

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:
Студентка группы № ИСП-22
Новикова Наталья
Проверил:
Преподаватель информатики
Базяк Г.В. Коновалов И. В.

Долгопрудный, 2024

ВВЕДЕНИЕ

В данном отчете представлены результаты комплексного анализа рынка недвижимости в Московской области, Новой Москве и Москве. Целью исследования было выявление ключевых тенденций, факторов, влияющих на цены.

Цель:

Собрать и проанализировать данные для дальнейшего использования в разработке прогностических моделей, которая будет оценивать цену квадратного метра недвижимости в Московском регионе (Москва, Новая Москва, Московская область).

Задачи:

1. **Сбор данных:** Используя открытые источники, собрать обширный набор данных о квартирах.
2. **Очистка и подготовка данных:**
 - Удалить ненужные данные.
 - Дополнить недостающую информацию.
 - Выявить и устранить аномалии.
3. **Визуализация данных:**
 - Выявить взаимосвязи между данными и усреднённые показатели.

ОСНОВНАЯ ЧАСТЬ

Сбор данных:

В качестве сбора информации был выбран интернет-ресурс «Циан». С помощью скриптов на Python и библиотеки CianParser было собрано 8393 строки

Очистка и подготовка данных:

1. **Объединение данных:** Все собранные данные были объединены в одну таблицу с использованием библиотеки Pandas.
2. **Колонки с данными:** author, author_type, url, location, deal_type, accommodation_type и тд. (рис.1)
3. **Определение типов данных:** Определены типы данных в таблице (рис.2).
4. **Удаление пропущенных значений:** Значения -1, обозначающие отсутствие данных, были заменены на Nan и удалены. Было принято решение удалить столбцы с большим количеством пропусков и не особо важные для нас столбцы (рис.3 и 3.1). Также пустые данные в таких столбцах как метро, лоджии и парковка были заменены на “Нет балкона/лоджии/парковки” (рис.4).
5. **Форматирование данных:** Ненужные данные были отсеяны, а некоторые столбцы отформатированы для удобства анализа (рис.5).

Визуализация данных:

Для визуализации данных использовались библиотеки `matplotlib`. Были созданы следующие графики:

1. Средняя цена за квартиру по городу (рис.6).
2. Цена за м² в зависимости от года постройки (рис.7).
3. Цена за м² по наличию парковки (рис.8).
4. график корреляций (рис.9).

Аналитика данных:

Анализ графиков позволил сделать следующие выводы:

Основные факторы, влияющие на цену квартиры: город, наличие парковки, площадь квартиры и год постройки.

Заключение

В результате проделанной работы были:

- Собраны, очищены и подготовлены данные о квартирах.
- Созданы удобные для анализа визуализации.
- Выявлены ключевые факторы, влияющие на стоимость недвижимости в Московской области, Москве и Новой Москве.

Полученные данные и выводы могут быть использованы для разработки прогностических моделей, помогающих оценивать стоимость квартир.

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

Рис.1

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 3150 entries, 1 to 8391
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype  
---  -
0   url                    3150 non-null   object  
1   location               3150 non-null   object  
2   floor                  3150 non-null   object  
3   floors_count           3150 non-null   object  
4   rooms_count            3150 non-null   int64   
5   total_meters           3150 non-null   object  
6   price                  3150 non-null   float64  
7   year_of_construction   3150 non-null   int64   
8   object_type            3150 non-null   object  
9   have_loggia            3150 non-null   object  
10  parking_type           3150 non-null   object  
11  living_meters           3150 non-null   float64  
12  kitchen_meters         3150 non-null   float64  
13  ceiling_height         3150 non-null   float64  
14  district               3150 non-null   object  
15  street                 3150 non-null   object  
16  underground            3150 non-null   object  
dtypes: float64(4), int64(2), object(11)
memory usage: 443.0+ KB
```

Рис. 2

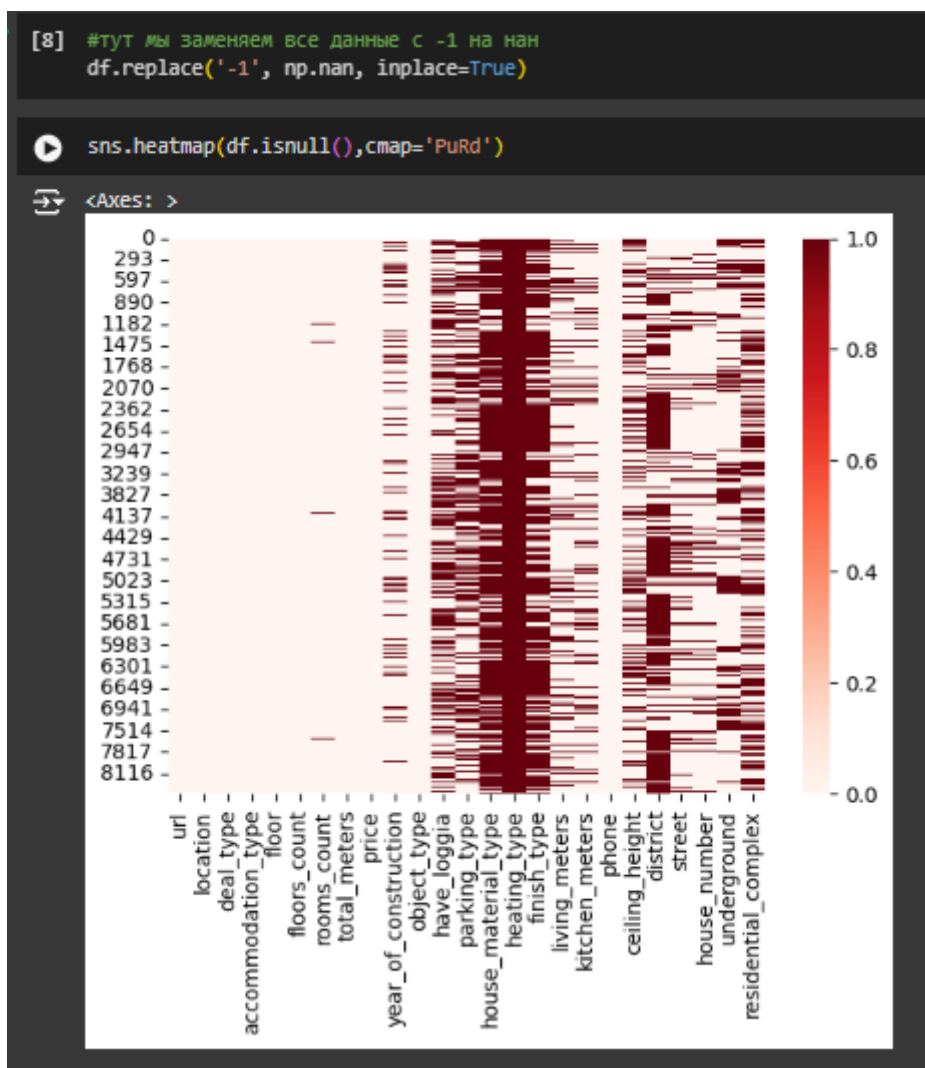


Рис.3

✓ **Удаляем и заменяем пустые значения**

Удаляем колонки в которых слишком много пропусков
heating_type, house_material_type, finish_type

Так же удаляем колонки которые нам не понадобятся в будущем
phone, deal_type, accommodation_type, house_number, residential_complex

Ещё мы видим, что в лодже, метро и парковке много пустых значений, но это просто их отсутствие

Заменяем нан на нет метро, парковки и лоджи

```
df = df.drop(['heating_type', 'house_material_type', 'finish_type'], axis=1)
df = df.drop(['phone', 'deal_type', 'accommodation_type', 'house_number', 'residential_complex'], axis=1)
```

Рис.3.1

```
11] df['have_loggia'] = df['have_loggia'].fillna('нет лоджи')
df['parking_type'] = df['parking_type'].fillna('нет парковки')
df['underground'] = df['underground'].fillna('нет метро')
```

Рис.4

```
<class 'pandas.core.frame.DataFrame'>
Index: 2820 entries, 1 to 8391
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  -
0   location              2820 non-null   object
1   floor                 2820 non-null   int64
2   floors_count          2820 non-null   int64
3   rooms_count           2820 non-null   int64
4   total_meters          2820 non-null   float64
5   price                 2820 non-null   float64
6   year_of_construction  2820 non-null   int64
7   object_type           2820 non-null   int64
8   have_loggia           2820 non-null   object
9   parking_type          2820 non-null   object
10  living_meters          2820 non-null   float64
11  kitchen_meters         2820 non-null   float64
12  ceiling_height         2820 non-null   float64
13  district              2820 non-null   int64
14  street                2820 non-null   int64
15  underground            2820 non-null   int64
16  square_price           2820 non-null   int64
dtypes: float64(5), int64(9), object(3)
memory usage: 396.6+ KB
```

Рис.5

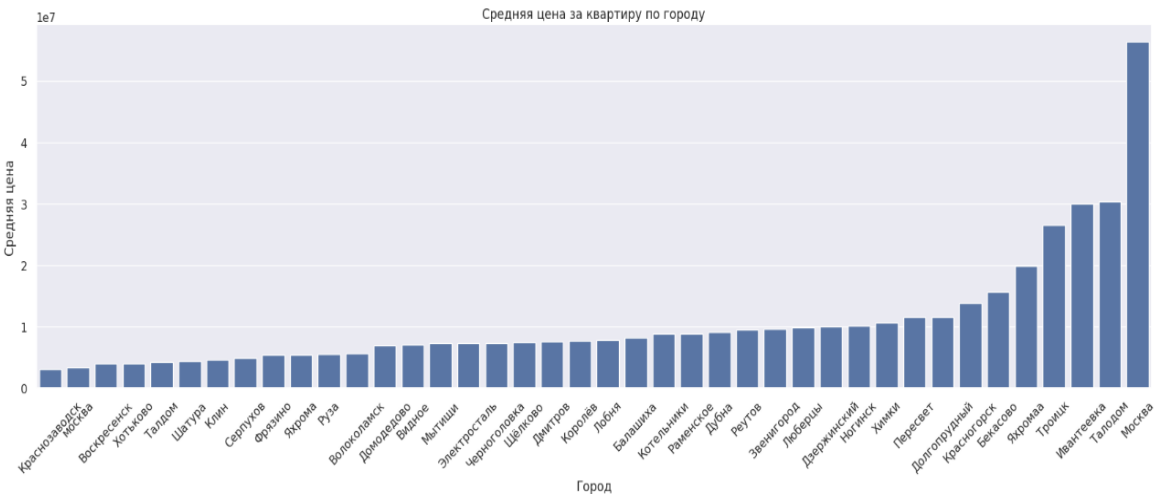


Рис.6



Рис.7

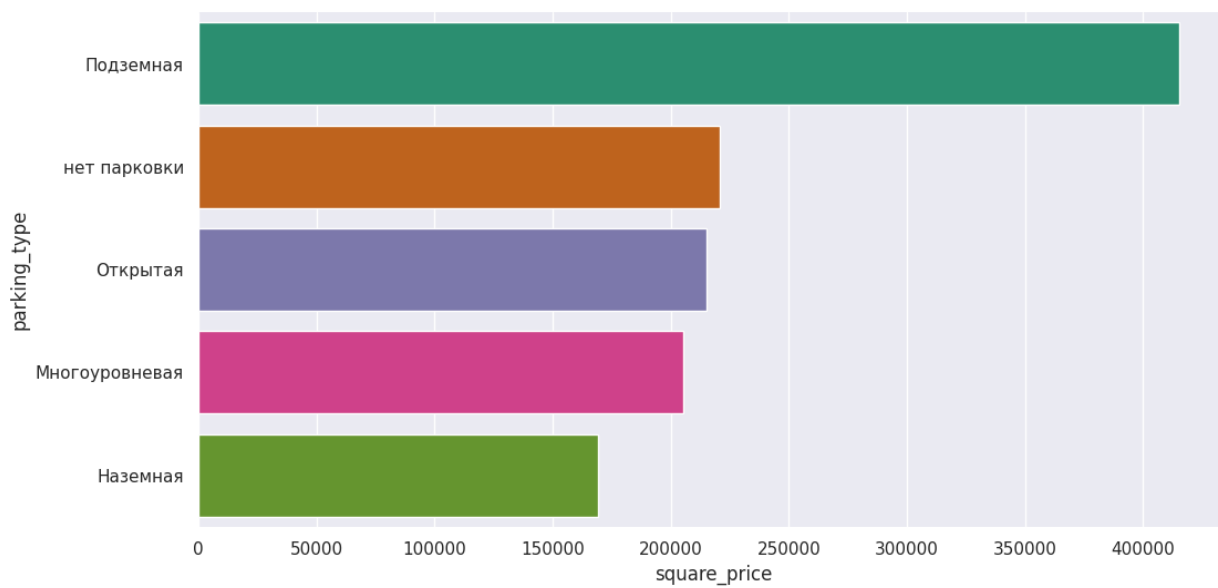


Рис.8

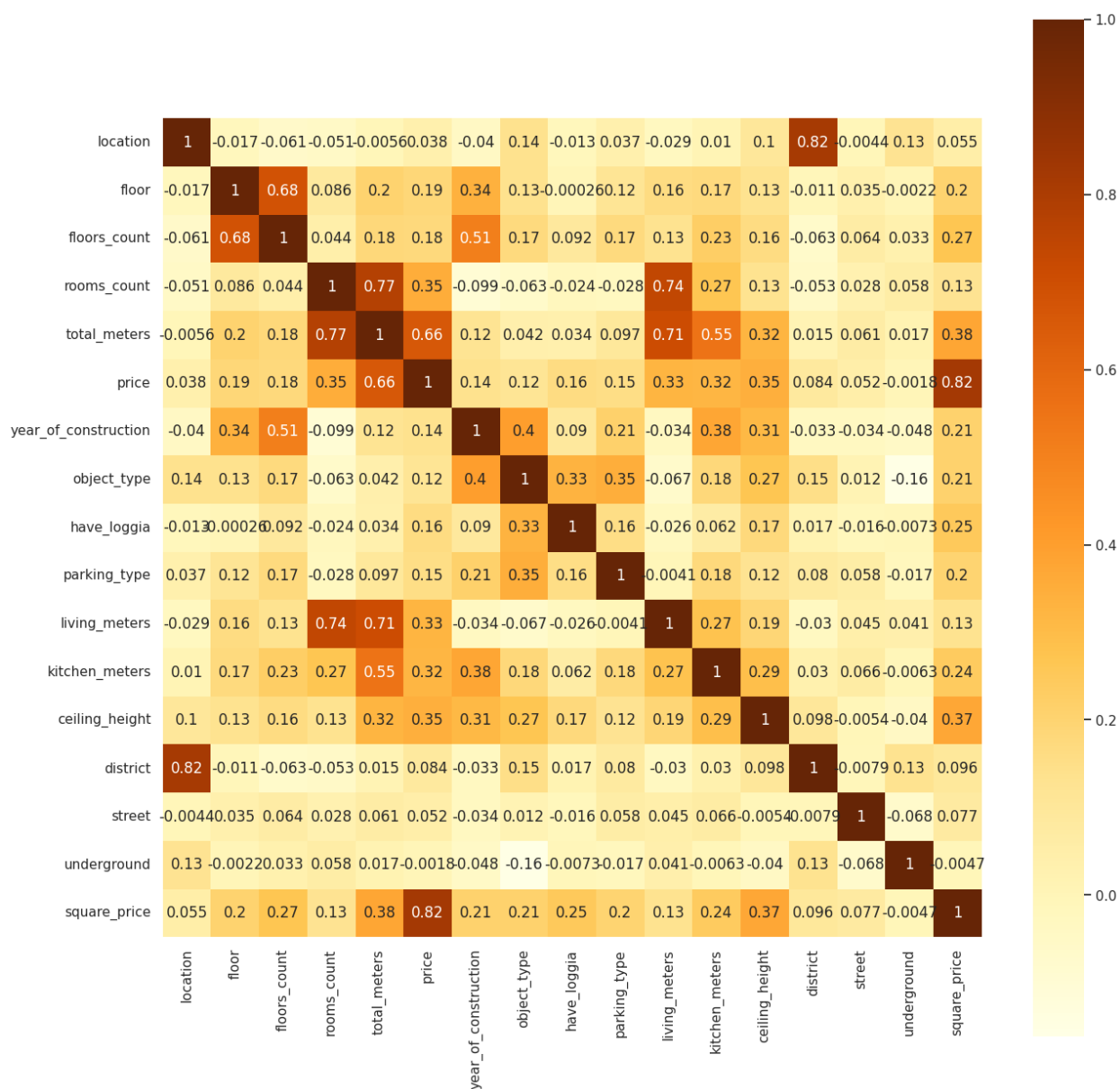


Рис.9