Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»



**Отчет по ЛР №3**

**по курсу «Технологии машинного обучения»**

**«Обработка пропусков в данных, кодирование категориальных признаков, масштабирование данных»**

**ИСПОЛНИТЕЛЬ:**

Кемерова А.М.

Группа ИУ5-61Б

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

**ПРЕПОДАВАТЕЛЬ:**

Гапанюк Ю.Е.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

## **Москва 2020**

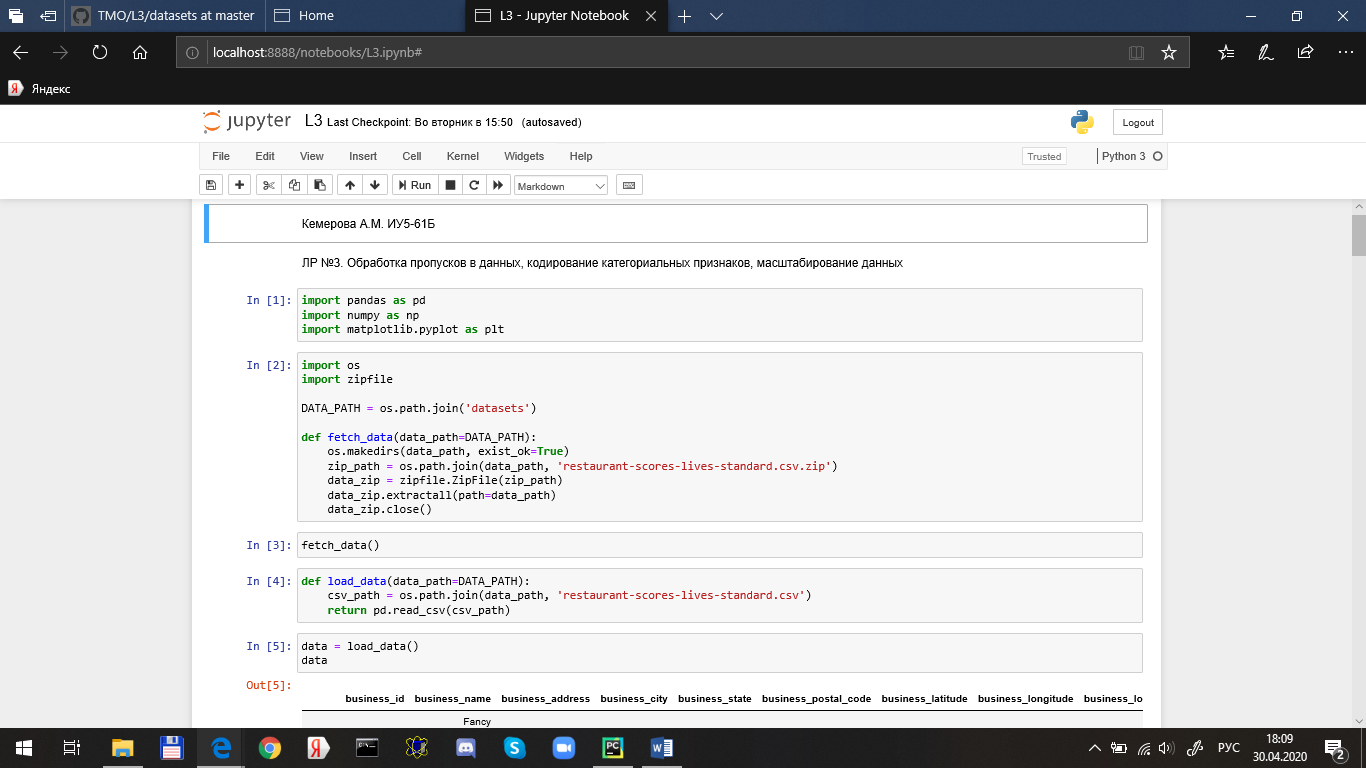
**Цель лабораторной работы:**

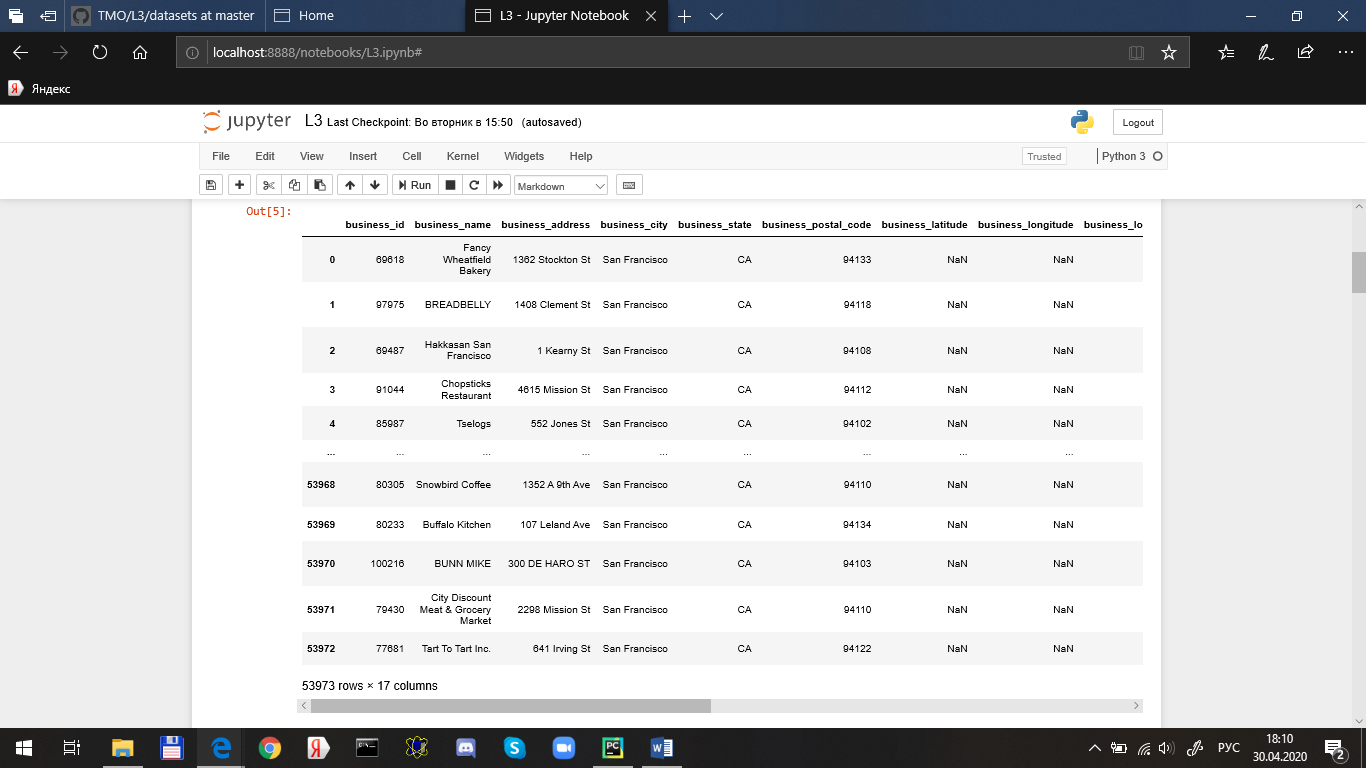
Изучение способов предварительной обработки данных для дальнейшего формирования моделей.

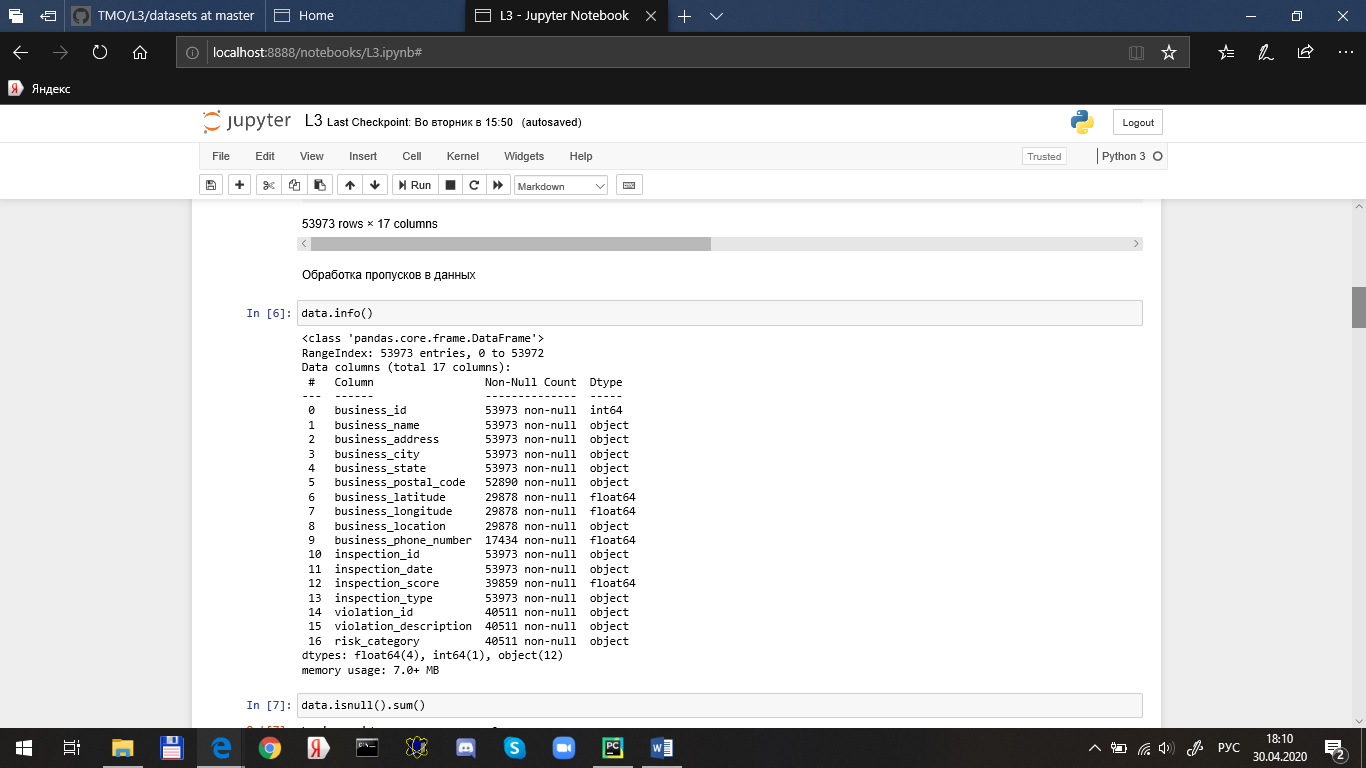
**Задание:**

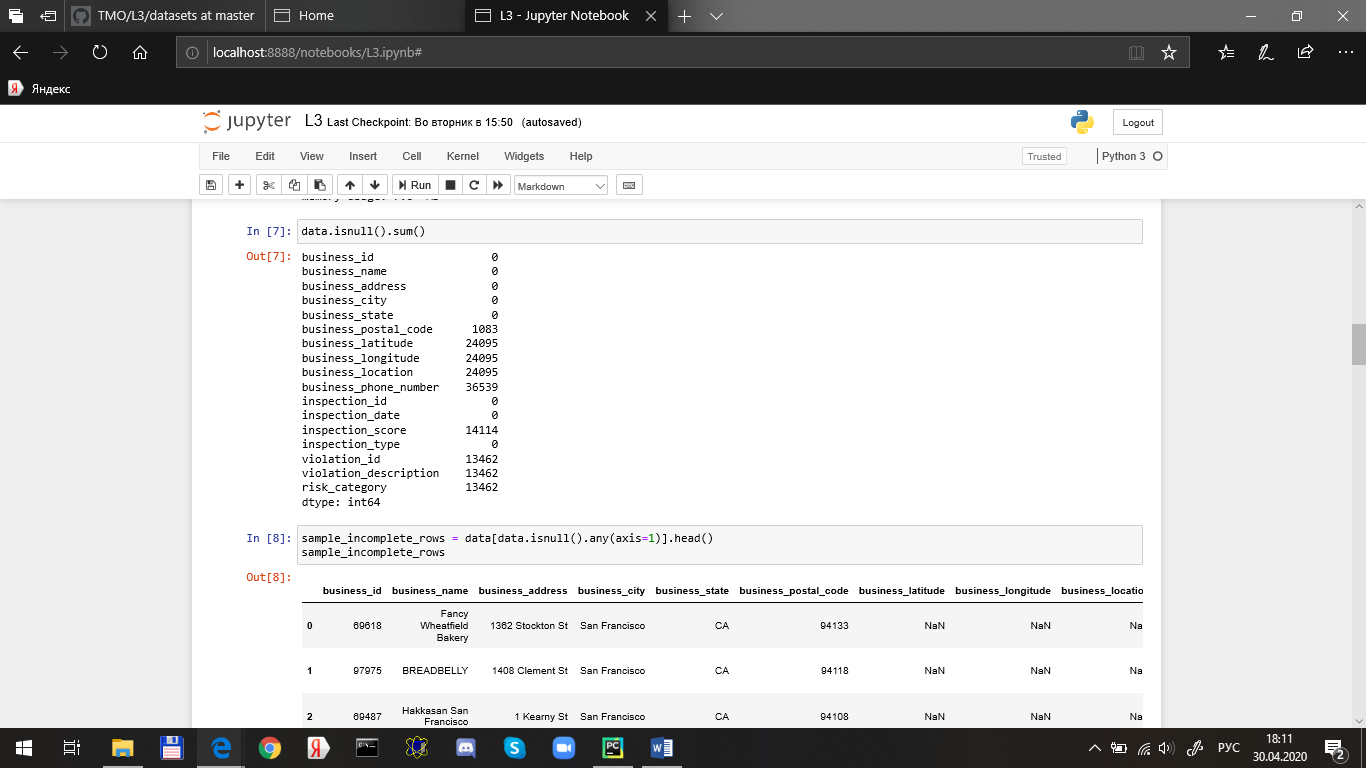
1. Выбрать набор данных (датасет: [restaurant-scores-lives-standard.csv](https://github.com/AngelikaKemerova/TMO/blob/master/L3/datasets/restaurant-scores-lives-standard.csv)), содержащий категориальные признаки и пропуски в данных. Для выполнения следующих пунктов можно использовать несколько различных 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:
   1. обработку пропусков в данных;
   2. кодирование категориальных признаков;
   3. масштабирование данных.

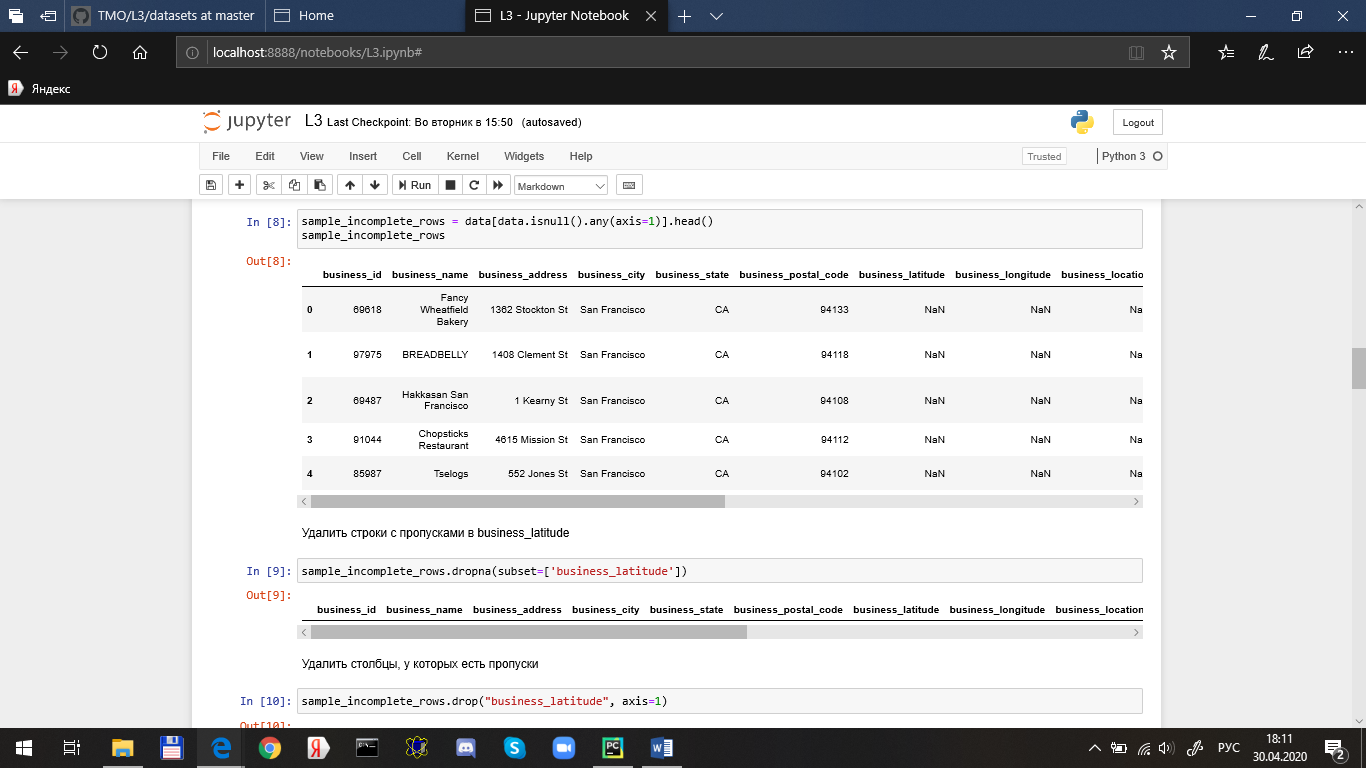
**Выполнение:**

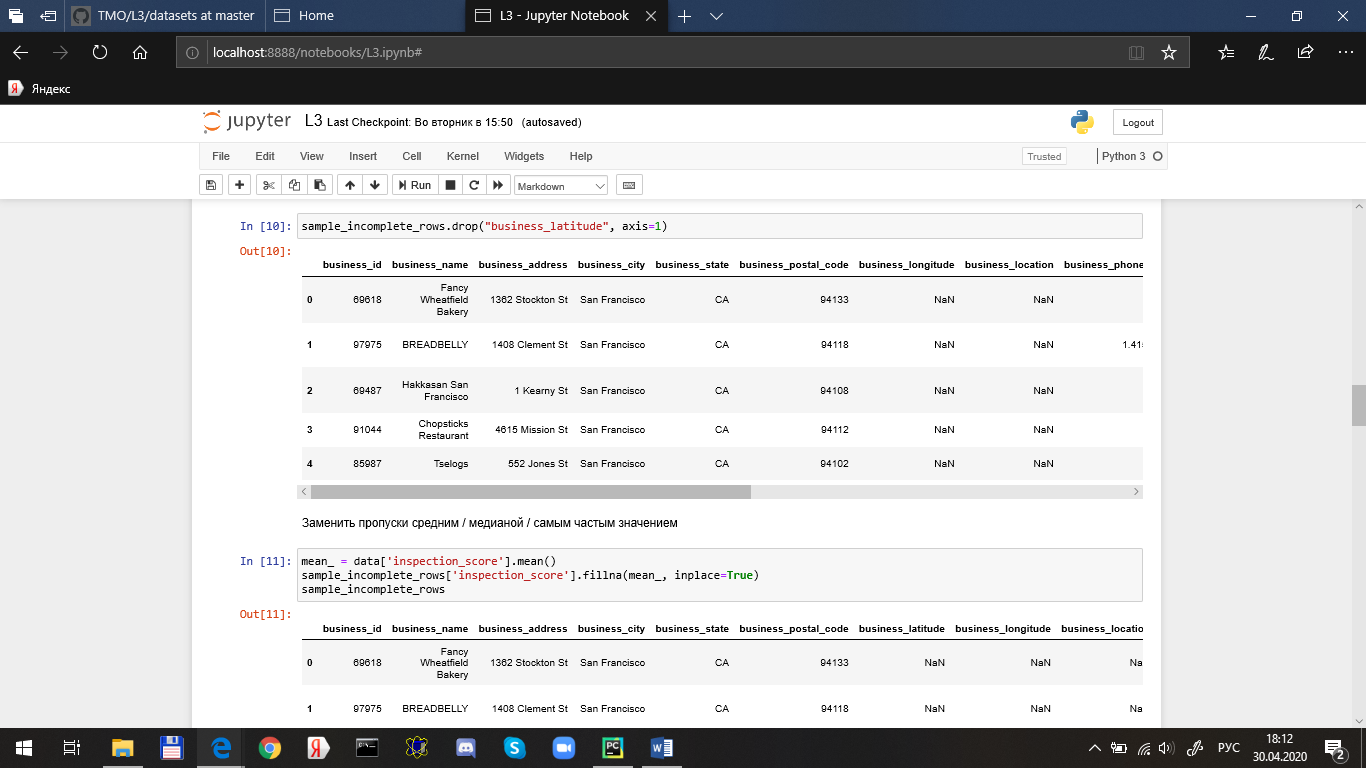


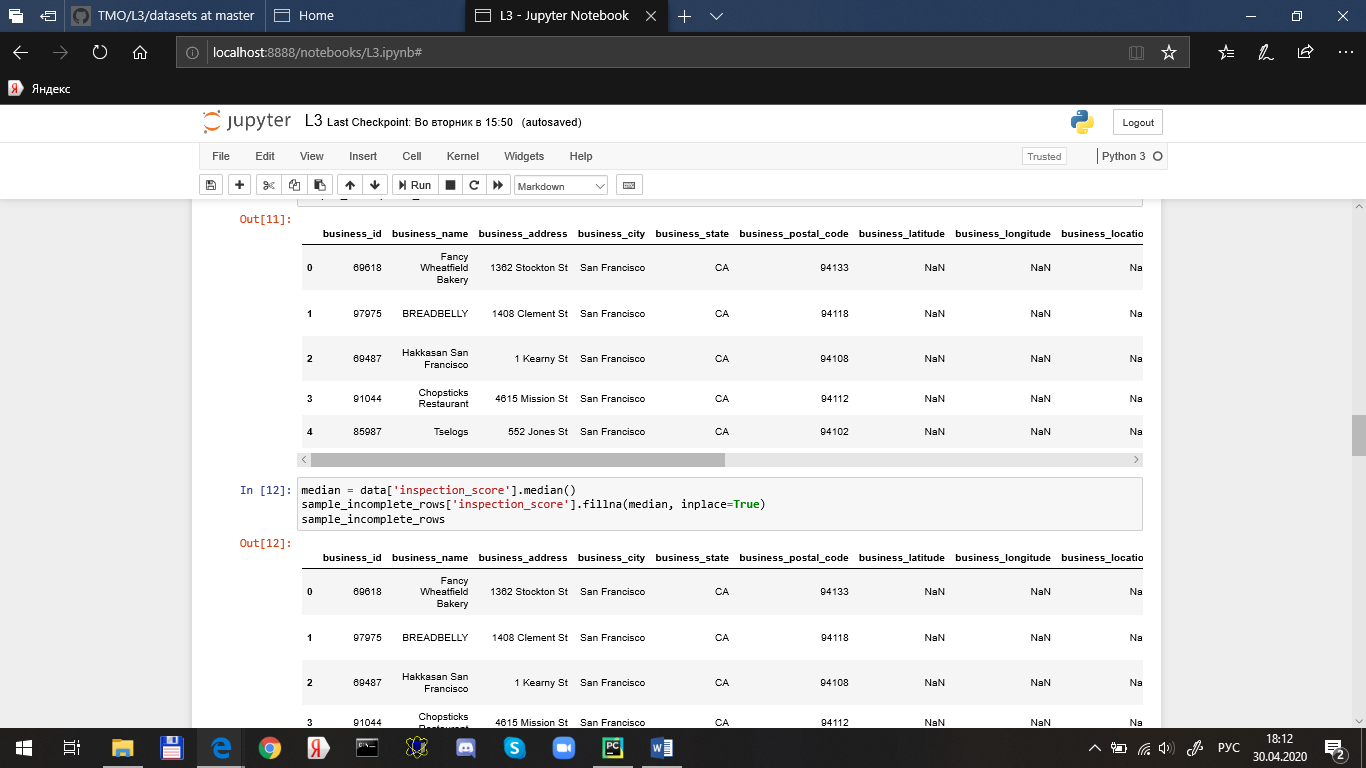


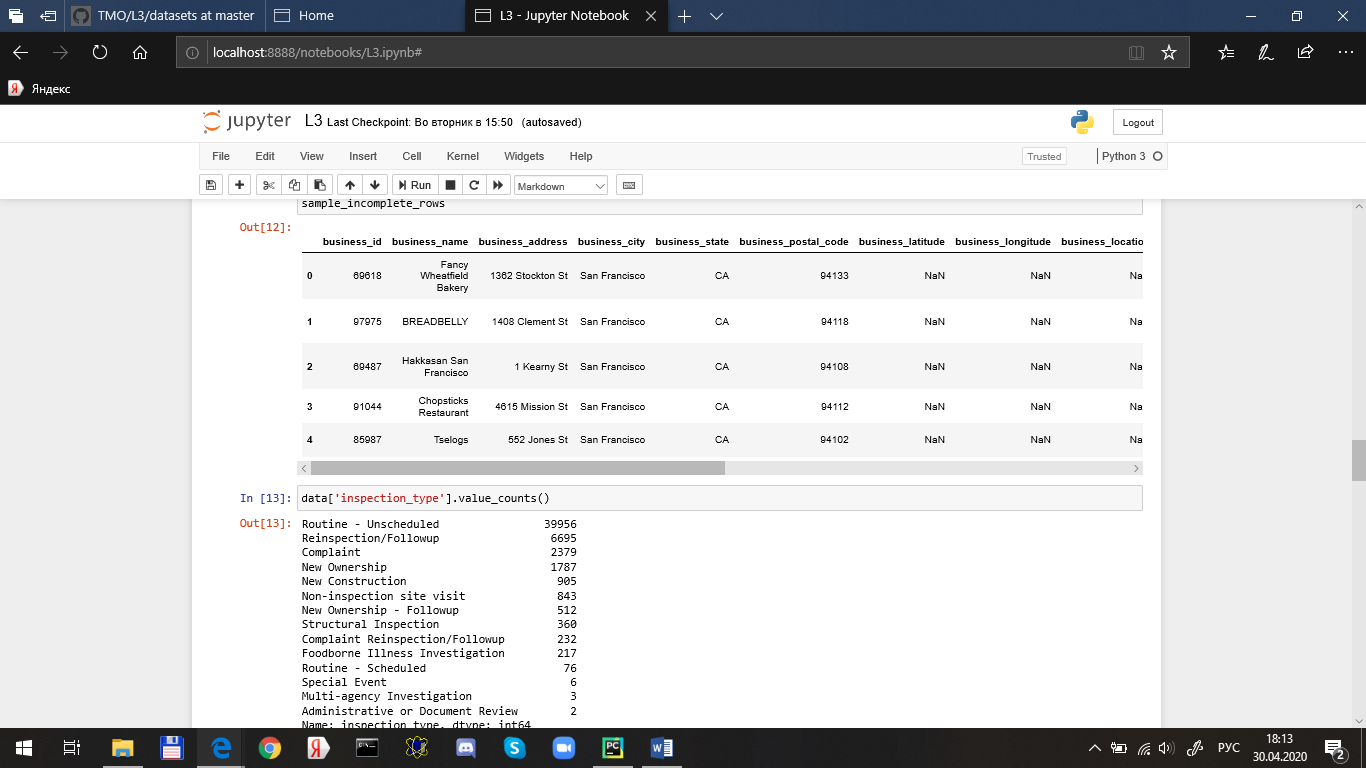


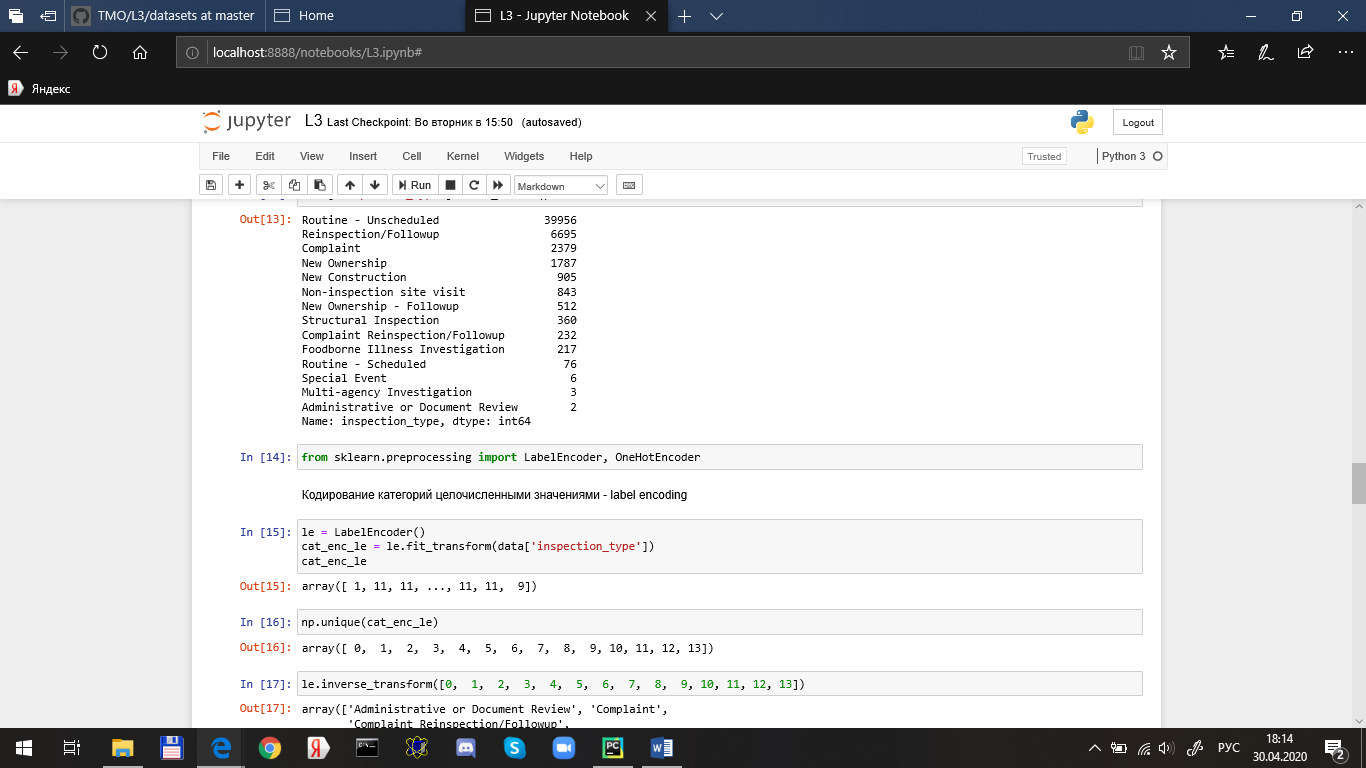


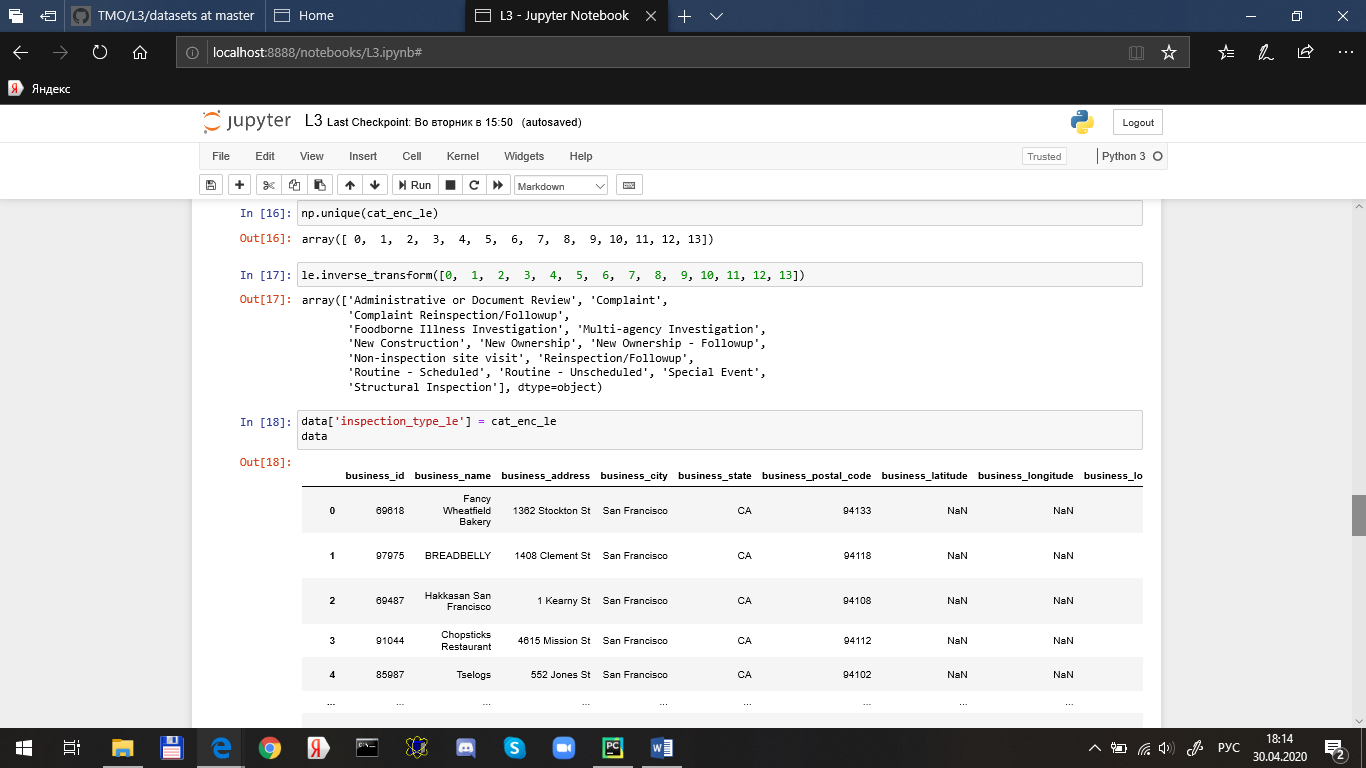


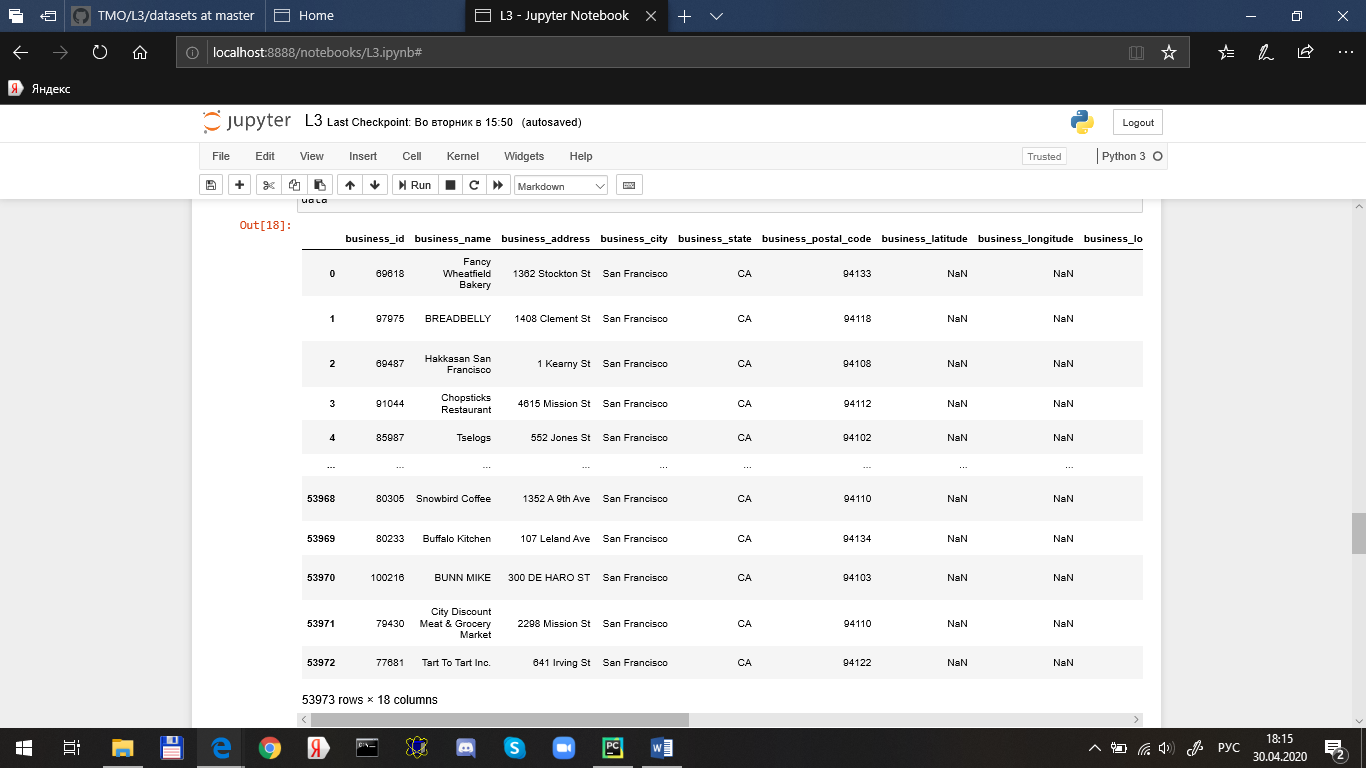


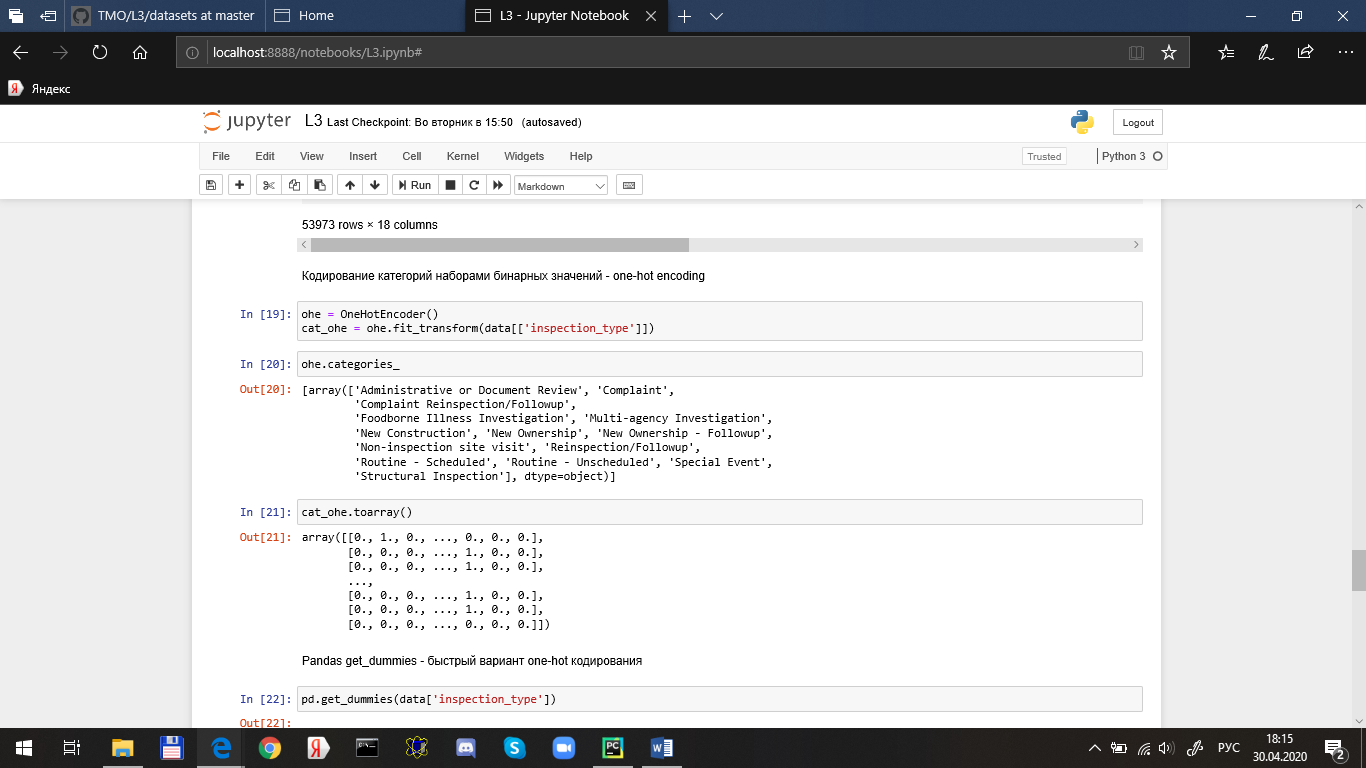


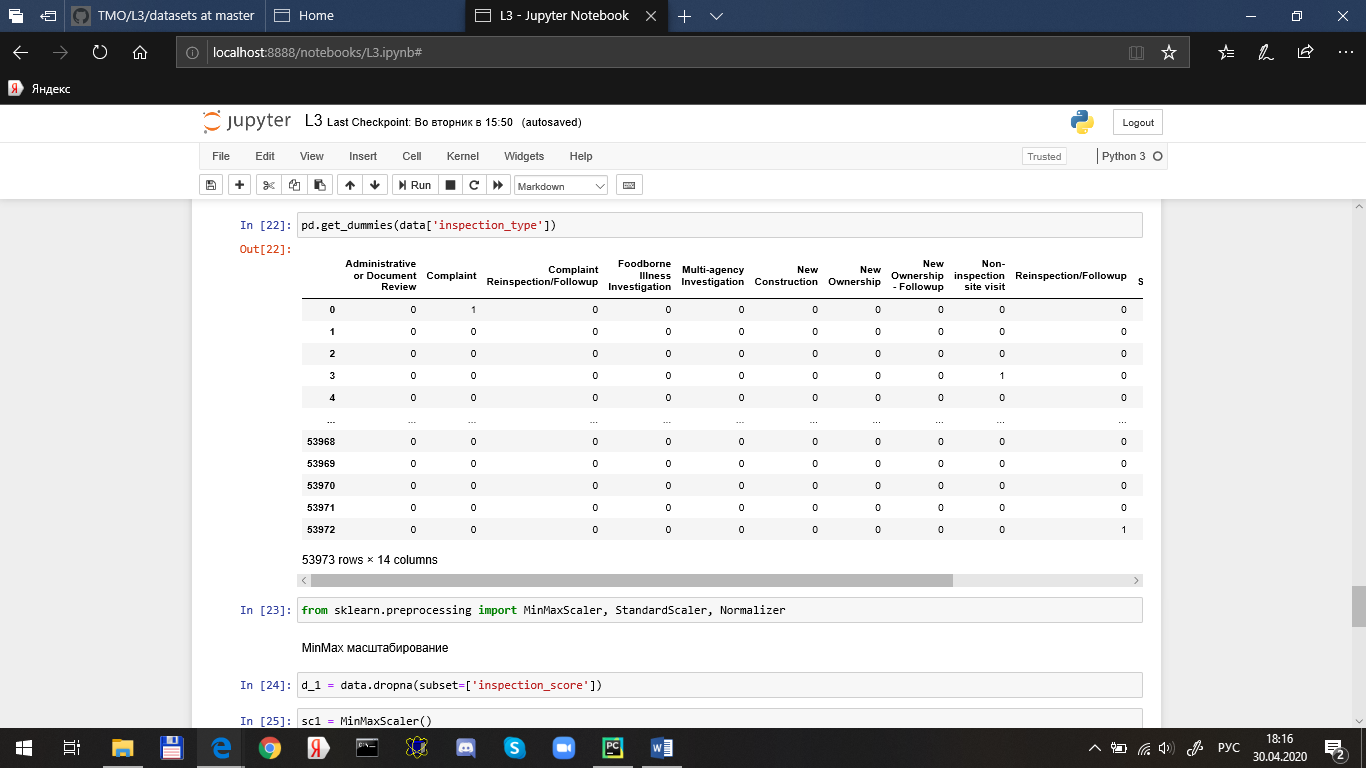


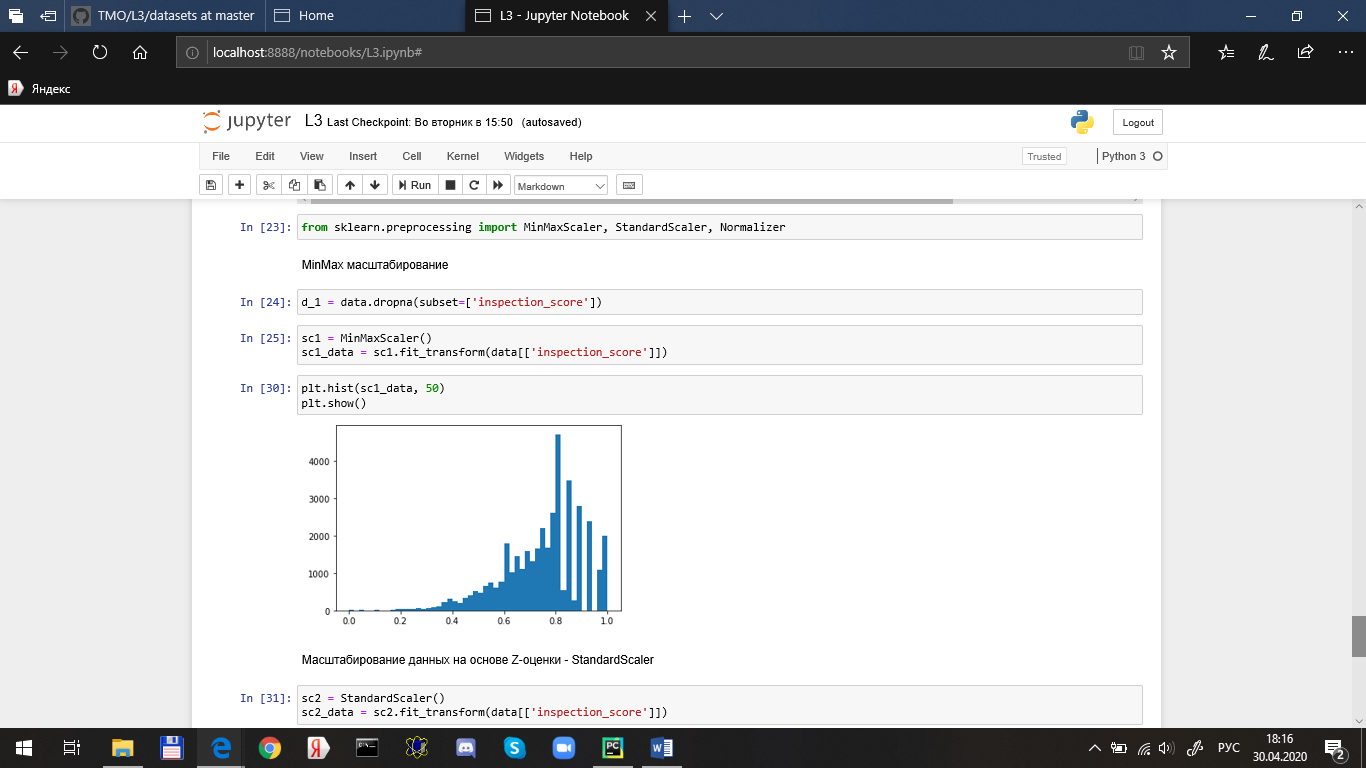


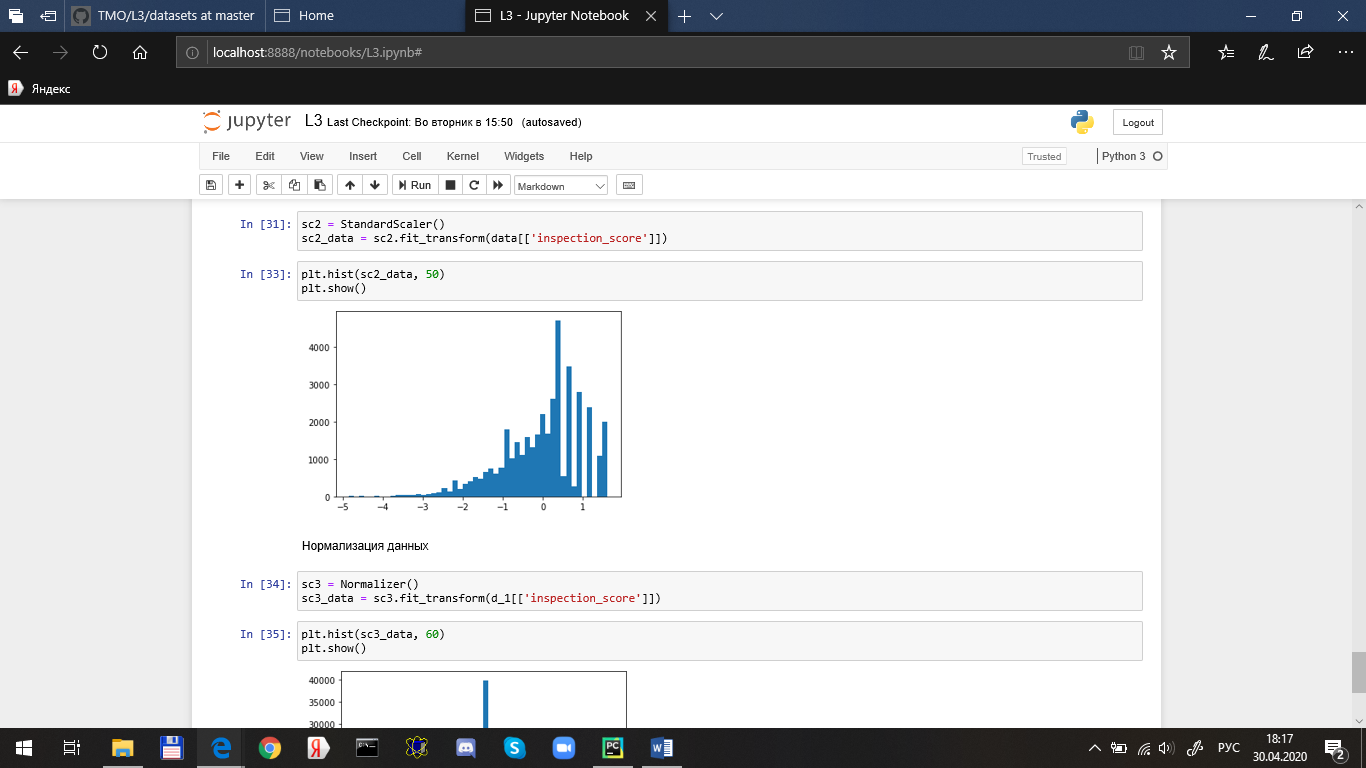


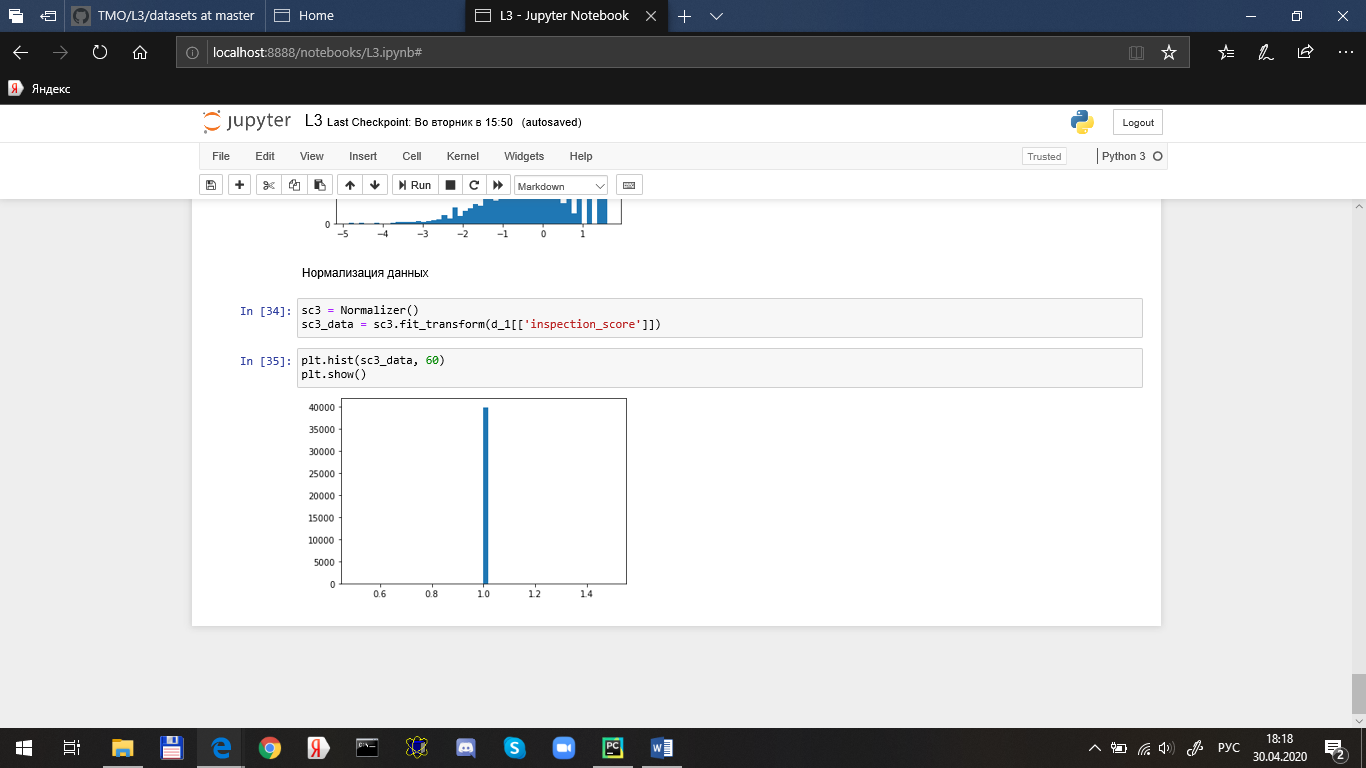












**Вывод:**

Изучила способы предварительной обработки данных для дальнейшего формирования моделей.