Государственное бюджетное профессиональное

образовательное учреждение Московской области

«Физико-технический колледж»

**Аналитический отчет**

***«*Модель оценки цены квартиры на вторичном рынке по Московскому региону: Москва, Новая Москва, Московская область*»***

Работу выполнила:

Студент группы ИСП-21

Замараева Виктоия

Долгопрудный 2024

**Введение**

В данном аналитическом отчете рассматривается модель оценки квартир по Московской области, Москве и Новой Москве. Московский регион является одним из крупнейших рынков недвижимости в России, привлекающий местных жителей и приезжих.

**Цель:**

Собрать данные и произвести анализ данных для построения модели, которая будет оценивать цену квадратного метра недвижимости в Московском регионе.

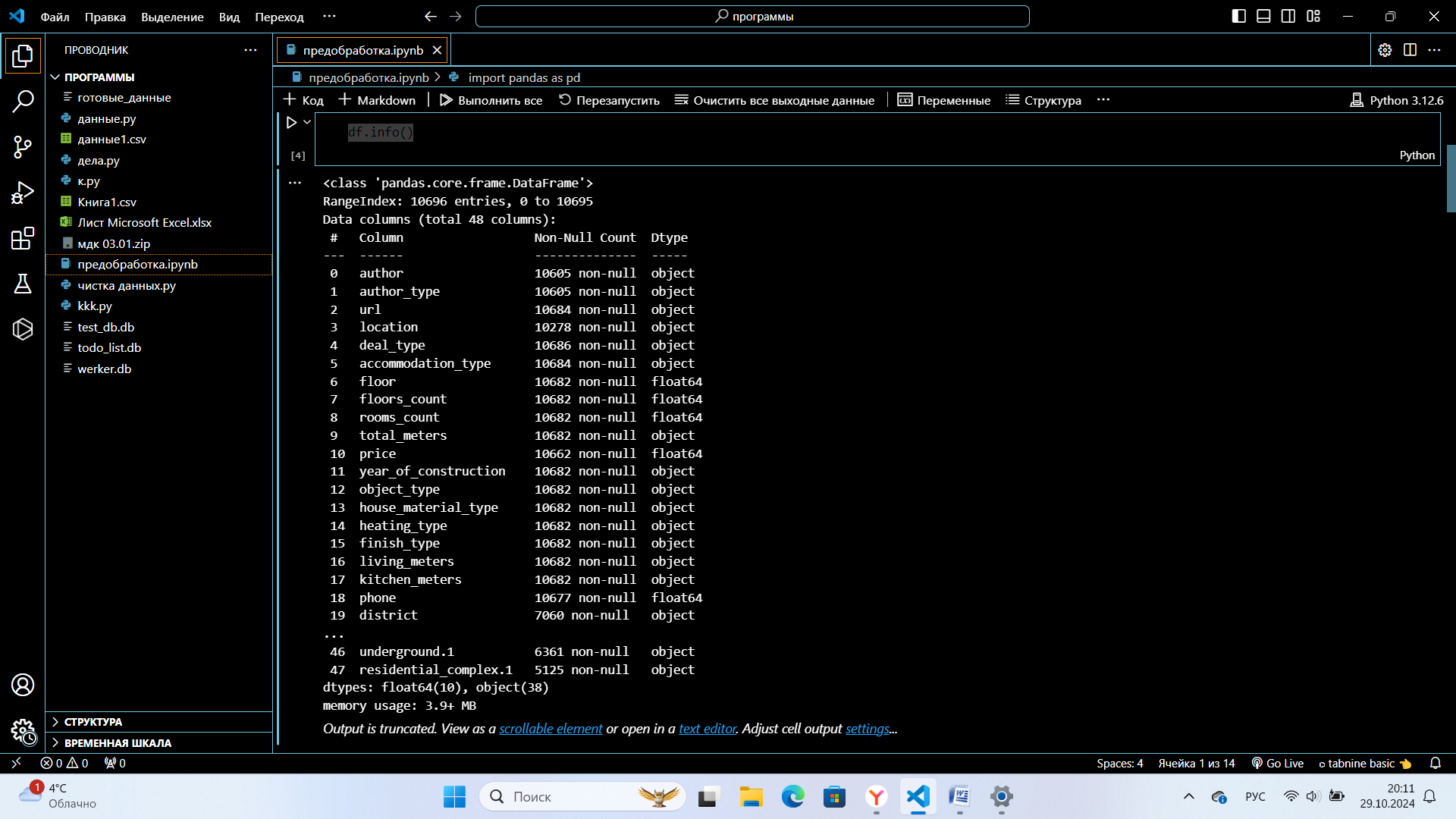
**Задачи:**

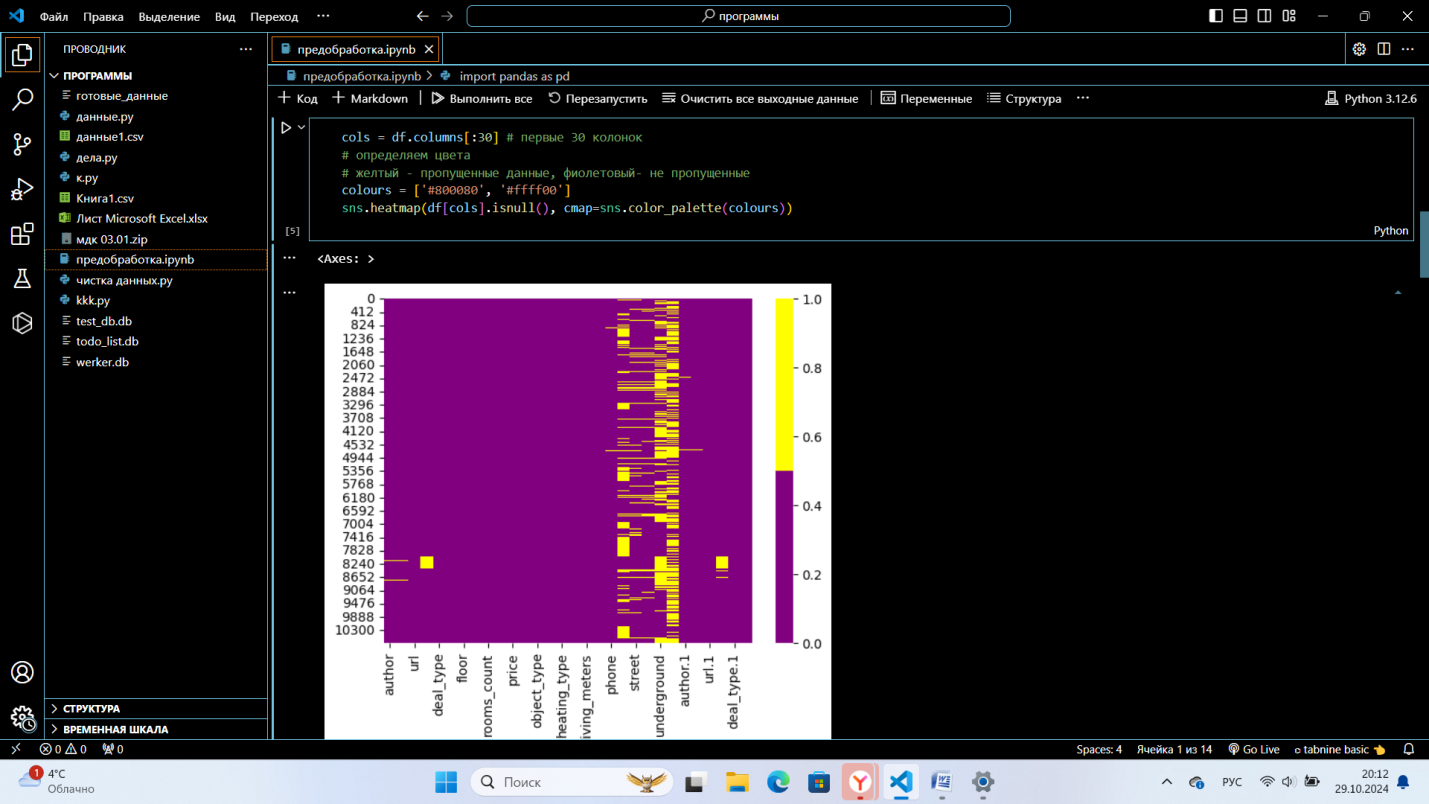
* Используя открытые источники (Циан) составить список параметров, значительно влияющих на цену квадратного метра жилой площади.
* С учетом выявленных факторов произвести парсинг данных по квартирам на продажу, используя различные парсеры.
* Произвести подготовку данных для анализа.
* Визуализировать взаимосвязь между ними; определить признаки, оказывающие наиболее сильное влияние.

**Основная часть**

Для выполнения задачи парсинга данных с сайта Циан, я сначала получила доступ к необходимой информации на сайте. Затем, мы с одногруппницей с помощью программы парсера cianparser извлекли все данные по квартирам. После этого выявила данные которые больше всех будут влиять на цену квартиры.

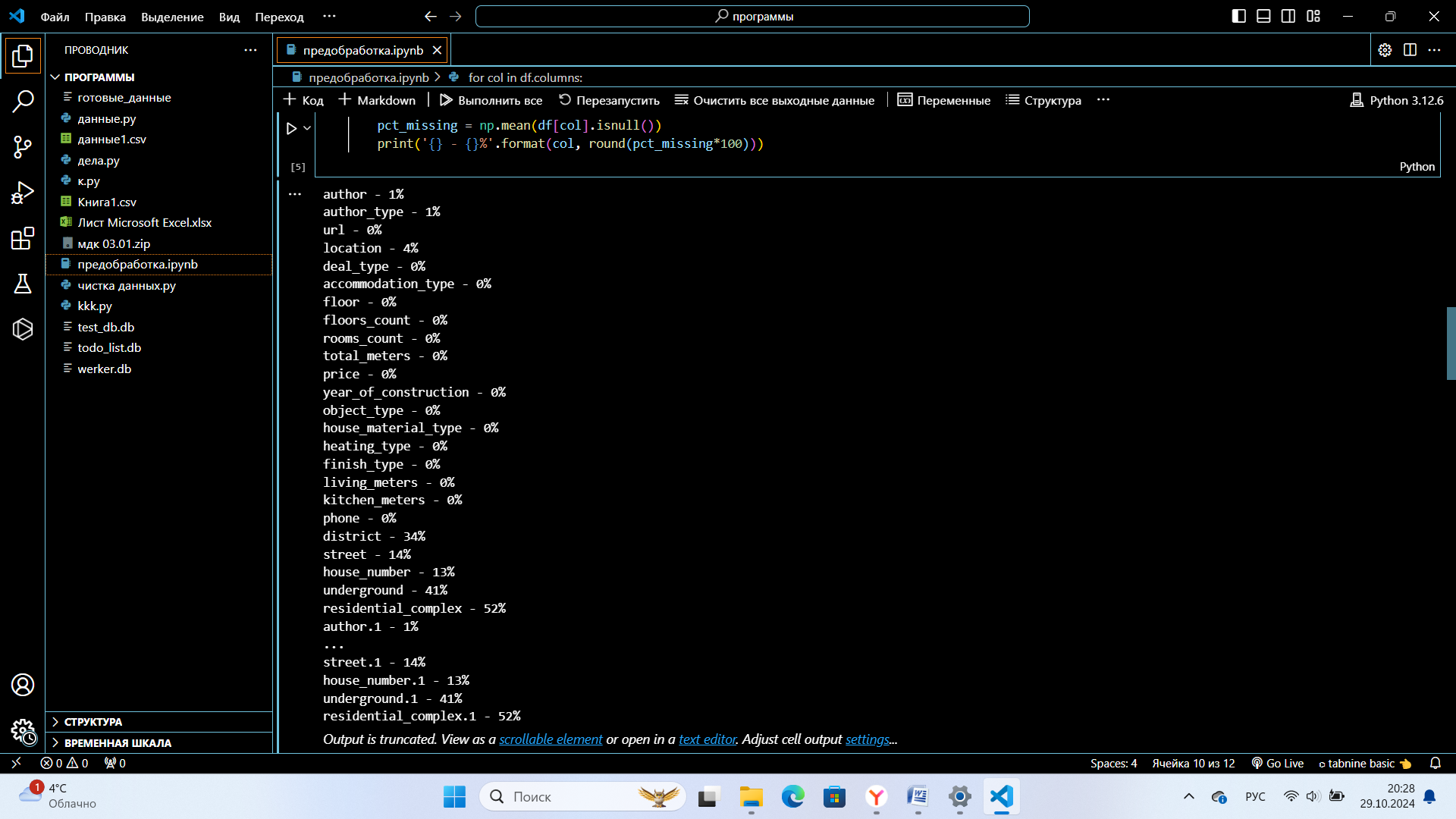
Для выполнения задачи подготовки данных для анализа я использовала несколько библиотек pandas, numpy и другие. В начале работы у меня возникли трудности с кодировкой данных, нос помощью онлайн редактора я решила эту проблему и начала работать с данными.

В начале я вывела какие данные у меня есть всего и какого они типа. Вывожу на тепловую карту пропущенных значений и смотрю какие столбцы или строки удалить.

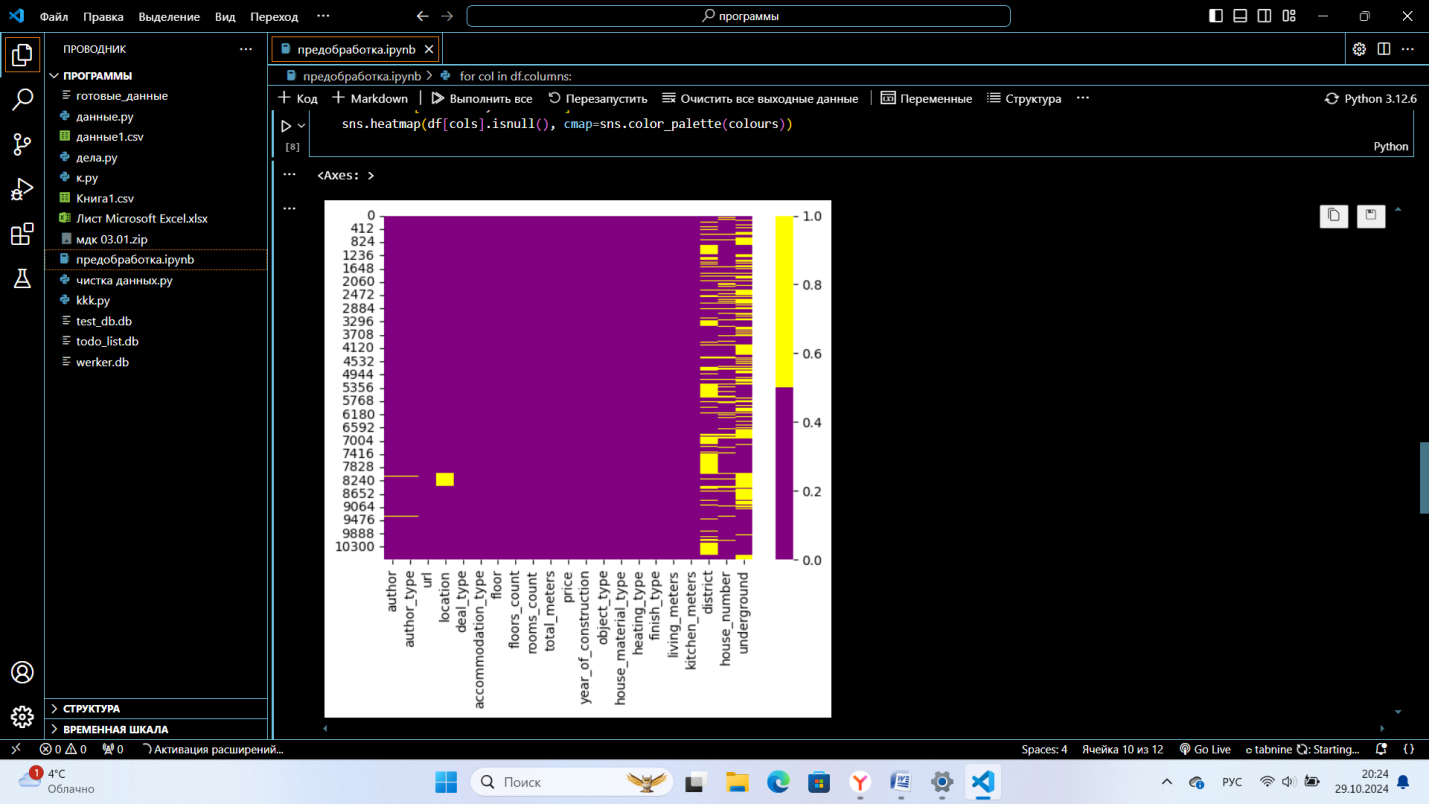


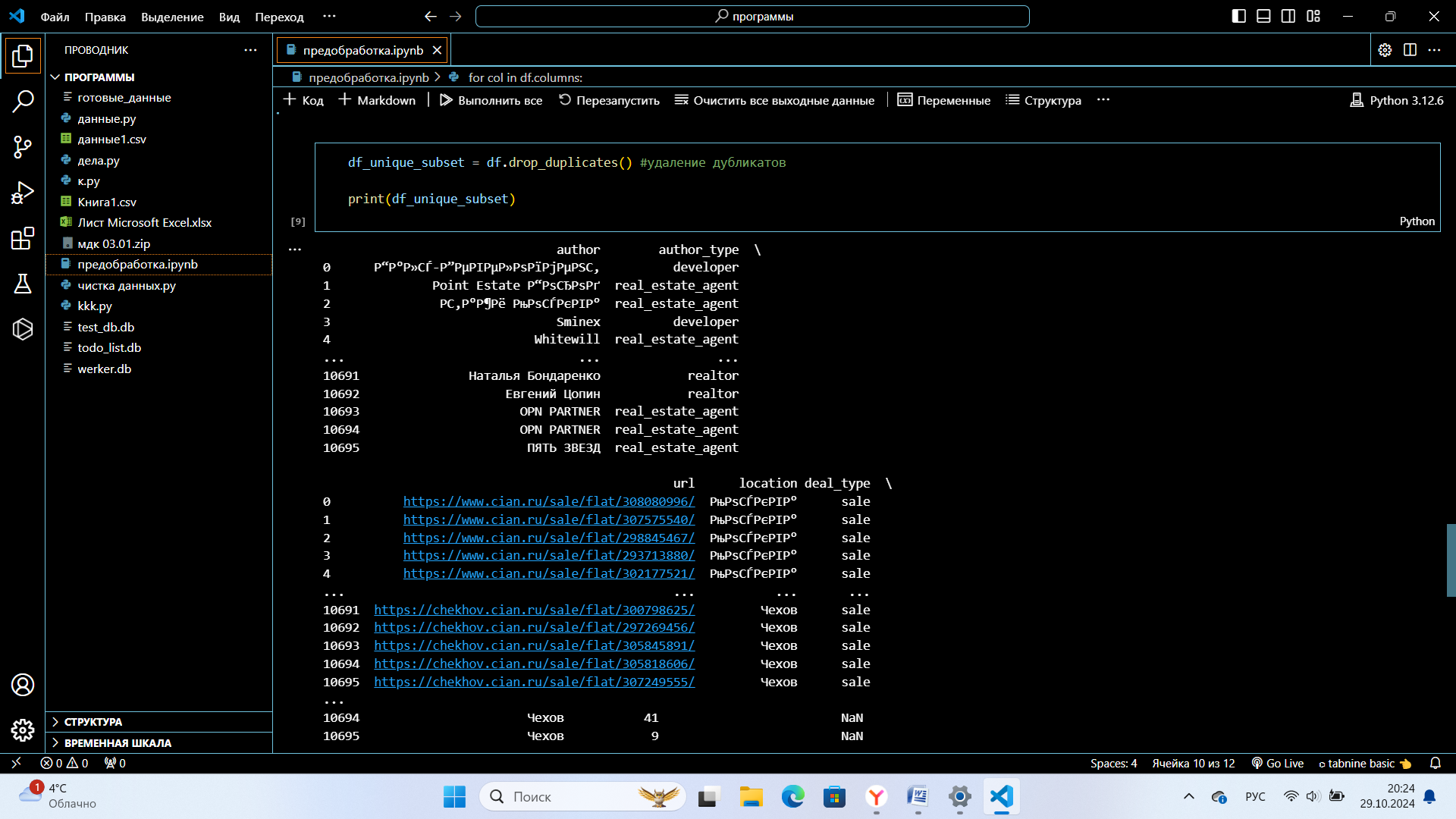
Больше всего пропущенных данных в колонке phone, street и underground.

Далее вывожу процентное содержание пропусков.

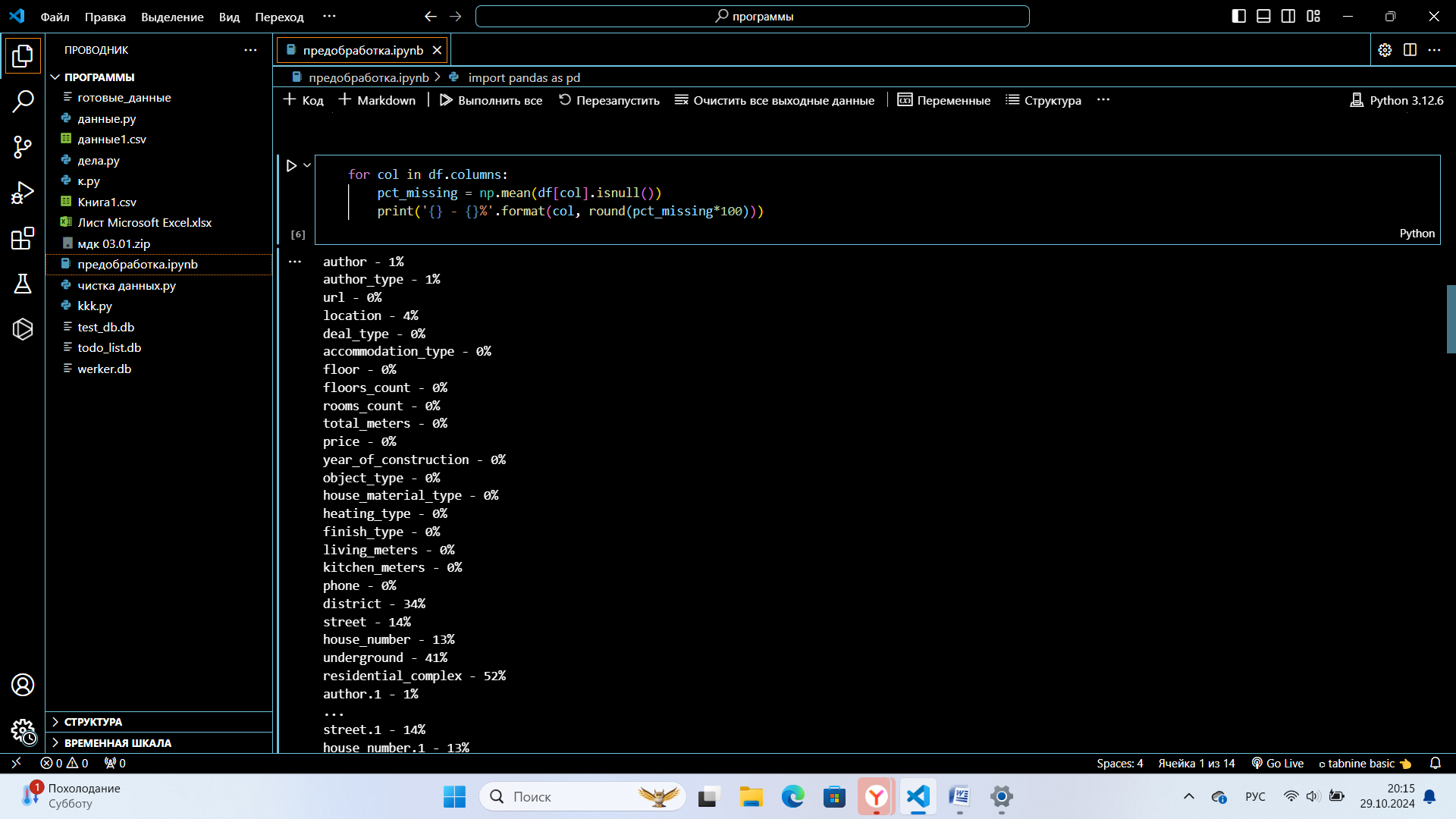


Потом удаляю выбранные столбцы и дубликаты столбцов и опять смотрю в тепловую карту и удаляю дубликаты.

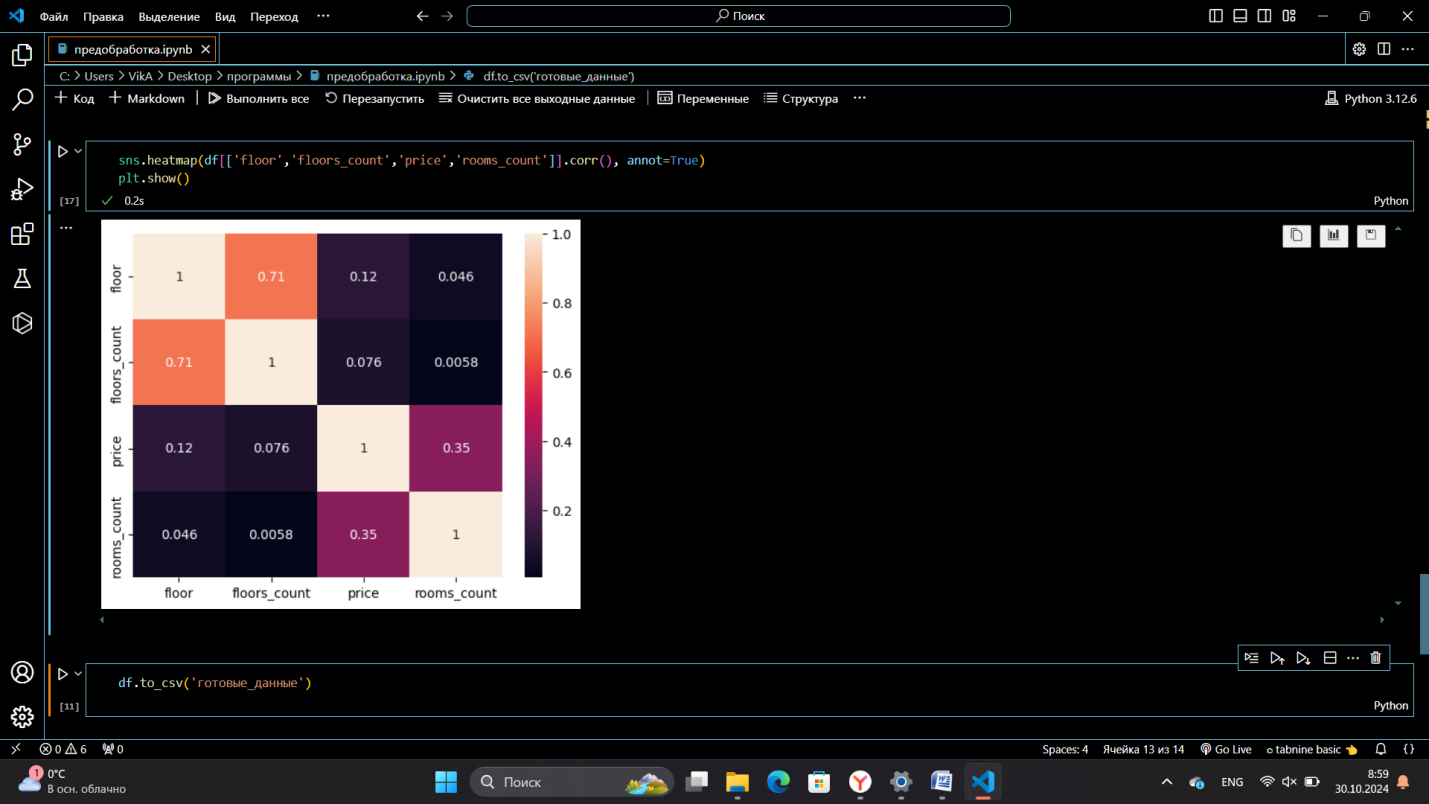




После удаление всех не нужных данных вывожу процентное содержание пропусков, их стало меньше.



Потом вывела корреляцию числовых значений. Больше всего влияют на цену квартиры параметры price и rooms\_count.

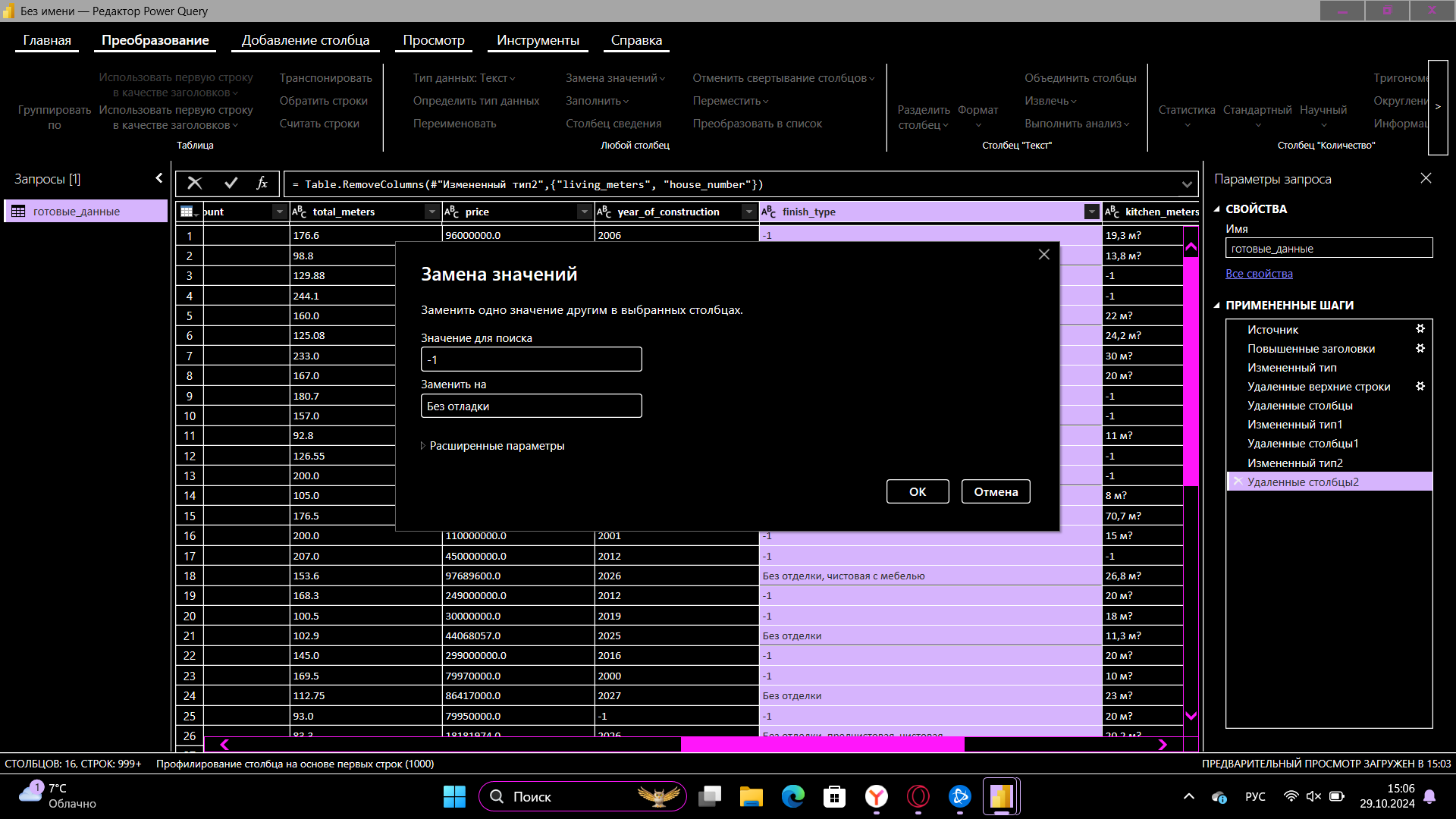


Теперь можно идти визуализировать.

Я использовала в визуализации инструмент Power Bi. С его помощью я создала интерактивный дашборд для выявления факторов, которые больше влияю на цену квартиры.

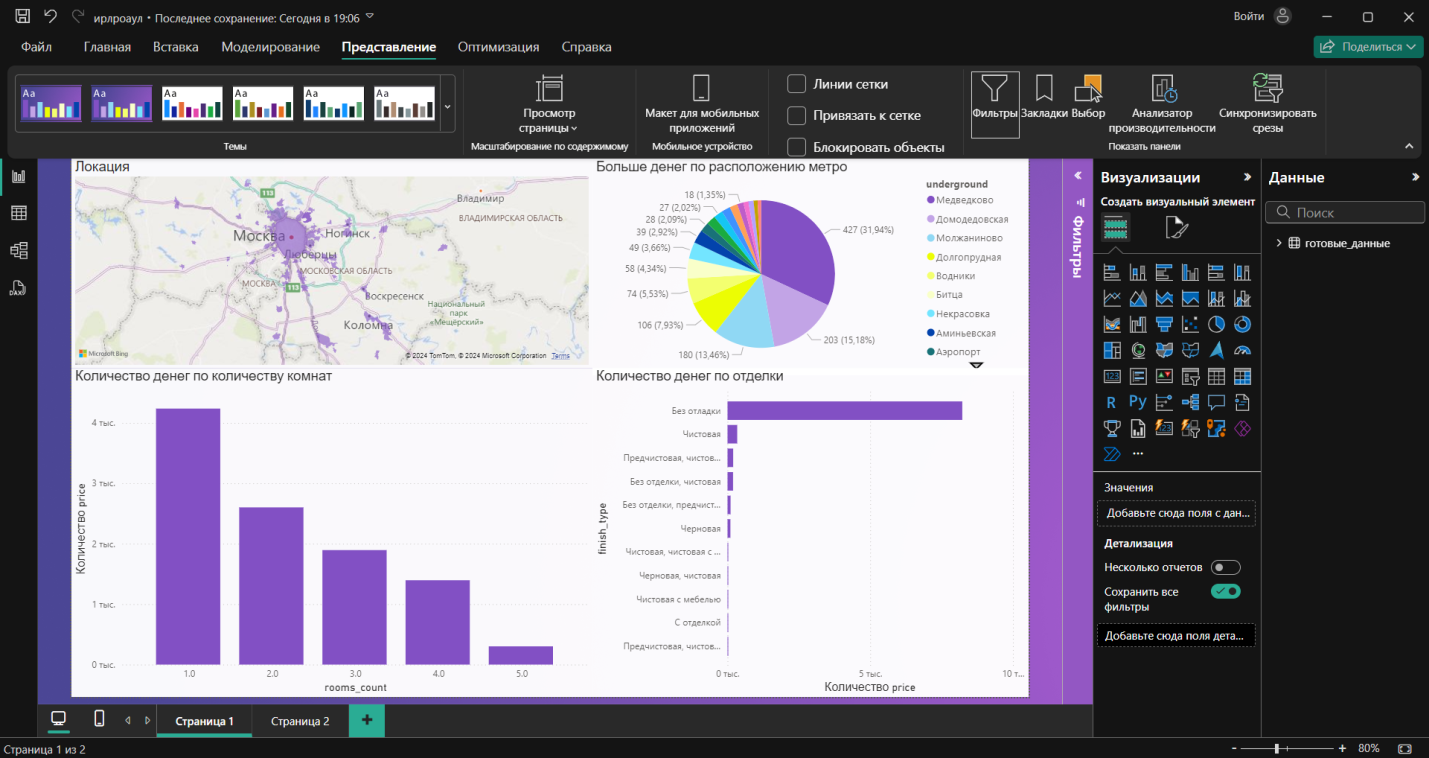
В начале работы я закинула .csv файл в редактор Power Qwery

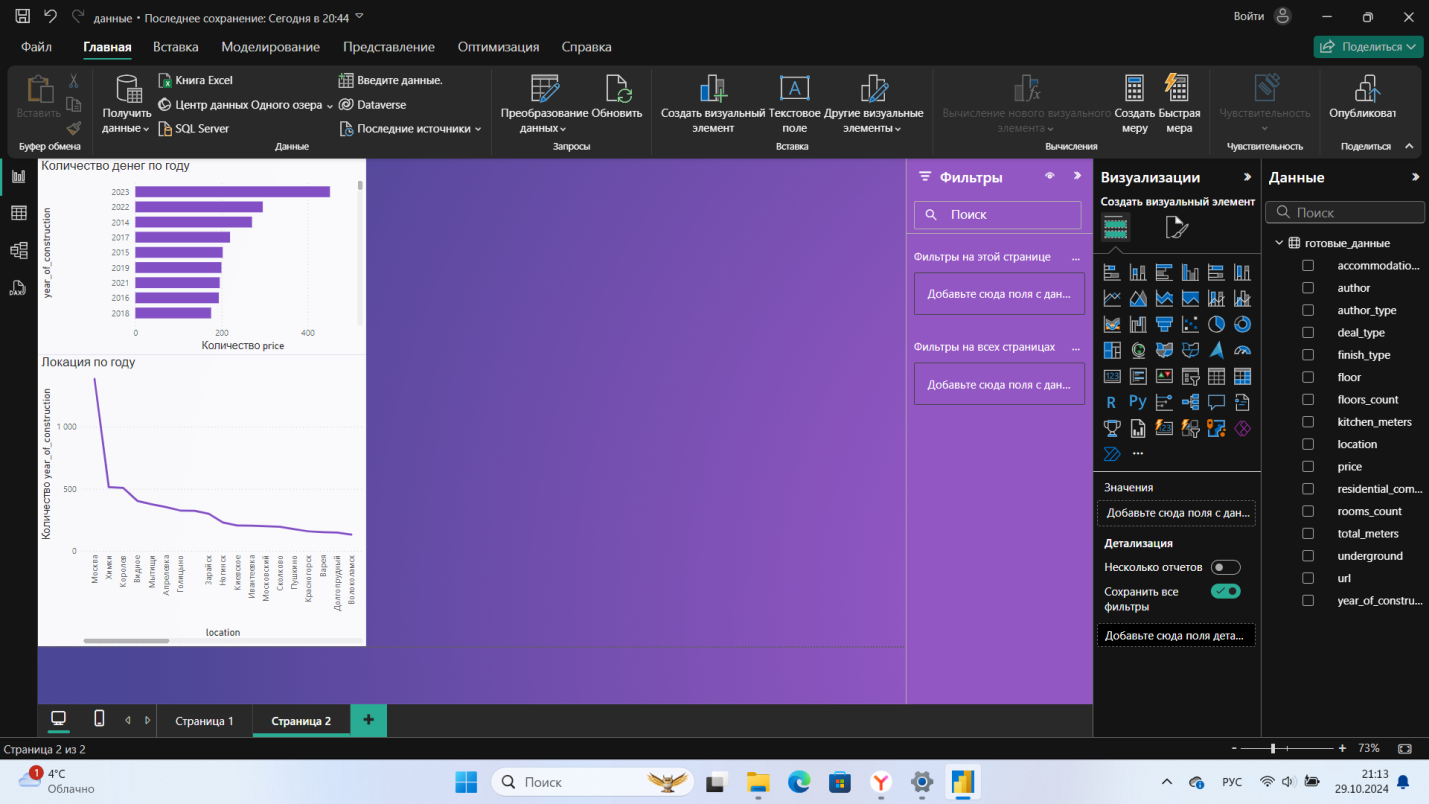
По моему мнению тип отделки квартиры очень важен, поэтому я заменила на среднее значение.



К сожалению из-за того что Power Bi иностранный инструмент для визуализации, price обозначается как текстовые тип данных. Но можно price можно сравнивать с другими данными только по количеству.

Вот такой дашборд получился.



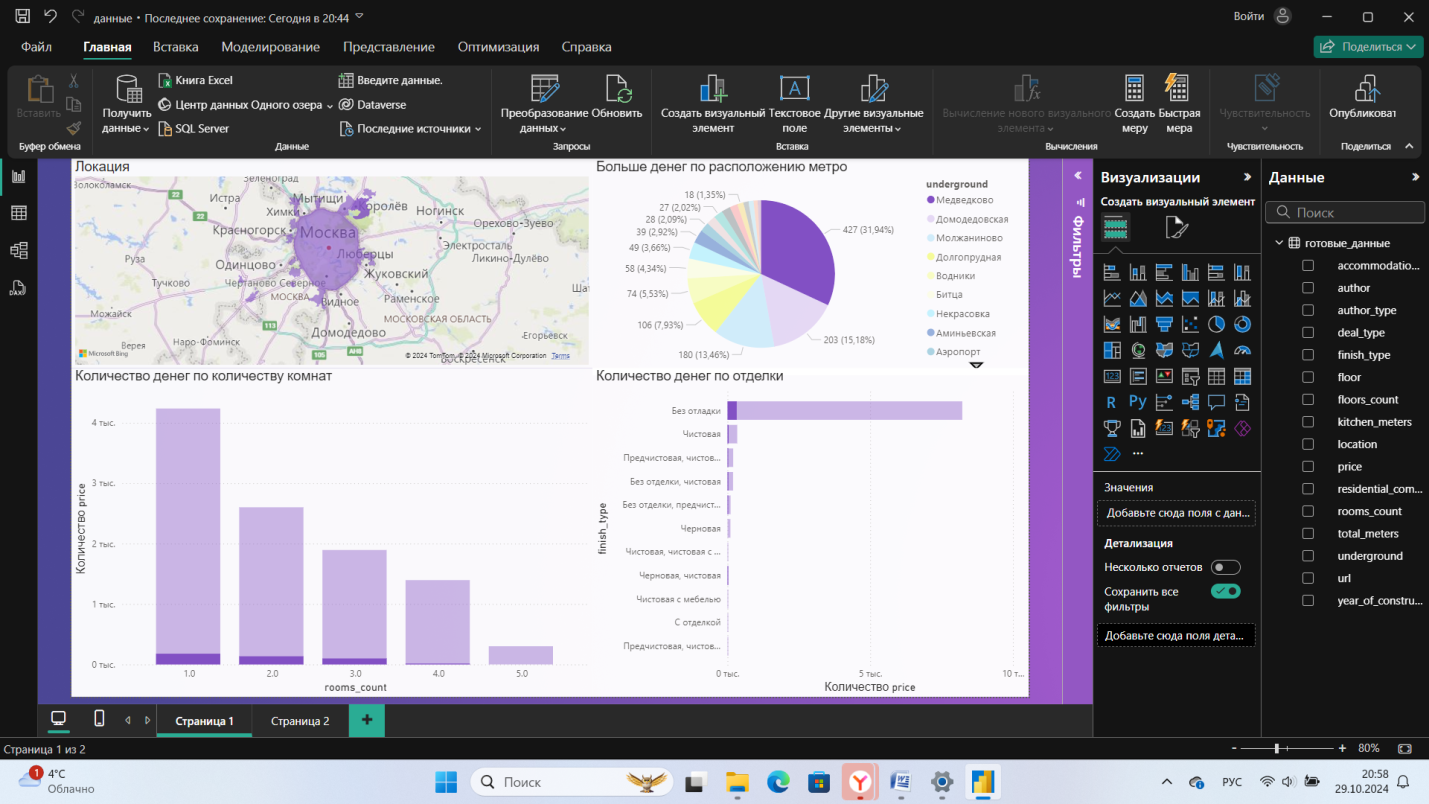


**Анализ данных**

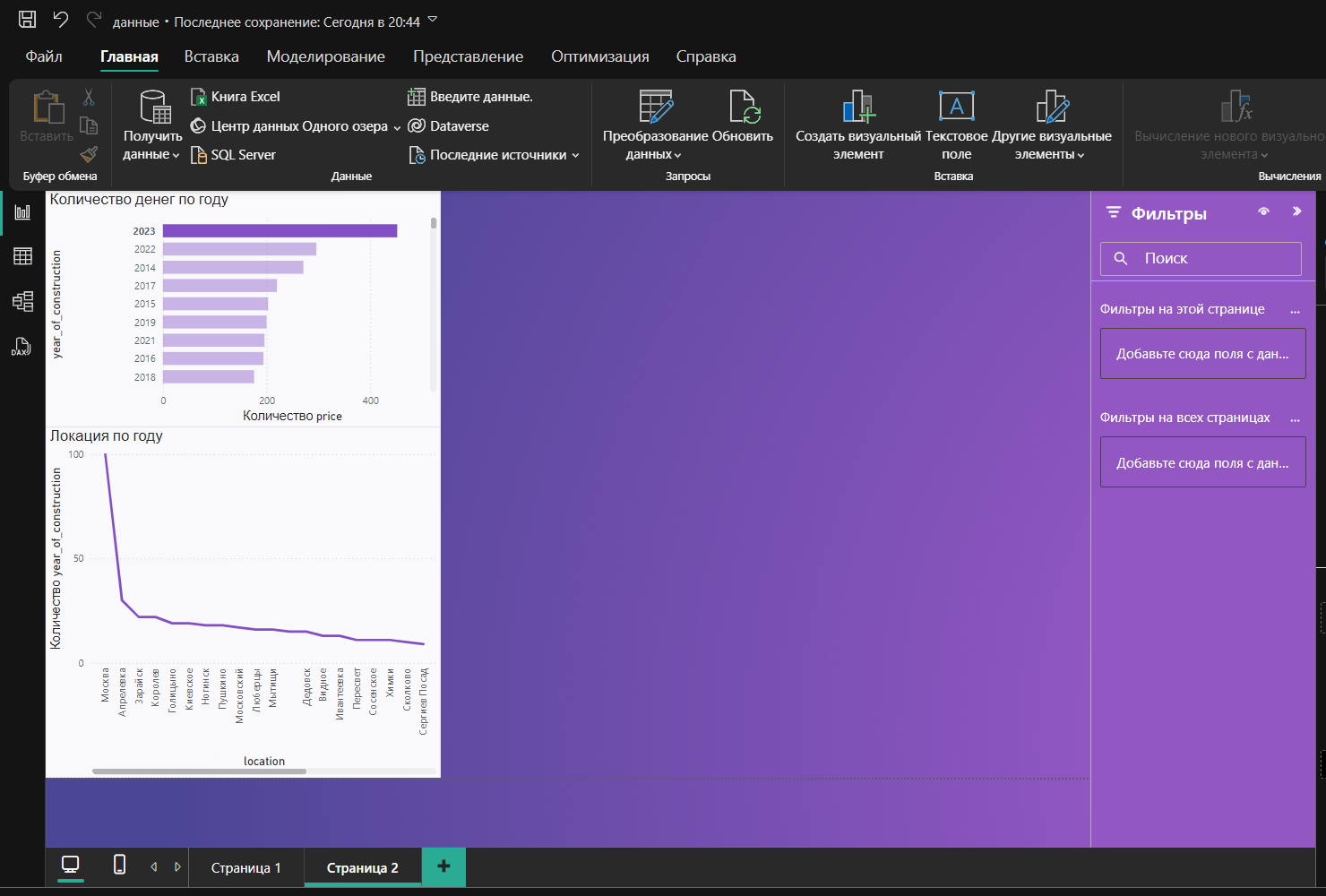
После полученного дашборда можно выделить важные критерии по цене квадратного метра квартиры.

Критерии:

1. Метро
2. Отделка
3. Нахождение в крупных городах
4. Количество комнат



Больше всего цена в центре Москвы, так как она соответсвует всем важным критериям. Есть рядом метро и большинство квартир уже с отделкой , среднее значение держится по количеству комнат.



Так же квартиры в 2023 году выросли в цене и видно сколько объявление было сделано с информацией годом постройки.

**Заключение**

В результате анализа были собраны данные, построена модель и визуализация, в анализе которой выявлены важные критерии в оценивании цены квадратного метра недвижимости в Московском регионе. Основными факторами, оказывающими влияние на стоимость жилья, оказались расположение около метро, нахождение в крупных городах, наличие отладки в квартире. Полученные результаты могут быть использованы для дальнейшего прогнозирования цен на недвижимость.