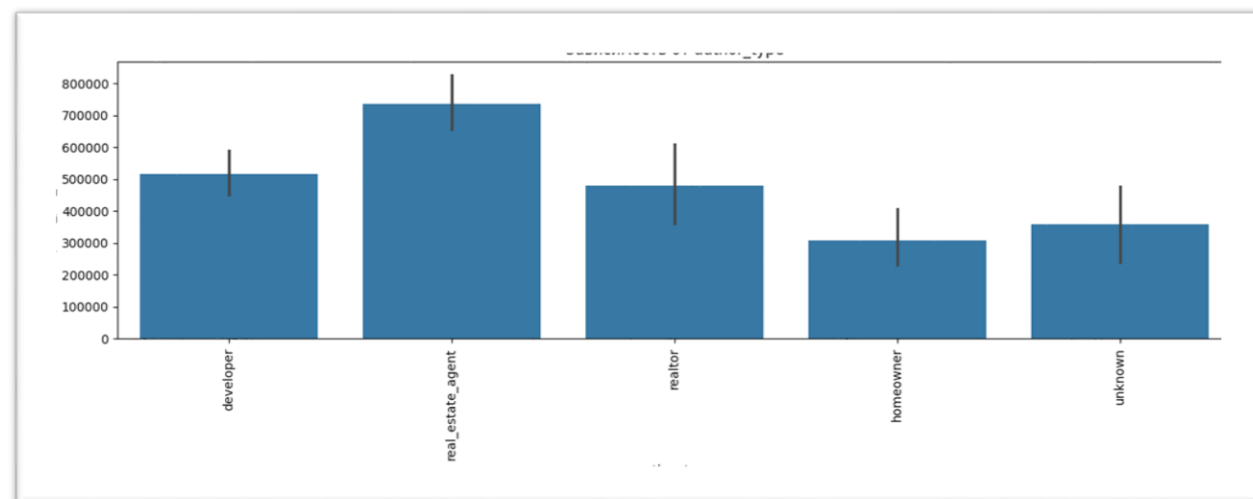
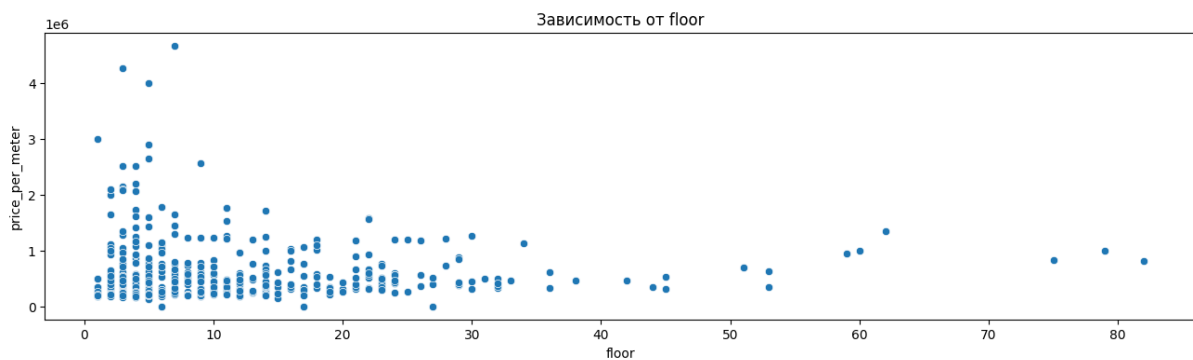
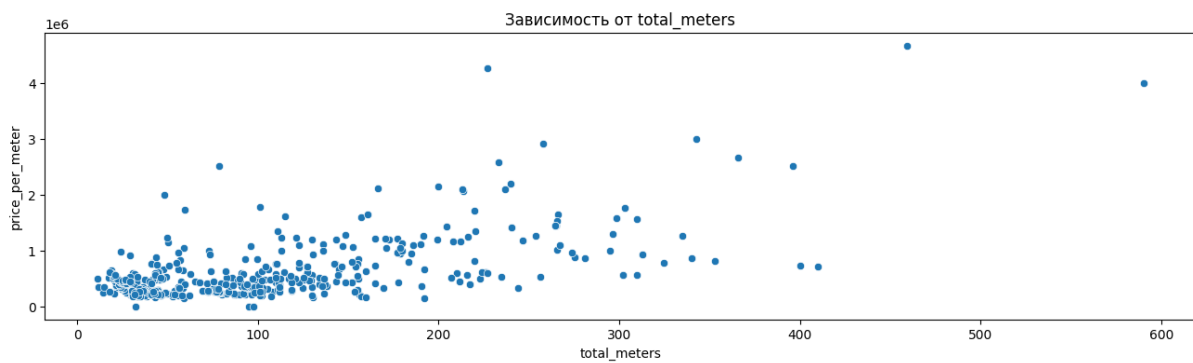
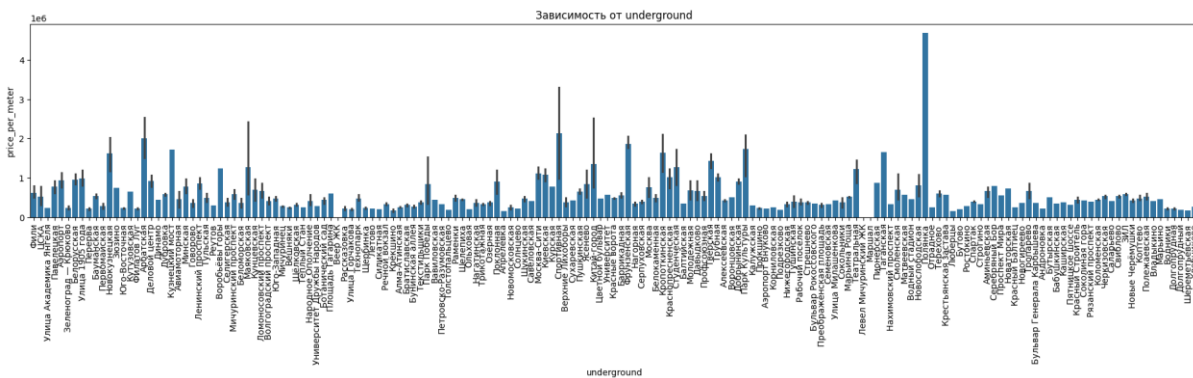
Проверка на пропущенные значения была выполнена с помощью кода

3.4 Визуализация данных

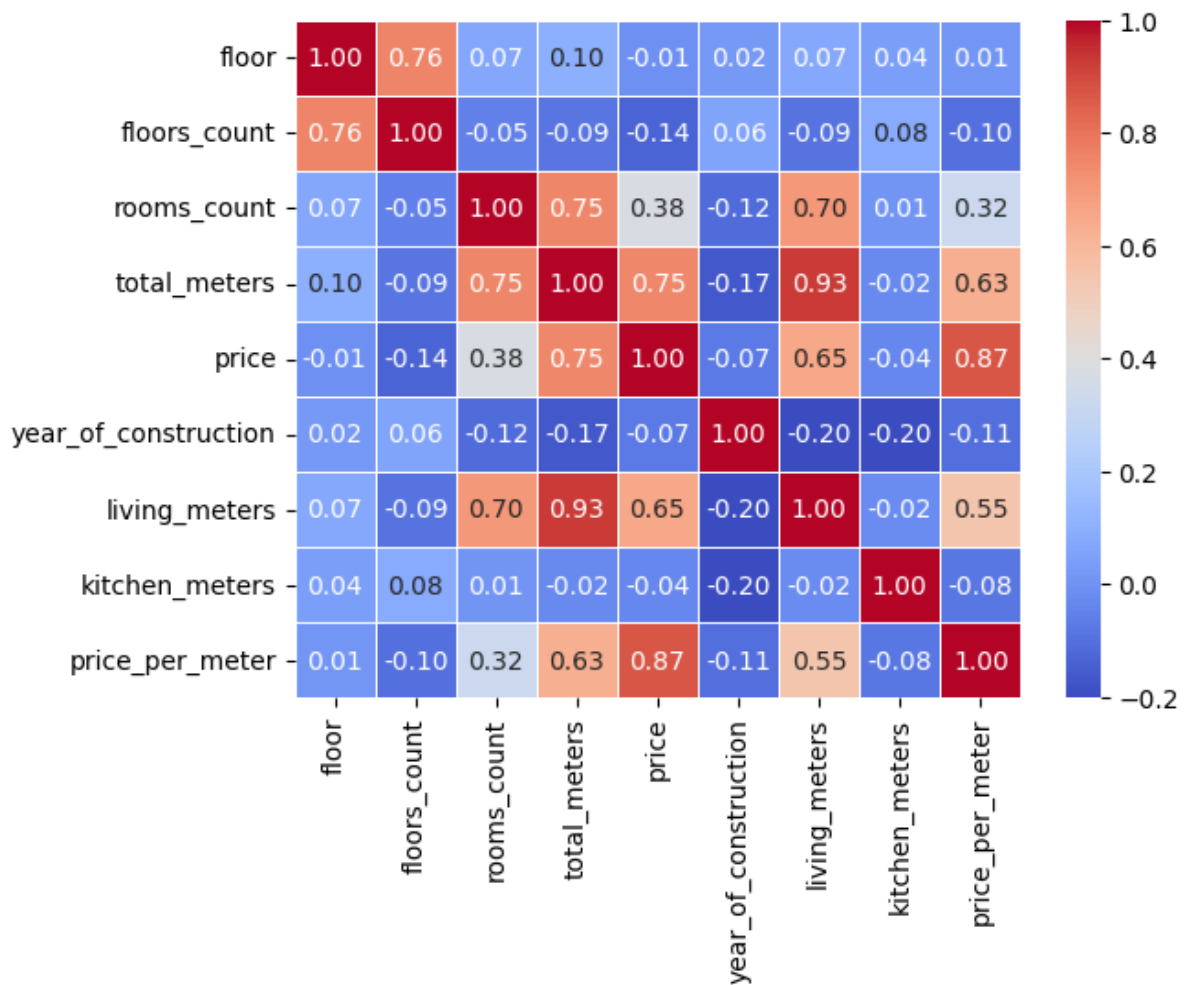
Для анализа взаимосвязи между ценой за квадратный метр и другими признаками были построены следующие графики:

- Диаграммы рассеяния:





- Тепловая карта корреляции, показывающая степень взаимосвязи:



4. Результаты и выводы

4.1 Анализ корреляции

Тепловая карта корреляции показала следующие ключевые зависимости:

- Цена за квадратный метр наиболее сильно коррелирует с общей площадью

6. Заключение

В ходе работы был проведен анализ и очистка данных о рынке недвижимости, полученных из Циан. Выполненная обработка позволила выявить ключевые зависимости между параметрами объектов и подготовить данные для дальнейшего использования в построении моделей предсказания цен. Данные готовы к применению для задач машинного обучения и мониторинга изменений рынка.