**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, искусственный интеллект и системы управления

КАФЕДРА Системы обработки информации и управления

**Лабораторная работа №2**

**«Обработка признаков»   
«Методы обучения с подкреплением»**

ИСПОЛНИТЕЛЬ: Ролдугин Е.В.

ФИО

группа ИУ5-25М \_\_\_\_\_\_\_\_\_\_\_

подпись

"\_\_"\_\_\_\_\_\_\_\_\_2023 г.

ПРЕПОДАВАТЕЛЬ: Гапанюк Ю.Е.

ФИО

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

подпись

"\_\_"\_\_\_\_\_\_\_\_\_2023 г.

Москва – 2023

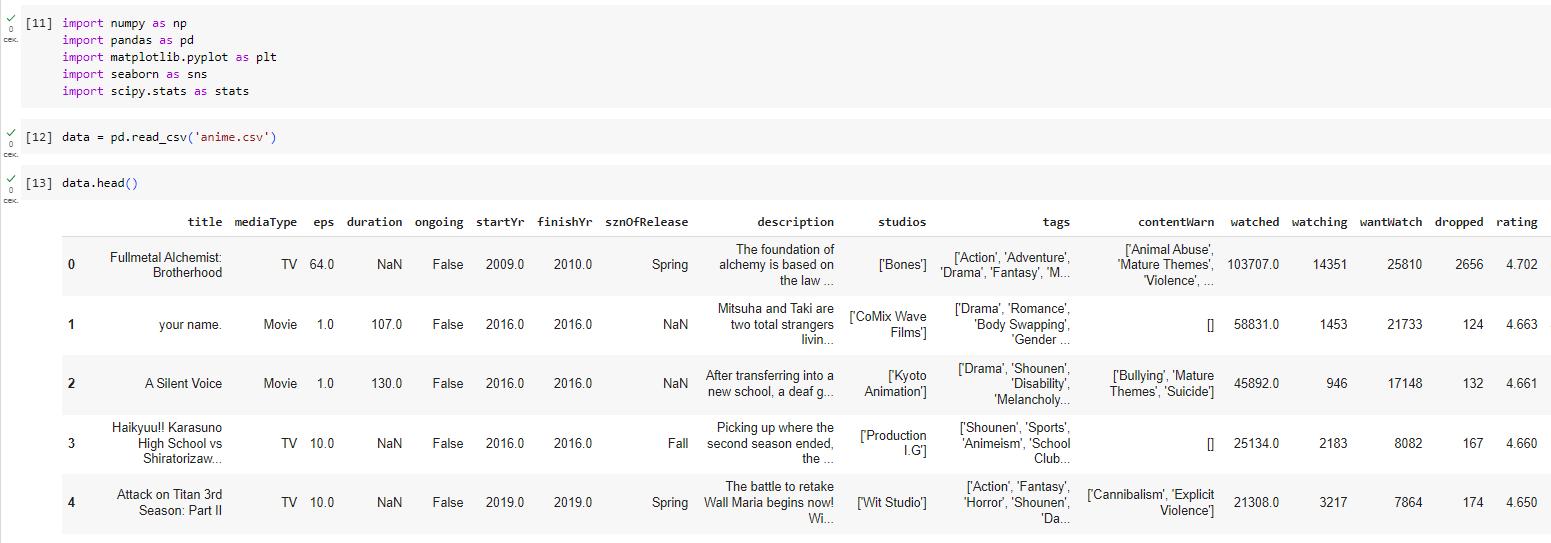
Выбрать набор данных (датасет), содержащий категориальные и числовые признаки и пропуски в данных. Для выполнения следующих пунктов можно использовать несколько различных наборов данных (один для обработки пропусков, другой для категориальных признаков и т.д.) Просьба не использовать датасет, на котором данная задача решалась в лекции.

Для выбранного датасета (датасетов) на основе материалов лекций решить следующие задачи:

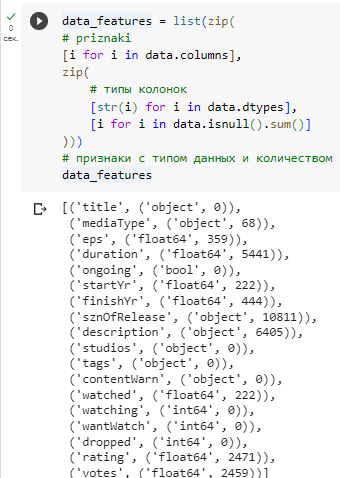
1. устранение пропусков в данных;
2. кодирование категориальных признаков;
3. нормализация числовых признаков;

Для данной лабораторной работы выберем датасет: anime.csv

Импортируем нужные нам библиотеки и выведем:

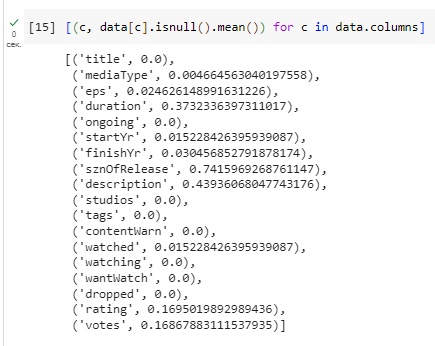


Проверим, есть ли пропущенные значения:

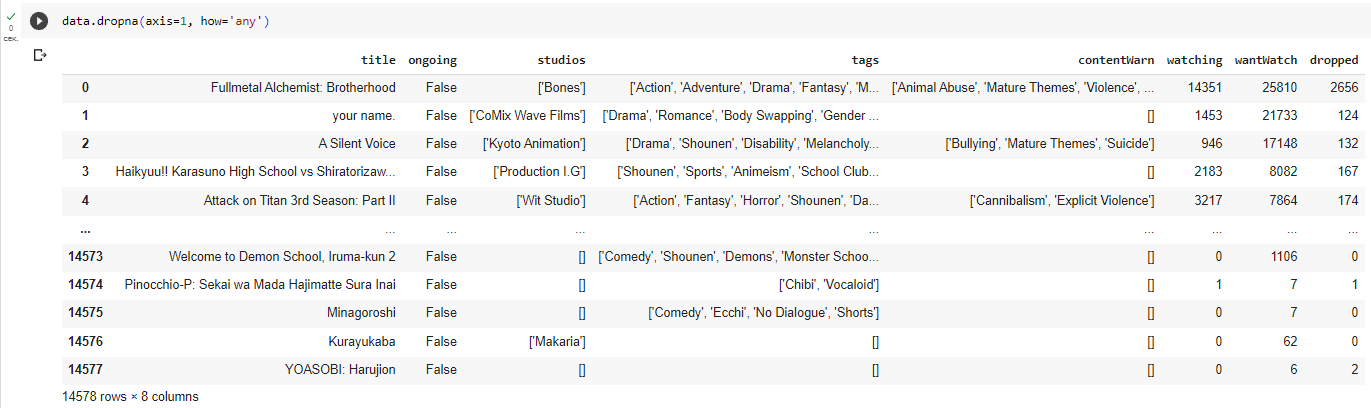


Устранение пропусков

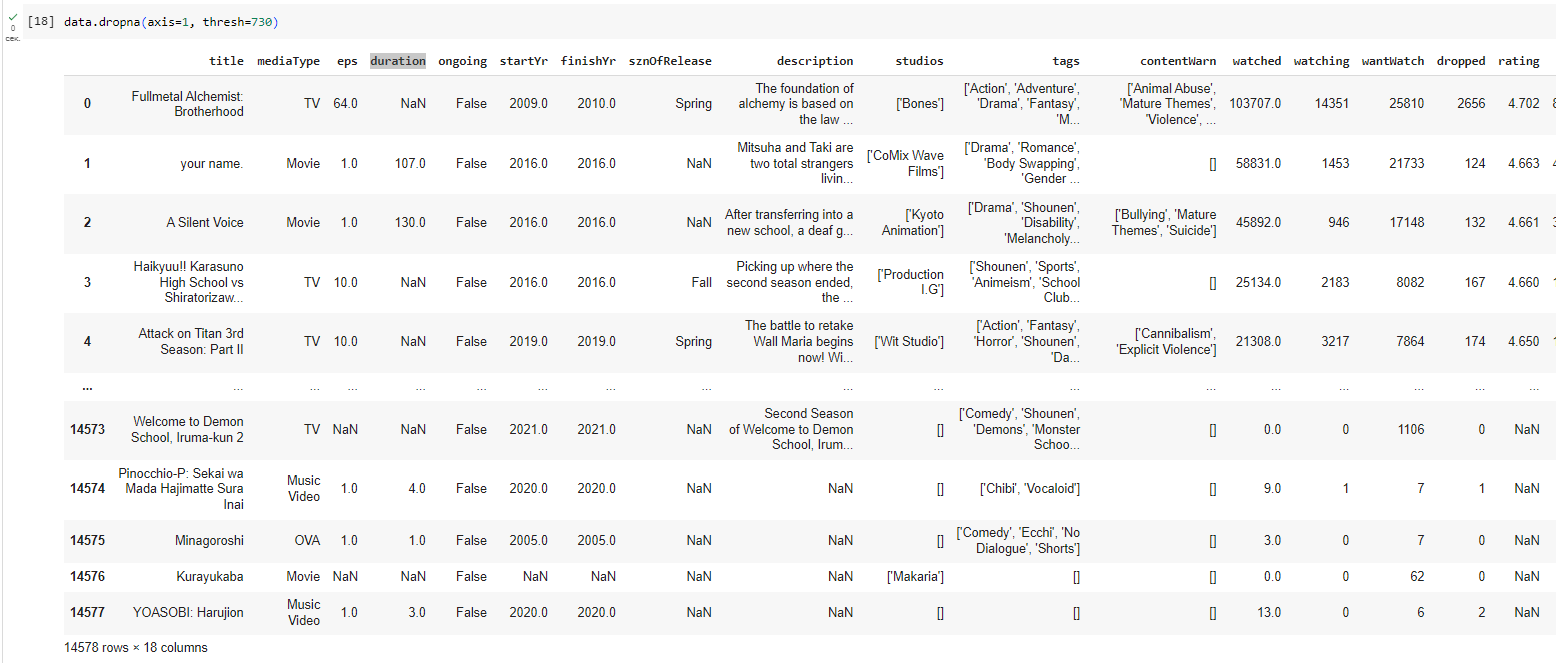
Доля пропусков (в процентах):



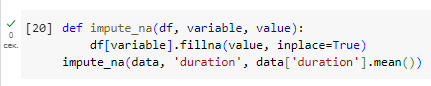
Удалим колонки, которые содержат пустые значения:



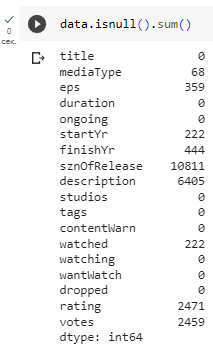
Удалим колонки с высоким процентом (более 50%) пропусков:



Заполним пропуски duration средними значениями:



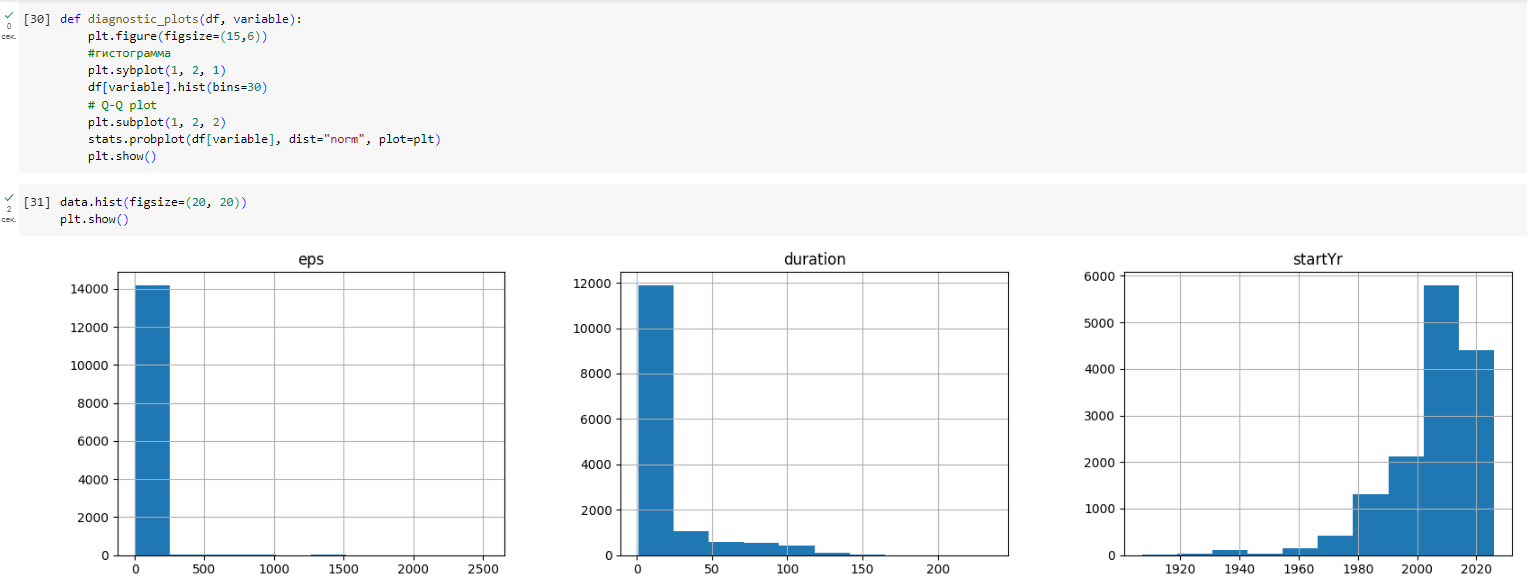
Удостоверимся, что признак duration не имеет пустых значений:

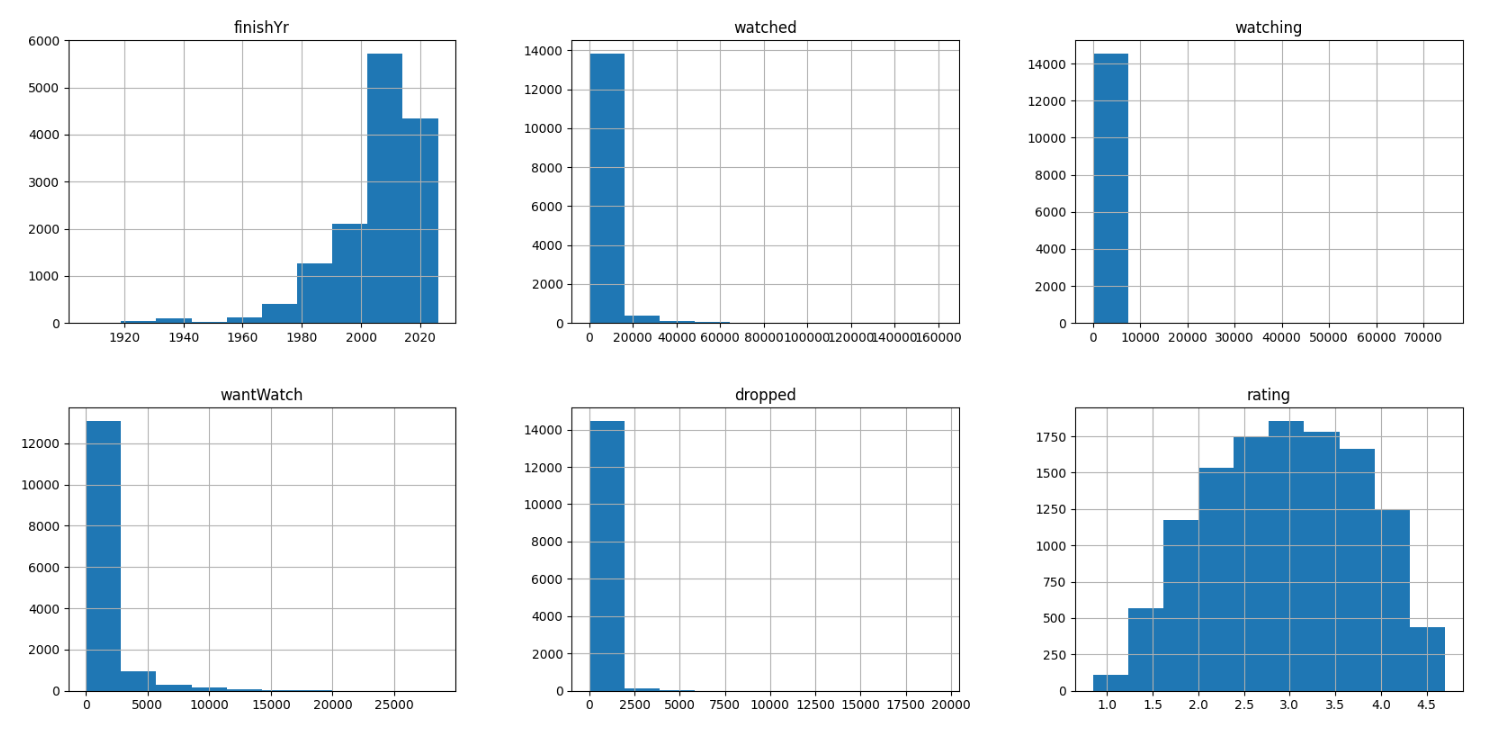


**Кодирование категориальных признаков**



**Нормализация числовых признаков**





Вывод: в рамках данной части лабораторной работы были решены следующие задачи:

* устранение пропусков в данных;
* кодирование категориальных признаков;
* нормализация числовых признаков;