

Университет ИТМО

Практическая работа №4

По дисциплине “Визуализация и моделирование”

Автор: Лайок Олег Владимирович

Поток: 1_2

Группа: K3242

Факультет: ИКТ

Преподаватель: Чернышева А.В.

Санкт-Петербург, 2021 г.

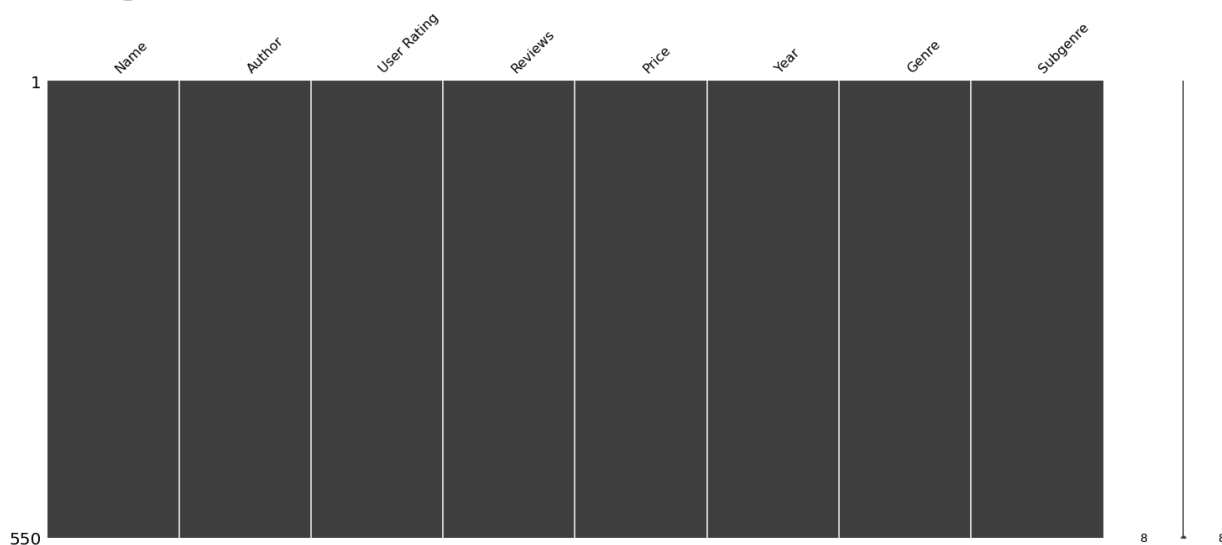
Ссылка на код

https://colab.research.google.com/drive/1JinaKNjPEPZMjCHyNdPxfX_awXCIx2DC#scrollTo=zGVzTwqkXiv2

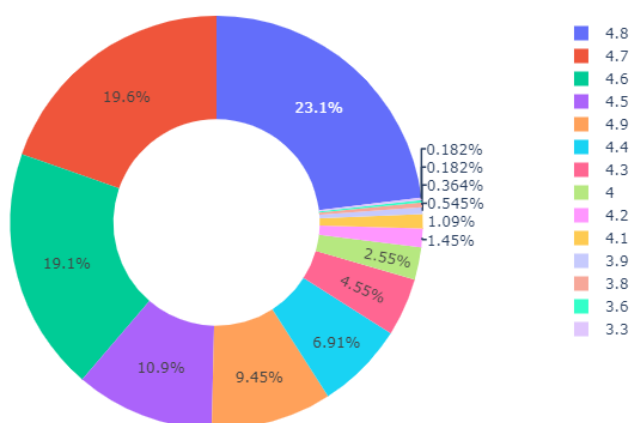
1. Проверка данных:

Т.к. в третьей работе данные никак не изменялись(за исключением добавления нового столбца subgenre) вся визуализация данных из второй работы будет такой же, если повторить ее здесь. Поэтому, просто проверим целостность датасета визуально, и для примера возьмем одну визуализацию из 2 работы и сделаем ее с помощью другой диаграммы:

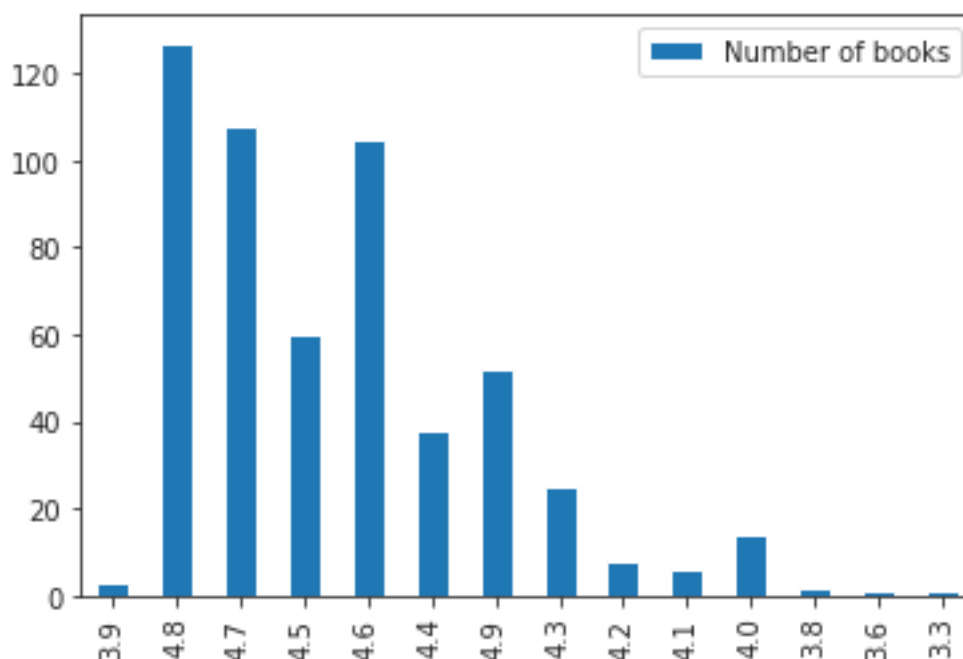
Посмотрим что в каждом столбце одинаковое число записей:



Попробуем визуализировать данные из второй работы, например посмотрим распределение рейтинга с помощью круговой диаграммы:



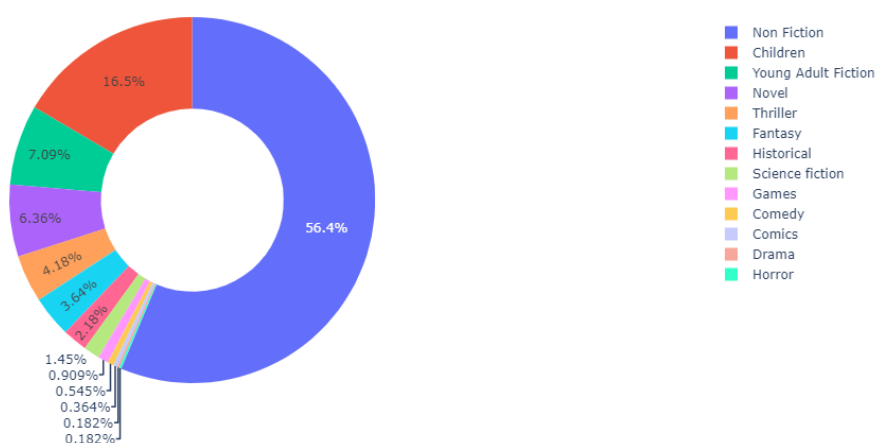
Данные аналогичны данным из второй работы:



2. Гипотезы:

- 1) Детская литература будет доминировать по количеству книг попавших в топ среди художественной литературы (во время проставления жанров попало очень много детских книг, нужно проверить их процентное соотношение)

Проверка: Гипотеза подтвердилась, Детские книги составляют 16% от общего числа книг (56% составляет не художественная литература), что почти на 10% больше второго лидера по числу книг, Young Adult Fiction у которых 7% от общего числа книг.

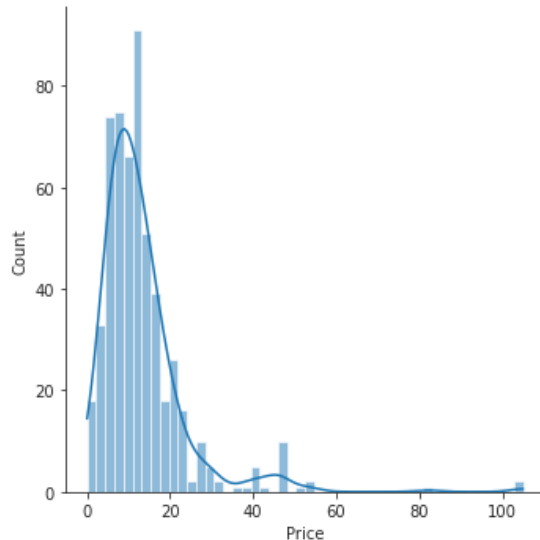


- 2) Средняя цена книг будет около 5\$ (т.к. в ходе этой работы обнаружилось, что в датасете хранится минимальная стоимость)

книги, нужно уточнить какая именно средняя минимальная цена у данных книг)

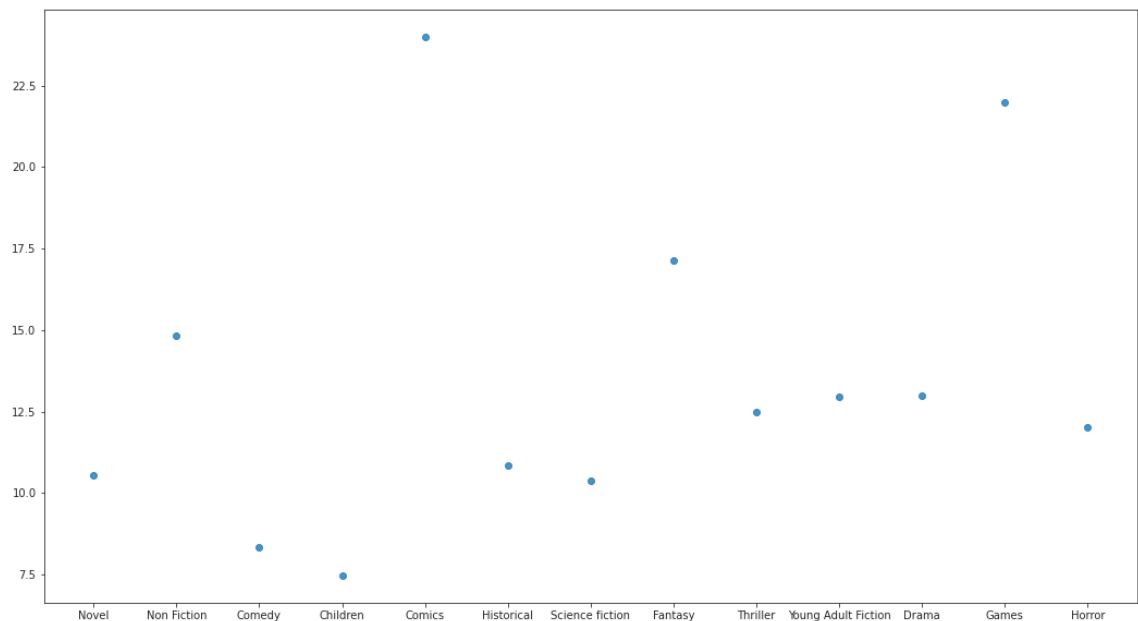
Проверка: В гипотезе я утверждал что средняя цена будет около 5, в ходе анализа было получено более точное значение – 13.

Так же можно заметить что цены и число рецензий распределяются по закону Пуассона (возможно гамма распределения).



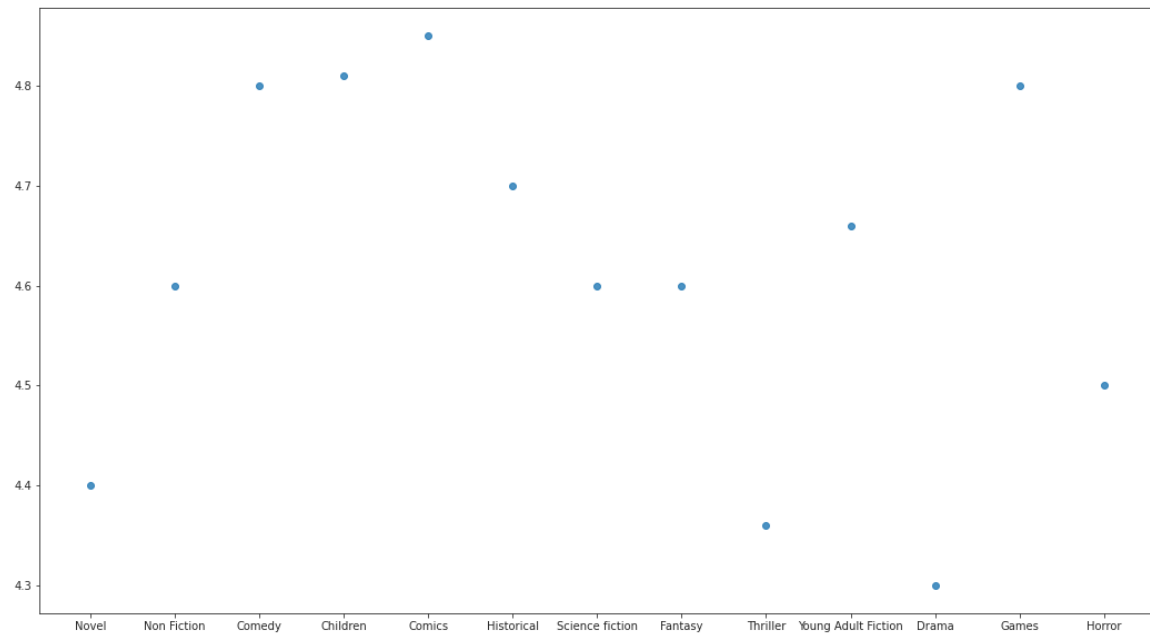
3) Средняя цена книг будет сильно отличаться от средней цены книги каждого поджанра (т.к. по логике нехудожественные книги и книги для подростков продают дороже)

Проверка: предположение было ошибочно, дороже всего продаются жанры Comics, Games и Fantasy. (на оси x названия поджанров, на оси y средняя цена).



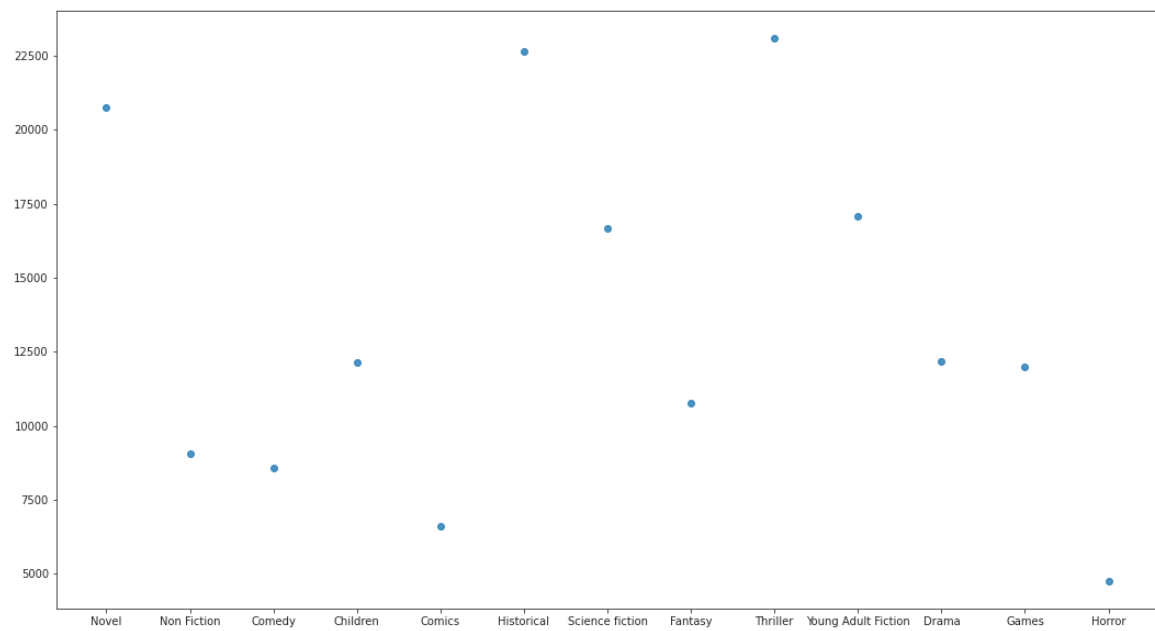
- 4) Т.к. появилась категория поджанров нужно проанализировать средний рейтинг в каждом поджанре, предварительное предположение: у детской литературы рейтинг максимальный, у Young Adult Fiction минимальный.

Проверка: Оказалось, предположение было довольно точным, Children идет на 2 месте сразу после Comics, однако рейтинг Young Adult Fiction не минимальный, а находится где-то посередине.



- 5) Аналогичная четвертому пункту гипотеза, но в отношении числа рецензий.

Проверка: Гипотеза не подтвердилась, мало того интересно заметить что поджанр Thriller имеет максимальное среднее число отзывов, хотя близкий к нему поджанр Horror наоборот минимальное.



3. Вывод:

В ходе работы были проверены гипотезы, сформулированные в третьей работе. Почти все гипотезы опроверглись, хотя некоторые предложения были близки.