Государственное бюджетное профессиональное

образовательное учреждение Московской области

«Физико-технический колледж»

**Отчёт по кейсу «Самолёт»:**

Работу выполнил:

Студент группы № ИСП-21

Городецкая Анастасия

Долгопрудный, 2024

Оглавление

[**Введение** 3](#_Toc181281206)

[**Анализ рынка недвижимости Москвы и Подмосковья** 3](#_Toc181281207)

[**Заключение** 5](#_Toc181281208)

[**Приложения** 6](#_Toc181281209)

# **Введение**

Данный аналитический отчет посвящен парсингу и анализу данных о квартирах в Москве и Подмосковье. В процессе исследования были собраны сведения о ценах, площадях, расположении объектов, а также других ключевых характеристиках. С помощью методов обработки и анализа данных, были выявлены закономерности и оценено влияние различных факторов на стоимость жилья.

Цель данного отчета – предоставить информацию о влиянии различных факторов на цену квадратного метра квартиры.

# **Анализ рынка недвижимости Москвы и Подмосковья**

Для выполнения поставленной задачи был выбран интернет-ресурс «Циан». Парсинг осуществлялся на языке Python при помощи библиотеки CianParser. В ходе работы была собрана информация о 8790 квартир (рис. 1) из различных городов Москвы и Московской области. В результате были получены такие параметры как:

• author - автор

• author\_type - тип автора

• url - URL (адрес в интернете)

• location - местоположение

• deal\_type - тип сделки

• accommodation\_type - тип жилья

• floor - этаж

• floors\_count - количество этажей

• rooms\_count - количество комнат

• total\_meters - общая площадь (в квадратных метрах)

• price - цена

• yearofconstruction - год постройки

• object\_type - тип объекта

• housematerialtype - тип материала дома

• heating\_type - тип отопления

• finish\_type - тип отделки

• living\_meters - жилая площадь (в квадратных метрах)

• kitchen\_meters - площадь кухни (в квадратных метрах)

• phone - телефон

• district - район

• street - улица

• house\_number - номер дома

• underground - метро (ближайшая станция)

• residential\_complex - жилой комплекс

Работа с данными началась с оформления данных в одну таблицу при помощи библиотеки Pandas. Сразу после этого последовало удаление дубликатов, но в данном случае их не оказалось. Чтобы узнать количество пропусков, все значения «-1» заменяются на NaN (рис. 2)

После этих действий выяснилось, что две колонки оказались полностью незаполненными, а еще несколько – частично пустыми. (рис. 3)

Затем началась чистка атрибутов – пустых, не дающих нужной информации, не влияющих на целевой параметр колонок; выведена тепловая карта (рис. 4) Далее смотрим матрицу пропусков в данных (рис. 5), от которых в следующих шагах необходимо избавиться.

Количество пропущенных данных в процентах:

author\_type 3.414394

rooms\_count 0.589536

price 0.528126

year\_of\_construction 21.456645

street 14.026038

underground 23.397200

Исходя из количества незаполненных ячеек в каждом столбце, были выбраны методы борьбы с пропусками:

author\_type – заполнен модой

price – удаление строк

rooms\_count – удаление строк

underground – заполнен городом

year\_of\_construction – заполнен медианой

street – удаление строк

Также были отфильтрованы столбцы living\_meters и kitchen\_meters. В датафрейм вошли значения больше 14 и 5 соответственно (рис. 6).

За очисткой и форматированием всех необходимых данных следует создание дополнительной колонки price\_per\_meter и вывод корреляции.

Для дальнейшего анализа и рассмотрения графиков (рис. 7-хх) был выбран инструмент Power BI.

Первый график– распределение цен на квартиры в зависимости от этажа.

Второй график – среднее значение цен по городу. На нем видно, что в Апрелевке и Голицыно присутствуют выбросы.

Третий график – максимальная цена квартиры в городах; также заметны выбросы

Четвертый график – соотношение этажа и года, когда было построено здание.

Пятый график – карта, на которой отмечены использованные в ходе работы города.

Шестой график – зависимость между годом постройки здания и этажом с объявления.

Седьмой график – зависимость цены от типа автора.

Восьмой график – соотношение средней цены за кв. метр и метро

# **Заключение**

Благодаря проделанной работе, можно сделать вывод о том, что цена за кв. метр частично зависит от локации, количества комнат, жилой площади квартиры, площади кухни и расположения ближайшей станции общественного транспорта.

Мое личное мнение о результатах: работа успешно выполнена, хоть и имеет много недочетов. Стоит собрать датасет с большего количества городов Московской области и уменьшить, возможно, количество данных, собранных с Москвы, т. к. цены там выше, что портит статистику среднего; либо разделить Москву и область и рассматривать отдельно. Также мною не были убраны выбросы (например, неприлично высокая цена на квартиры в Голицыно и Апрелевке) за неимением достаточного количества знаний и времени, а также были риски сокращения датафрейма до малого количества строк, если чистить по всем городам, а не только Голицыно и Апрелевку. Кроме того, возникли некоторые проблемы с пропусками – их можно было заполнить по-другому, но, опять же не хватило времени и ума.

# **Приложения**

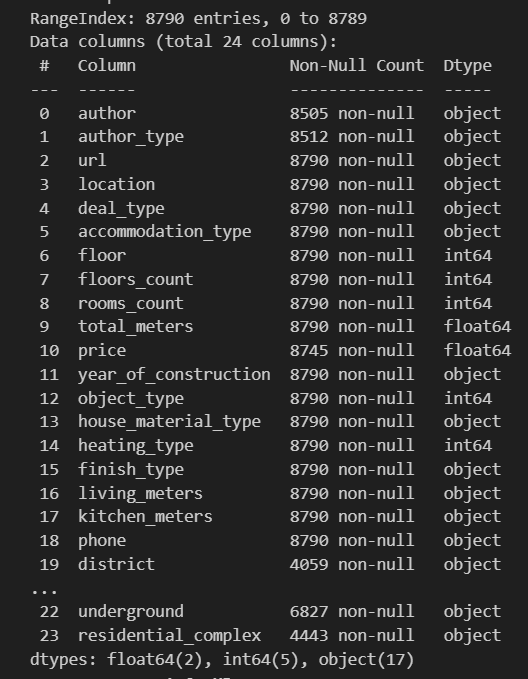


Рис. 1

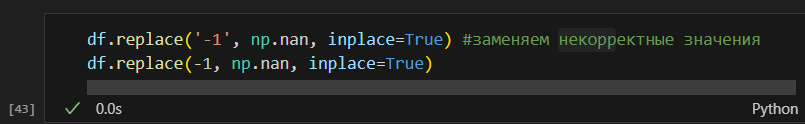


Рис. 2

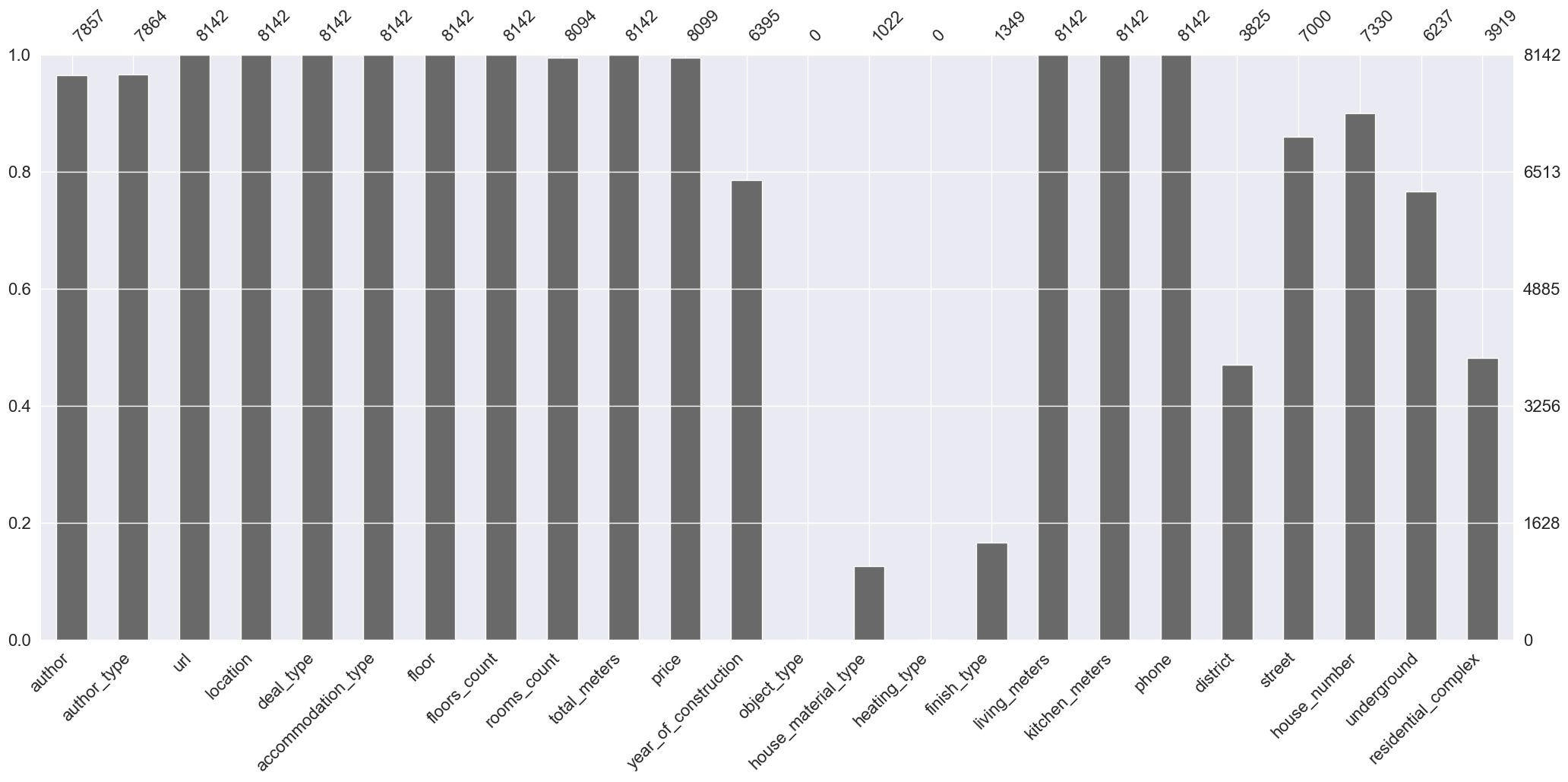


Рис. 3

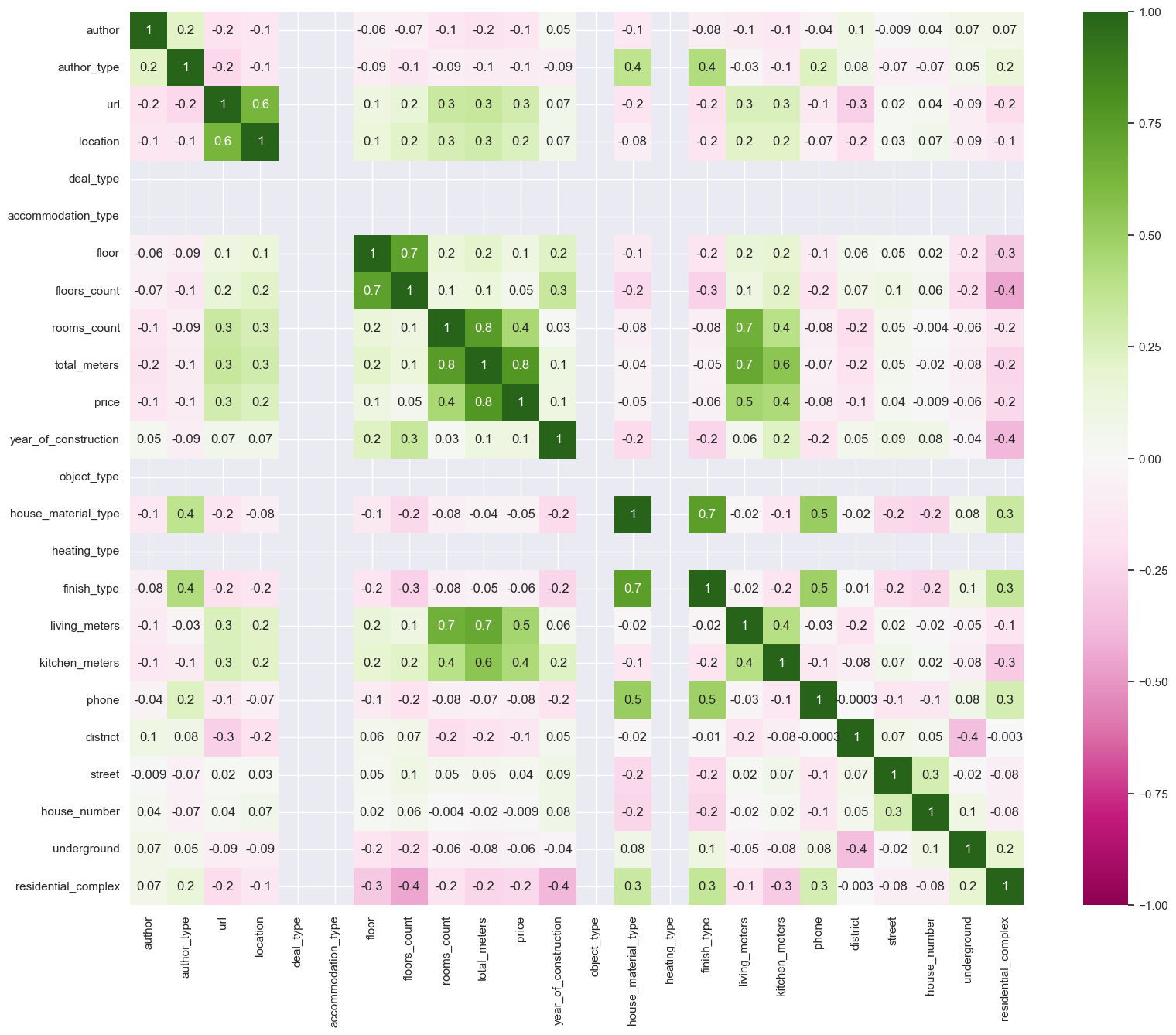


Рис. 4

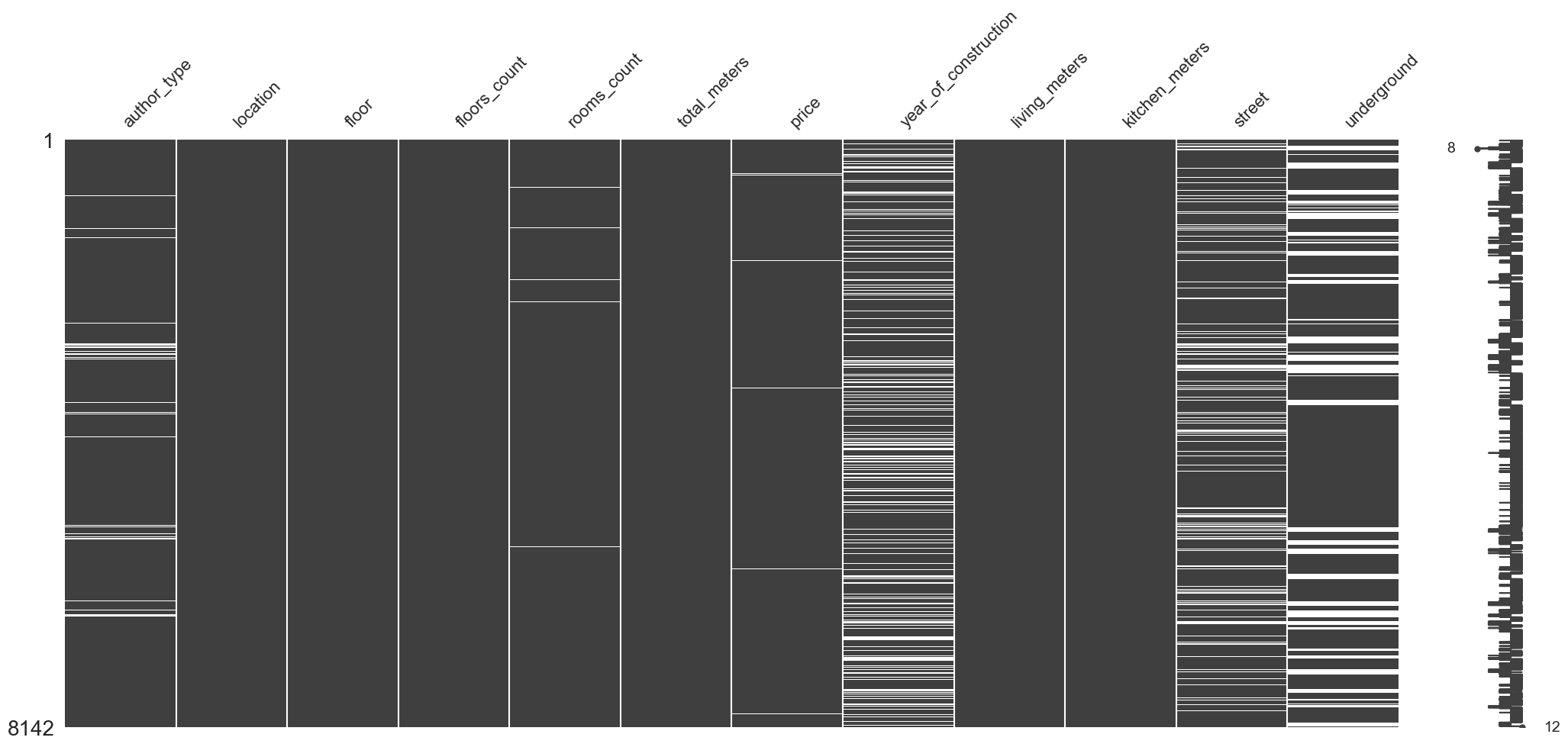


Рис. 5

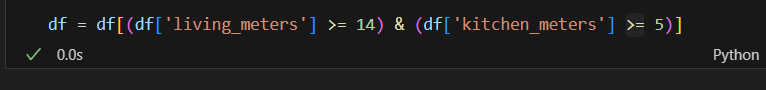


Рис. 6

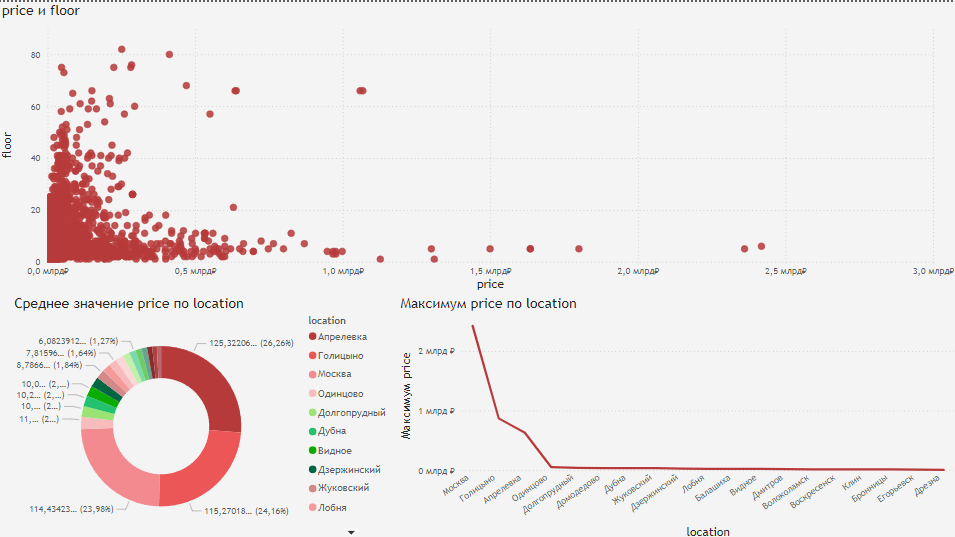


Рис. 7

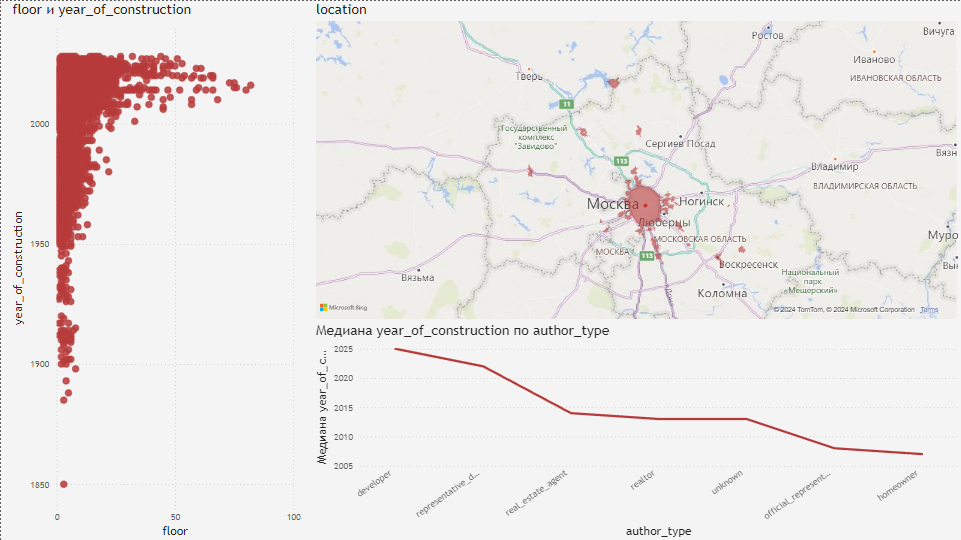


Рис. 8

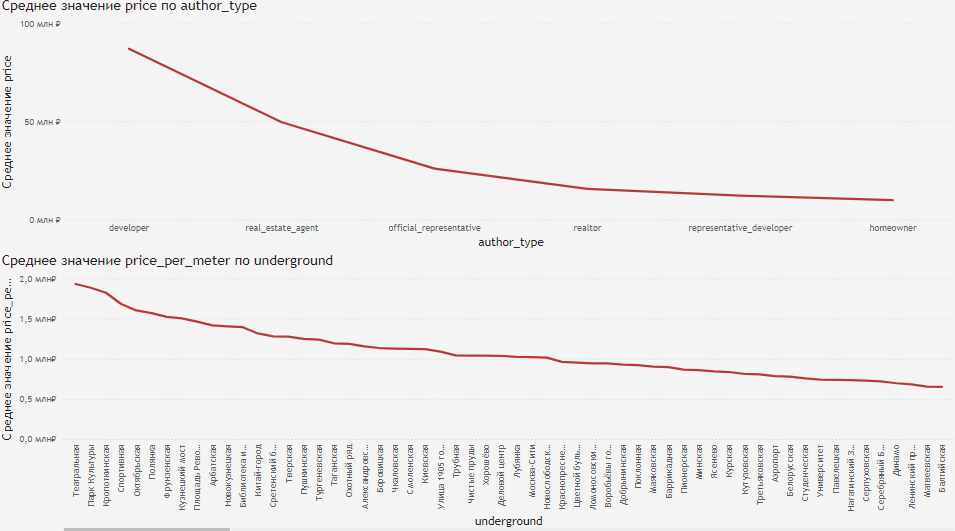


Рис. 9