Московский государственный технический

университет им. Н.Э. Баумана.

Факультет «Информатика и управление»

Кафедра «Системы обработки информации и управления»

Курс «Технологии машинного обучения»

Отчет по рубежному контролю №1

Вариант 15

Выполнила: Проверил:

студентка группы ИУ5-64Б преподаватель каф.ИУ5

Носова Элина Гапанюк Ю. Е.

Подпись и дата: Подпись и дата:

Москва, 2022 г.

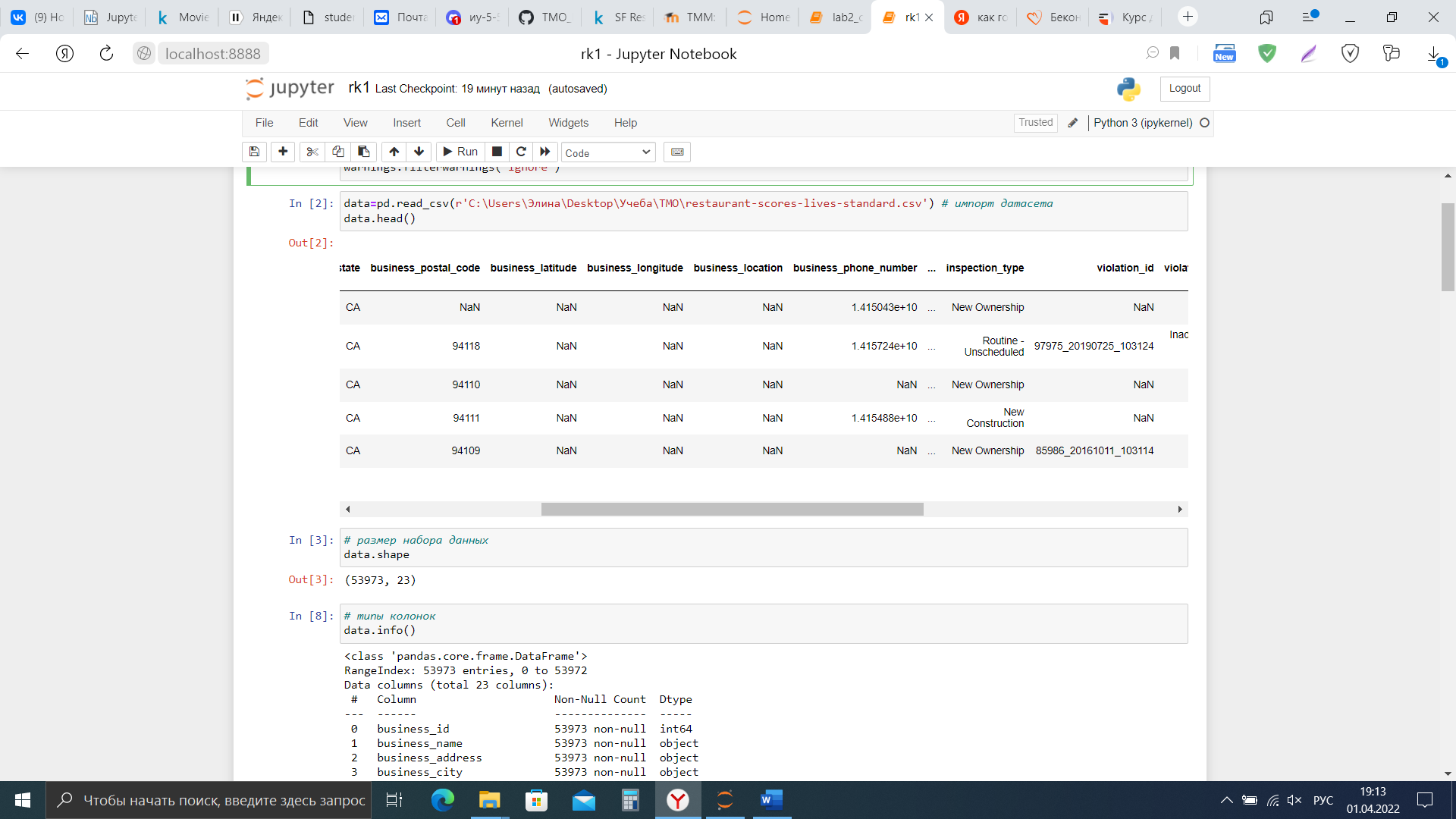
Условие (задача 2):

Для заданного набора данных проведите обработку пропусков в данных для одного категориального и одного количественного признака. Какие способы обработки пропусков в данных для категориальных и количественных признаков Вы использовали? Какие признаки Вы будете использовать для дальнейшего построения моделей машинного обучения и почему?

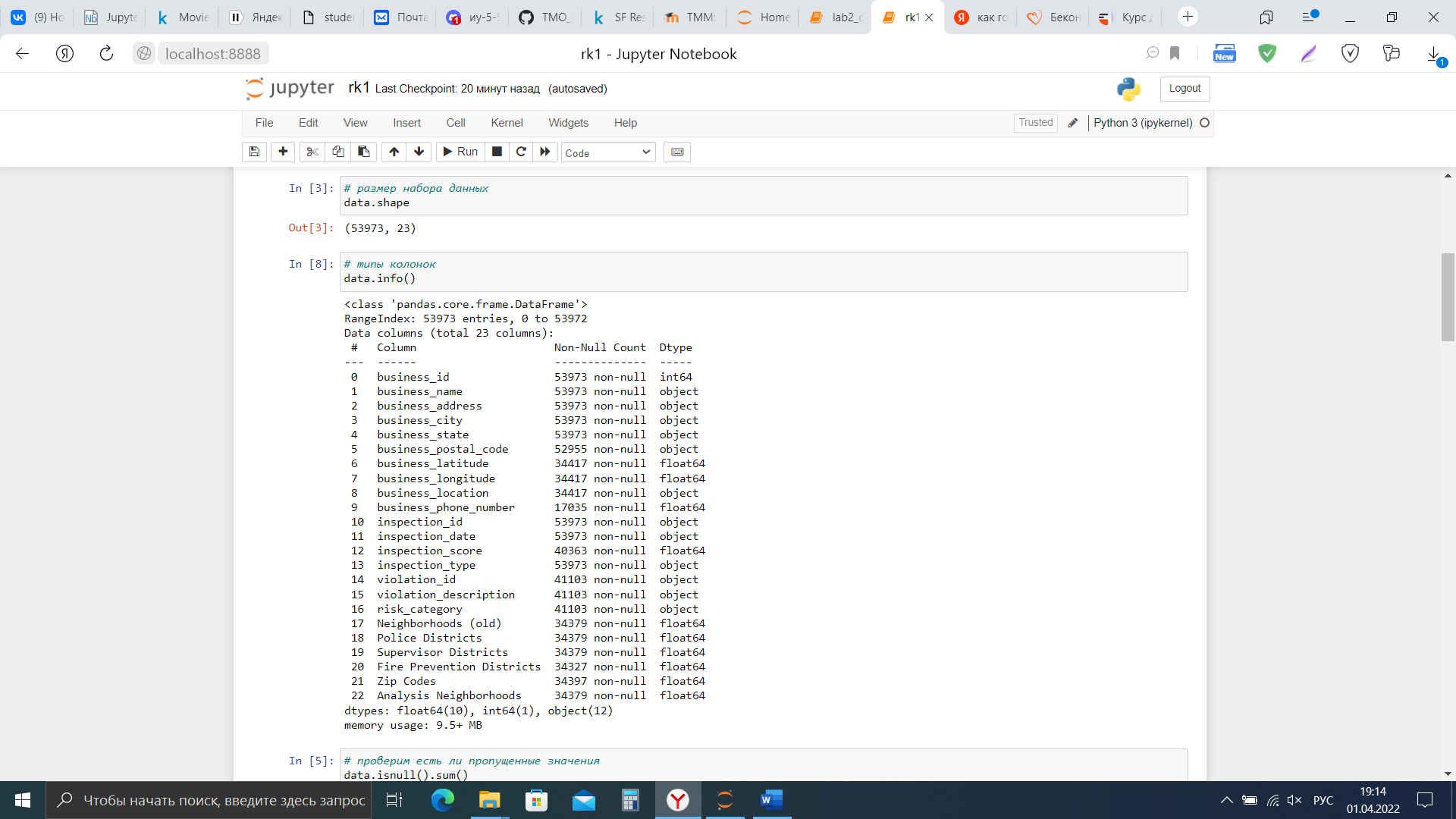
Набор данных (7):

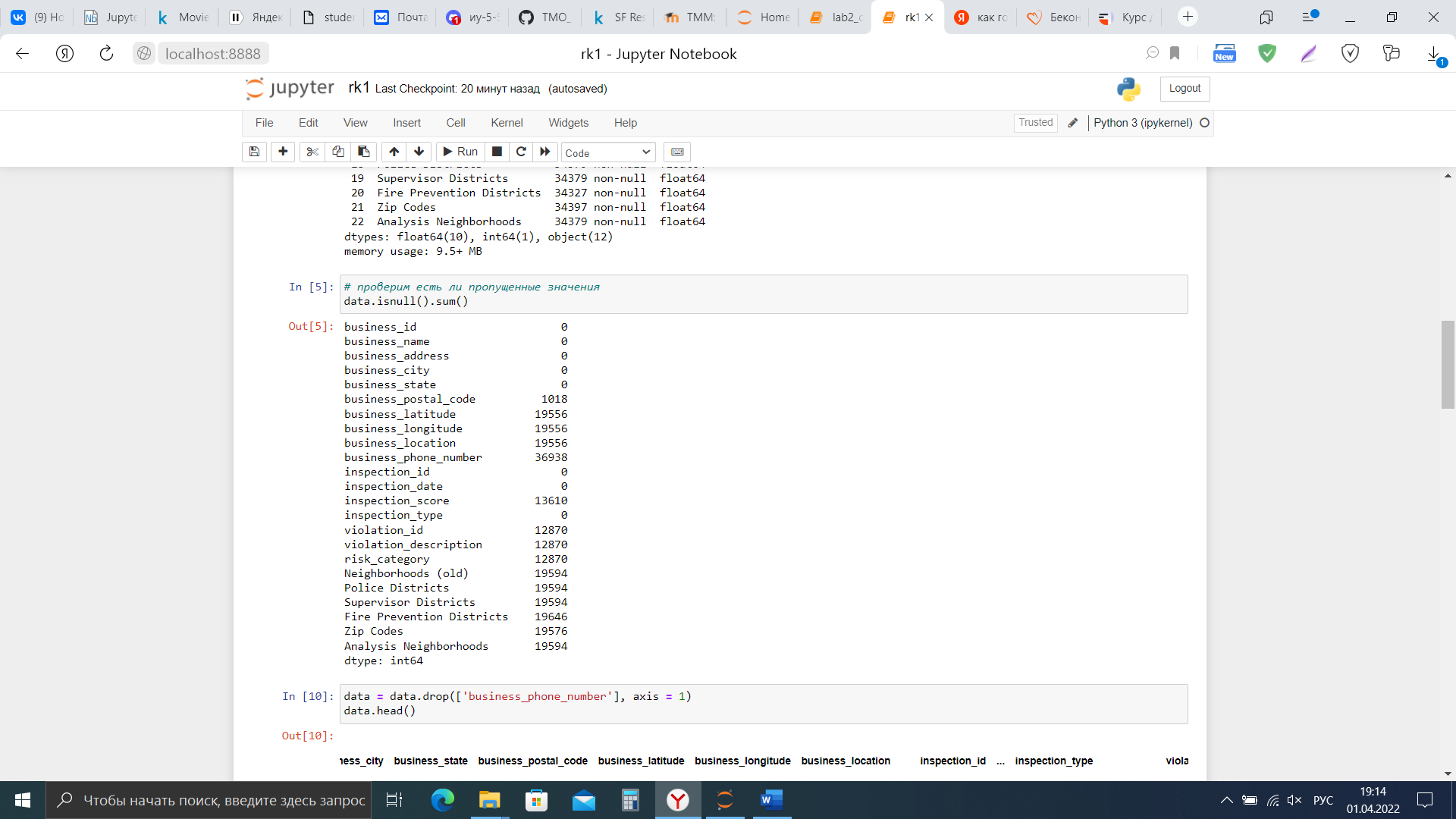
<https://www.kaggle.com/san-francisco/sf-restaurant-scores-lives-standard>

Загрузка и проверка данных:

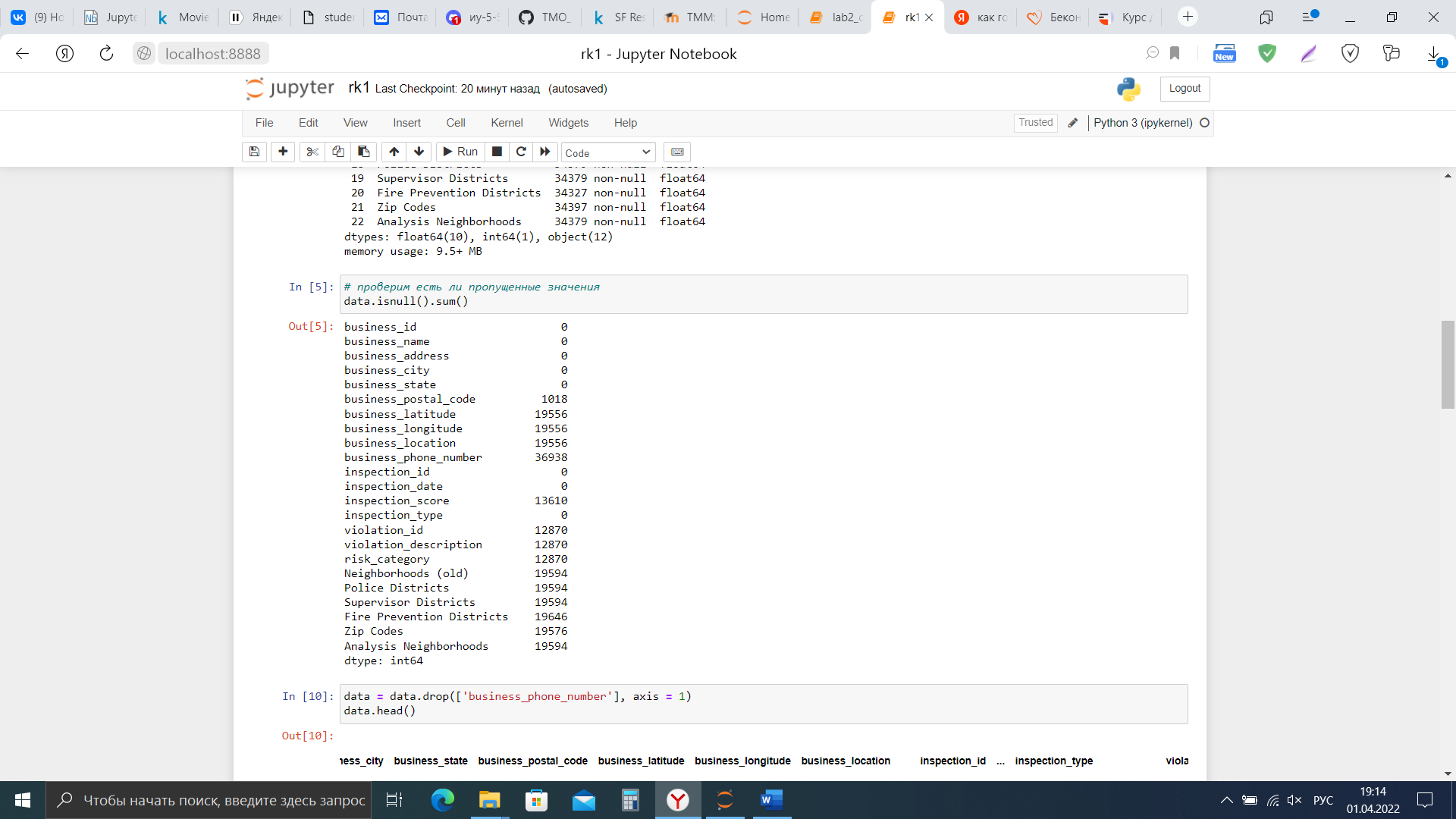


Информация о данных:

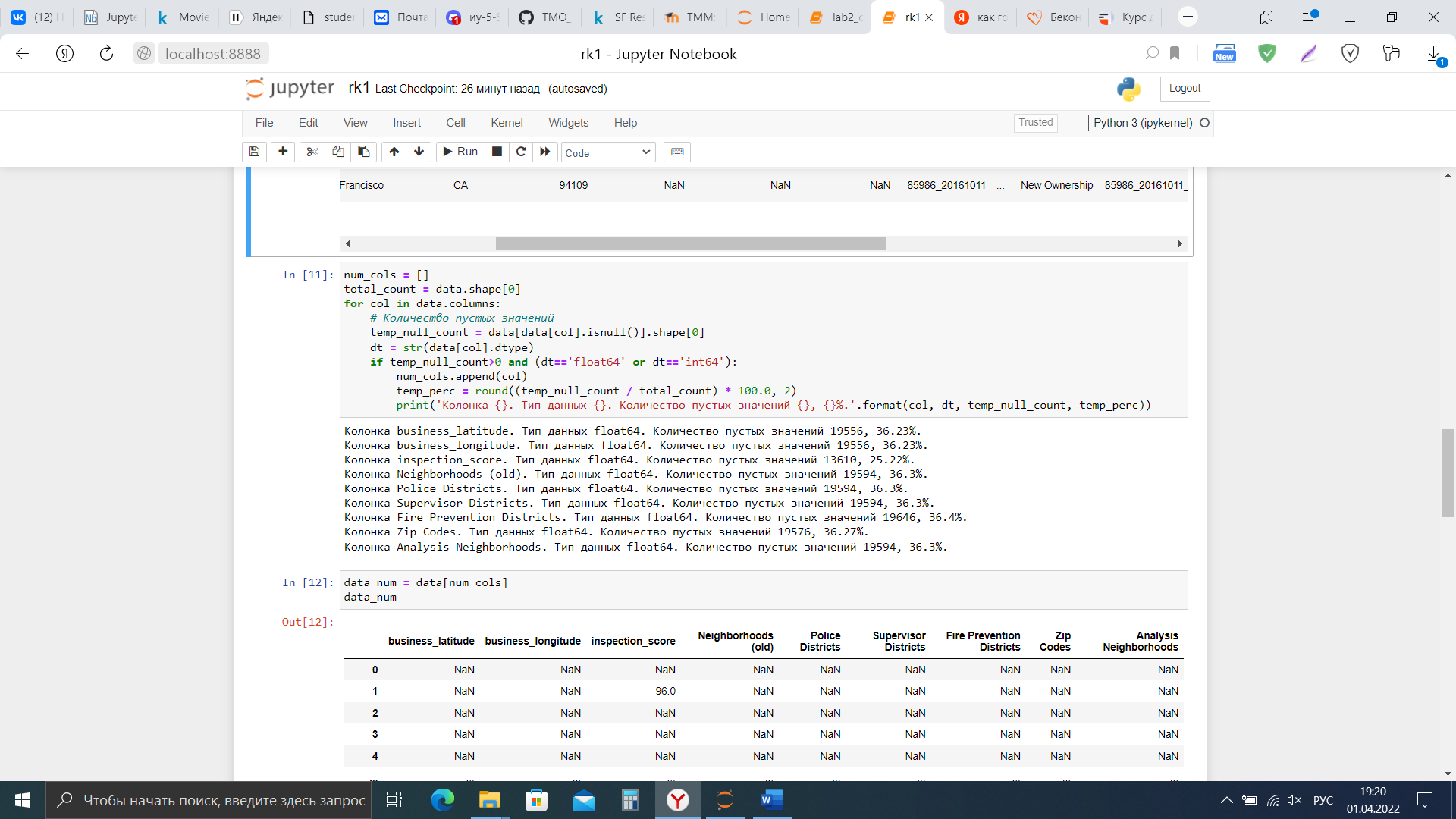




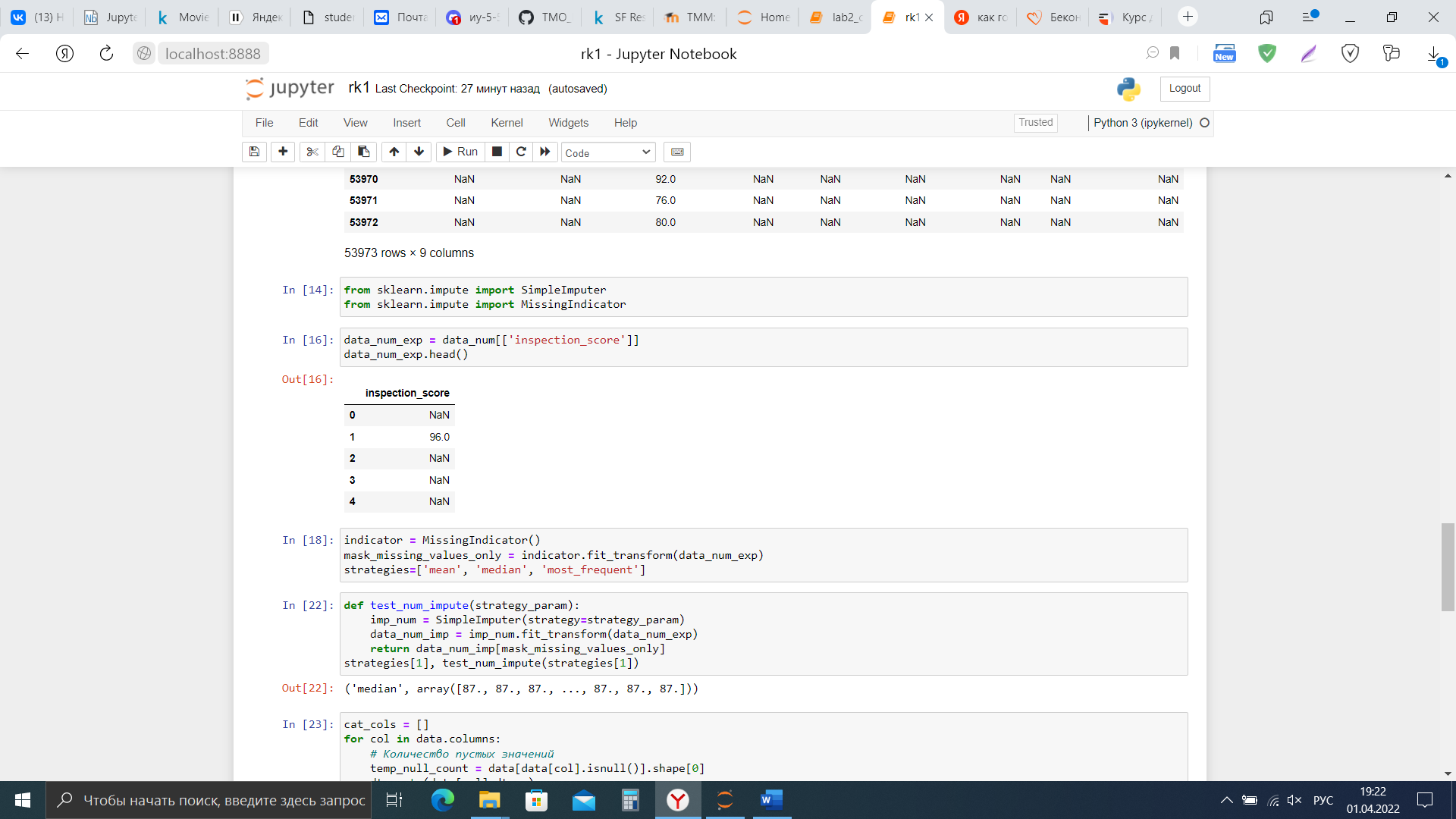
Так как в столбце business\_phone\_number много пропусков удалим его.



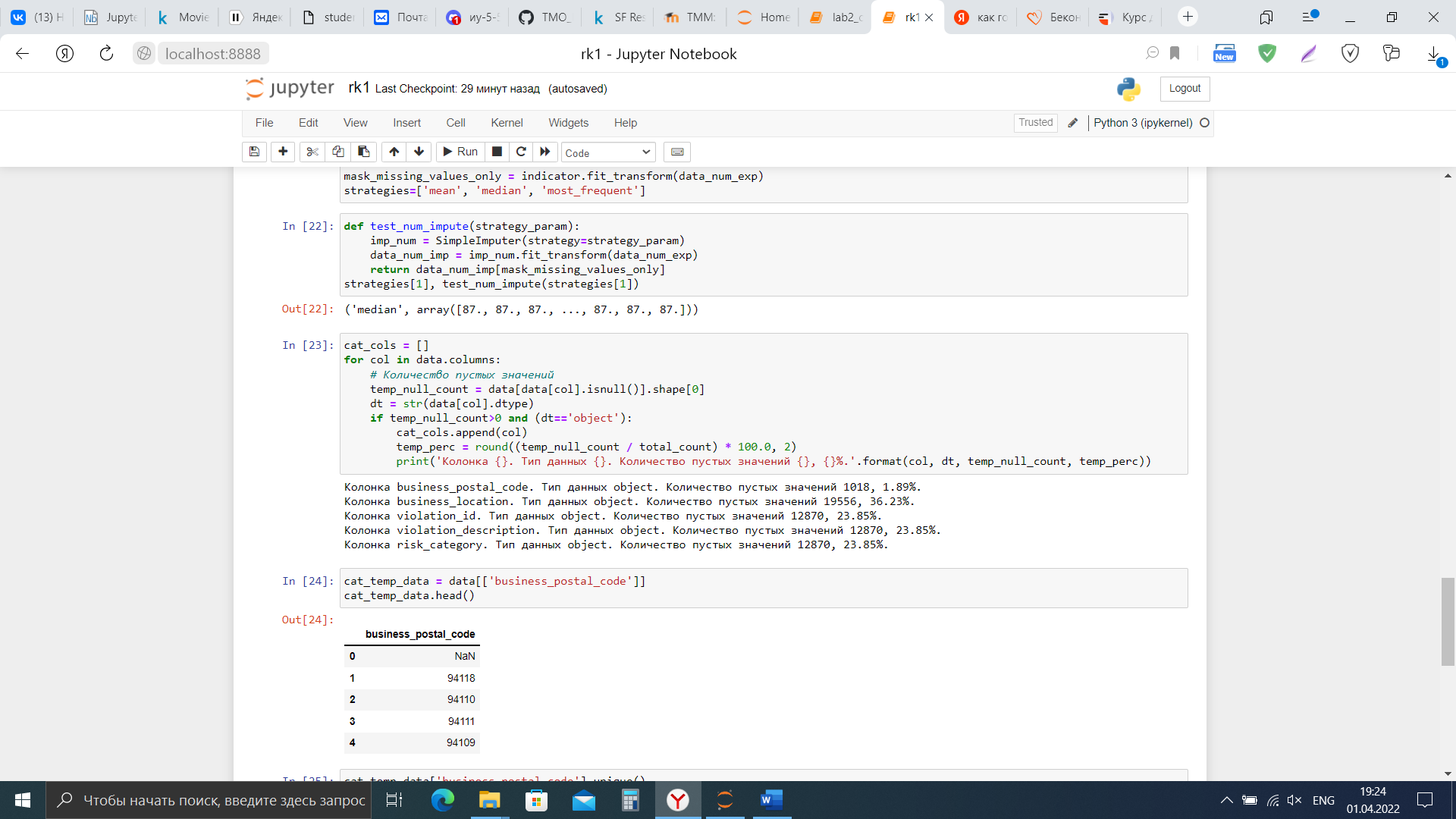
Определим количество пропусков по количественному признаку



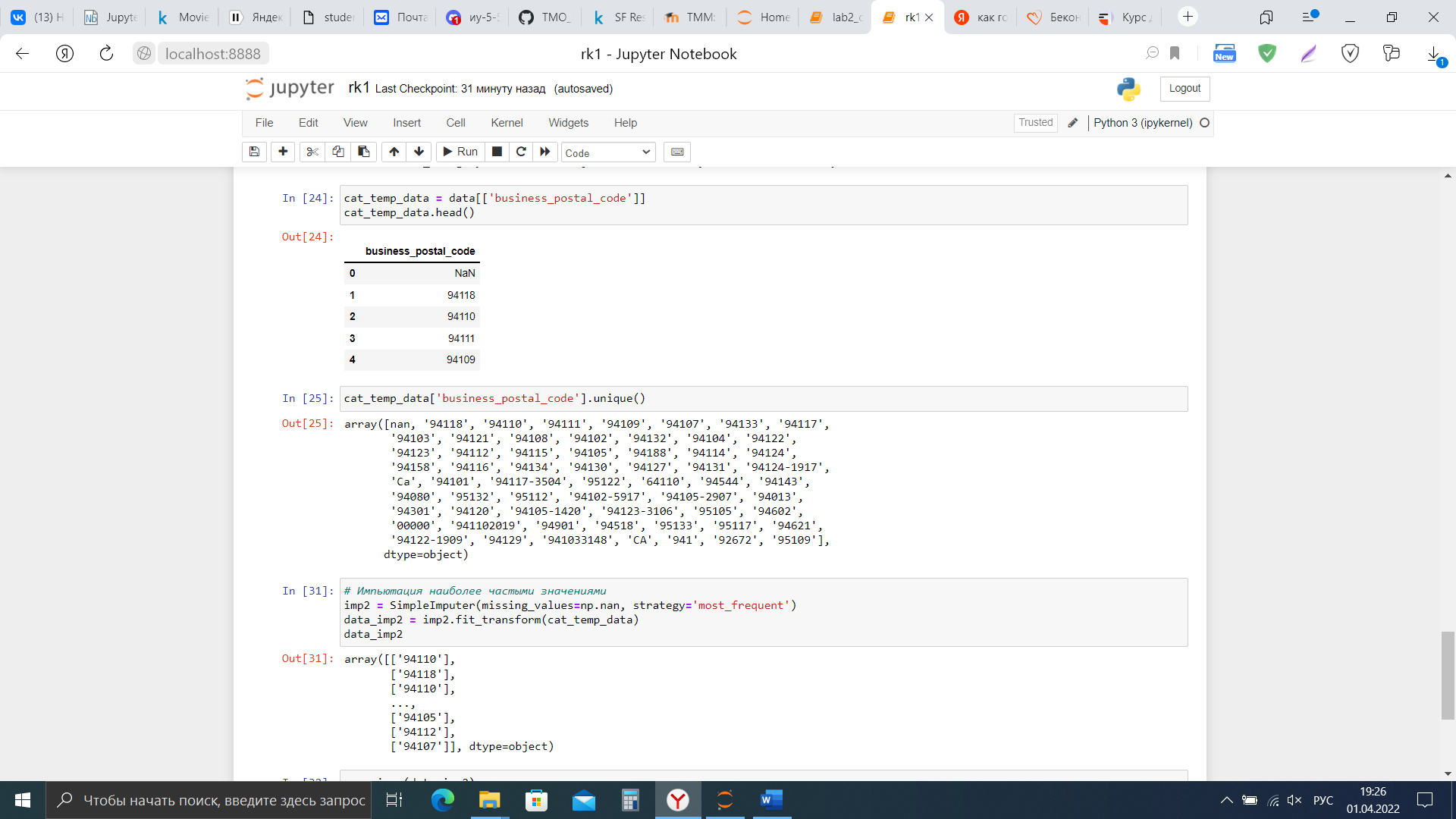
Возьмем столбец inspection\_score и заполним пропуски в нем медианным значением



Далее определим количество пропусков по категориальному признаку



Возьмем столбец business\_postal\_code и заполним пропуски в нем самым часто встречающимся значением



Для дальнейшего построения моделей машинного обучения можно взять признаки: business\_name, inspection\_date, inspection\_score, inspection\_type, violation\_description, risk\_category, т.к. они имеют наибольшую значимость.