

Университет ИТМО

## Практическая работа №2

По дисциплине “Визуализация и моделирование”

Автор: Лайок Олег Владимирович

**Поток:** 1\_2

**Группа:** K3242

**Факультет:** ИКТ

**Преподаватель:** Чернышева А.В.

Санкт-Петербург, 2021 г.

1. Описание датасета: топ 50 книг бестселлеров на сайте Amazon, в период с 2009 по 2019 год, с краткой характеристикой каждой книги в топе.
2. Содержание датасета:

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

3. Описательная статистика Pandas:

В ходе проведения описательной статистики и визуализации были полученные следующие данные:

- a. Получен график книг, которые чаще всего появлялись в топе
- b. Получен график авторов, которые чаще всего появлялись в топе
- c. Выяснили что соотношение non-fiction/fiction 56.4% к 43.6%, построена круговая диаграмма.
- d. Был проанализирован рейтинги книг. Для начала для этого столбца были подсчитаны некоторые меры центральной тенденции и разброса.
- e. Затем, я построил график распределения количества книг. Рейтинг 4.8 показывает максимум, по общему числу книг. Однако я решил проверить как этот параметр меняется для топа книг каждого конкретного года.
- f. Построен график распределения рейтинга в 2019 и 2009 году. К сожалению, по ним не сильно видно как изменяется рейтинг с течением времени.
- g. Построили график распределения рейтинга для каждого года, но этот графк тоже не очень хорошо отражает общую картину.
- h. Простроили столбчатую диаграмму для количества книг каждого рейтинга в каждом году. Здесь данные уже интереснее: можно заметить что в 2019 году с огромным отрывом лидируют книги с

рейтингом 4.8, а в 2012 году график распределения рейтингов практически идеально подходит под график нормального распределения.

- i. Был посчитан средний рейтинг книг в каждом году и построен график изменения среднего рейтинга с течением времени. Тут уже можно сделать несколько выводов: В 2012 году средний рейтинг книг вошедших в топ был минимальным. В 2019 году средний рейтинг книг, вошедших в топ, стал максимальным. По графику можно понять, что средний рейтинг увеличивается в последнее время.
- j. Далее было проанализировано число рецензий на книги и получен топ 10 книг с самым большим числом рецензий.
- k. Построена столбчатая диаграмма распределения числа рецензий в каждом году. На диаграмме отчетливо видно, что чем позже в списке книга, тем больше на нее рецензий, однако, чтобы убедиться лучше построить график изменения среднего числа рецензий с течением времени. На диаграмме отчетливо видно, что чем позже в списке книга, тем больше на нее рецензий, однако, чтобы убедиться в этом был построен график изменения среднего числа рецензий с течением времени.
- l. На данном графике мы можем убедиться, что проверка была не зря: в самом деле, минимальное число рецензий было в 2009 году, однако максимальное было в 2014 и в 2019 году.
- m. Были получены данные книг, которые повлияли на такой резкий скачок в эти года. Интересно заметить, что в 2014 году вышли две суперуспешные художественные книги, на которые очевидно будет много рецензий, а вот в 2019 году в топе по отзывам мотивационная книга и две детских.

#### 4. Вывод:

По полученным данным мы можем получить представление о том, какое качество у книг в топе было в каждом году и как изменялось с течением времени. Метриками качества книг могут служить рейтинг книг и количество рецензий на них. Именно по этим параметрам и проводился анализ