

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

Отчёт по кейсу «Самолёт»

Работу выполнил:
Студент группы № ИСП-22
Рыжков Артём Русланович

Долгопрудный, 2024

Введение:

В данном отчёте рассматриваются выводы, полученные с аналитической работы над данными в области «Квартиры в Московской области, и Москве».

Цель:

Собрать данные и произвести аналитическую работу над ними для будущих работ, например, создание модели на основе выводов.

Задачи:

- Используя открытые источники собрать список данных.
- На основе полученной информации произвести удаление ненужных данных, дополнение необходимых, выявление аномалий и их отчистка.
- Визуализация данных при помощи, инструментов для подобных задач. Нахождение взаимосвязей между данными или их полное отсутствие, усреднённых показателей для уверенного отчёта.

Процесс проделанной работы:

Для выполнения основной задачи, существует небольшой выбор источников, откуда собирать данные, при помощи скриптов, написанных на языке Python, было получено около десяти тысяч объявлений в нужных регионах (Рисунок 1-2).

```
import cianparser
parser = cianparser.CianParser(location="Одинцово")# здесь указывается город для сбора информации
# переменная sale не менялась нам нужен только такой тип
data = parser.get_flats(deal_type="sale", rooms=(1), additional_settings={"start_page":1, "end_page": 54}, with_extra_data=True, with_saving_csv=True)
#rooms=1) вместо 1 можно указать количество комнат для сбора
```

Рисунок 1

| | author | author_type | url | location | deal_type | accommodation_type | floor | rooms_count | rooms_count | total_meters | ... | block_type | living_meters | kitchen_meters | phone | ceiling_height | district | street | house_number | underground | residential_complex | |
|---|------------------|-------------------|---|----------|-----------|--------------------|-------|-------------|-------------|--------------|-----|-------------|---------------|----------------|-------------|----------------------|-----------------|------------------|--------------|------------------|---------------------|------------|
| 0 | ANT Development | developer | https://www.cian.ru/sale/flat/303519396/ | Moscow | sale | flat | 11 | 12 | 5 | 265.8 | - | Частная | -1 | | 74951146288 | 3 м | Дорогоминовское | Nah | 3a1 | Парк Победы | Восточный Район | Решение XX |
| 1 | ID 18178647 | real_estate_agent | https://www.cian.ru/sale/flat/301430189/ | Moscow | sale | flat | 26 | 31 | 5 | 246.7 | - | 1482.9 м² | 28 м² | 79166462390 | 3.1 м | Очаковское-Митинское | Нежинская | 3a1 | Дачинское | Кузьминская Роща | | |
| 2 | Monumental Group | real_estate_agent | https://www.cian.ru/sale/flat/306031323/ | Moscow | sale | flat | 4 | 8 | 5 | 117.0 | - | -1 | 10 м² | 79672096336 | 3.2 м | Пресненский | Большая Садовая | 3C3 | Маяковская | | Nah | |
| 3 | Галснорд | developer | https://www.cian.ru/sale/flat/296031233/ | Moscow | sale | flat | 24 | 24 | 5 | 172.5 | - | Без отделки | 104.6 м² | 103.5 м² | 74951379308 | 3.2 м | Финансовый парк | Береговой проезд | 2 | Вилла | Береговой-2 | |
| 4 | Ремис | real_estate_agent | https://www.cian.ru/sale/flat/307903940/ | Moscow | sale | flat | 3 | 9 | 5 | 234.8 | - | -1 | 140 м² | 23 м² | 79663231016 | -1 | Тверской | Краснопроектная | 7 | Новослободская | Ласточкино гнездо | |

5 rows x 27 columns

Рисунок 2

Далее мы убираем часть лишней информации из наших данных. Затем выводим количество отсутствующей информации визуально (Рисунок 3).

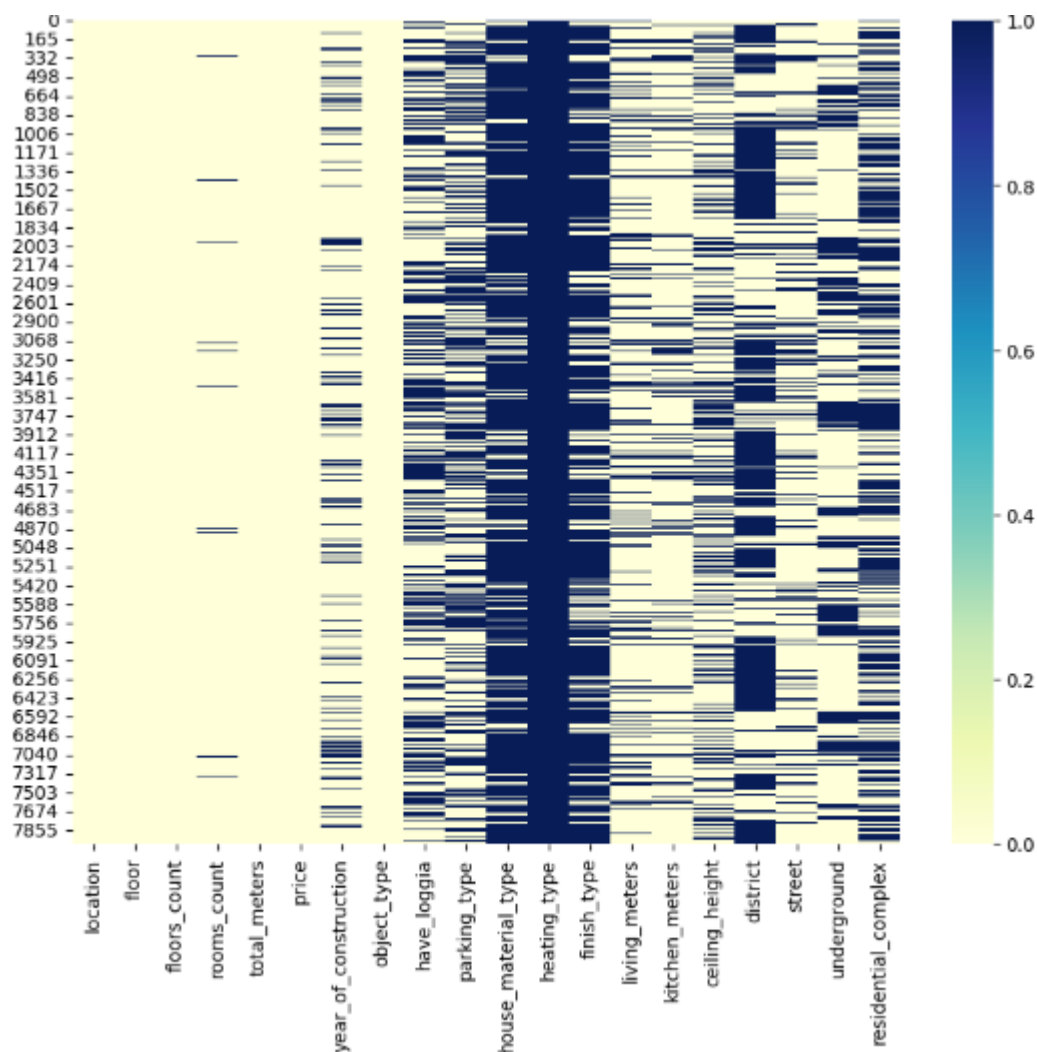


Рисунок 3

После этого убираем полностью пустые строки, повторяющуюся информацию и заполняем пустые значения 0, а затем выводим количество пустот в столбцах (Рисунок 4-5).

```

1 null_counts = df.isnull().sum().reset_index()
2 null_counts.columns = ['Column', 'Null Count']
3
4 print(null_counts)

```

| | Column | Null Count |
|----|----------------------|------------|
| 0 | location | 0 |
| 1 | floor | 0 |
| 2 | floors_count | 0 |
| 3 | rooms_count | 0 |
| 4 | total_meters | 0 |
| 5 | price | 0 |
| 6 | year_of_construction | 0 |
| 7 | object_type | 0 |
| 8 | have_loggia | 0 |
| 9 | parking_type | 0 |
| 10 | living_meters | 0 |
| 11 | kitchen_meters | 0 |
| 12 | ceiling_height | 0 |
| 13 | street | 0 |
| 14 | underground | 0 |

Рисунок 4

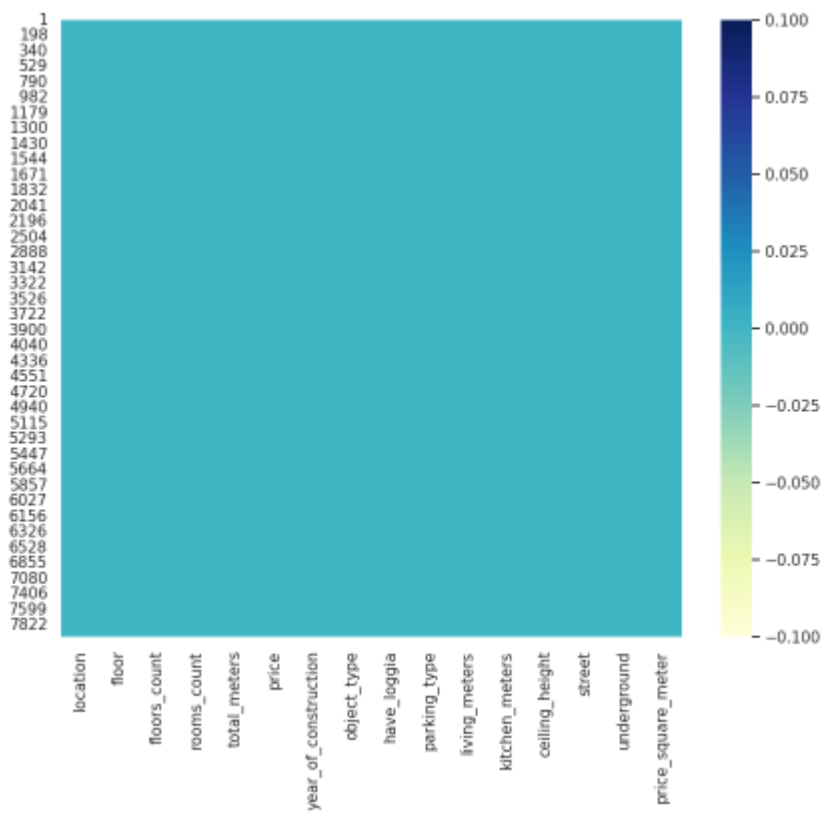


Рисунок 5

Меняем все возможные колонки с типа object на int или float и удаляем аномалии (Рисунок 6).

```
df.info()

<class 'pandas.core.frame.DataFrame'>
Index: 3972 entries, 0 to 7977
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   location               3972 non-null   int64
1   floor                  3972 non-null   int64
2   floors_count           3972 non-null   int64
3   rooms_count            3972 non-null   int64
4   total_meters            3972 non-null   float64
5   price                  3972 non-null   int64
6   year_of_construction    3972 non-null   float64
7   object_type             3972 non-null   int64
8   have_loggia             3972 non-null   int64
9   parking_type            3972 non-null   int64
10  living_meters           3147 non-null   float64
11  kitchen_meters          3491 non-null   float64
12  ceiling_height           3972 non-null   float64
13  street                  3972 non-null   int64
14  underground             3972 non-null   int64
dtypes: float64(5), int64(10)
memory usage: 496.5 KB
```

Рисунок 6

Окончательная часть создание матрицы корреляции (Рисунок 7)

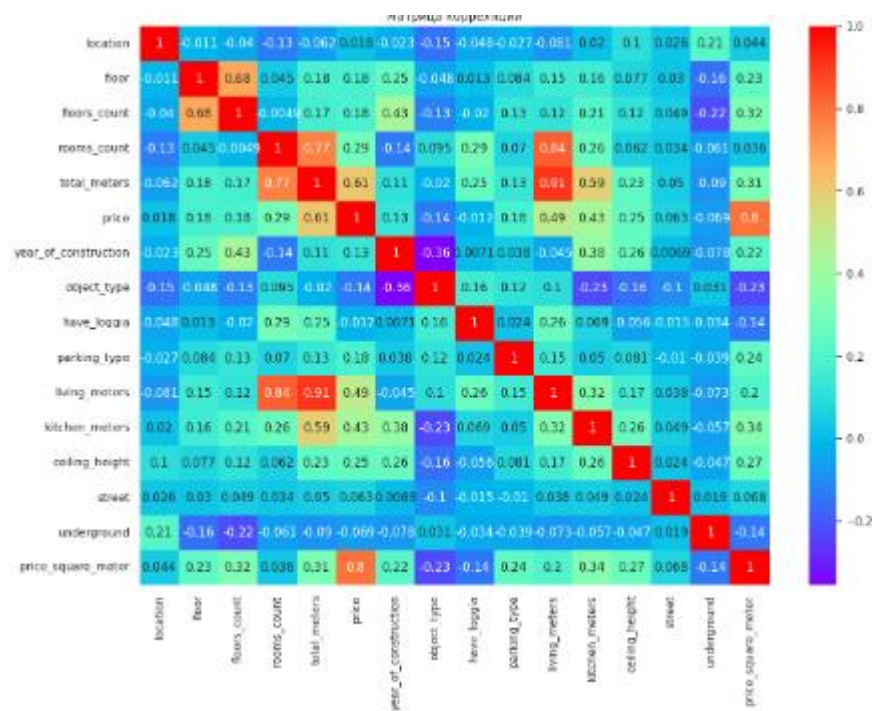


Рисунок 7

Аналитика данных:

Благодаря выведенным графикам, можно сделать выводы, что цена в основном зависит от типа отделки, материала дома, города. От года постройки зданий зависит лишь их количество на рынке и количество комнат.

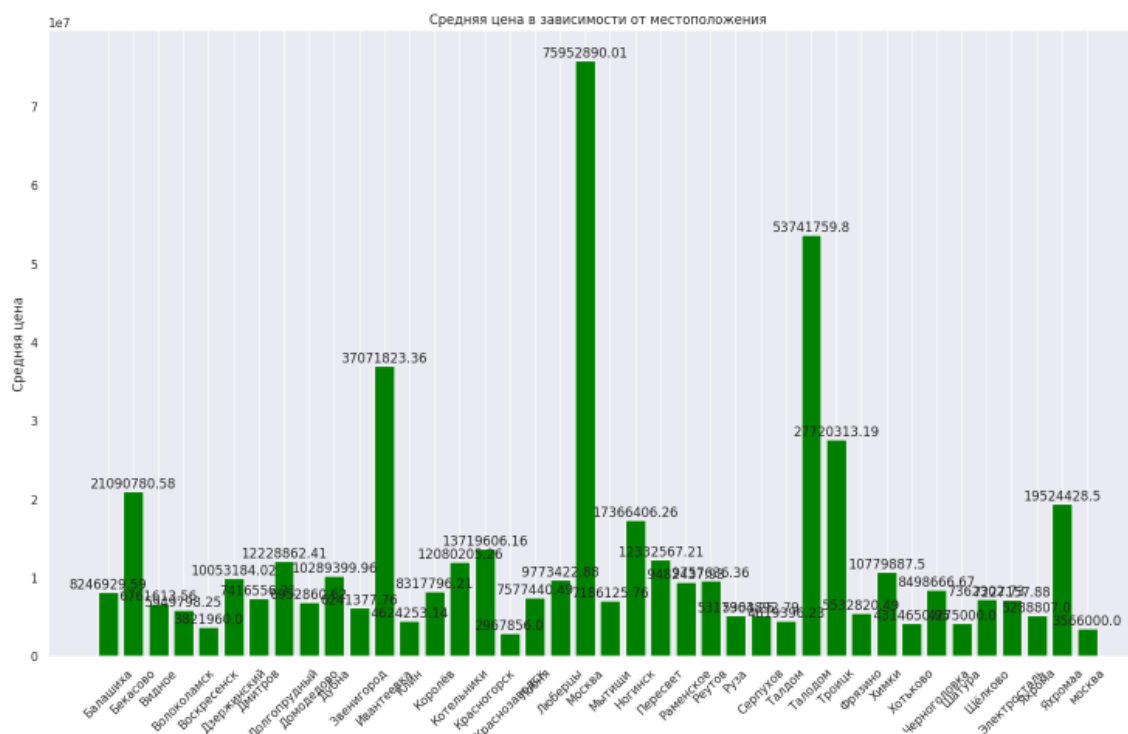


Рисунок 8



Рисунок 9

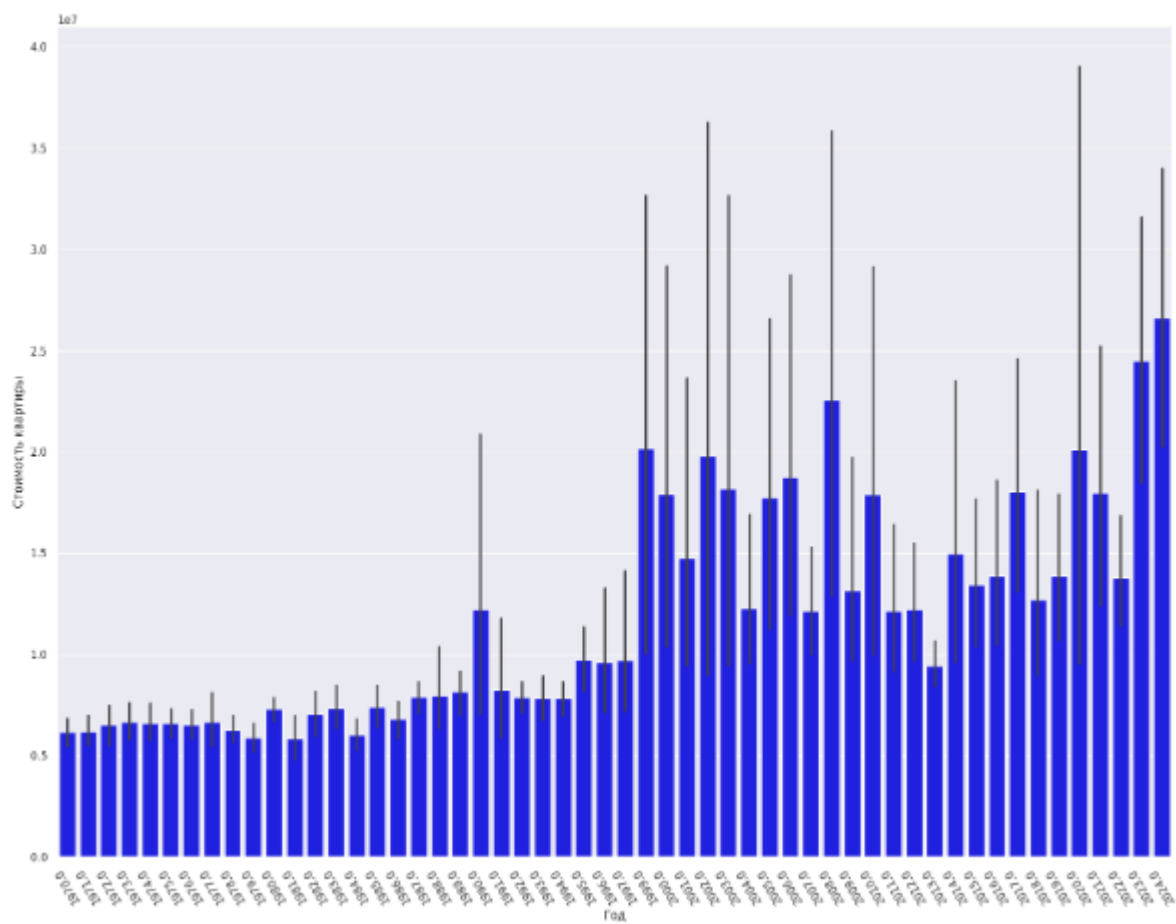


Рисунок 10

Заключение:

В результате аналитической работы были собраны, отсортированы, почищены данные, построены удобные для анализа графики, благодаря которым получилось выявить не маловажные критерии в оценивании стоимости недвижимости в Московской Области и Москве. Основными факторами, оказывающими влияние на стоимость, выявились общая площадь, жилая площадь, высоты потолков и количества комнат. Полученные данные могут быть использованы для дальнейшей разработки прогностических моделей.