

**Аналитический отчет по анализу данных**  
**недвижимости Автор: [Алексей Костюк Дмитриевич]**  
**Группа: [ИСП-23В]**

---

## 1. Введение

### 1.1 Цель исследования

Целью данной работы является анализ данных о недвижимости, полученных с [помощью API cian](#), для выявления факторов, влияющих на стоимость квадратного метра жилья, 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. Методология и 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 [API cian](#). Используются запросы к API для получения информации о квартирах в Москве по различным параметрам (количество комнат, площадь, этаж и т.д.).

---

### 3. Этапы работы

#### 3.1 Загрузка данных через API cian

Для загрузки данных был разработан класс `cianApi`, который выполняет автоматизированные запросы к [API cian](#) с необходимыми параметрами (тип недвижимости, регион, количество комнат и т.д.). На каждом этапе выводились промежуточные данные для проверки корректности полученной информации.

#### 3.2 Предварительная обработка данных

После загрузки данных из API был выполнен следующий процесс:

- Создан DataFrame с нужными колонками: `price`, `area`, `rooms`, `square_price`, `subways`, `monthly_payment`.
- Обнаружены пропущенные значения в колонке `subways` и `monthly_payment`, которые были обработаны с использованием [pandas](#) и других методов.

#### 3.3 Выявление столбцов с пропущенными значениями

Проверка на пропущенные значения была выполнена с помощью кода:

В результате были обнаружены пропуски в колонках `subways`, `monthly_payment` и других, которые были заполнены с помощью [pandas](#) и других методов.

#### 3.4 Визуализация данных

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

- **Диаграммы рассеяния** для колонок `price`, `area`, `rooms` относительно `square_price`.
- **Тепловая карта корреляции**, показывающая степень взаимосвязи между числовыми признаками (см. прилагаемый график).

---

### 4. Результаты и выводы

#### 4.1 Анализ корреляции

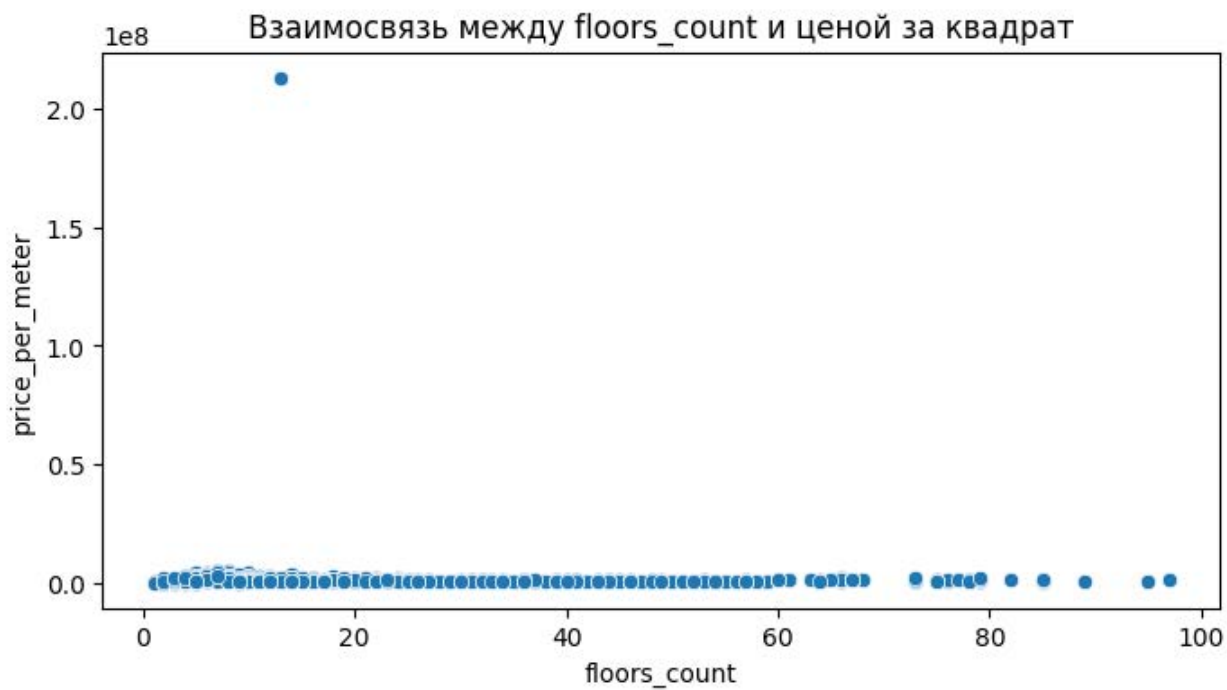
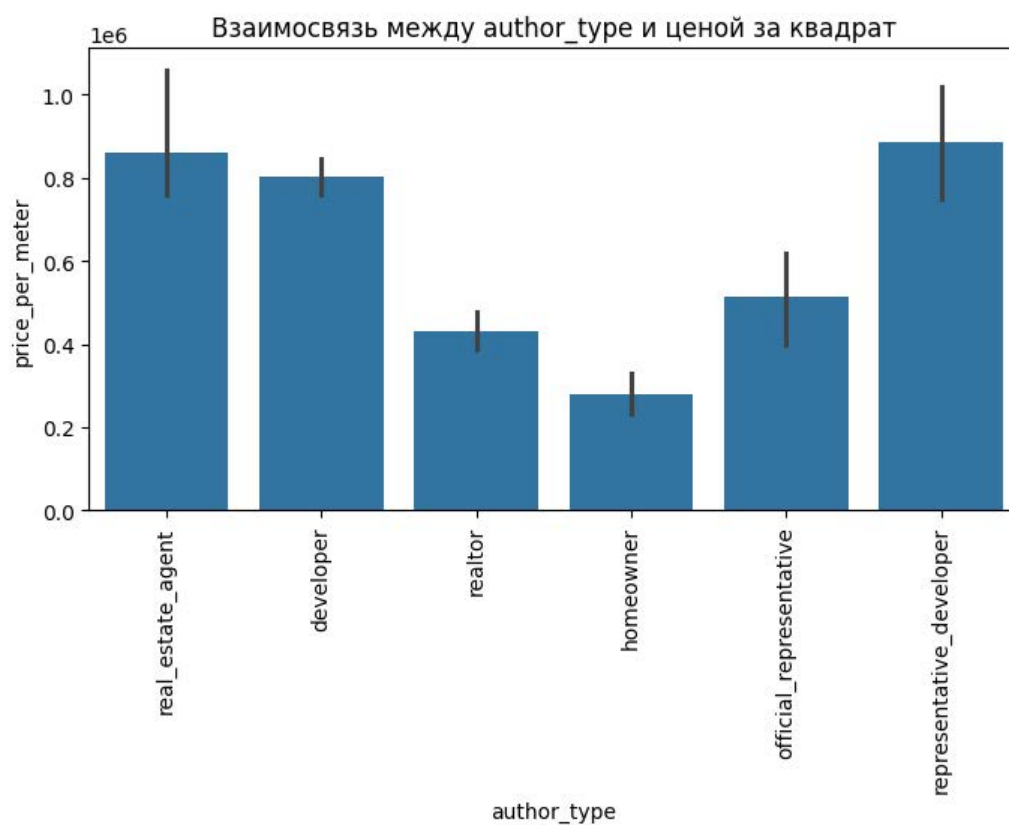
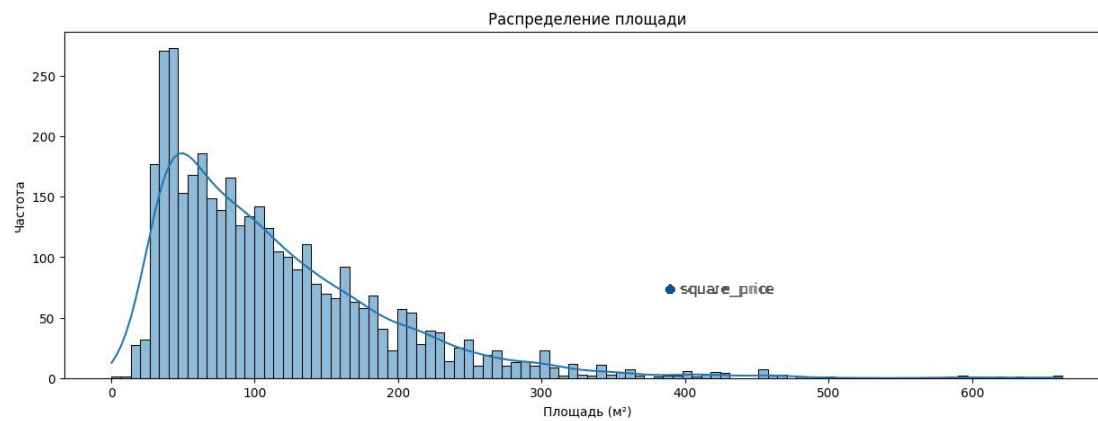
Тепловая карта корреляции показала следующие ключевые зависимости:

- **Цена за квадратный метр (`square_price`)** наиболее сильно коррелирует с общей ценой (`price`) и площадью квартиры (`area`).
- Количество комнат (`rooms`) показало слабую корреляцию с ценой за квадратный метр, что говорит о меньшем влиянии этого параметра на стоимость в сравнении с общей площадью и общей ценой.

#### 4.2 Обработка пропущенных значений

Использование [pandas](#) позволило корректно заполнить пропуски в колонке `subways`, основываясь на схожих значениях соседних объектов. Данный метод обеспечил более точное восстановление данных по сравнению с простыми статистическими методами (среднее, медиана и т.д.).

---



## 5. Рекомендации

1. **Использование обработанных данных для построения модели ценообразования:**
    - o Данные готовы для обучения модели машинного обучения, которая может предсказывать стоимость квартиры на основе признаков price, area, rooms, subways.
  2. **Регулярное обновление данных через API:**
    - o Для актуальности модели рекомендуется периодически обновлять данные через [API cian](#), чтобы учесть изменения на рынке недвижимости.
  3. **Дальнейший анализ категориальных переменных:**
    - o Рекомендуется детально изучить влияние других категориальных признаков, таких как renovation и placement\_type, которые могут оказать влияние на цену.
- 

## 6. Заключение

В ходе работы был проведен анализ и очистка данных о рынке недвижимости, полученных из [cian API](#). 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.

---

## Приложения

1. Графики зависимости и корреляции.
- 

Отчет составлен на основе данных, полученных из [cian API](#).