|  |  |
| --- | --- |
| Gerb-BMSTU_01 | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

ФАКУЛЬТЕТ Информатика и системы управления

КАФЕДРА Системы обработки информации и управления

**Отчёты по лабораторной работе №3**

По дисциплине:

«Технологии машинного обучения»

Выполнил:

Студент группыИУ5-62 **\_\_\_ \_ \_ \_\_\_\_ \_ Ким Р.И. \_\_**

(Подпись, дата) (Фамилия И.О.)

Проверил:

**\_\_\_\_\_\_\_\_\_ \_ Гапанюк Ю. Е.\_**

(Подпись, дата) (Фамилия И.О.)

Москва, 2020

**Описание задания**

Цель лабораторной работы: изучение способов предварительной обработки данных для дальнейшего формирования моделей.

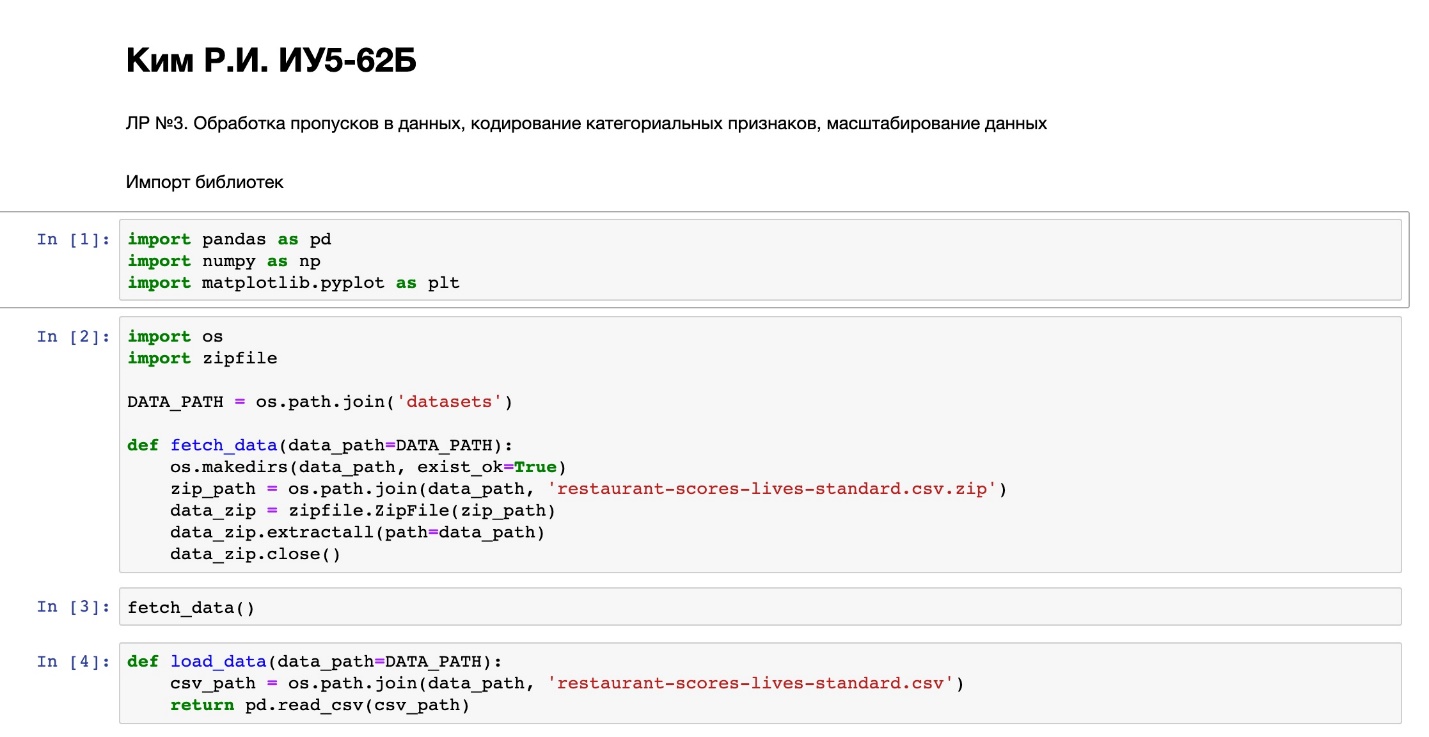
Задание:

1. Выбрать набор данных (датасет), содержащий категориальные признаки и пропуски в данных. Для выполнения следующих пунктов можно использовать несколько различных наборов данных (один для обработки пропусков, другой для категориальных признаков и т.д.)
2. Для выбранного датасета (датасетов) на основе материалов [лекции](https://github.com/ugapanyuk/ml_course_2020/blob/master/common/notebooks/missing/handling_missing_norm.ipynb) решить следующие задачи:

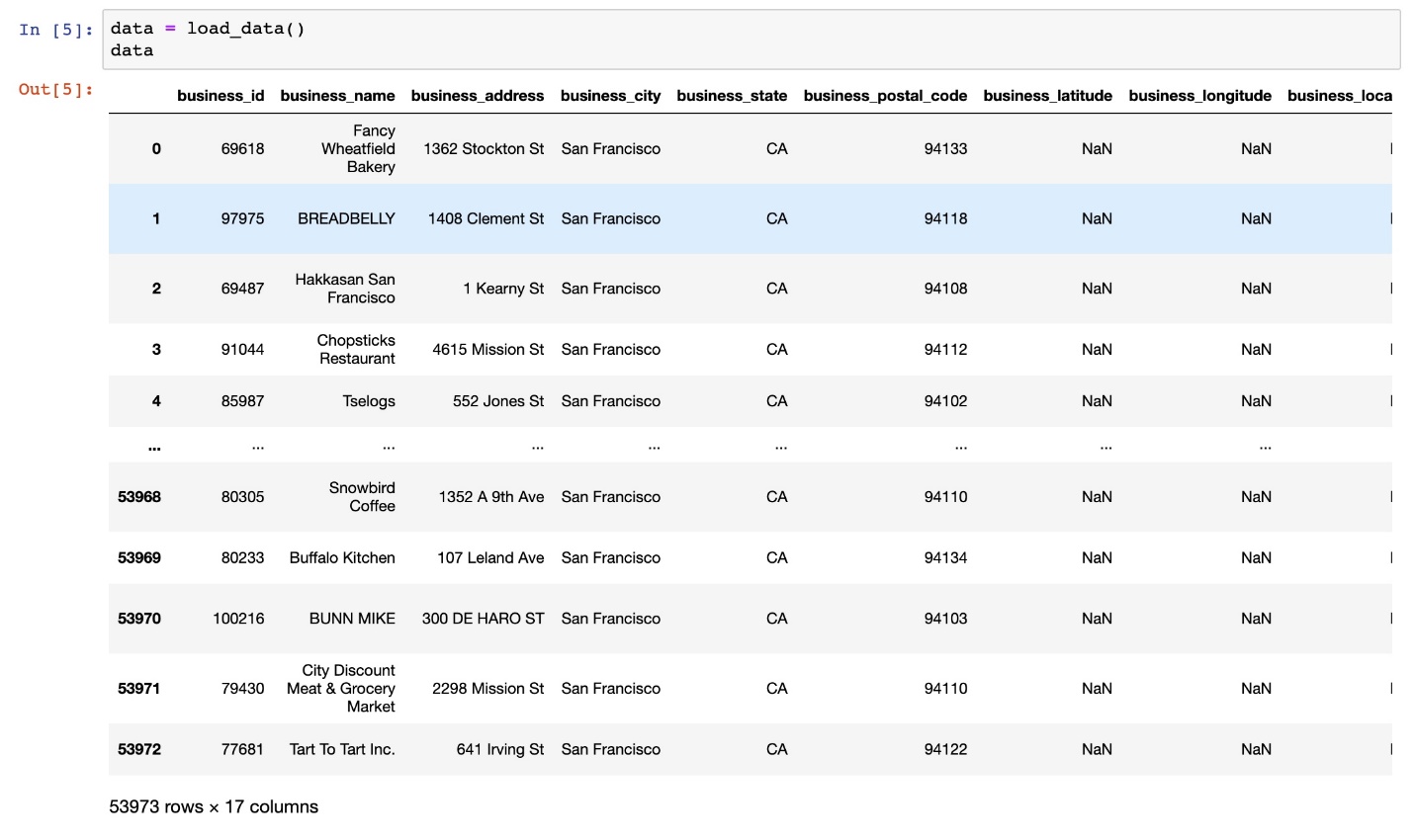
* обработку пропусков в данных;
* кодирование категориальных признаков;
* масштабирование данных.

**Текст программы и экранные формы с примерами выполнения программы**

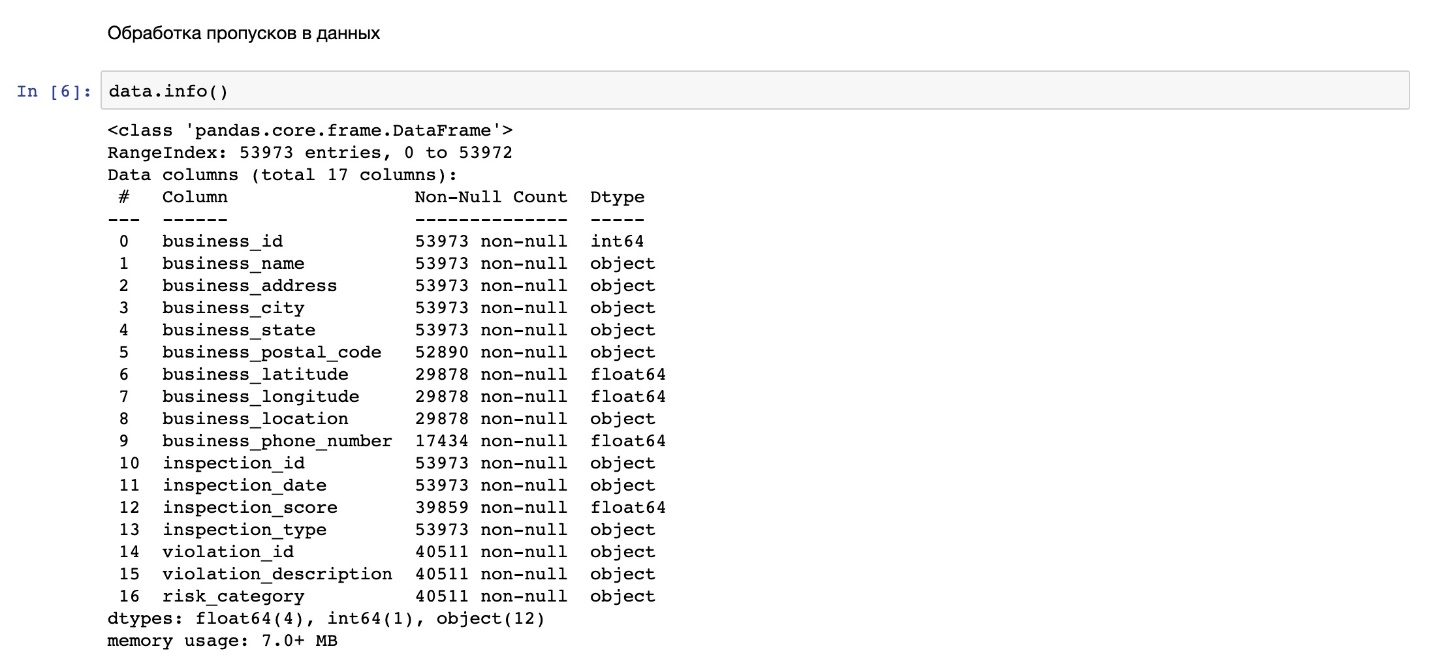
**Импорт библиотек**

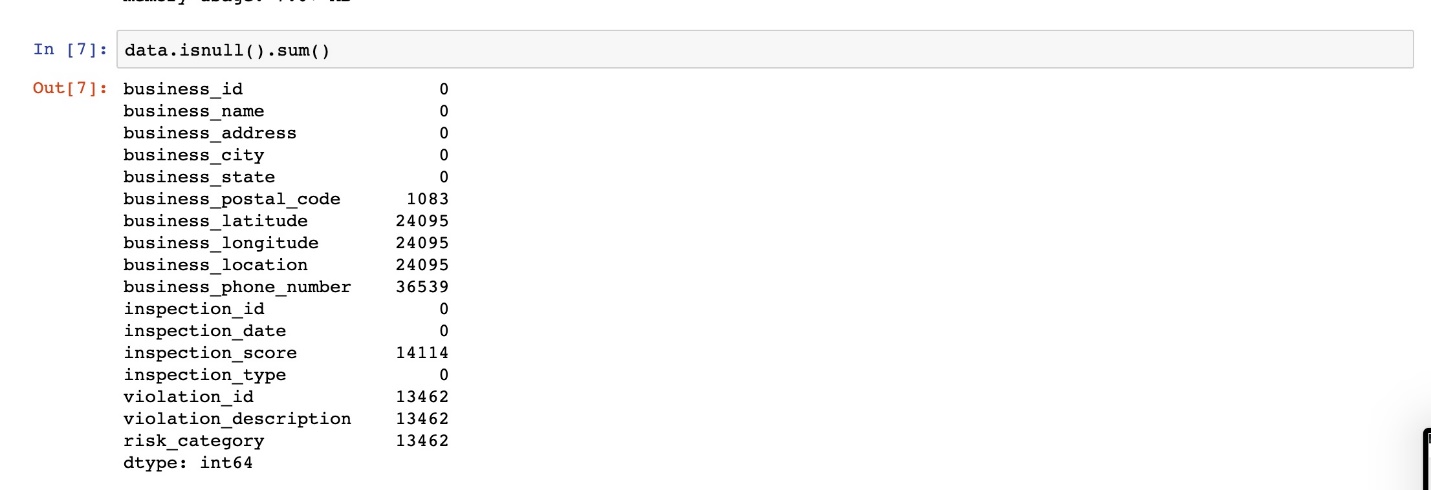


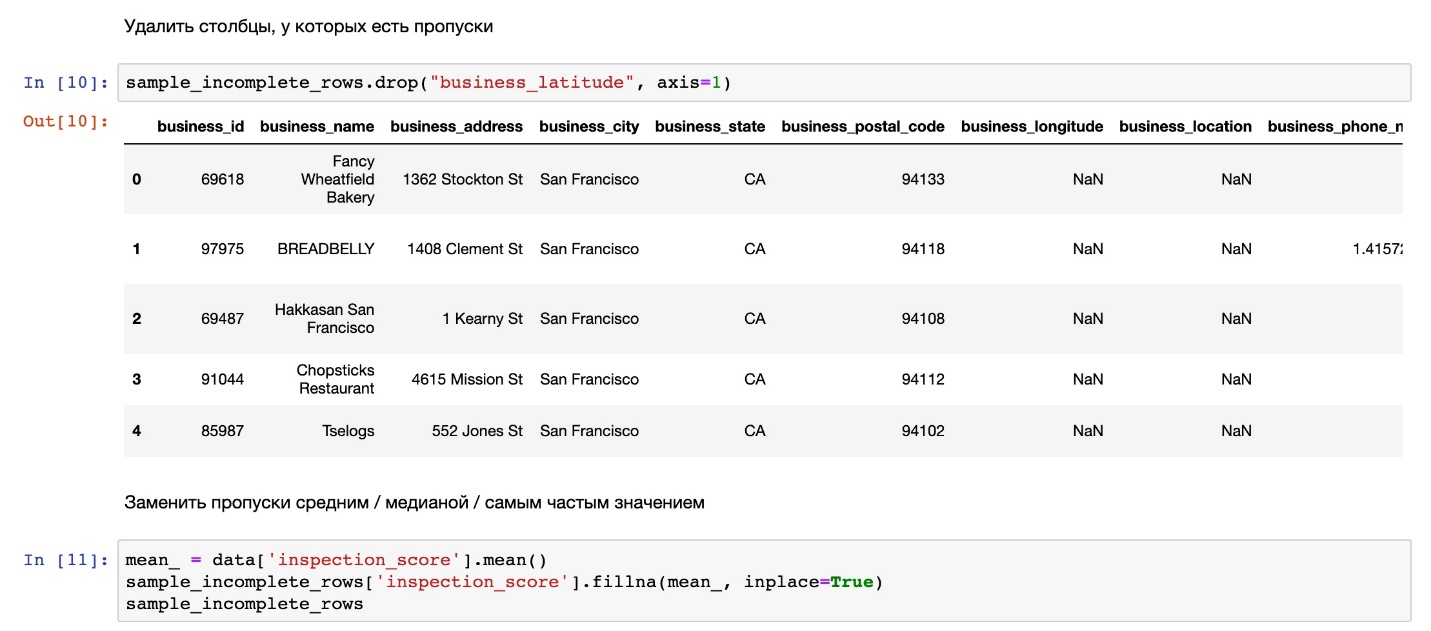
**Загрузка данных**

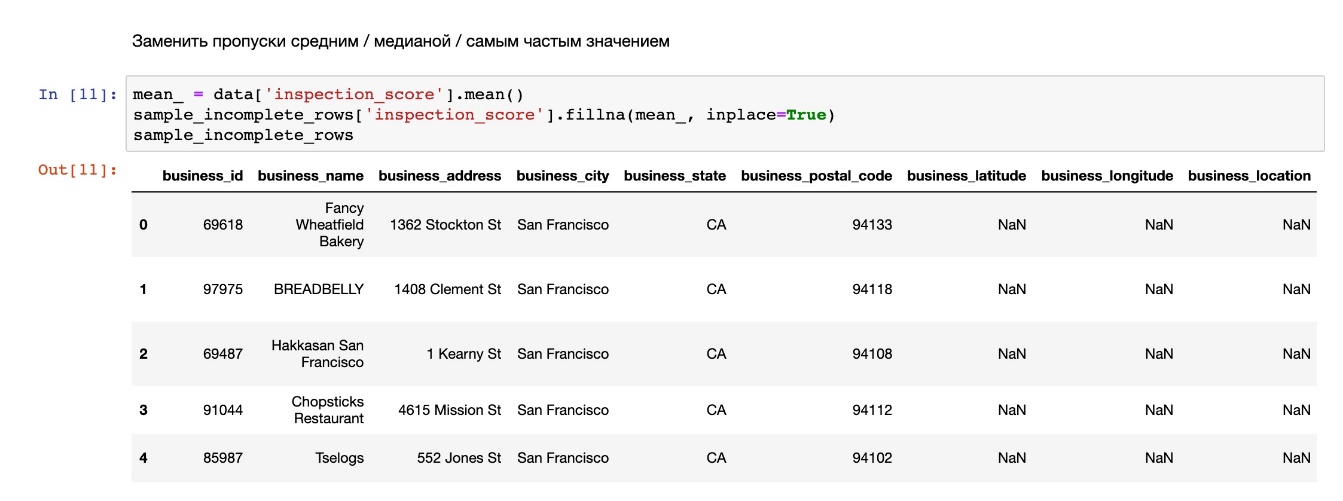


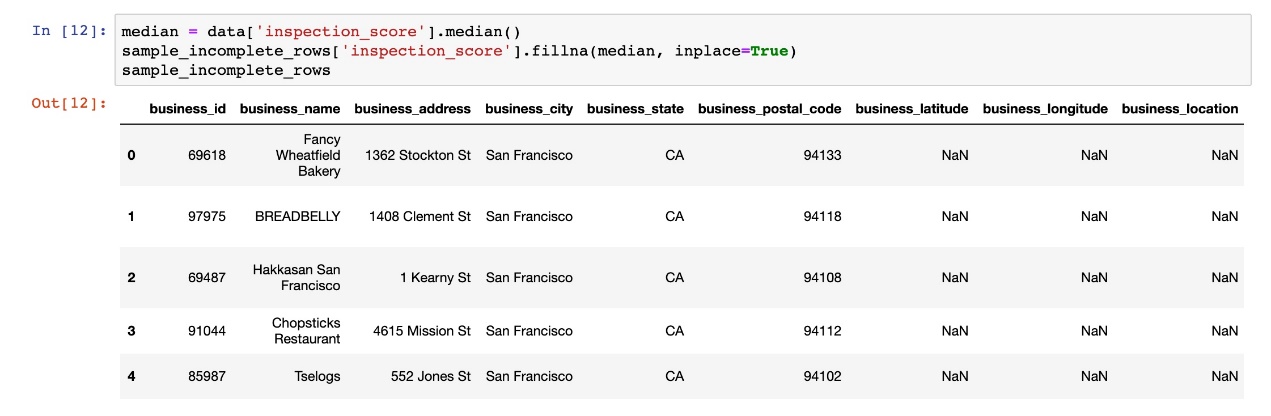
**Обработка пропусков в данных**



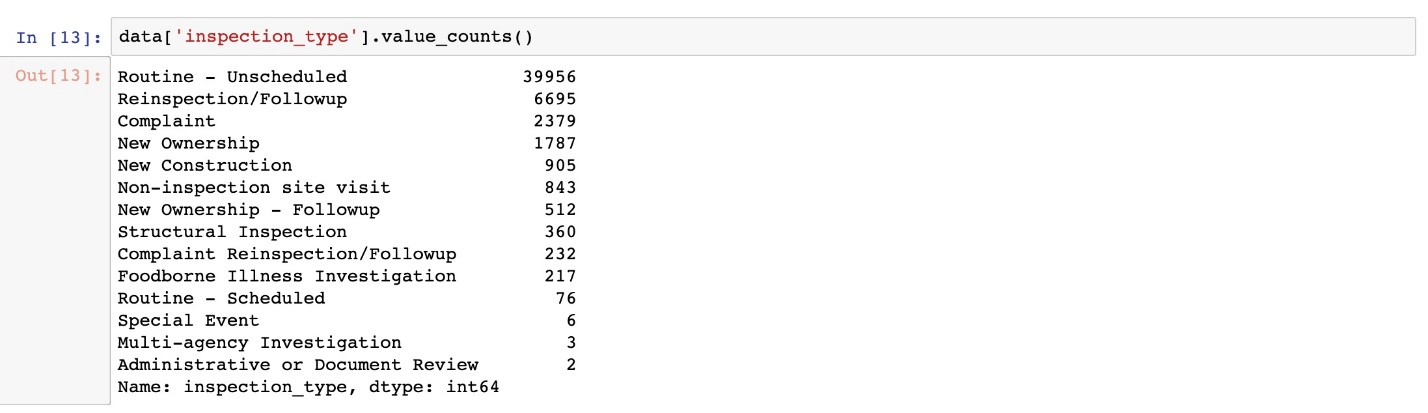




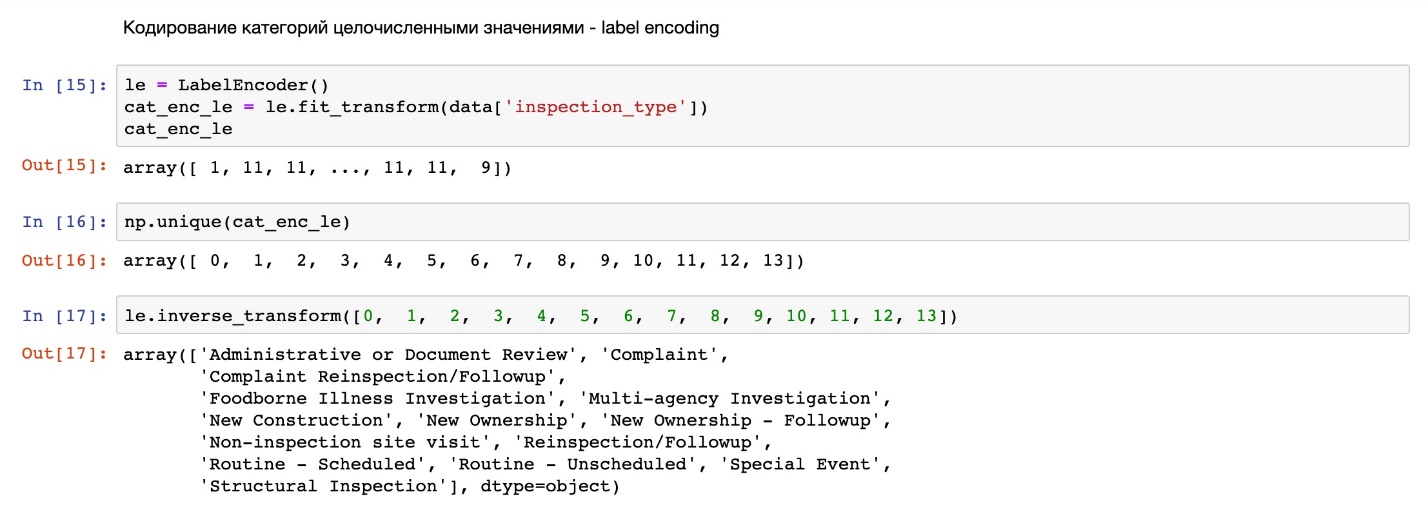


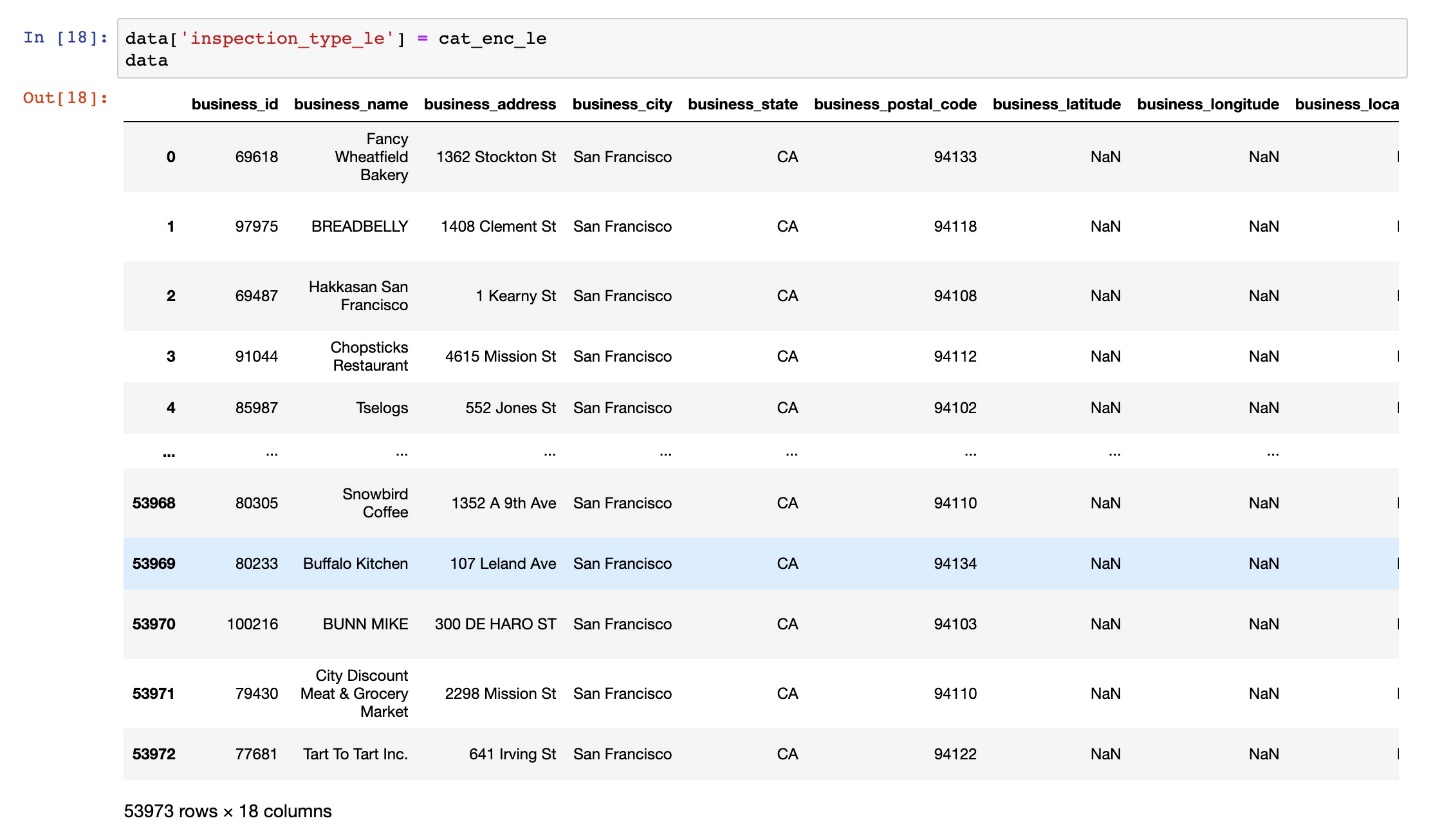


**Проверим, что получилось**

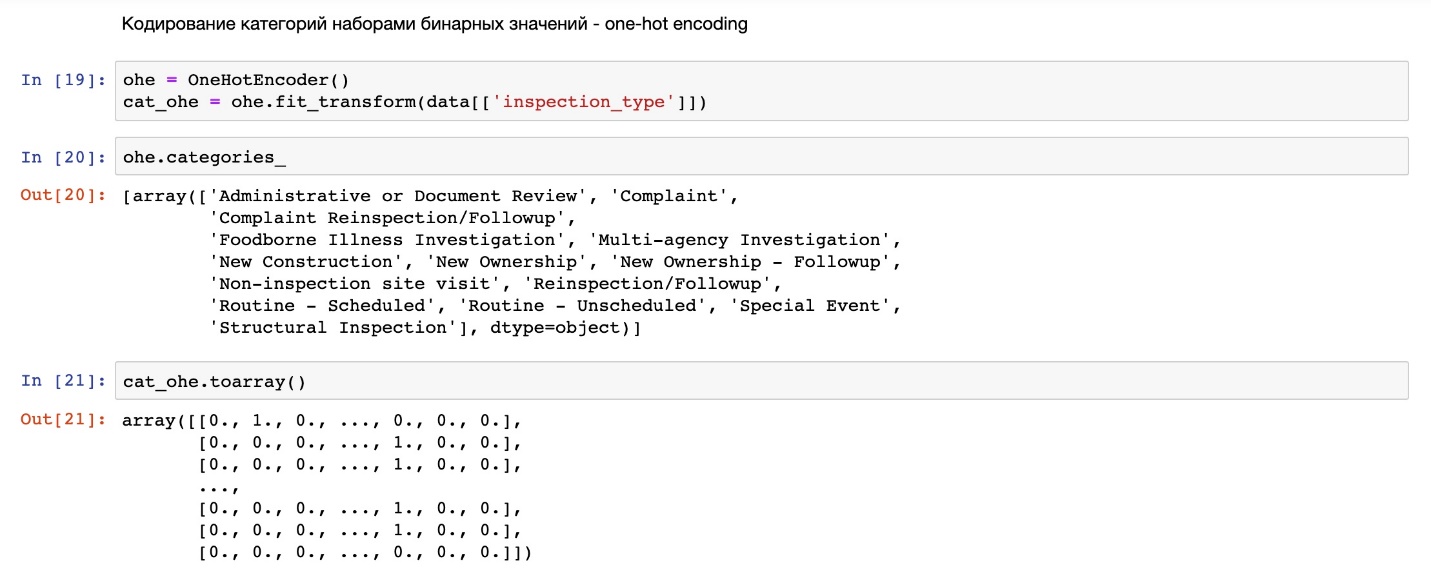


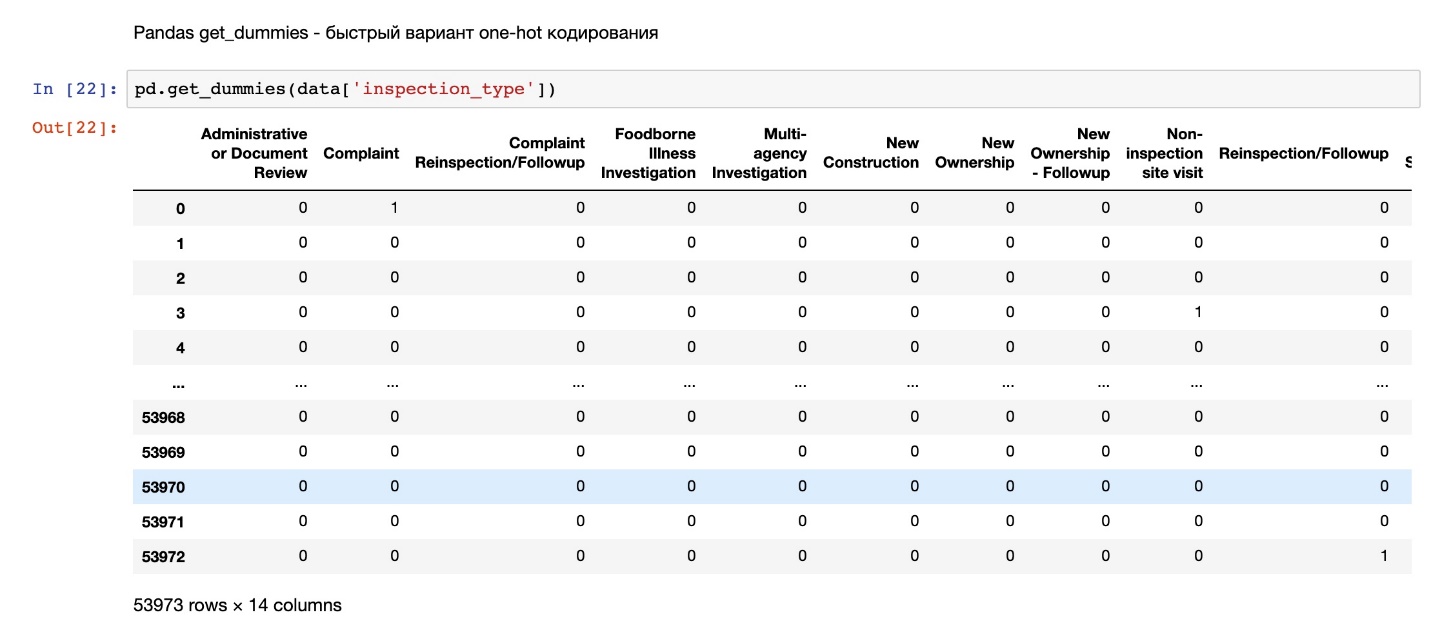
**Кодирование категорий целочисленными значениями**





**Кодирование категорий наборами бинарных значений**





**Масштабирование**

