Московский государственный технический университет им. Н.Э. Баумана

Факультет «Информатика и системы управления»

Кафедра «Системы обработки информации и управления»



**Отчет**

**Лабораторные работа № 3**

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

**По курсу «Технологии машинного обучения»**

**ИСПОЛНИТЕЛЬ:**

Бакланов Н.В.

Группа ИУ5-61Б

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

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

**ПРЕПОДАВАТЕЛЬ:**

Гапанюк Ю.Е.

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

"\_\_"\_\_\_\_\_\_\_\_\_\_\_2020 г.

**Москва 2020**

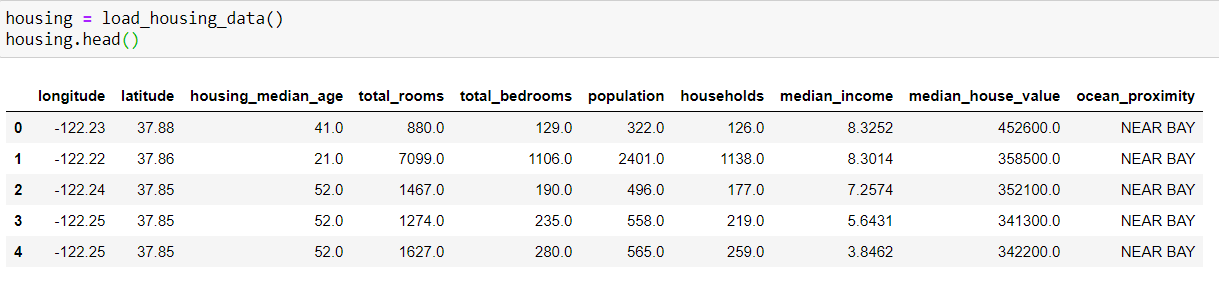
**Цель лабораторной работы:** изучение способов предварительной обработки данных для дальнейшего формирования моделей.

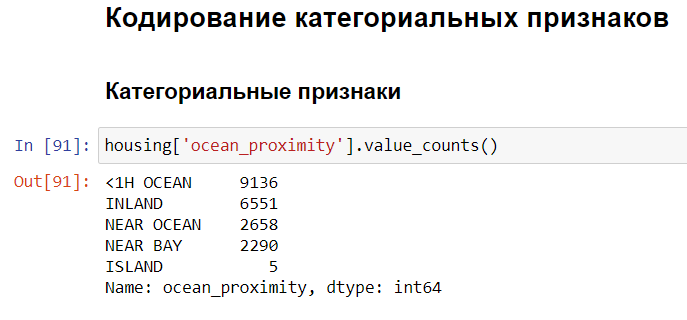
### Задание:

1. Выбрать набор данных (датасет), 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:

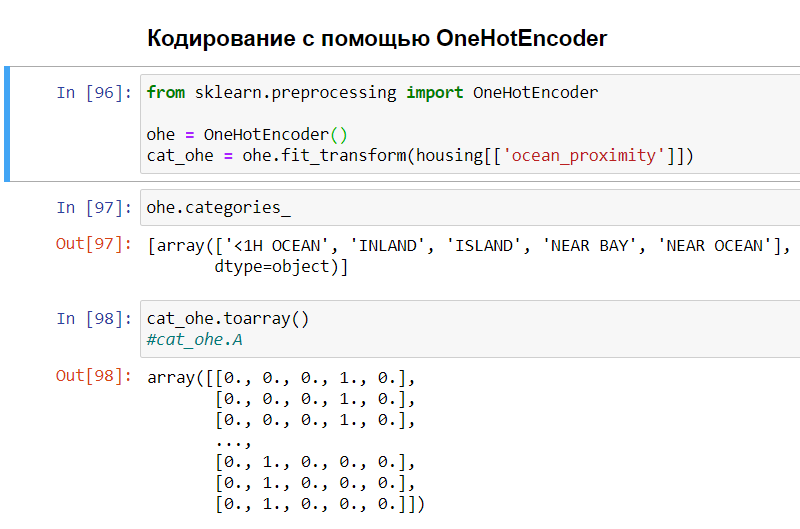
* обработку пропусков в данных;
* кодирование категориальных признаков;
* масштабирование данных.

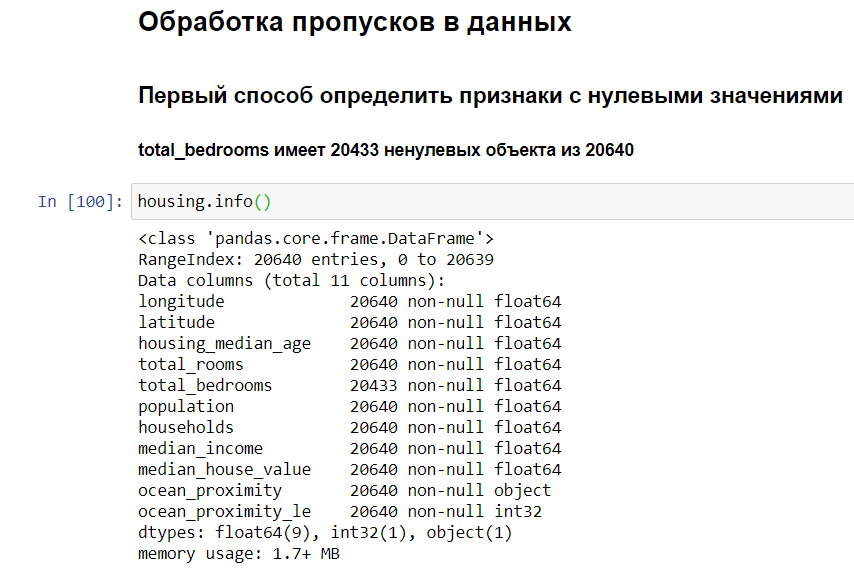
**Решение:**

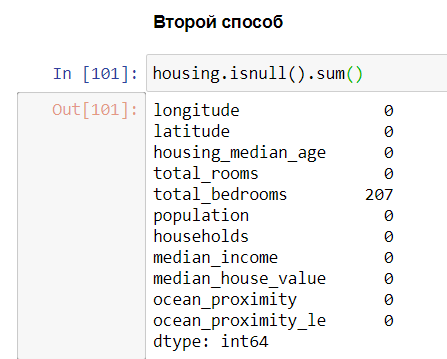
****

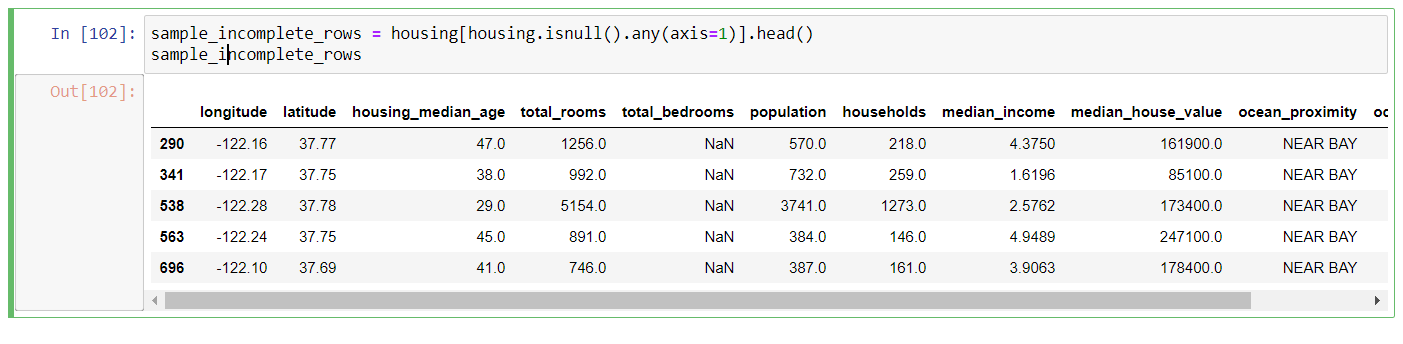
****

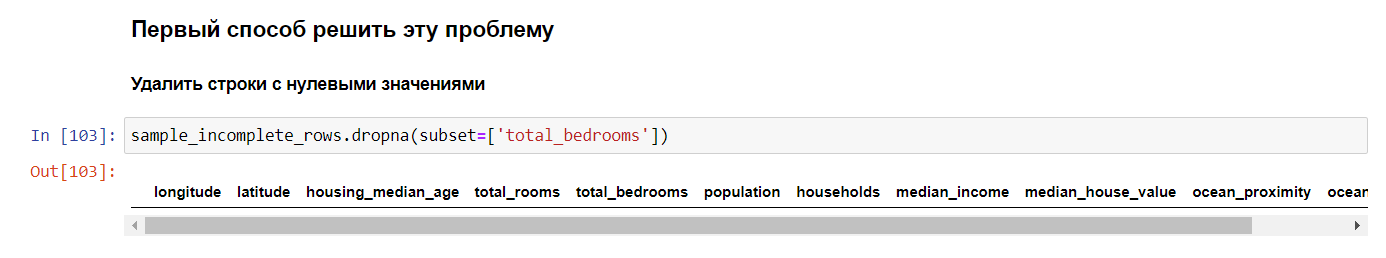
****

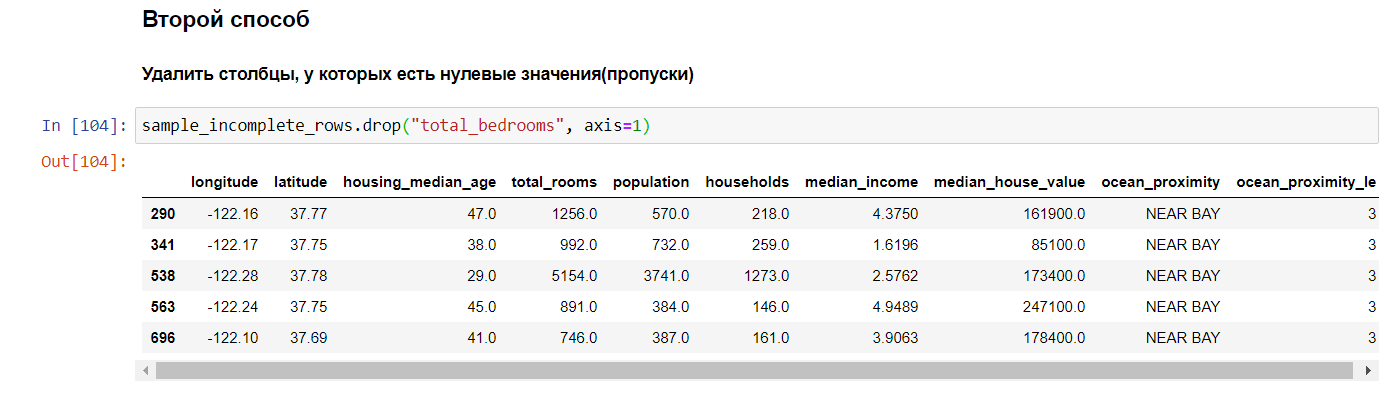
****

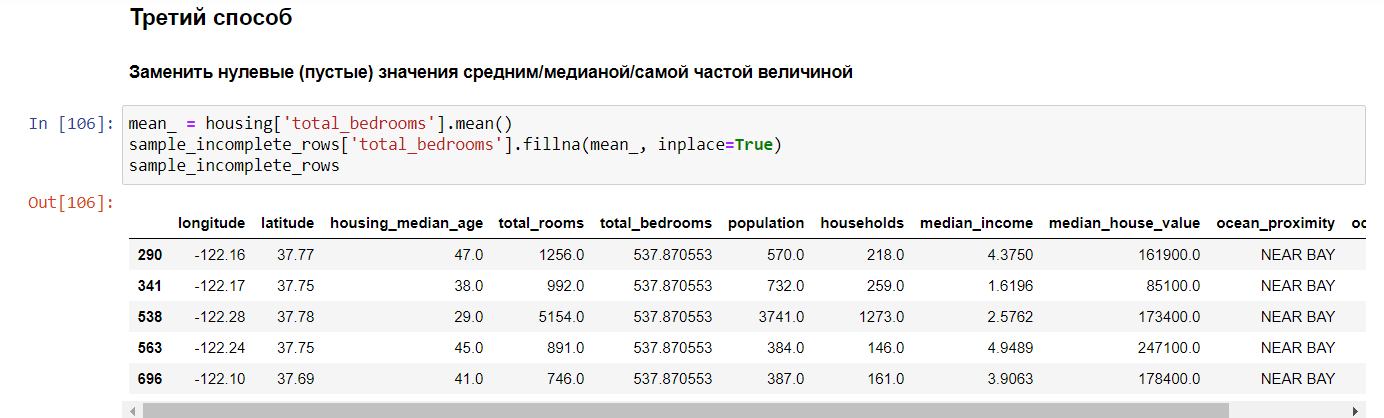
****

****

****

****

****

****

****

Гитхаб: <https://github.com/nikbizkit/TMO>