Государственное бюджетное профессиональное

образовательное учреждение Московской области

«Физико-технический колледж»

**Отчёт по кейсу «Самолёт»:**

Работу выполнил:

Студент группы № ИСП-22

Серый Александр

**Введение:**

В данном отчете рассматриваются выводы по первому интенсиву по сбору и анализу данных о продаваемых квартирах в Москве и Московской области.  
Для выполнения работы использовался язык програмирования Python.

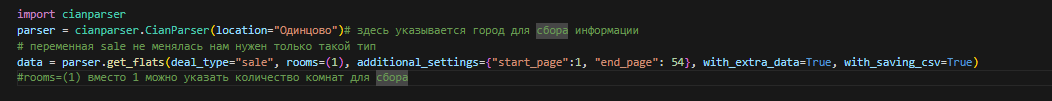
**Цель:**

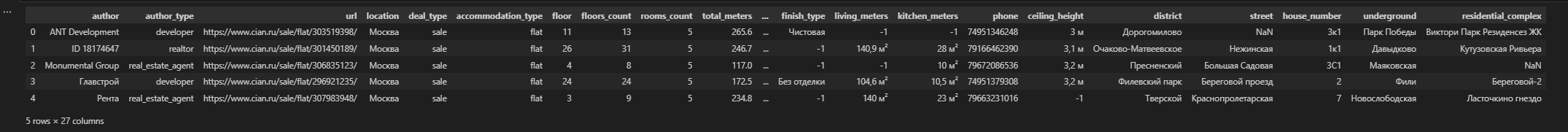
Собрать данные о продаваемых квартирах в Москве и Московской области, провести работу над ними, проанализировать и заполнить пропуски для дальнейших задач.

**Задачи:**

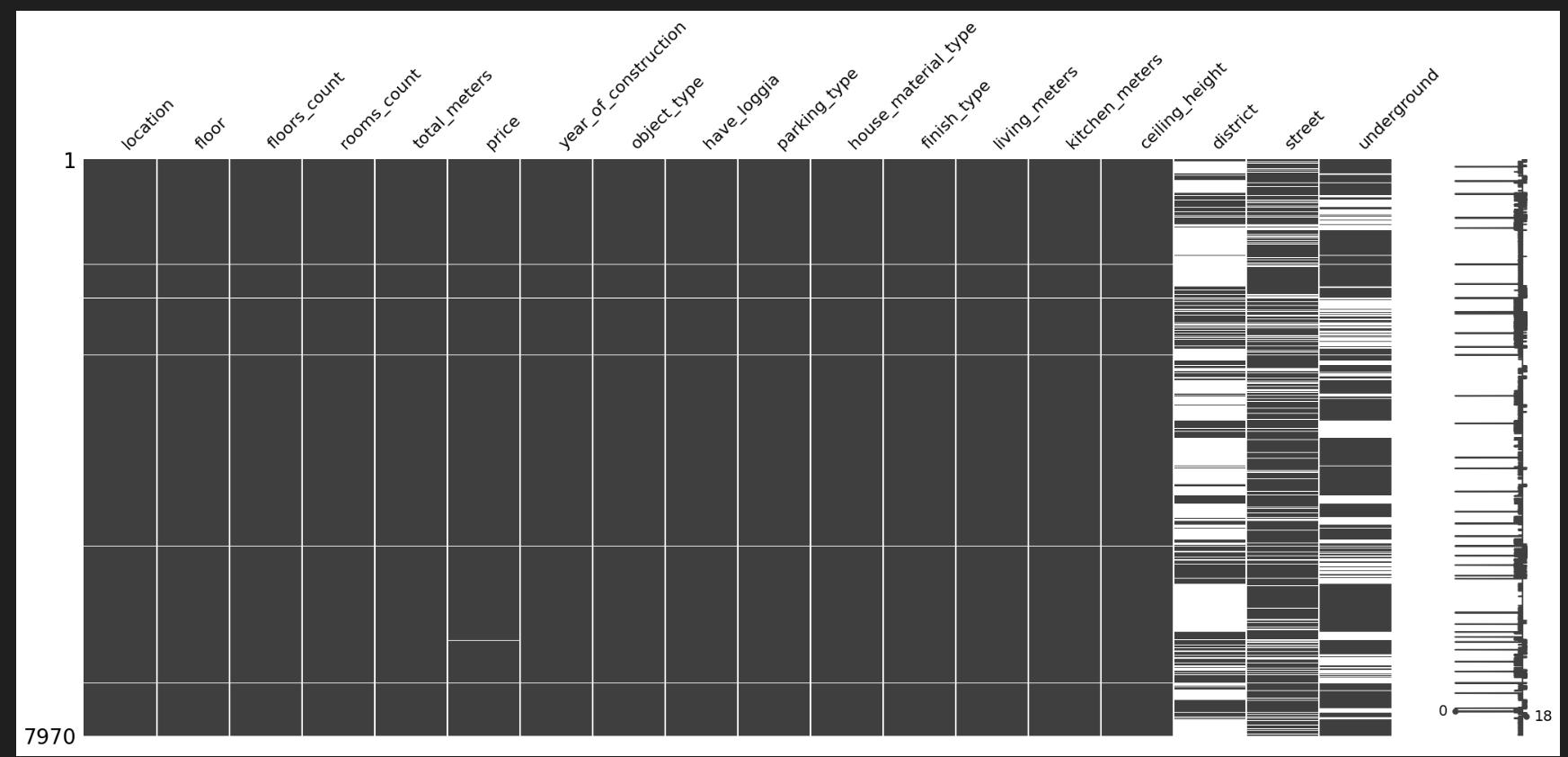
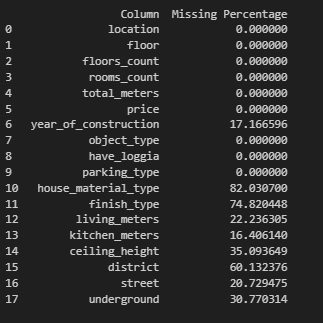
1. Собрать данные используя открытые источники, например такие как Циан или ДомКлик
2. Обработка собранной информации, очистка от ненужной или неверной информации
   1. Визуализация в ходе выполнения обработки информации.

**Ход выполнения работы.**

Первый этап — это сбор данных, который был выполнен с помощью языка программирования Python и библиотеки CianParser. В процессе работы пришлось изменить некоторые части кода библиотеки для более корректного вывода и уменьшения количества пропусков при сборе информации. Сбор данных производился по городам Москвы и Московской области с указанием города, типа продажи и количества комнат.

Второй этап — это обработка и анализ собранных данных. Вначале мы смотрим, все ли мы верно сделали и данные отображаются.  


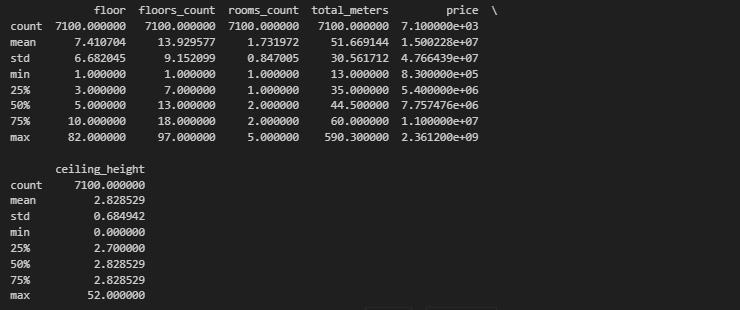
Далее мы убираем часть лишней информации из наших данных. Затем выводим количество отсутствующей информации визуально

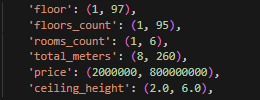


После этого убираем полностью пустые строки и повторяющуюся информацию, а затем выводим количество пустот в столбцах

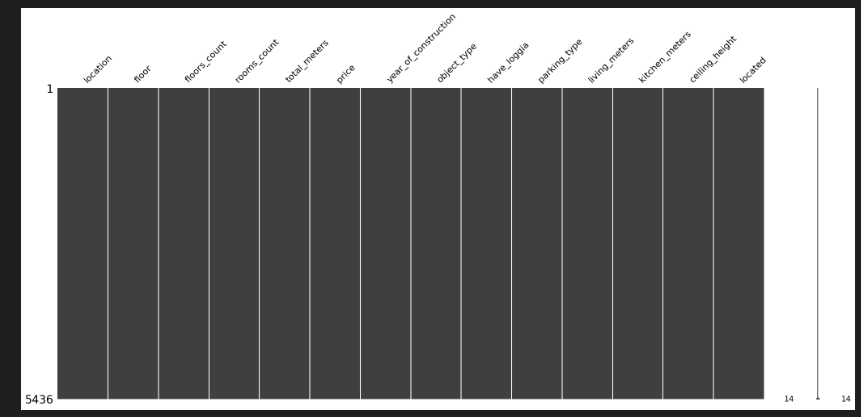
После этого удаляем столбцы, в которых более 70% значений являются пропусками.

Далее убираем еще в некоторых столбцах пропуски заполнением средним значение где это возможно

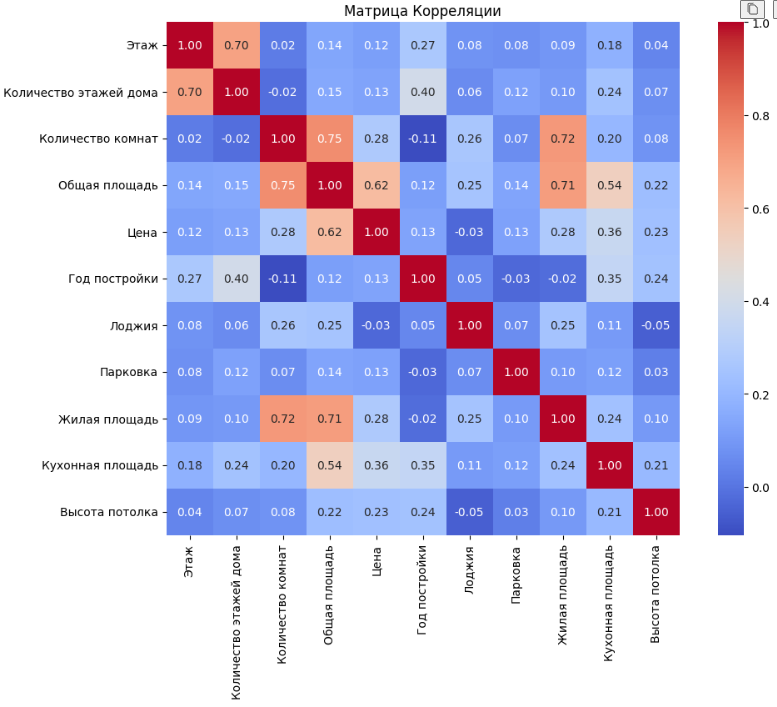
Выводит статистическую сводку числовых данных

После чего ставим ограничения, чтобы не было недостоверной информации

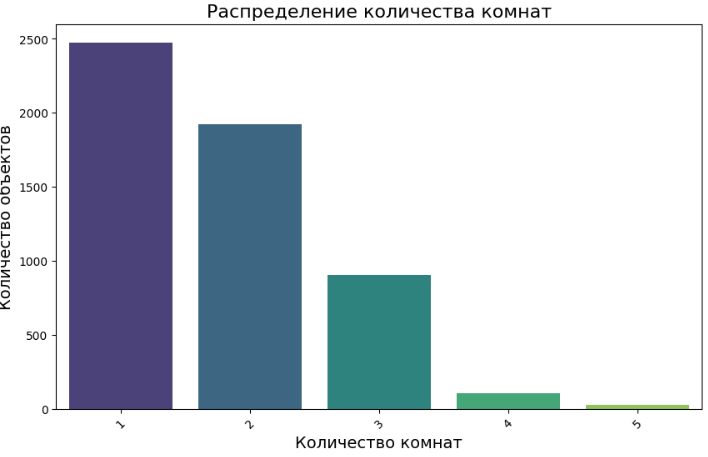
Далее удаляются все возможные пропуски, и выводится визуально для проверки.



Окончательная часть включает просмотр зависимостей и исправление оставшейся некорректной информации.

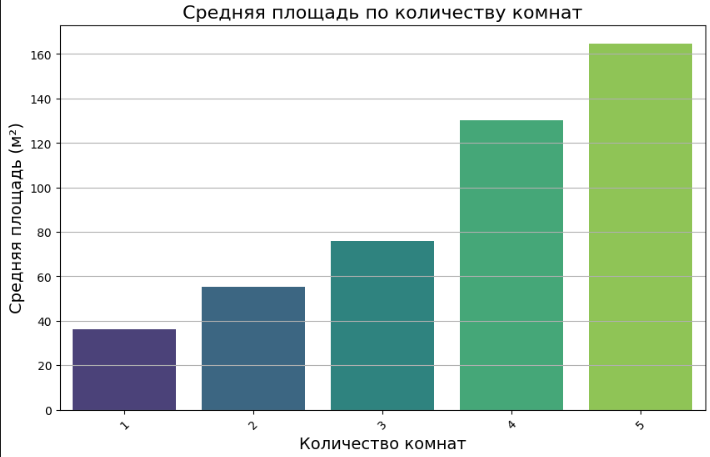
Вывод таблицы зависимостей

Отображение зависимости цены от года

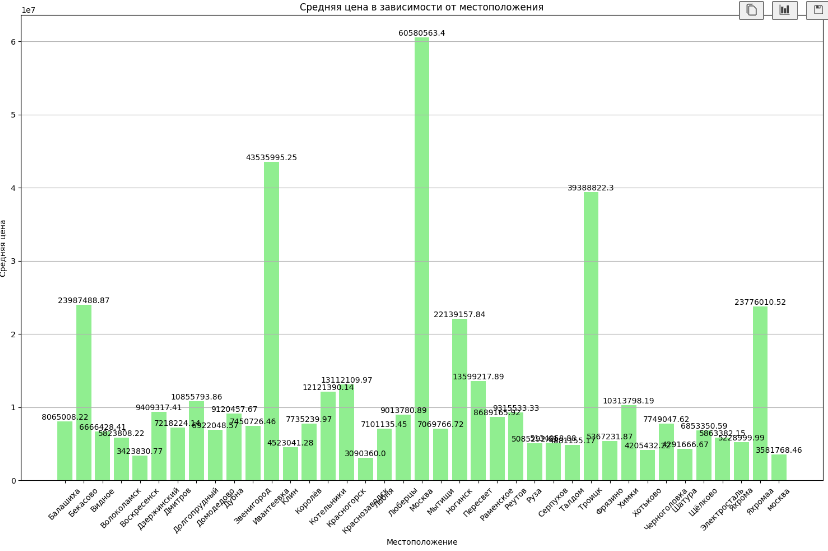


Количество информации после обработки по количеству комнат

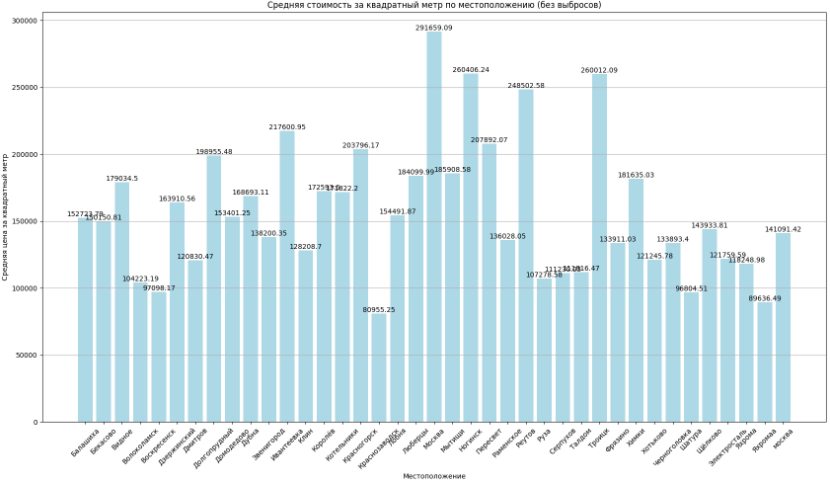
Зависимость площади от количества комнат.

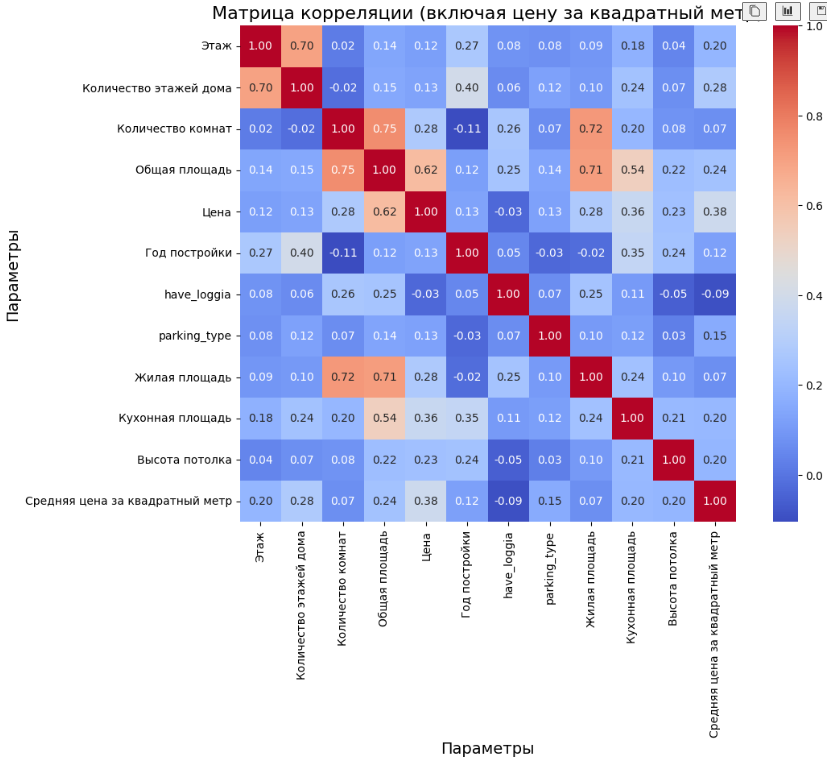


Средняя цена в (рублях) зависамости от города



Средняя цена на мтрах квадратный в разных городах

Также отображение зависимостей с стоимостью за квадратный метр.



Ссылки на источники сбора информации для проекта.

<https://www.cian.ru>

Ссылки на сам проект.

Выполнявший работу <https://github.com/Sr123Saha>

Ссылка на выполненную работу https://github.com/Sr123Saha/1intensive1

ссылка на работу в googl colab <https://colab.research.google.com/drive/1z1g0W5xrtg4gZ5MXBhbfsrJbOrVXsrB6>