Московский Государственный Технический Университет

им. Н.Э. Баумана

Отчет по лабораторной работе №3

по курсу

Технологии Машинного Обучения

# Выполнил:

## Муравьев О.М.

ИУ5-62

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

# Проверил:

## Гапанюк Ю.Е.

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

Москва, 2019

# 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/blob/master/common/notebooks/missing/handling_missing_norm.ipynb) решить следующие задачи:

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

# Код и результаты выполнения

Выбираем набор данных. В качестве набора данных возьму набор данных в котором описаны все пассажиры Титаника.

1. Подключим библиотеки:

import numpy as np

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.impute import SimpleImputer

from sklearn.impute import MissingIndicator

from sklearn.preprocessing import LabelEncoder, OneHotEncoder

from sklearn.preprocessing import MinMaxScaler, StandardScaler, Normalizer

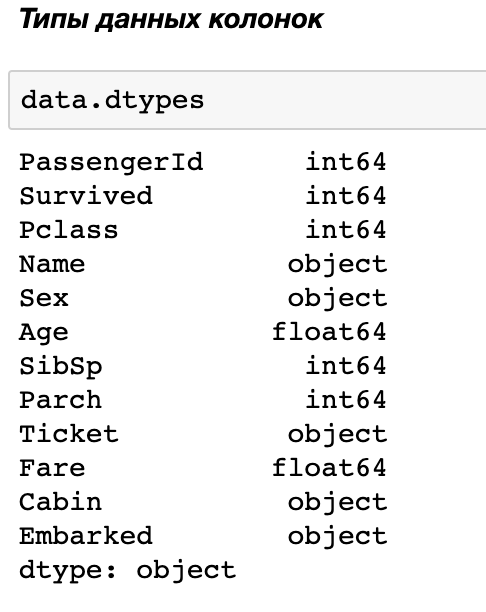
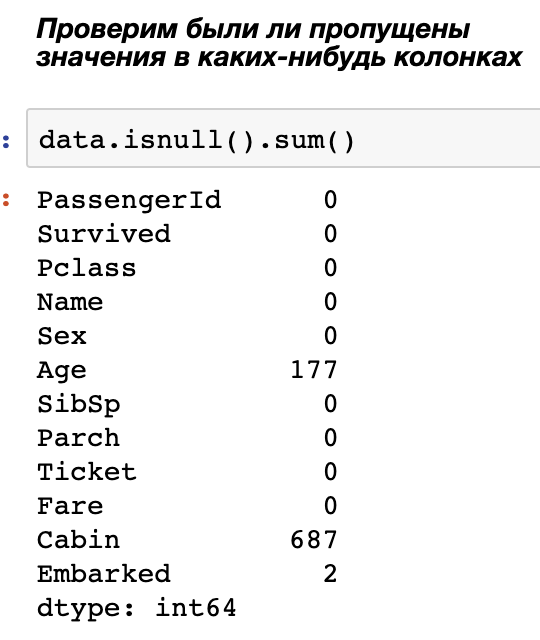
%matplotlib inline

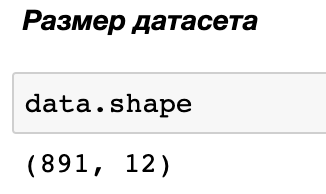
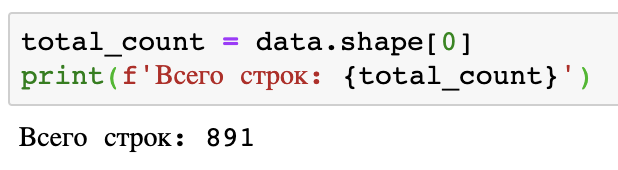
sns.set(style="ticks")

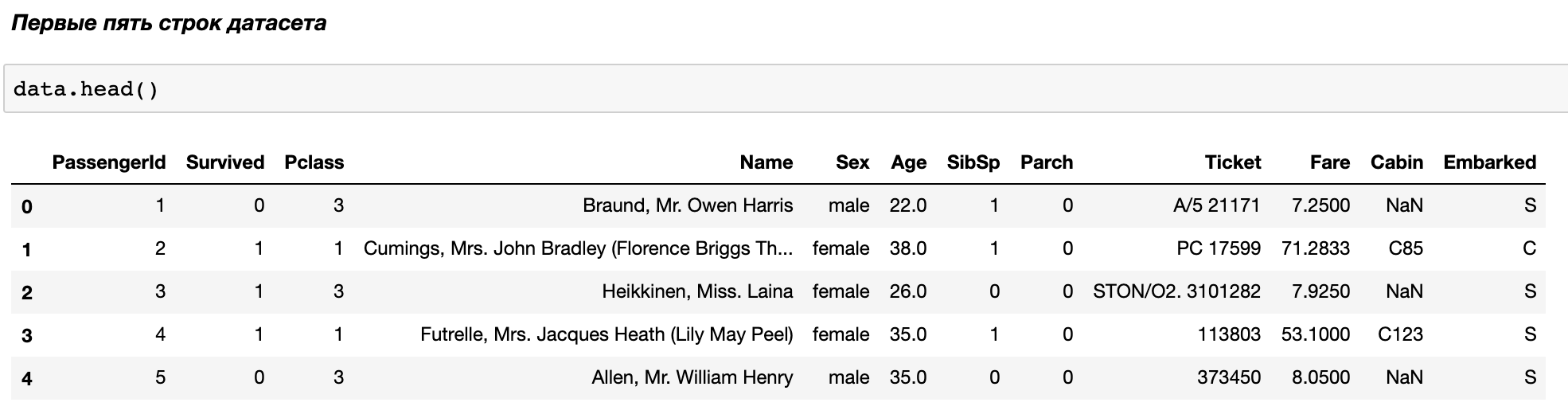
1. Подключаем набор данных

data = pd.read\_csv('train.csv', sep=",")

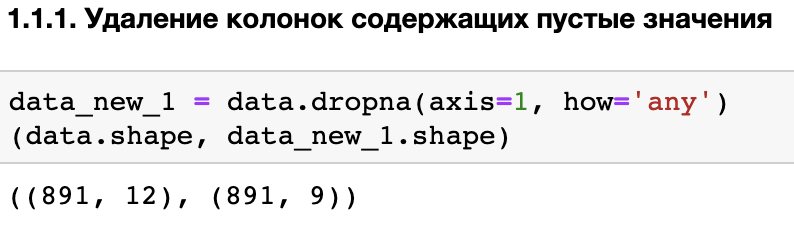
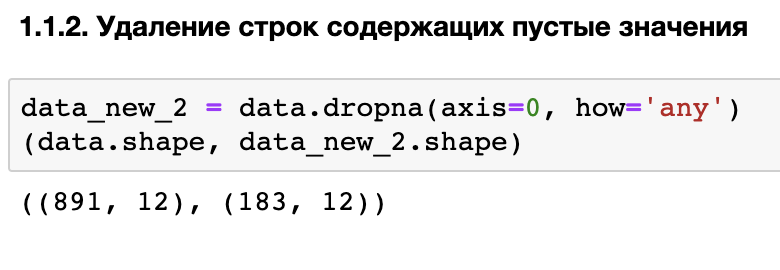
1. Описание датасета:

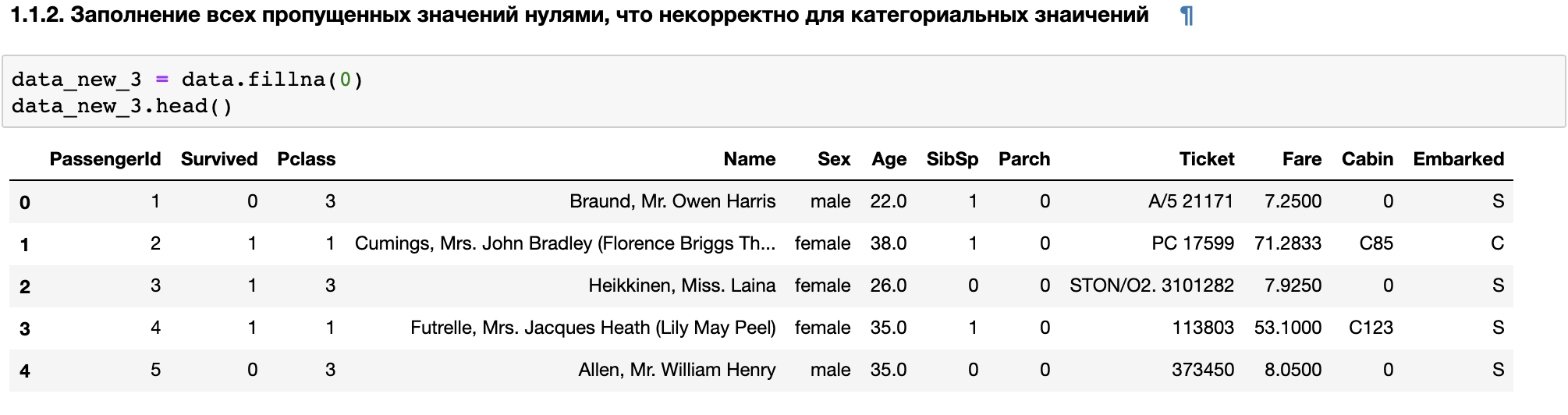
 

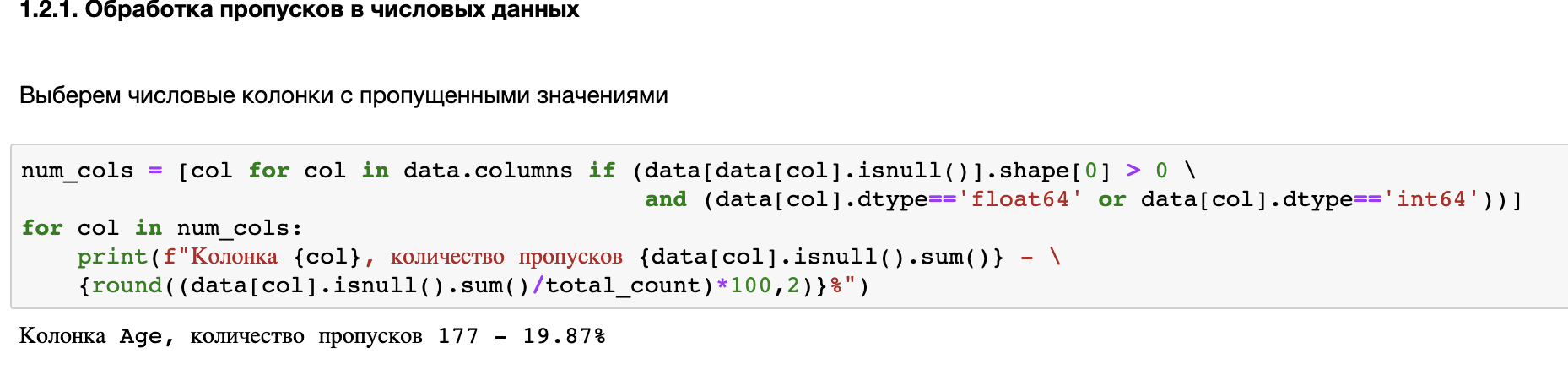
 

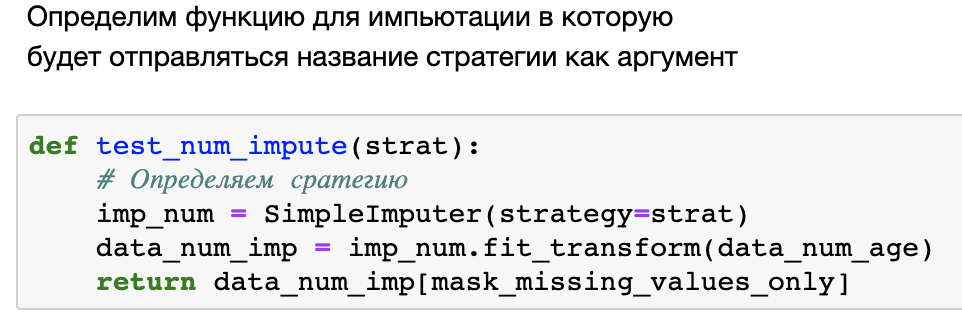
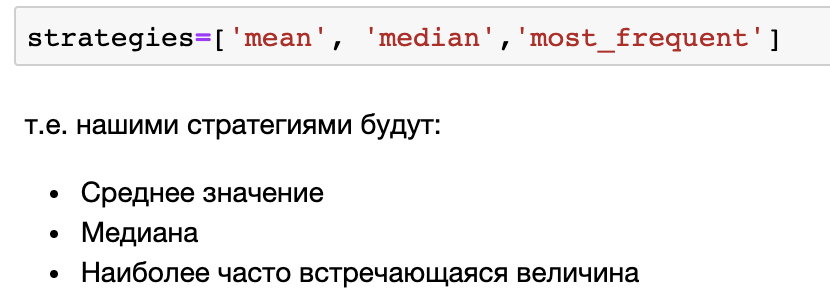


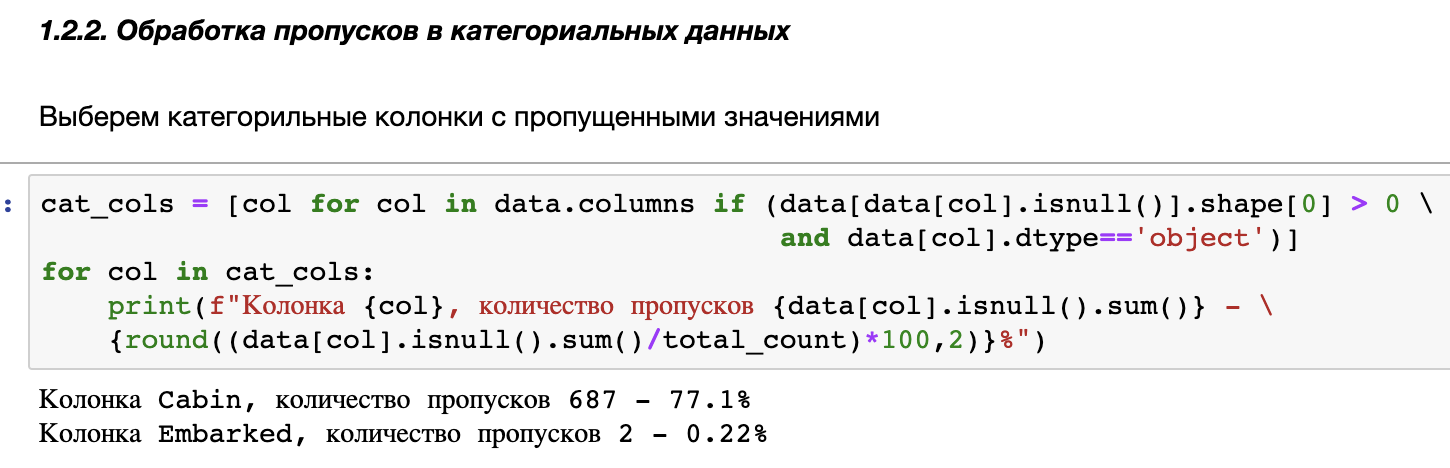
1. Обработка пропущенных данных:

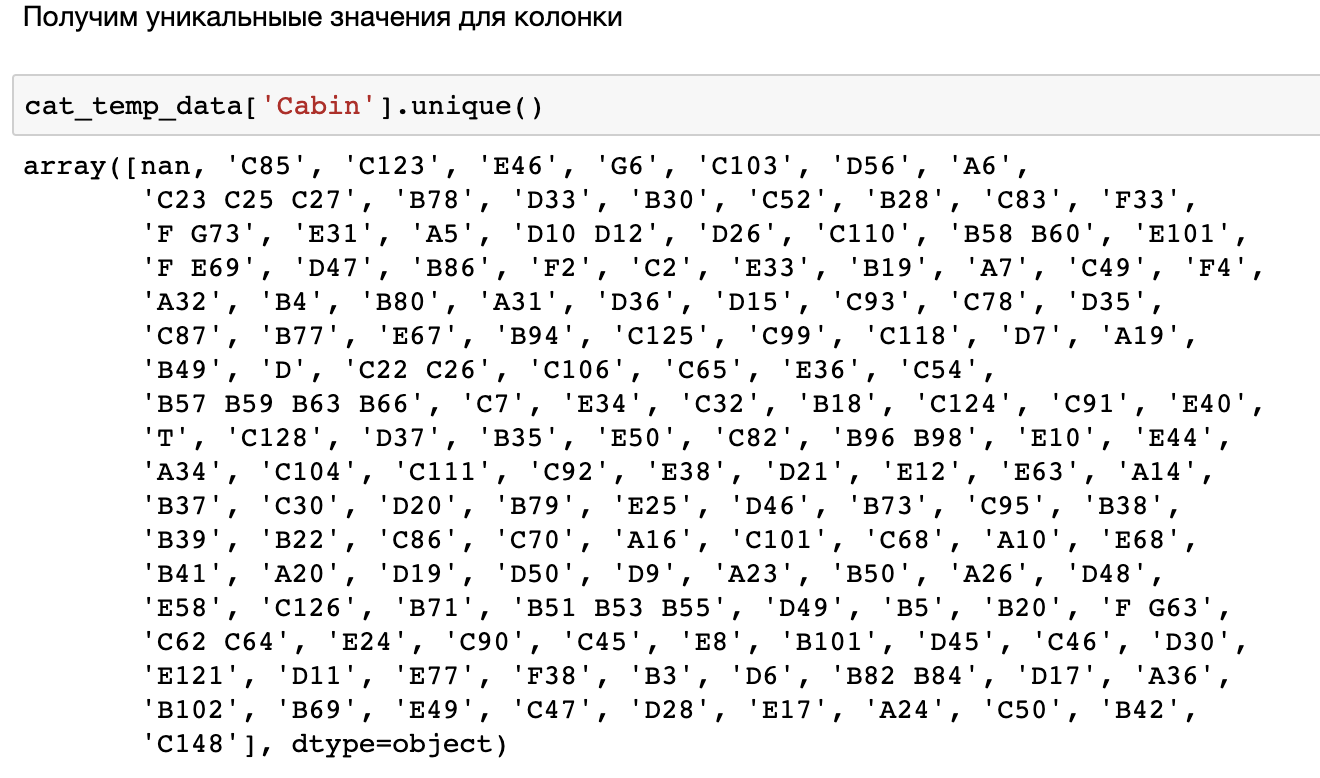
 

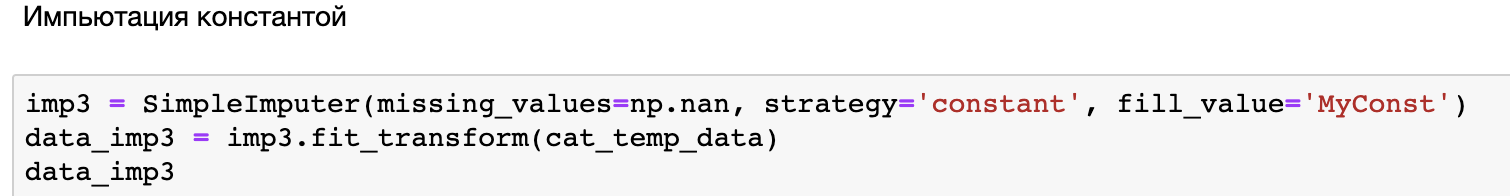
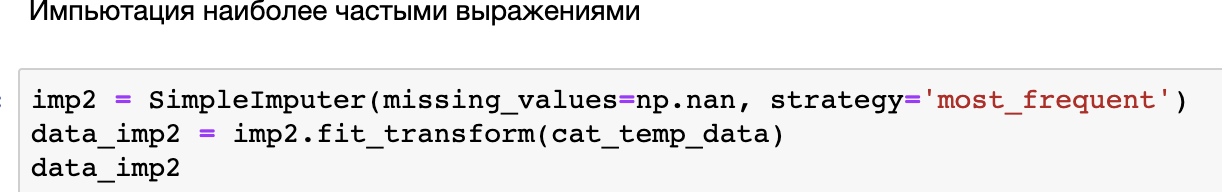


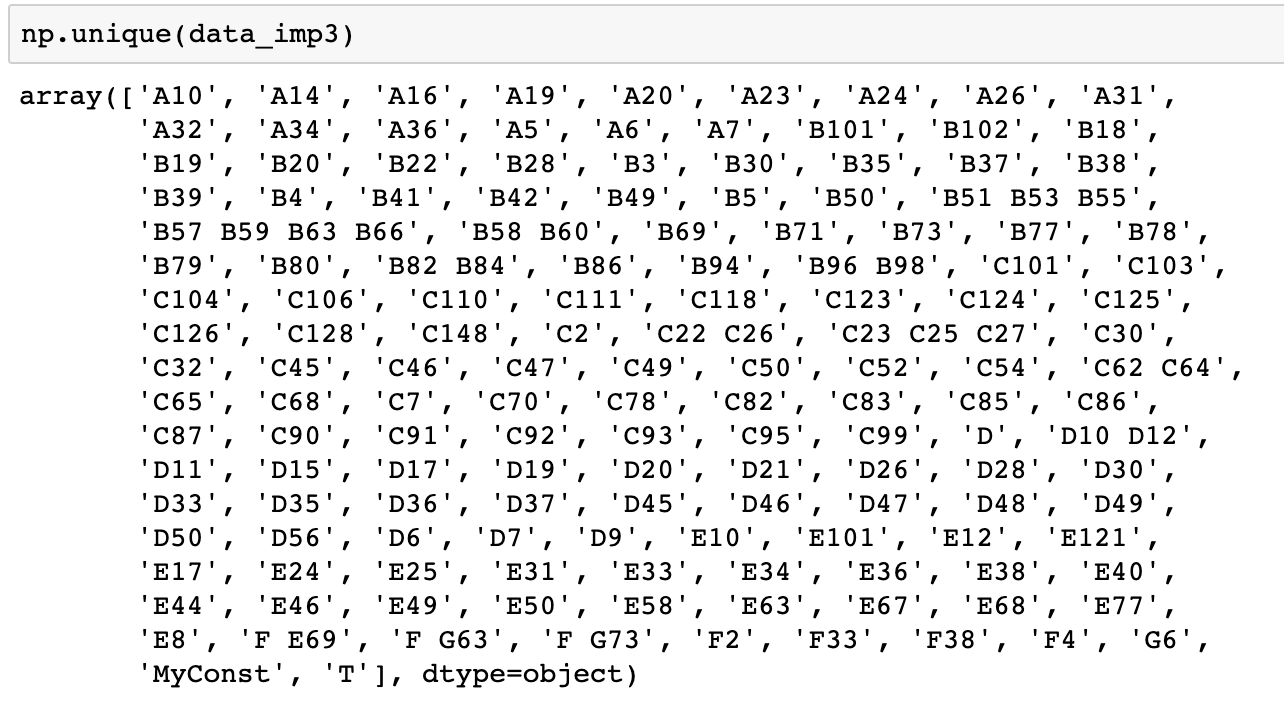




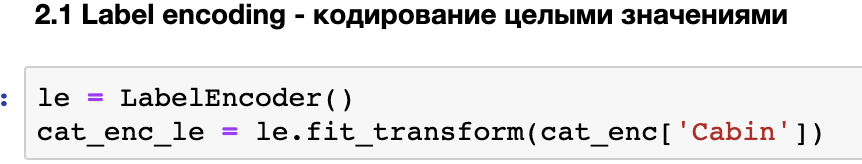


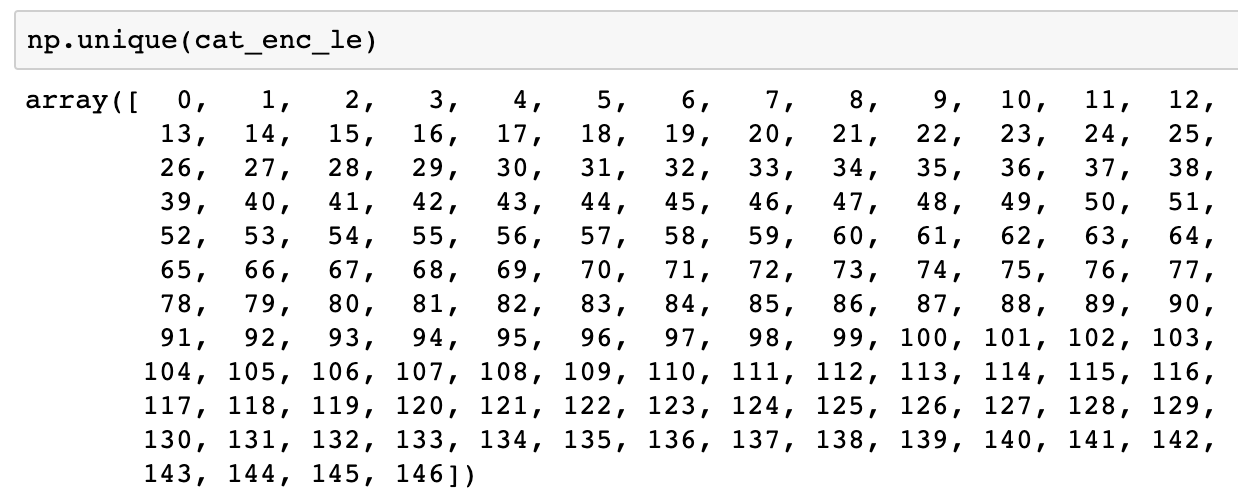
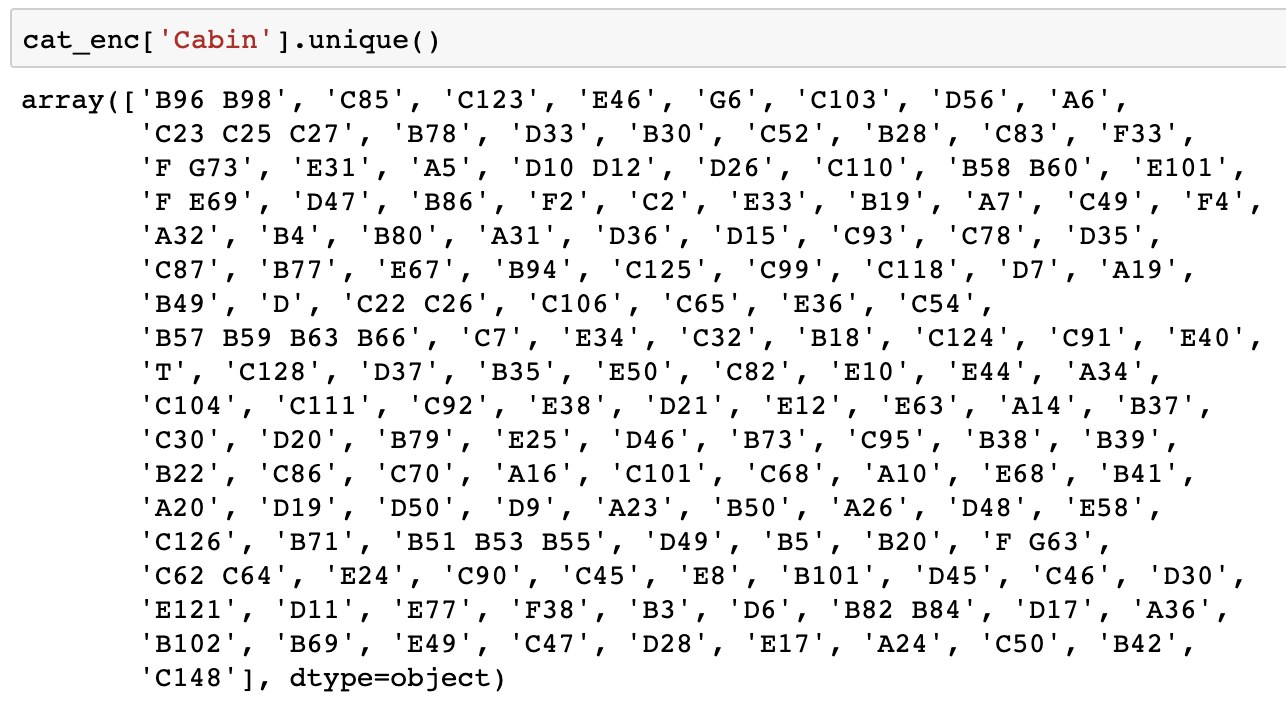


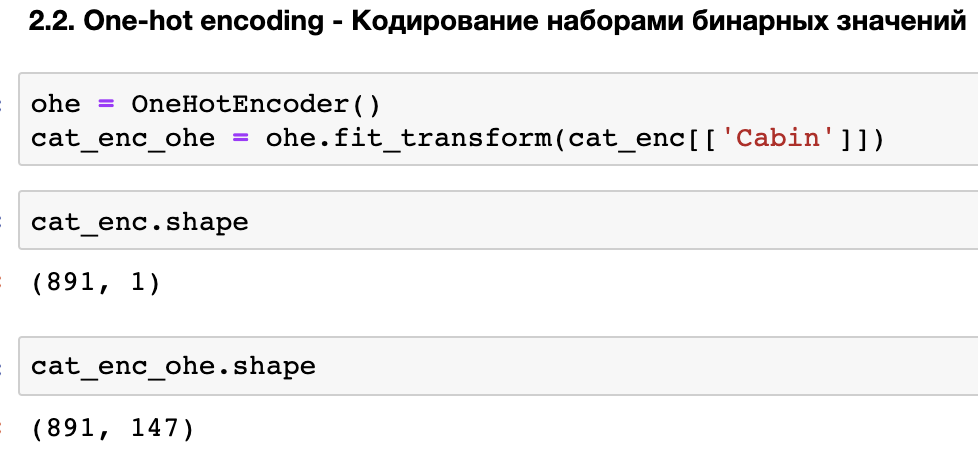




1. Замена категориальных данных на числовые:

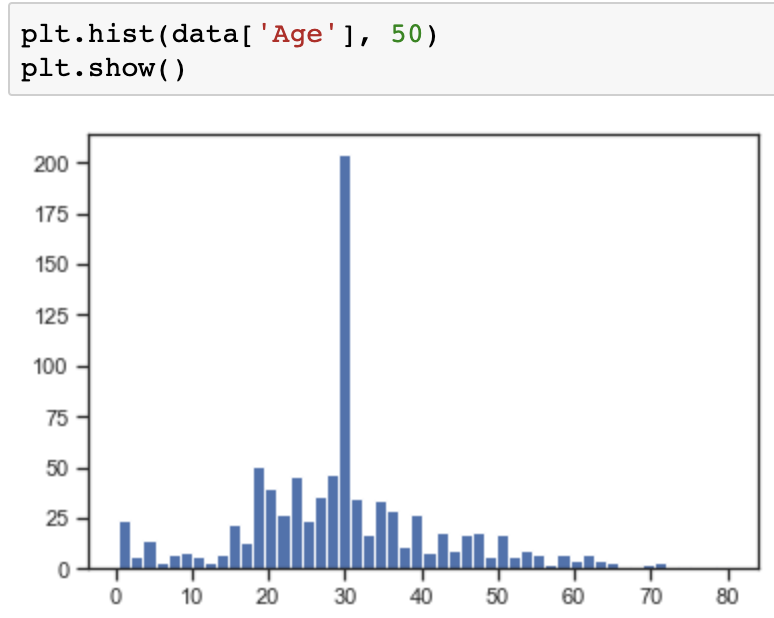
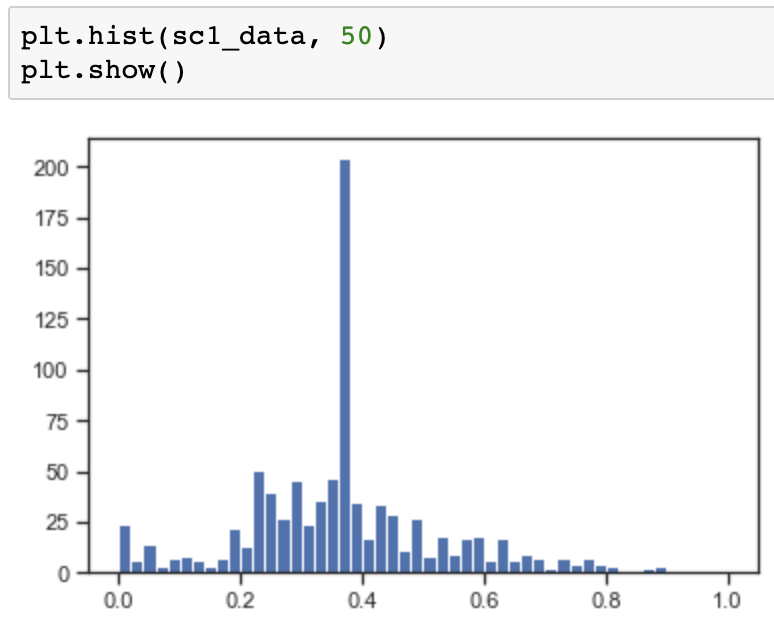




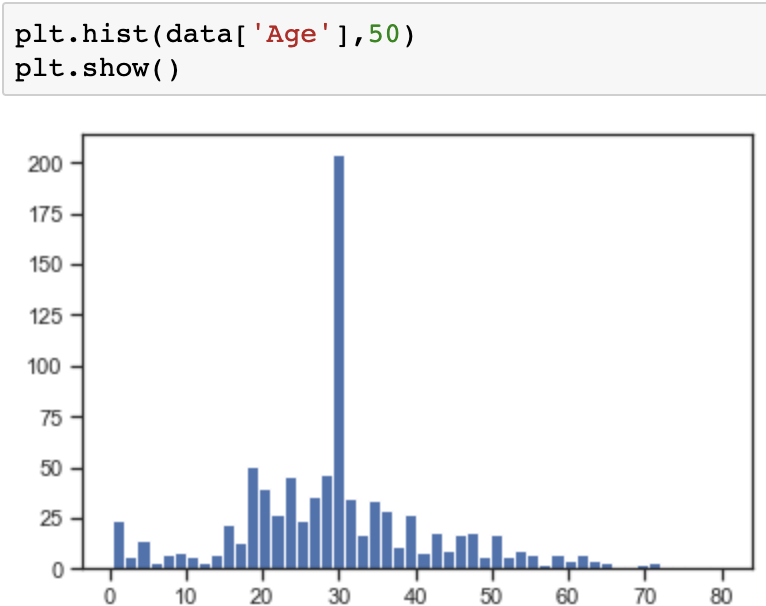
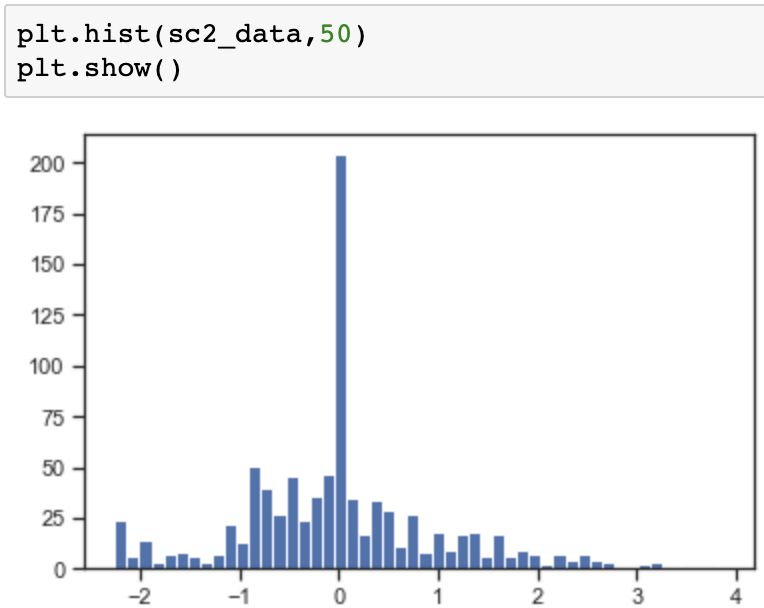


1. Масштабирование данных





1. Нормализация



