Университет ИТМО

Практическая работа №1

По дисциплине “Визуализация и моделирование”

Автор: Лайок Олег Владимирович

**Поток:** 1\_2

**Группа:** К3242

**Факультет:** ИКТ

**Преподаватель:** Чернышева А.В.

Санкт-Петербург, 2021 г.

Задача: найти датасет и кратко описать его, описать данные датасета и сформулировать задачи, которые можно решить с использованием этого датасета.

1. Я подобрал свой датасет на сайте Kaggle <https://www.kaggle.com/sootersaalu/amazon-top-50-bestselling-books-2009-2019> . Он включает себя топ 50 книг бестселлеров на сайте Amazon в период с 2009 по 2019 год. В нем представлены краткая характеристика книг с сайта Amazon.

|  |  |  |
| --- | --- | --- |
| Название колонки в датасете | Данные | Тип данных |
| Name | Название книги | Текстовые |
| Author | Имя автора | Текстовые |
| User Rating | Рейтинг пользователей | Числовые |
| Reviews | Количество отзывов | Числовые |
| Price | Цена | Числовые |
| Year | Год, когда книга в топе | Числовые |
| Genre | Жанр (только два – fiction/non-fiction) | Текстовые |

1. Какие задачи можно решить с этим датасетом: для начала, я заметил, что эти данные можно дополнить, чтобы расширить их репрезентативную способность, т.к. потенциал этой выборки очень большой. Хотелось бы добавить номер в рейтинге у каждой из книг, расширить количество жанров (художественную литературу подразделить на фантастика, фентези, драма и тд). После этого можно провести гораздо больше исследований, например: как изменялись вкусы людей в жанрах за последние 10 лет, с чем связанна популярность тех или иных книг(например сравнить информацию с википедии если в том же году, в каком книга оказалась в топе, были фильмы или сериалы по этой книге), какой количество раз та или иная книга или автор оказывались в топе, проанализировать рейтинг книги на сайте и место в топе.