Государственное бюджетное профессиональное

образовательное учреждение Московской области

«Физико-технический колледж»

**Аналитический отчёт**

Работу выполнила:

Студент группы № ИСП-21

Колпакова Ангелина

Долгопрудный, 2024

**Оглавление**

[Введение 3](#_Toc181362877)

[Основная часть 3](#_Toc181362878)

[Заключение 8](#_Toc181362879)

# **Введение**

Данный аналитический отчет фокусируется на парсинге и анализе данных о квартирах в Москве и Московской области.

**Цель:** Собрать и проанализировать данные для разработки модели, основанную на анализе влияния различных факторов.

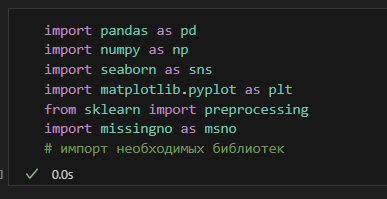
**Задачи:**

* Изучение сайта Циан для выявления ключевых параметров, влияющих на стоимость жилья.
* Парсинг данных.
* Произвести подготовку данных для анализа.
* Проанализировать и визуализировать взаимосвязь между параметрами и ценой.

# **Основная часть**

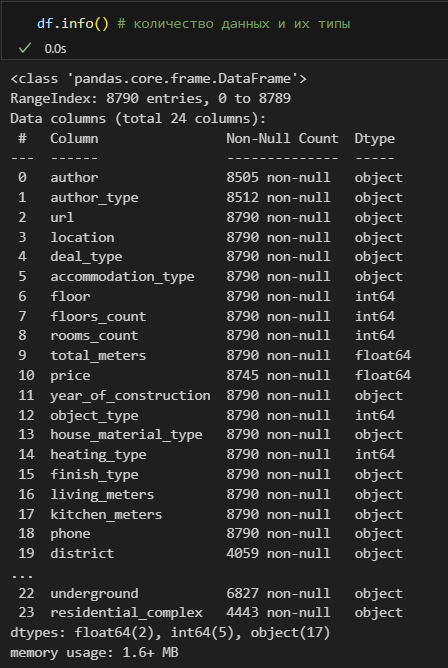
Для выполнения задачи сбора данных был выбран сайт «Циан». С помощью библиотеки cianparser собиралась информация о кввартирах. В результате парсинга был получен набор данных, содержащий информацию о 8791 квартире из Москвы и Московской области.

Для выполнения задачи подготовки данных для анализа я использовала несколько библиотек:

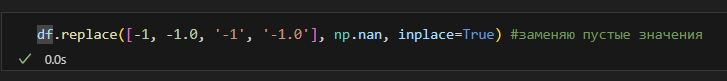


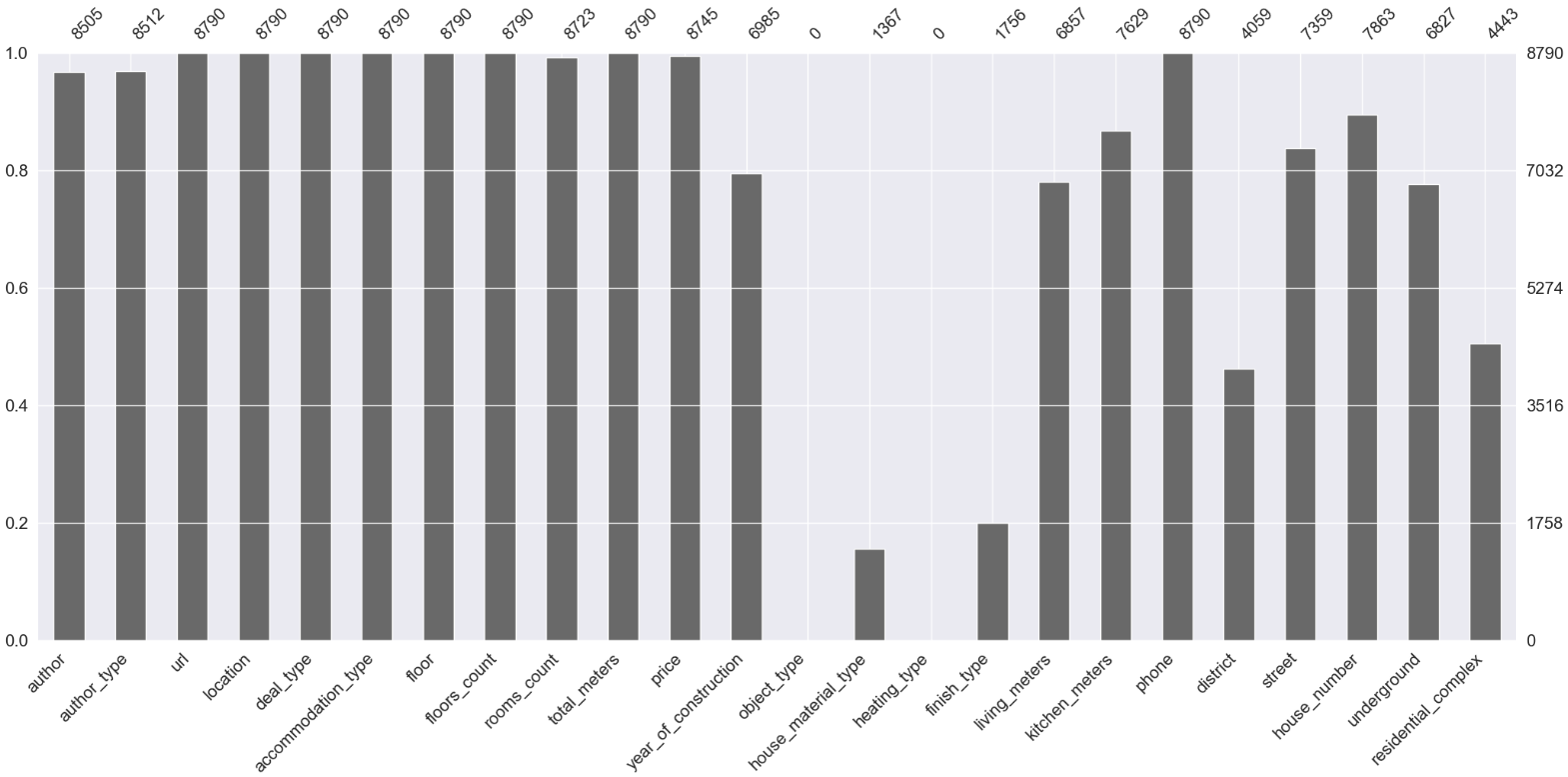
Собранные данные были объединены в единую таблицу с помощью библиотеки Pandas. После этого была проведена проверка на наличие дубликатов, но дубликаты в полученном наборе данных отсутствовали.

В начале я вывела какие данные у меня есть всего и какого они типа.

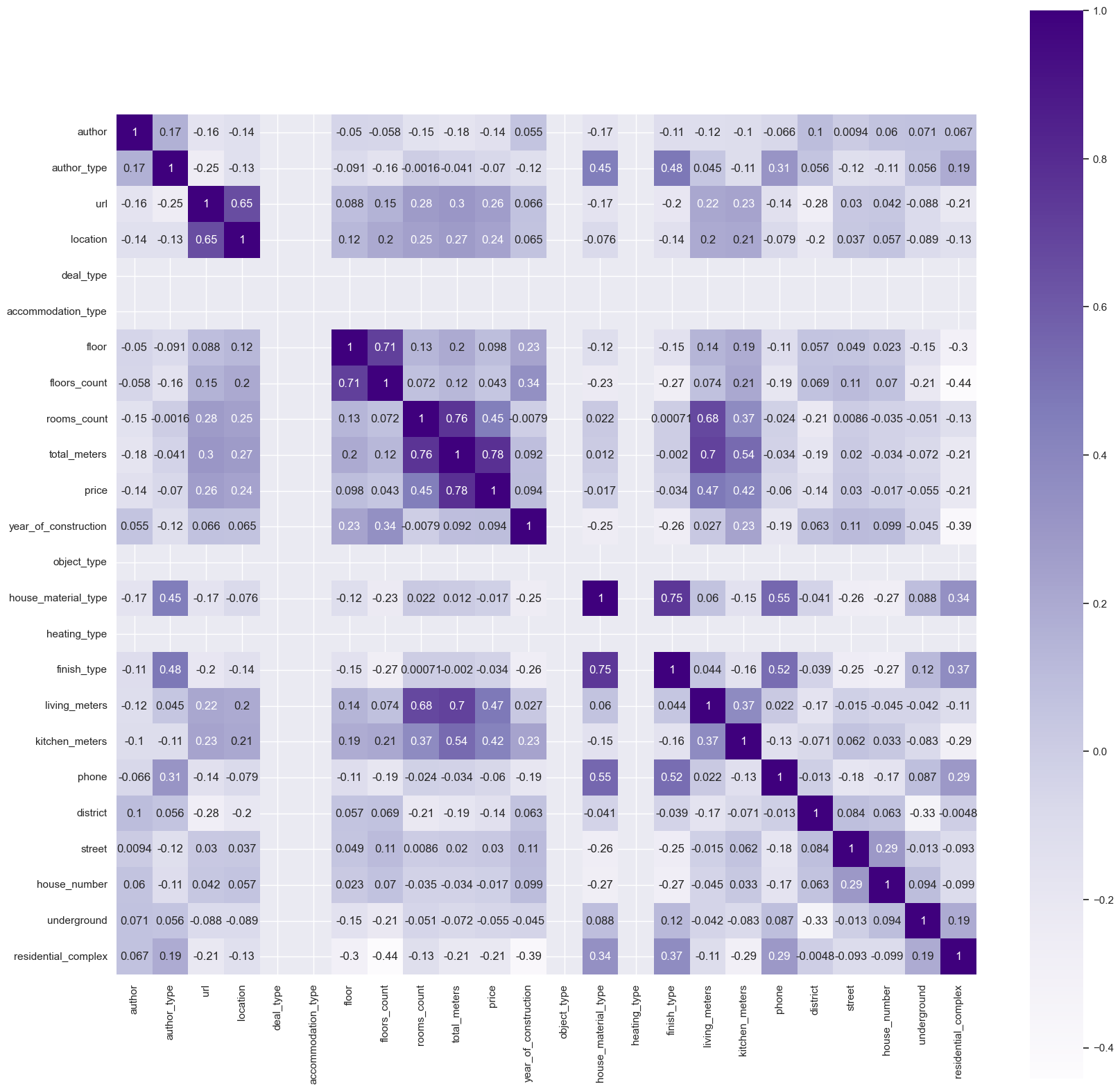


Заменяю пустые значения, чтобы узнать количество пропусков

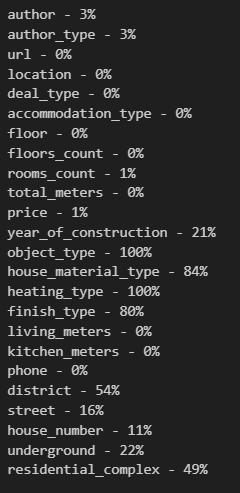




Можем заметить, что две колонки оказались полностью незаполненными, а еще некоторые частично пустыми.

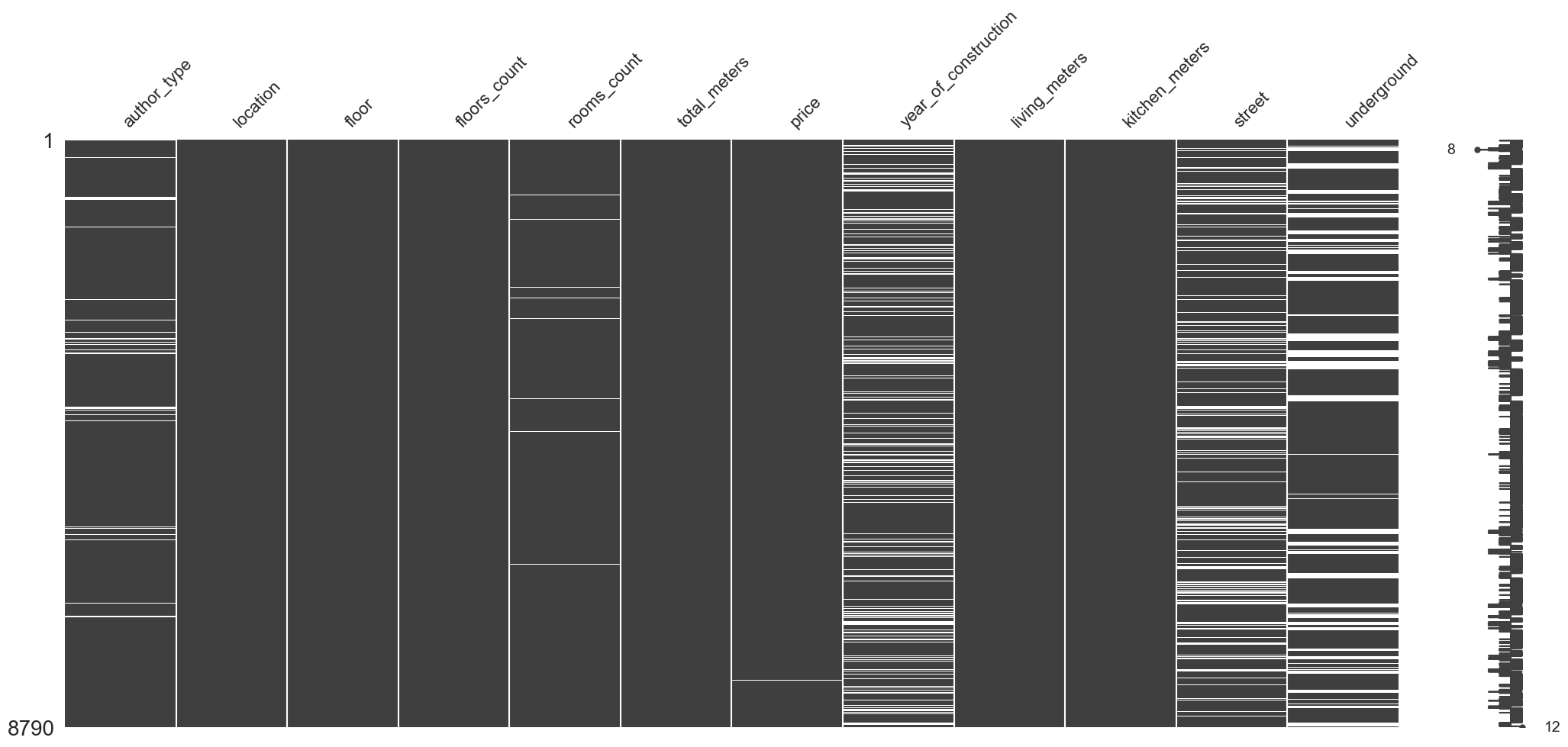
Затем вывожу тепловую карту и смотрю какие столбцы или строки удалить

Далее смотрим на количество пропущенных данных в процентах:

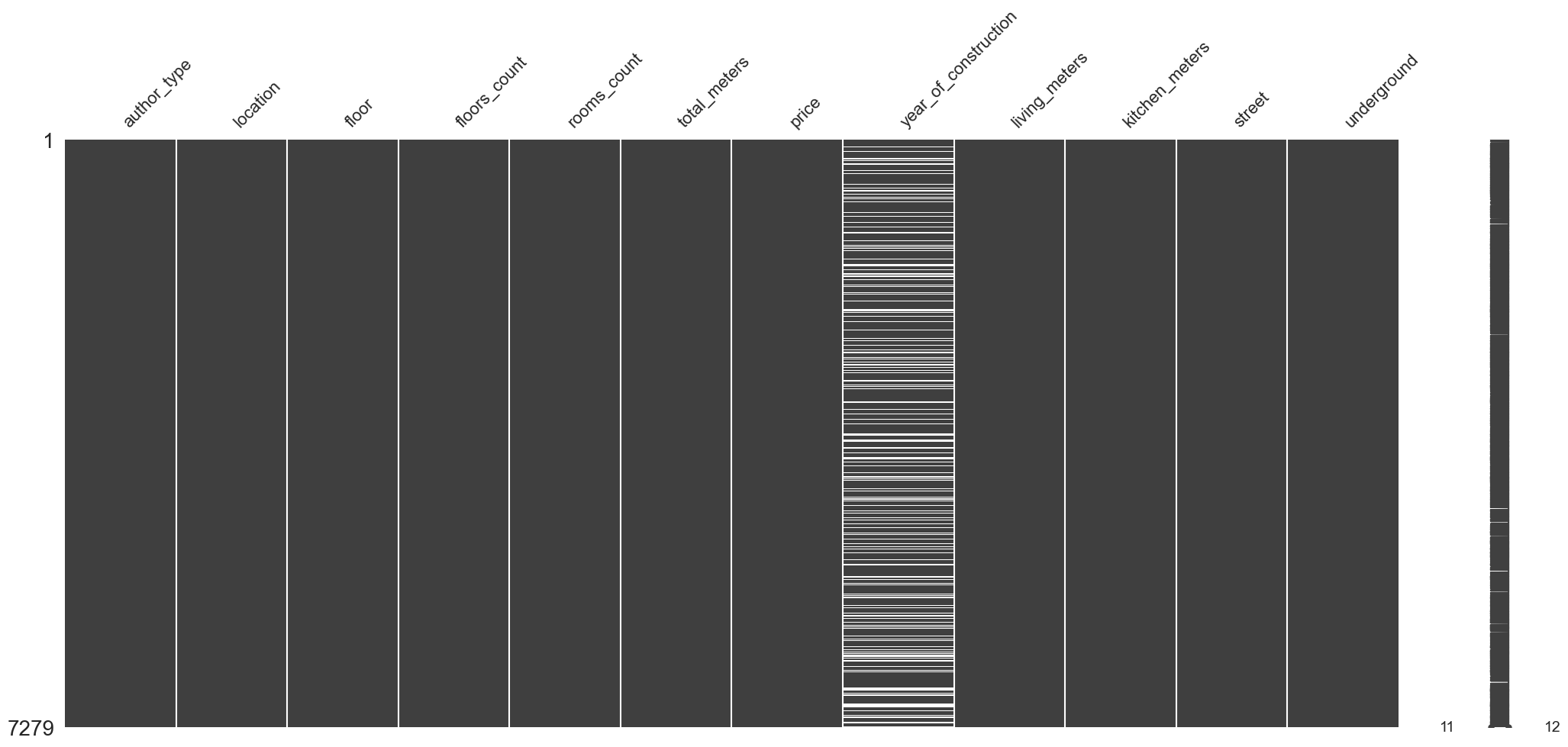


Затем я удаляю ненужные колонки (пустые, с малым количеством данных, с уникальными значениями и то, что не влияет на цену).

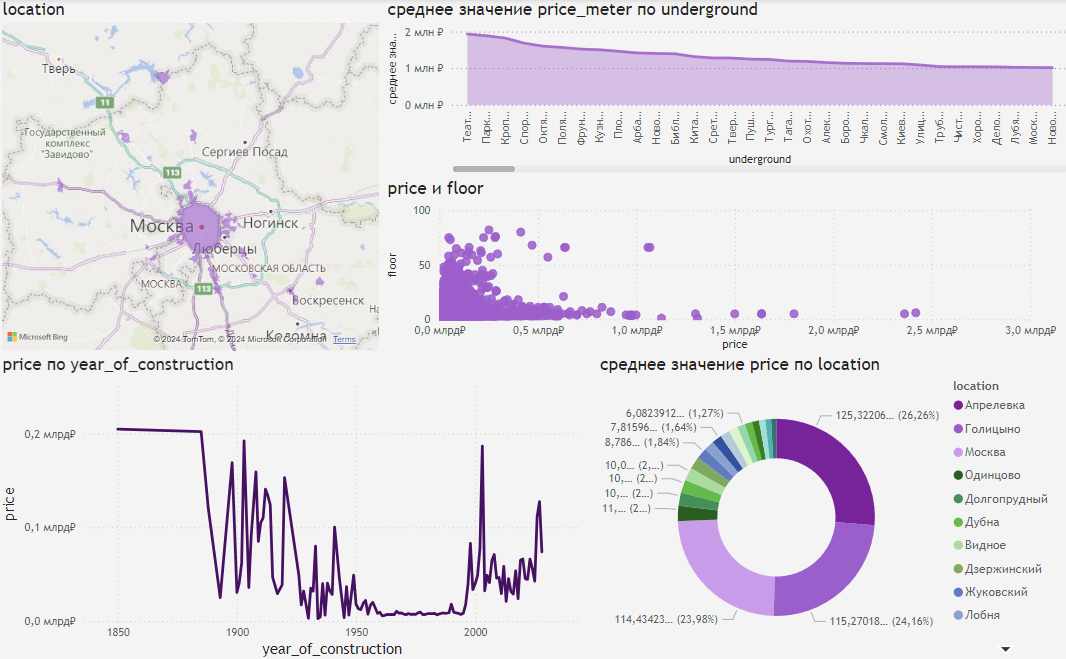
Далее я заполняю пропуски и удаляю строки



living\_meters и kitchen\_meters – заполняю средним значением; author\_type – заполняю модой; price, rooms\_count и street – удаляю строки; underground – заполняю городом.



Далее я продолжила анализ в Power BI.



1. карта, на которой отмечены города, которые использовались в ходе работы.

2. соотношение средней цены за кв. метр и метро.

3. распределение цен на квартиры в зависимости от этажа.

4. соотношение цены и года, когда было построено здание.

5. среднее значение цен по городу.

# **Заключение**

Анализ полученных данных показал, что на стоимость квадратного метра жилья в Московском регионе оказывают влияние такие факторы, как расположение (локация), количество комнат, жилая площадь квартиры, а также близость к станции общественного транспорта.