Государственное бюджетное профессиональное

образовательное учреждение Московской области

«Физико-технический колледж»

**Аналитический отчет**

***«Обработка данных и моделирование цен на квартиры сайта циан»***

Работу выполнил:

Студент группы ИСП-21

Басенко Евгения

Проверила:

преподаватель

Базяк Г.В.

**Введение.**

Цель данного проекта провести анализ рынка недвижимости в Москве и городах Подмосковья, выявить, что влияет на ценовую политику и установить главные критерии оценки.

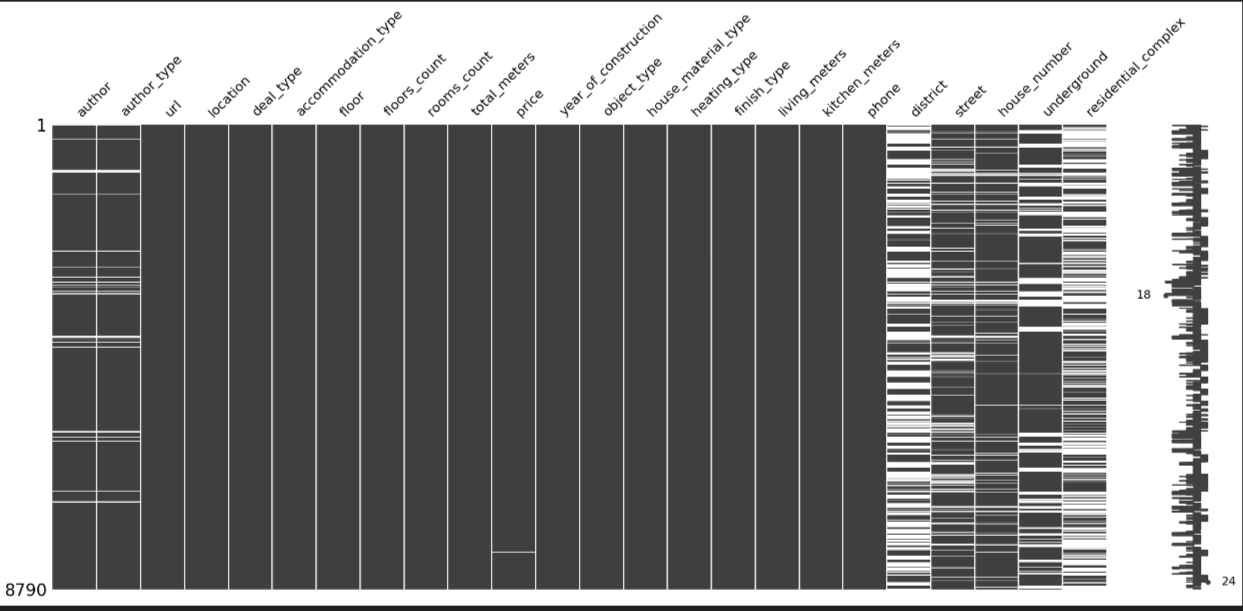
**Задачи.**

* Получить данные по рынку при помощи циана
* Привести данные к однородному виду
* Провести анализ
* Подготовить дашборд

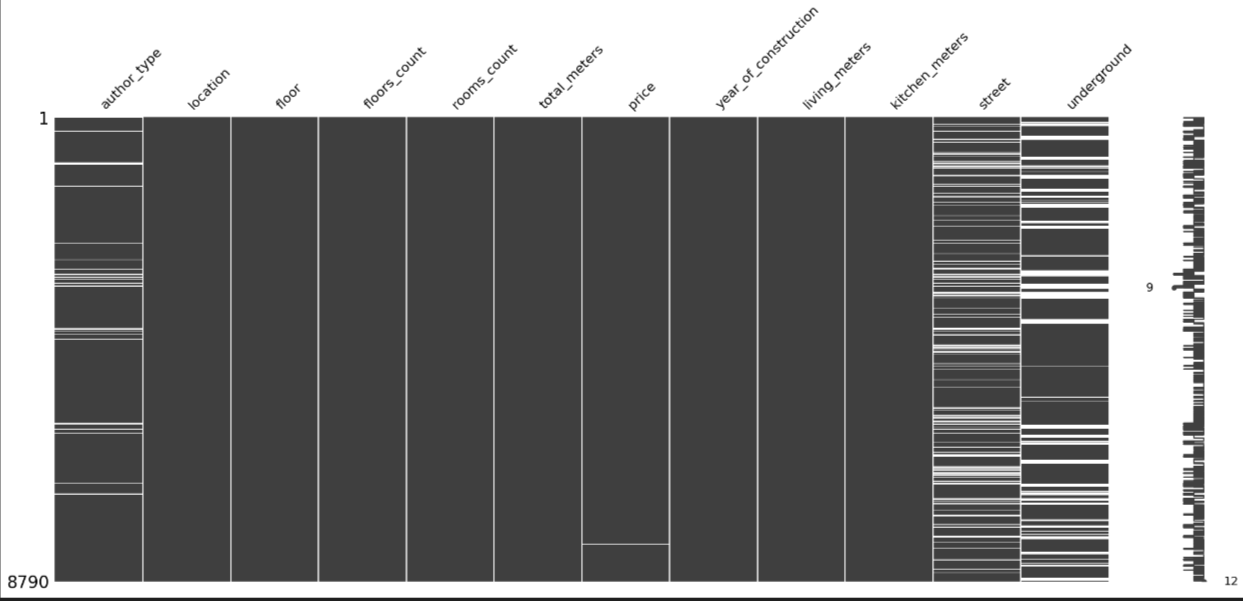
**Ход работы.**

Для начала соберем датасет с помощью парсинга. В моем случае удалось собрать 12 тысяч строк. Далее я загрузила все отдельные таблицы в одну большую и удалила повторяющиеся строки, в итоге у меня осталось 8,7 тысяч строк.

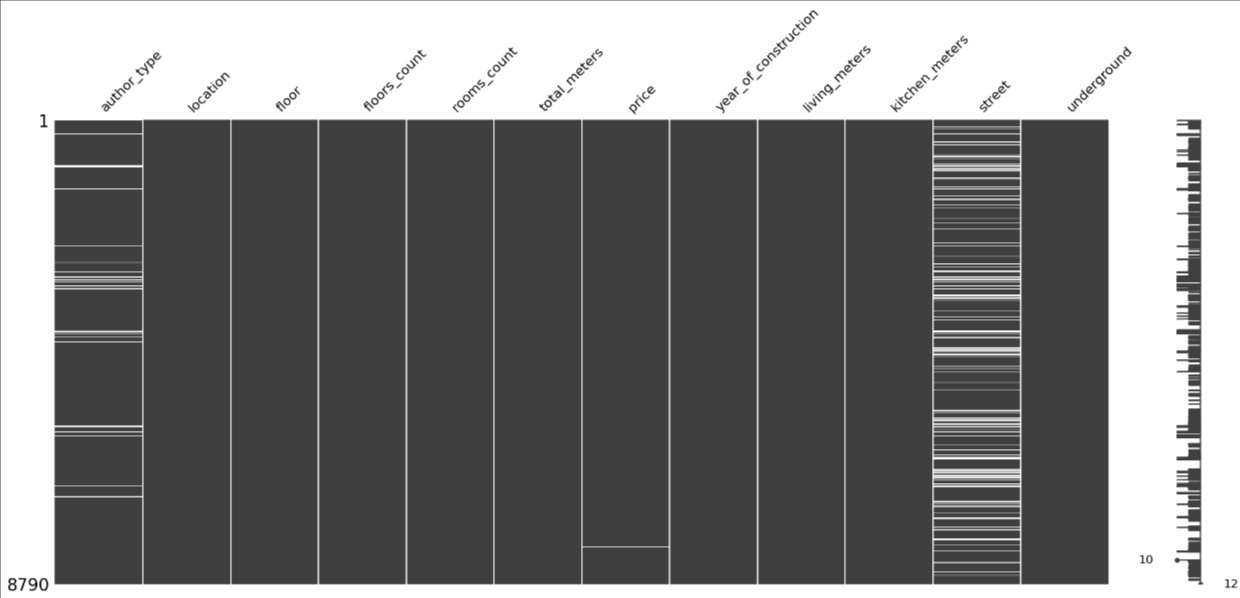
Далее начала работу с самим датасетом. Начала я непосредственно с выявления пустых строк, а также посмотрела процент заполненности.



Как мы видим, в датасете много пропусков. У некоторых колонок даже больше 40%, их сразу на удаление. Также в датасете есть колонки, которые не несут никакой информации – либо везде одинаковое значение, либо колонка не несет никакого вклада в информацию о цене. Удаляем колонки: finish\_type, house\_number, house\_material\_type, heating\_type, deal\_type, accommodation\_type, phone, url, author, district, residential\_complex, object\_type.

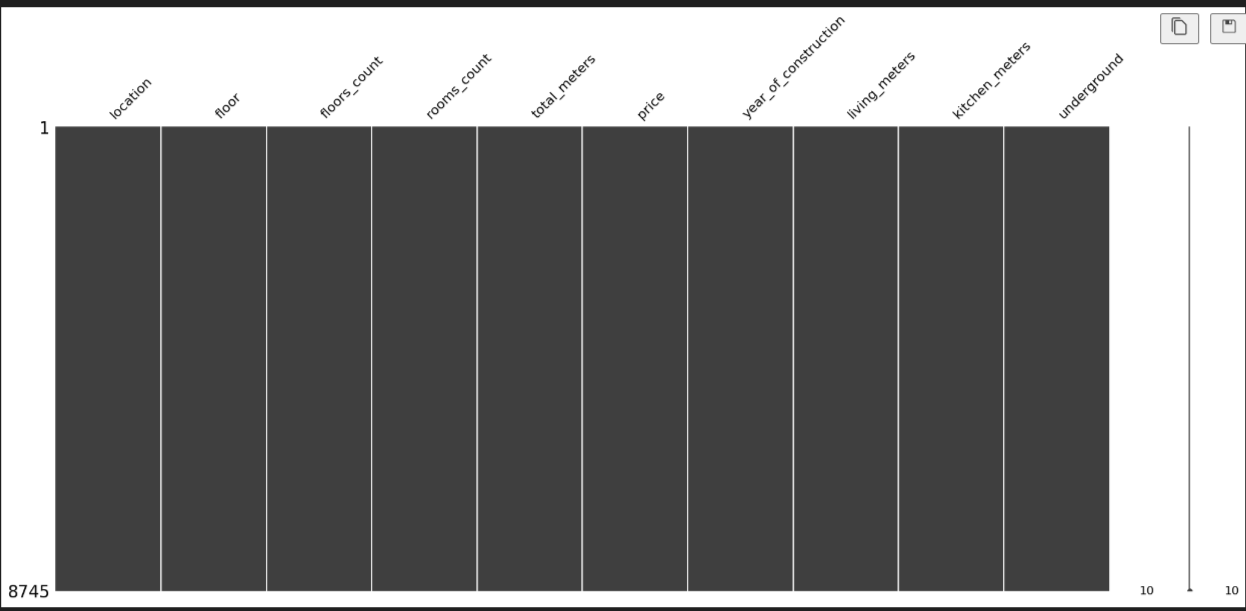


Далее преобразуем колонки living\_mereкs и kitсhen\_meters в численные значения, но для этого надо удалить три последних символа (м2). Еще дозаполним колонку underground. Так как не везде в Подмосковье есть метро, перенесем в пустые строчки значения из location.

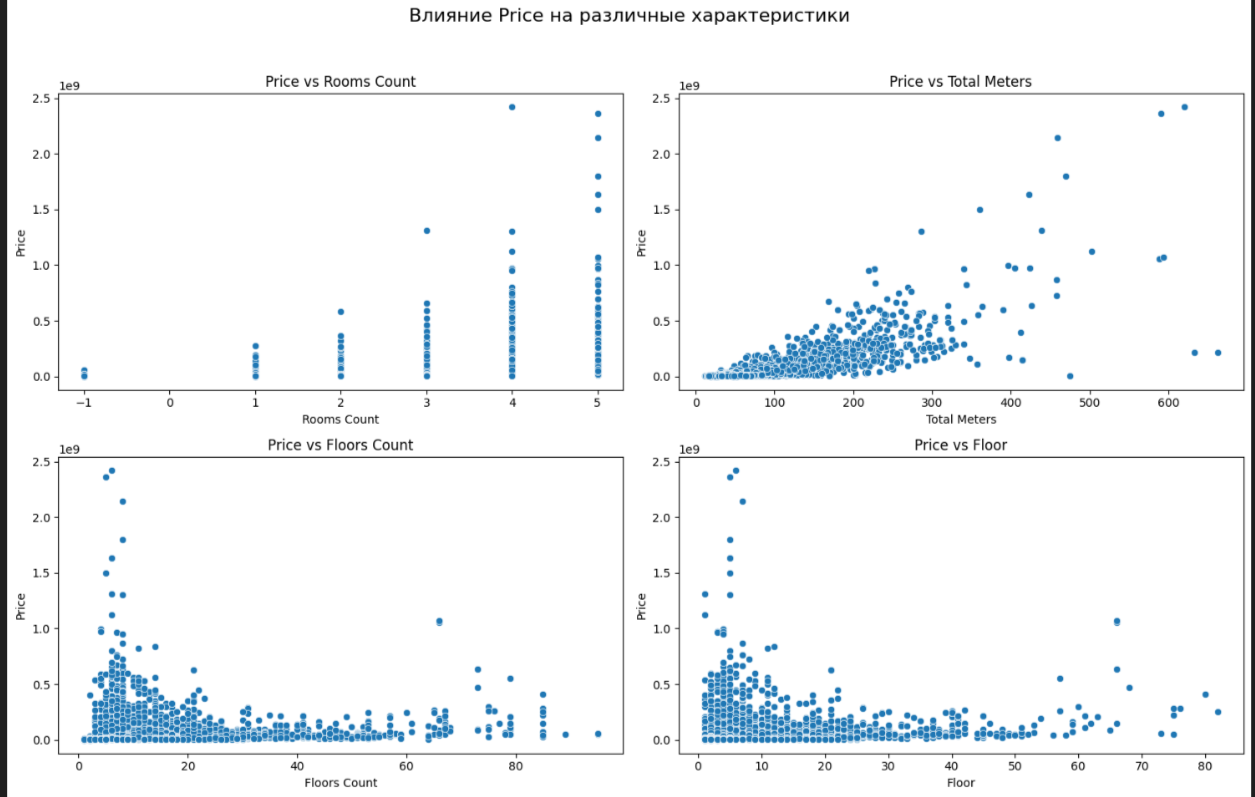


На данном этапе я удалила street и author\_type, а также удалила строки, где были пропуски по колонке price.

Получаем чистый массив.

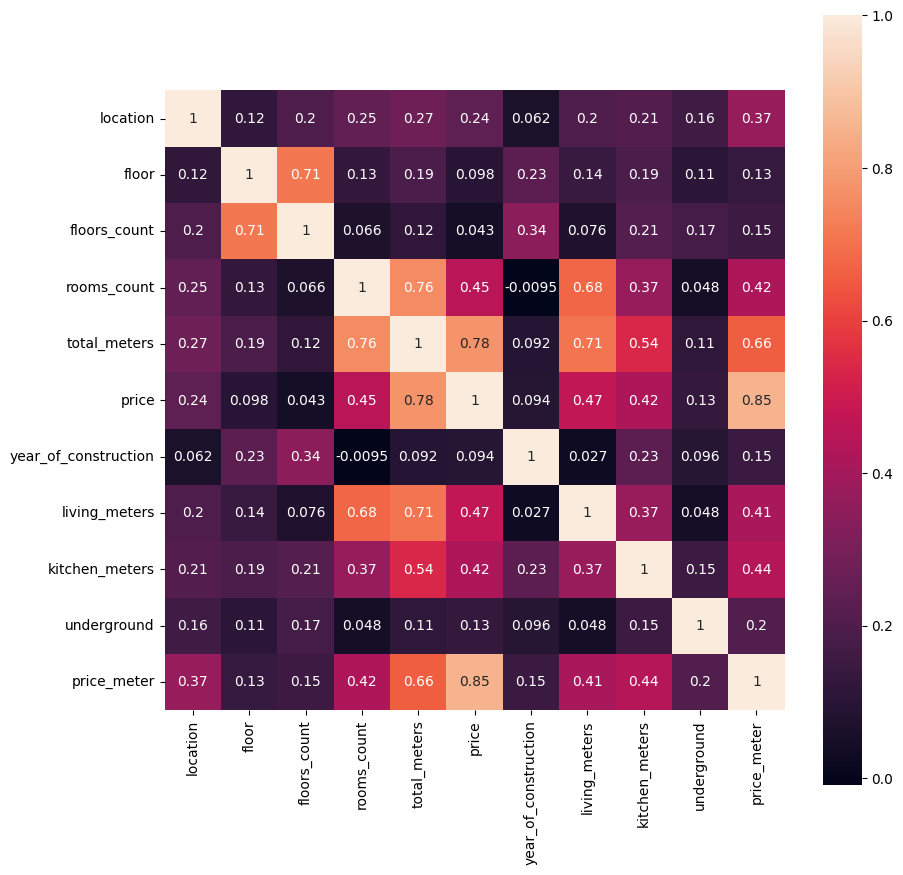


Заодно посмотрим в промежуточном этапе пару диаграмм



Далее заменяю все -1 на нан чтобы не портилась статистка. И завершаю анализ созданием новой колонки price\_meter, которая является ценой за квадратный метр.

Теперь можем строить конечную тепловую карту.



**Вывод**

Как видим, на наш главный параметр «цена за кв. метр» больше всего влияют: цена, общее количество метров, площадь кухни и жилая площадь, а также локация.

Представление с помощью PowerBI:

