

# **Исследовательский проект “Популярные книги”**

Выполнен Дроздовой Алёной БПИ229

## **Цель исследования:**

Проанализировать популярные в данный момент книги на LiveLib.ru и выявить, от чего зависит их популярность, также определить взаимосвязи между данными

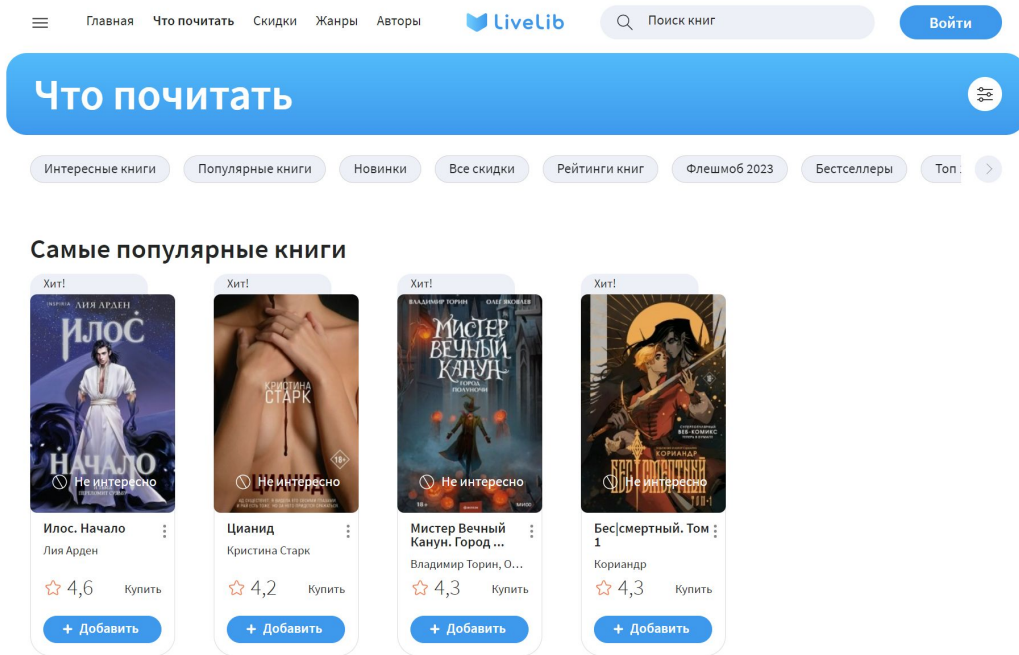
## **Задачи исследования:**

- сформировать датасет с помощью веб-скрейпинга
- сформулировать гипотезы
- выбрать переменные для исследования и преобразовать их при необходимости
- найти и обосновать возможные зависимости величин

# Объект исследования

Для анализа мной была выбрана подборка “Что почитать”, содержащая самые популярные книги сайта LiveLib.ru.

Для этого я сгенерировала 20 ссылок вида:  
"<https://www.livelib.ru/books/movers-and-shakers/listview/biglist/~{i+1}>"




Затем с каждой страницы мною были получены внутренние ссылки на все книги, которые были дополнены до внешней

	urls
0	<a href="https://www.livelib.ru/book/1007304604-ilos-nachalo-liya-arden?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007304604-ilos-nachalo-liya-arden?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
1	<a href="https://www.livelib.ru/book/1007394649-tsianid-kristina-stark?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007394649-tsianid-kristina-stark?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
2	<a href="https://www.livelib.ru/book/1007485342-mister-vechnyj-kanun-gorod-polunochoi-vladimir-torin-oleg-yakovlev?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007485342-mister-vechnyj-kanun-gorod-polunochoi-vladimir-torin-oleg-yakovlev?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
3	<a href="https://www.livelib.ru/book/1007493583-bessmertnyj-tom-1-koriandr?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007493583-bessmertnyj-tom-1-koriandr?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
4	<a href="https://www.livelib.ru/book/1007280463-chelovekbenzopila-kniga-4-vo-sne-nastoyaschaya-zhest-tatsuki-fudzimoto?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007280463-chelovekbenzopila-kniga-4-vo-sne-nastoyaschaya-zhest-tatsuki-fudzimoto?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
5	<a href="https://www.livelib.ru/book/1007146420-pesn-sorokoputa-renessans-frensis-kel?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007146420-pesn-sorokoputa-renessans-frensis-kel?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
6	<a href="https://www.livelib.ru/book/1007564878-vy-najdete-et-o-v-biblioteke-mitiko-aoyama?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007564878-vy-najdete-et-o-v-biblioteke-mitiko-aoyama?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
7	<a href="https://www.livelib.ru/book/1007024758-moya-temnaya-koroleva-kristina-te?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007024758-moya-temnaya-koroleva-kristina-te?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
8	<a href="https://www.livelib.ru/book/1007568157-poehali vse-cto-nuzhno-znat-nachinayuschim-voditeliyam-moryahin-sergej?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007568157-poehali vse-cto-nuzhno-znat-nachinayuschim-voditeliyam-moryahin-sergej?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
9	<a href="https://www.livelib.ru/book/1007338282-nekromans-opus-1-evgeniya-safonova?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007338282-nekromans-opus-1-evgeniya-safonova?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
10	<a href="https://www.livelib.ru/book/1007351431-dar-krougi-anna-pushkina?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007351431-dar-krougi-anna-pushkina?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
11	<a href="https://www.livelib.ru/book/1007546047-paradoks-severa-viktoria-pobedinskaya?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007546047-paradoks-severa-viktoria-pobedinskaya?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
12	<a href="https://www.livelib.ru/book/1007361097-ne-draznite-nekromanta-nika-vejmar?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007361097-ne-draznite-nekromanta-nika-vejmar?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
13	<a href="https://www.livelib.ru/book/1007351434-kogda-zakanchivayutsya-skazki-elena-kondratskaya?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007351434-kogda-zakanchivayutsya-skazki-elena-kondratskaya?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
14	<a href="https://www.livelib.ru/book/1007300539-stalnoj-alhimik-kniga-17-hiroma-arakava?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007300539-stalnoj-alhimik-kniga-17-hiroma-arakava?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
15	<a href="https://www.livelib.ru/book/1007401063-tanamor-nezirmyj-vrag-ekaterina-sobol?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007401063-tanamor-nezirmyj-vrag-ekaterina-sobol?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
16	<a href="https://www.livelib.ru/book/1007458243-holodnye-glaza-islam-hanipaev?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007458243-holodnye-glaza-islam-hanipaev?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>
17	<a href="https://www.livelib.ru/book/1007332186-sharlatany-robin-kuk?reclist=movers-and-shakers_block&amp;rvid=4157392131">https://www.livelib.ru/book/1007332186-sharlatany-robin-kuk?reclist=movers-and-shakers_block&amp;rvid=4157392131</a>

Затем с этих ссылок были получены следующие данные

	heading	author	edition	rating	read	planned	reviews	quotes	genre	ISBN	year	pages	language	age_limit
0	Илос. Начало	Лия Арден	Потомки Первых, книга №3.5	4,6	1460	3300	46	83	Русскоефэнтези	978-5-04-164474-1	2022	448	Русский	16+
1	Цианид	Кристина Старк	Mainstream	4,2	2069	3920	92	270	Любовныероманы	978-5-17-146746-3	2022	448	Русский	
2	Мистер Вечный Канун. Город Полуночи	Владимир Торин, Олег Яковлев	Мистер Вечный Канун, книга №1	4,3	1321	3829	74	72	Русскоефэнтези	978-5-00195-790-4	2022		Русский	18+
3	Бес смертный. Том 1	Кориандр	Эксмо	4,3	1111	2030	23	14	Комиксы,манга,гرافическиероманы	978-5-04-110975-2	2022	176	Русский	18+
4	Человек-бензопила. Книга 4. Во сне. Настоящая жесьть	Тацуки Фудзимото	4 из 6	4,6	797	378	1	3	Комиксы,манга,гرافическиероманы	978-5-389-21080-6	2022	392	Русский	18+
5	Песнь Сорокопута. Ренессанс	Фрэнсис Кель	Песнь Сорокопута, книга №2	4,3	819	3093	30	27	Русскоефэнтези	978-5-04-164581-6	2022	480	Русский	18+
6	Вы найдете это в библиотеке	Митико Аояма	Миф	4,4	649	3011	47	106	Современнаязарубежнаялитература	978-5-00195-754-6	2022	256	Русский	16+
7	Моя темная королева	Кристина Тэ	Манн	4	1020	2752	71	102	Фэнтези	978-5-00195-487-3	2022	320	Русский	
8	Поехали! Все, что нужно знать начинающим водителям	Моряхин Сергей	Издательский дом «Питер»	4,9	297	436	5	0	АвтомобилииПДД	978-5-4461-2331-5	2022	192	Русский	16+
9	Некроманс. Opus 1	Евгения Сафонова	Мир Топракейм, книга №3	4,3	549	2224	32	58	Русскоефэнтези	978-5-04-168484-6	2022	480	Русский	16+
10	Дар Кроуги	Анна Пушкина	АСТ	4,3	519	1589	87	16	Фэнтези	978-5-17-147104-0	2022	352	Русский	16+
11	Парадокс Севера	Виктория Побединская	CLEVER	4,4	404	601	45	63	Книгидляподростков	978-5-00211-002-5	2022		Русский	16+
12	Не дразните некроманта!	Ника Веймар	Фесские истории, книга №2	4,2	524	539	112	35	Романтическоефэнтези	978-5-04-168483-9	2022	320	Русский	16+
13	Когда заканчиваются сказки	Елена Кондрацкая	Дивные Берега, книга №2	4,2	477	1629	33	14	Русскоефэнтези	978-5-04-164574-8	2022	416	Русский	18+

Этих данных мне показалось недостаточно, поэтому был выполнен поиск в каталоге chitai-gorod.ru по ISBN через генерацию ссылок вида "https://www.chitai-gorod.ru/search?q="+isbn



Каталог

978-5-04-164474-1

×

🔍

Войти

Заказы

Закладки

Корзина

АкцииПодборкиЧитай-журналКнижные циклы

## РЕЗУЛЬТАТЫ ПОИСКА

В наличии

☐

Книги ^

☐ Художественная литература

Цена ^

690

690

Раздел

Фэнтези

Цикл

Потомки Первых

Автор


Лия Арден

Серия

Universum. Потомки Первых. Миры Лии Арден

По запросу «978-5-04-164474-1» найден 1 товар

По совпадению ▾





Затем с помощью веб-скрейпинга полученных ссылок на книги был сформирован следующий датасет

	heading	author	ISBN	edition	age_limit	genre	language	cover_type	year	price	old_price	pages	weight	num_of_copies	rating	reads	planned	reviews	quotes
0	Илос. Начало	Лия Арден	978-5-04-164474-1	INSPIRIA	16+	Русскоефэнтези	Русский	Твердыйпереплёт	2022	690	690	448	541	18000	4,6	1460	3300	46	83
1	Цианид	Кристина Старк	978-5-17-146746-3	АСТ	18+	Любовныероманы	Русский	Твердыйпереплёт	2022	549	675	448	400	20000	4,2	2069	3920	92	270
2	Мистер Вечный Канун. Город Полуночи	Владимир Торин, Олег Яковлев	978-5-00195-790-4	Манн,ИвановиФербер	18+	Русскоефэнтези	Русский	Твердыйпереплёт	2022	723	925	576	556	3000	4,3	1321	3829	74	72
3	Человек-бензопила. Книга 4. Во сне. Настоящая жесьть	Тацуки Фудзимото	978-5-389-21080-6	Азбука	18+	Комиксы,манга,графическиероманы	Русский	Твердыйпереплёт	2022	944	944	392	530	40000	4,6	797	378	1	3
4	Песнь Сорокопута. Ренессанс	Фрэнсис Кель	978-5-04-164581-6	Эксмо	18+	Русскоефэнтези	Русский	Твердыйпереплёт	2022	608	745	480	439	5000	4,3	819	3093	30	27
5	Вы найдете это в библиотеке	Митико Аояма	978-5-00195-754-6	Манн,ИвановиФербер	16+	Современнаязарубежнаялитература	Русский	Твердыйпереплёт	2023	622	795	256	500	4000	4,4	649	3011	47	106
6	Моя темная королева	Кристина Тэ	978-5-00195-487-3	Манн,ИвановиФербер	16+	Фэнтези	Русский	Твердыйпереплёт	2022	684	875	352	449	4000	4	1020	2752	71	102
7	Поехали! Все, что нужно знать начинающим водителям	Моряхин Сергей	978-5-4461-2331-5	Питер	16+	АвтомобилииПДД	Русский	Твердыйпереплёт	2023	610	756	192	325	2000	4,9	297	436	5	0
8	Некроманс. Opus 1	Евгения Сафонова	978-5-04-168484-6	Эксмо	16+	Русскоефэнтези	Русский		2022	549	675	480	460	7000	4,3	549	2224	32	58
9	Дар Кроуги	Анна Пушкина	978-5-17-147104-0	АСТ	16+	Фэнтези	Русский	Твердыйпереплёт	2022	419	645	352	379	4000	4,3	519	1589	87	16
10	Не дразните некроманта!	Ника Веймар	978-5-04-168483-9	Эксмо	16+	Романтическоефэнтези	Русский	Твердыйпереплёт	2022	443	545	320	260	2000	4,2	524	539	112	35
11	Когда заканчиваются сказки	Елена Кондрацкая	978-5-04-164574-8	Эксмо	18+	Русскоефэнтези	Русский	Твердыйпереплёт	2022	549	675	416	439	4000	4,2	477	1629	33	14
12	Стальной алхимик. Книга 17	Хирому Аракава	978-5-389-21140-7	Азбука	16+	Комиксы,манга,графическиероманы	Русский	Твердыйпереплёт	2022	1076	1076	264	497	8000	4,8	262	224	3	0

Найдём количество пустых и уникальных значений  
в каждом столбце

heading	0	heading	271
author	4	author	247
ISBN	0	ISBN	271
edition	0	edition	30
age_limit	1	age_limit	5
genre	4	genre	73
language	0	language	2
cover_type	14	cover_type	3
year	0	year	5
price	0	price	113
old_price	0	old_price	86
pages	0	pages	64
weight	0	weight	124
num_of_copies	8	num_of_copies	25
rating	0	rating	20
read	0	read	144
planned	0	planned	235
reviews	0	reviews	41
quotes	0	quotes	55



Преобразуем столбцы edition, age\_limit, language, cover\_type, year в категориальные переменные и создадим новый столбец discount, вычисляемый по формуле

$$\text{discount} = \frac{\text{old\_price} - \text{price}}{\text{old\_price}} * 100\%$$

Также создадим столбец sale, показывающий наличие скидки по столбцу discount

# Получаем следующий датасет, с которым и будем работать

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 271 entries, 0 to 270
Data columns (total 21 columns):
 #   Column              Non-Null Count  Dtype  
---  -
 0   heading             271 non-null   object 
 1   author              267 non-null   object 
 2   ISBN                271 non-null   object 
 3   edition             271 non-null   category
 4   age_limit           270 non-null   category
 5   genre               267 non-null   object 
 6   language            271 non-null   category
 7   cover_type          257 non-null   category
 8   year                271 non-null   category
 9   price               271 non-null   int64  
10  old_price           271 non-null   int64  
11  pages               271 non-null   int64  
12  weight              271 non-null   int64  
13  num_of_copies       263 non-null   float64 
14  rating              271 non-null   float64 
15  read                271 non-null   int64  
16  planned             271 non-null   int64  
17  reviews             271 non-null   int64  
18  quotes              271 non-null   int64  
19  discount            271 non-null   float64 
20  sale                271 non-null   category
dtypes: category(6), float64(3), int64(8), object(4)
memory usage: 45.6+ KB
```

## Строковые данные

- heading - название книги
- author - автор книги
- ISBN - уникальный код
- genre - жанр книги

## Категориальные данные

- edition - издательство
- age\_limit - ограничение по возрасту
- language - язык книги
- cover\_type - тип обложки
- year - год издания
- sale - наличие скидки

## Количественные данные

- price - стоимость книги (руб)
- old\_price - исходная стоимость книги (отличается, если есть скидка)
- pages - количество страниц
- weight - вес книги (г)
- num\_of\_copies - тираж издания (шт)
- rating - средняя оценка на LiveLib.ru
- read - количество отметок о прочтении на LiveLib.ru
- planned - количество планирующих прочитать на LiveLib.ru
- reviews - количество рецензий на LiveLib.ru
- quotes - количество цитат на LiveLib.ru
- discount - размер скидки

Выведем транспонированный срез первых 5 строк

	0	1	2	3	4
heading	Илос. Начало	Цианид	Мистер Вечный Канун. Город Полночи	Человек-бензопила. Книга 4. Во сне. Настоящая ...	Песнь Сорокопута. Ренессанс
author	Лия Арден	Кристина Старк	Владимир Торин, Олег Яковлев	Тацуки Фудзимото	Фрэнсис Кель
ISBN	978-5-04-164474-1	978-5-17-146746-3	978-5-00195-790-4	978-5-389-21080-6	978-5-04-164581-6
edition	INSPIRIA	АСТ	Манн,ИвановиФербер	Азбука	Эксмо
age_limit	16+	18+	18+	18+	18+
genre	Русскоефэнтези	Любовныероманы	Русскоефэнтези	Комиксы,манга,графическиероманы	Русскоефэнтези
language	Русский	Русский	Русский	Русский	Русский
cover_type	Твердыйпереплёт	Твердыйпереплёт	Твердыйпереплёт	Твердыйпереплёт	Твердыйпереплёт
year	2022	2022	2022	2022	2022
price	690	549	723	944	608
old_price	690	675	925	944	745
pages	448	448	576	392	480
weight	541	400	556	530	439
num_of_copies	18000.0	20000.0	3000.0	40000.0	5000.0
rating	4.6	4.2	4.3	4.6	4.3
read	1460	2069	1321	797	819
planned	3300	3920	3829	378	3093
reviews	46	92	74	1	30
quotes	83	270	72	3	27
discount	0.0	18.67	21.84	0.0	18.39
sale	False	True	True	False	True

Выведем первые 20 строк в отсортированном по убыванию тиража и цены датасете

	heading	author	num_of_copies	price
54	Тьма после рассвета	Александра Маринина	55000.0	690
215	Рыжий доктор	Екатерина Вильмонт	45000.0	525
3	Человек-бензопила. Книга 4. Во сне. Настоящая ...	Тацуки Фудзимото	40000.0	944
92	Тут мой дом	Эльчин Сафарли	40000.0	705
45	Двойник Запада	Лия Арден	40000.0	642
68	Магическая битва. Книга 3. Командный бой. Черн...	Гэгэ Акутами	30000.0	828
65	Naruto. Наруто. Книга 7. Наследие	Масаси Кисимото	25000.0	975
41	Магическая битва. Книга 4. Начало повиновения....	Гэгэ Акутами	20000.0	828
1	Цианид	Кристина Старк	20000.0	549
227	Дневник пофигиста: Тонкое искусство пофигизма ...	Марк Мэнсон	20000.0	531
40	Naruto. Наруто. Книга 9. День, когда их пути р...	Масаси Кисимото	18000.0	975
0	Илос. Начало	Лия Арден	18000.0	690
266	Моё настоящее имя. Истории с биографией	Людмила Улицкая	17000.0	797
114	Тэррлисс	Лия Стеффи	17000.0	608
181	Осколки Нашей Реальности	Медина Мирай	17000.0	507
121	Моя геройская академия. Книга 10. Культурный ф...	Кохэй Хорикоси	15000.0	823
104	Дом без воспоминаний	Донато Карризи	15000.0	641
195	Секреты «Голубого рифа»	Катя Брандис	12000.0	525
178	Великая иллюзия	Татьяна Степанова	12000.0	443
67	Пандемониум. Тьма в твоих глазах	Евгений Гаглов	12000.0	337

# 1. Рейтинг книги зависит от жанра

genre	rating	read	planned	reviews	quotes
Комиксы,манга,графическиероманы	4.664706	138.000000	124.823529	1.352941	0.470588
Детскаяфантастика	4.500000	13.000000	114.800000	2.000000	3.400000
Книгипсихологии	4.360000	41.400000	296.600000	10.400000	39.400000
Современныелюбовныероманы	4.300000	89.857143	383.285714	7.142857	10.285714
Отечественныедетективы	4.280000	103.000000	228.200000	8.200000	5.000000
Русскоефэнтези	4.218750	401.312500	1403.187500	21.625000	23.000000
Зарубежноефэнтези	4.166667	112.500000	792.333333	5.416667	7.583333
Современнаярусскаялитература	4.138889	72.833333	323.500000	6.500000	12.444444
Современнаяпроза	4.133333	94.916667	529.000000	10.083333	21.083333
Биографиииемуары	4.125000	42.375000	256.500000	3.500000	14.375000
Фэнтези	4.070000	273.600000	890.100000	31.100000	26.800000
Саморазвитие,личностныйрост	4.042857	42.857143	222.857143	3.142857	9.571429
Романтическоефэнтези	4.016667	264.166667	427.666667	23.833333	14.666667
Современнаязарубежнаялитература	4.010526	149.894737	742.947368	11.736842	13.105263
Зарубежныелюбовныероманы	4.010000	69.500000	589.400000	4.300000	5.300000
Зарубежныедетективы	3.960000	39.800000	334.400000	5.000000	5.200000
Современныедетективы	3.880000	38.200000	221.600000	2.600000	3.400000

Сводная таблица  
по средним значениям

Самые высокооцениваемые  
жанры - это комиксы и манги  
и детская фантастика

## 2. Статистика книги зависит от жанра

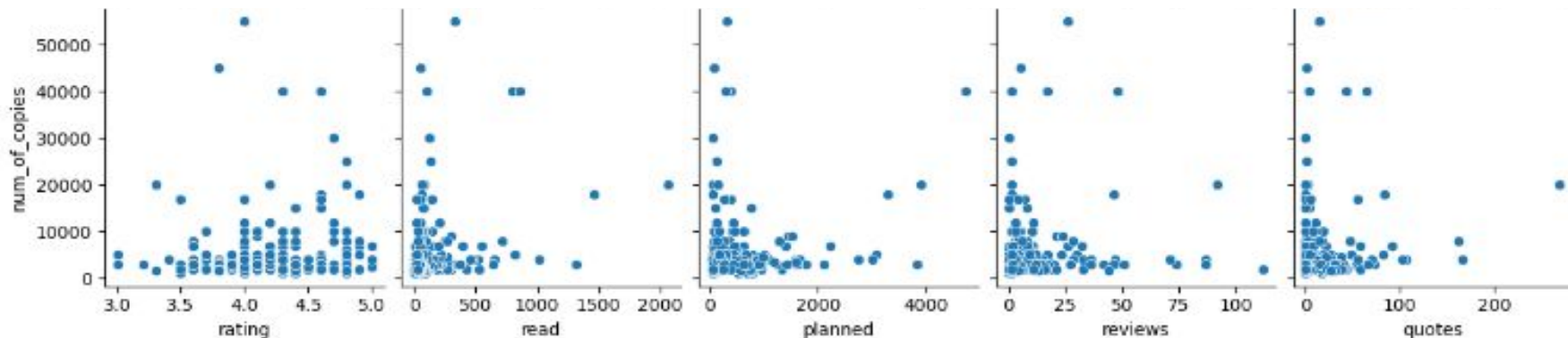
Сводная таблица по сумме значений

	genre
Современнаязарубежнаялитература	19
Современнаярусскаялитература	18
Комиксы,манга,графическиероманы	17
Русскоефэнтези	16
Современнаяпроза	12
Зарубежноефэнтези	12
Книгипопсихологии	10
Фэнтези	10
Зарубежныелюбовныероманы	10
Биографиииемуары	8
Современныелюбовныероманы	7
Саморазвитие,личностныйрост	7
Романтическоефэнтези	6
Современныедетективы	5
Детскаяфантастика	5
Отечественныедетективы	5
Зарубежныедетективы	5

	read	planned	reviews	quotes
genre				
Русскоефэнтези	6421	22451	346	368
Современнаязарубежнаялитература	2848	14116	223	249
Фэнтези	2736	8901	311	268
Комиксы,манга,графическиероманы	2346	2122	23	8
Романтическоефэнтези	1585	2566	143	88
Зарубежноефэнтези	1350	9508	65	91
Современнаярусскаялитература	1311	5823	117	224
Современнаяпроза	1139	6348	121	253
Зарубежныелюбовныероманы	695	5894	43	53
Современныелюбовныероманы	629	2683	50	72
Отечественныедетективы	515	1141	41	25
Книгипопсихологии	414	2966	104	394
Биографиииемуары	339	2052	28	115
Саморазвитие,личностныйрост	300	1560	22	67
Зарубежныедетективы	199	1672	25	26
Современныедетективы	191	1108	13	17
Детскаяфантастика	65	574	10	17

С большим отрывом первое место занимает русское фэнтези, хотя книг этого жанра меньше, чем современной литературы или комиксов. Но они также находятся в топе рейтинга, вместе с обычным, романтическим и зарубежным фэнтези

### 3. Чем больше тираж книги, тем больше значения данных с сайта LiveLib.ru

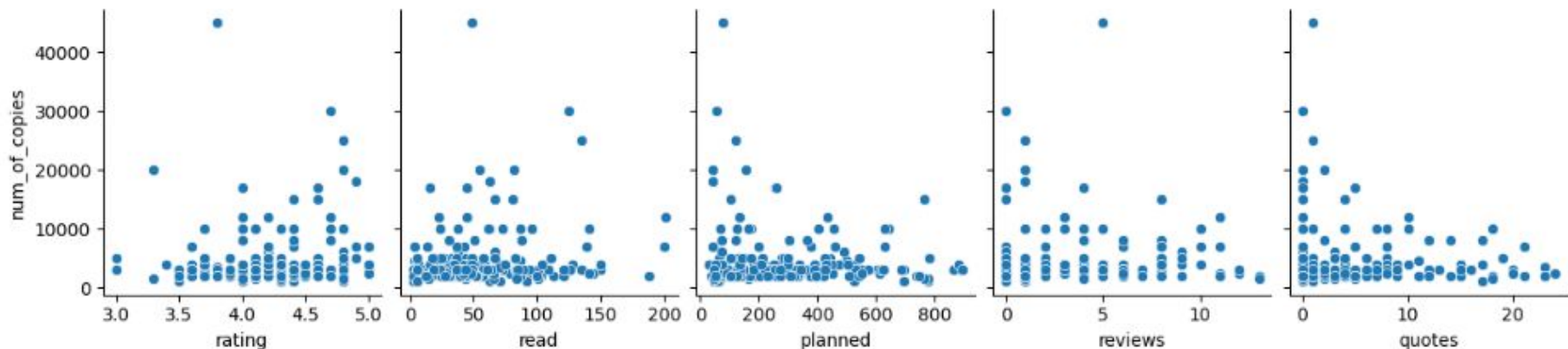


```
rating    0.113285
read      0.289922
planned   0.162214
reviews   0.102827
quotes    0.133342
dtype: float64
```

Коэффициенты корреляции довольно незначительные, но можно заметить, что большая часть данных расположена в 30% плоскости. Попробуем убрать выбросы



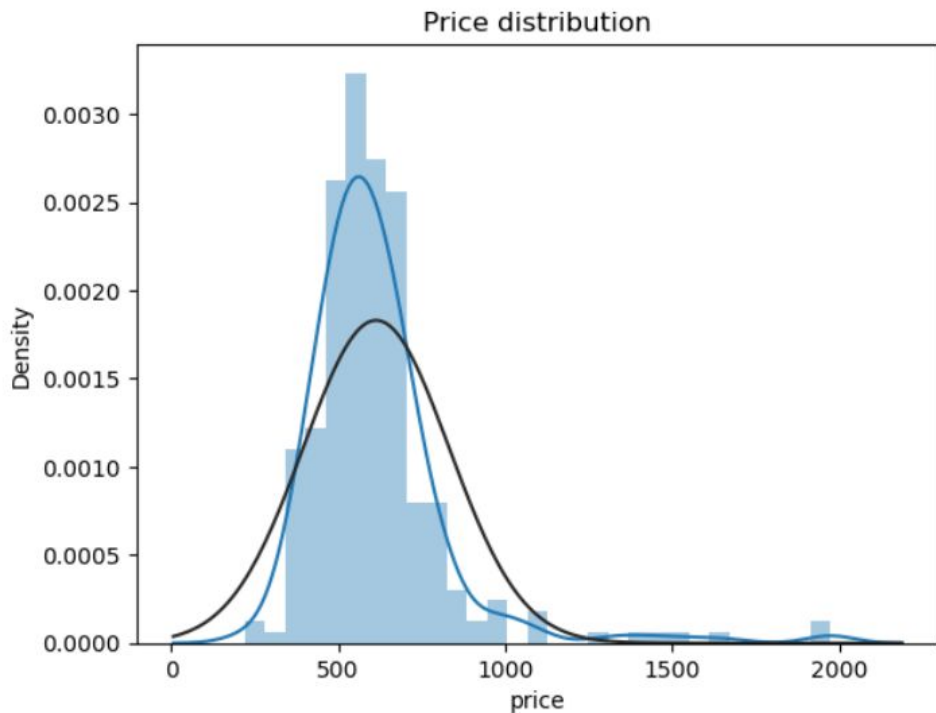
Данные были отфильтрованы с помощью межквартильного диапазона



```
rating      0.119796
read        0.187297
planned     -0.131501
reviews     -0.026147
quotes      -0.093979
dtype: float64
```

Как видно, коэффициент корреляции стал ещё меньше по модулю, значит данные величины не взаимосвязаны.

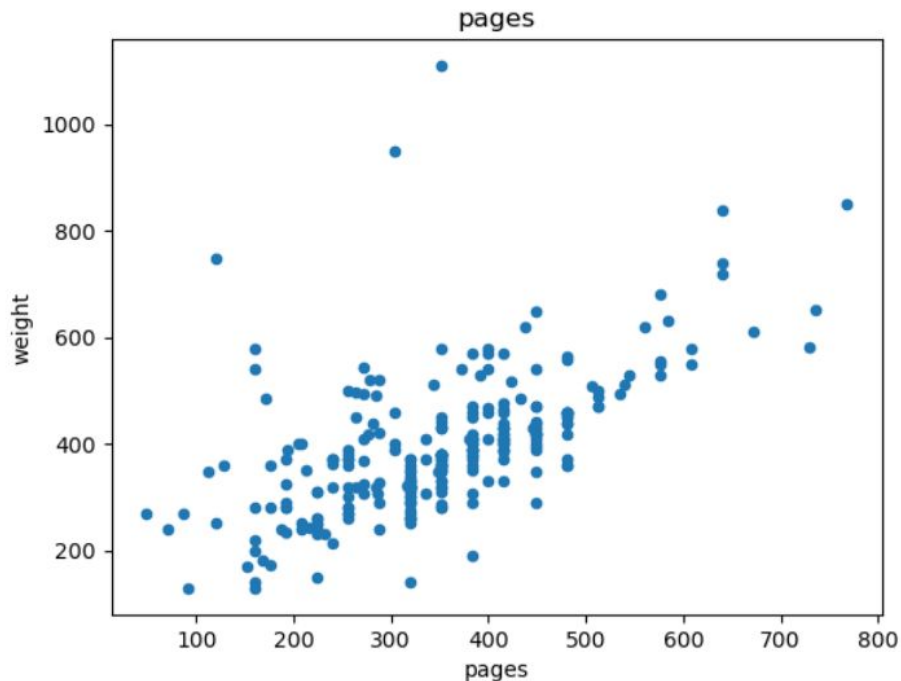
## 4. Стоимость книг подвержена нормальному распределению



В нашем случае "колокол" немного смещён влево и вытянут, но можно сказать, что стоимость книг подчиняется закону нормального распределения

## 5. Вес книги зависит от количества страниц

```
pages    0.594709  
dtype: float64
```



Коэффициент корреляции больше 0.5, что говорит о том, что присутствует средняя прямая корреляция (что в принципе логично, потому что чем больше бумаги, тем больше её вес)

## 6. Распределение ограничений по возрасту книг зависит от издательства.

Эксмо	85
АСТ	64
Азбука	23
Манн,ИвановиФербер	19
Альпинанон-фикшн	9
Питер	7
АльпинаПаблишер	7
INSPIRIA	7
Бомбора	6
РОСМЭН	6
Freedom	5
КоЛибри	5
Livebook	4
Городец	3
Синдбад	3
Стрекоза	2
Иностранка	2
КомпасГид	2
ФантомПресс	1
Фабрикакомиксов	1
Рипол-Классик	1
AdMarginemPress	1
Махаон	1
ИздательствоИванаЛимбаха	1
AltGraph	1
Альпина.Проза	1
PorcownBooks	1
LikeBook	1
Fanzon	1
Дримбук	1

edition	АСТ	Азбука	Манн,ИвановиФербер	Эксмо	Total
age_limit					
0+	0.000000	0.043478	0.052632	0.000000	0.010471
12+	0.031250	0.043478	0.157895	0.094118	0.073298
16+	0.578125	0.695652	0.578947	0.670588	0.633508
18+	0.390625	0.217391	0.210526	0.235294	0.282723

Отберём 4 популярных издательства и составим сводную таблицу.

В целом распределение довольно равномерное, но книги 0+ и 12+ в большинстве своём изданы издательством Манн,ИвановиФербер, 16+ - Азбукой и Эксмо, а книг 18+ больше всего у издательства АСТ.

## 7. Стоимость книги зависит от типа обложки

cover_type		Интегральныйперепле	Мягкийпереплет	Твердыйпереплёт
price	count	3.000000	19.000000	235.000000
	mean	436.000000	533.473684	623.029787
	std	180.277564	149.292542	227.182716
	min	236.000000	220.000000	337.000000
	median	486.000000	520.000000	586.000000
	max	586.000000	788.000000	1974.000000

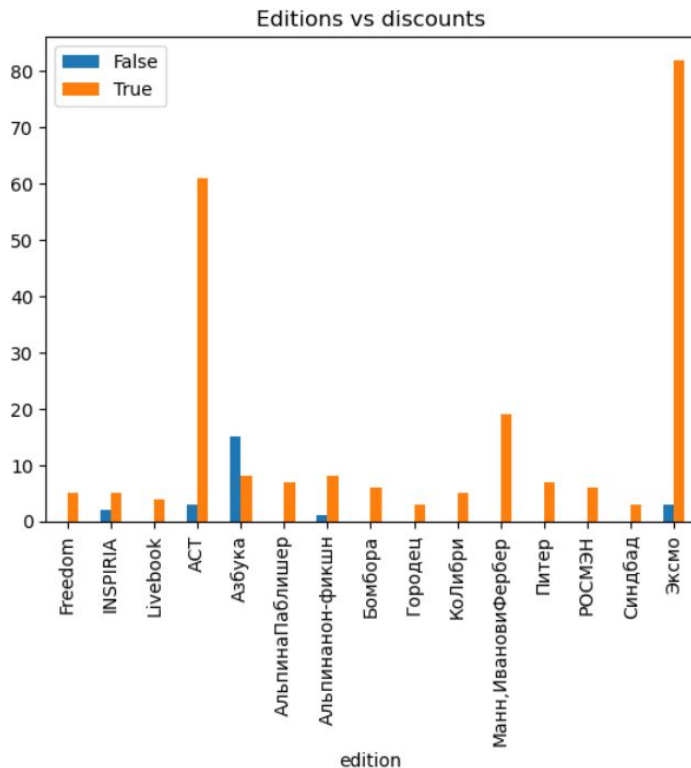
cover_type		Интегральныйперепле	Мягкийпереплет	Твердыйпереплёт
pages	count	3.000000	19.000000	235.000000
	mean	242.666667	272.210526	357.468085
	std	150.164354	98.760979	116.538290
	min	152.000000	128.000000	49.000000
	median	160.000000	240.000000	352.000000
	max	416.000000	448.000000	768.000000

Твёрдый переплёт дороже мягкого, который в свою очередь дороже интегрального. Проверим, не влияет ли на стоимость количество страниц

Получили аналогичный результат распределения типа обложек по количеству страниц. Коэффициент корреляции между стоимостью и страницами равен - 0.02.

Получаем, что зависимости стоимости от количества страниц нет, значит, стоимость книги на самом деле зависит от типа переплёта

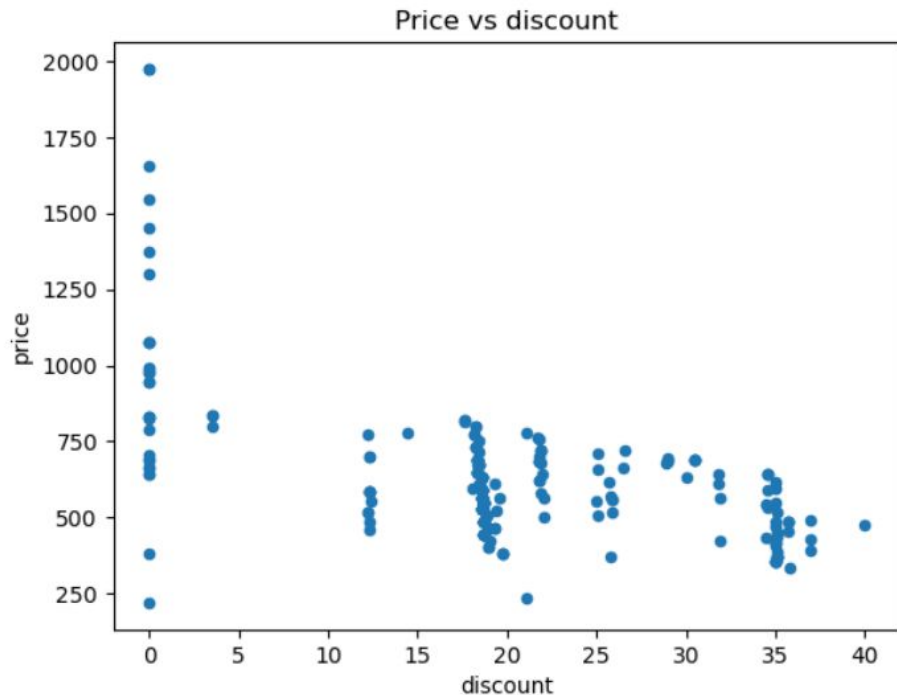
## 8. Наличие скидки зависит от издательства



Можно заметить, что книги десяти издательств все со скидкой, четырёх - почти всех, за исключением буквально 1-3, а вот у издательства Азбука буквально треть книг со скидкой. Значит, наличие скидки на сайте читай города зависит от издательства

## 9. Процент скидки зависит от первоначальной стоимости

```
discount    -0.563201  
dtype: float64
```



Модуль коэффициента корреляции больше 0.5, что говорит о средней корреляции, а знак отрицательный, поэтому величины обратно зависимы. То есть чем больше цена, тем меньше итоговая скидка



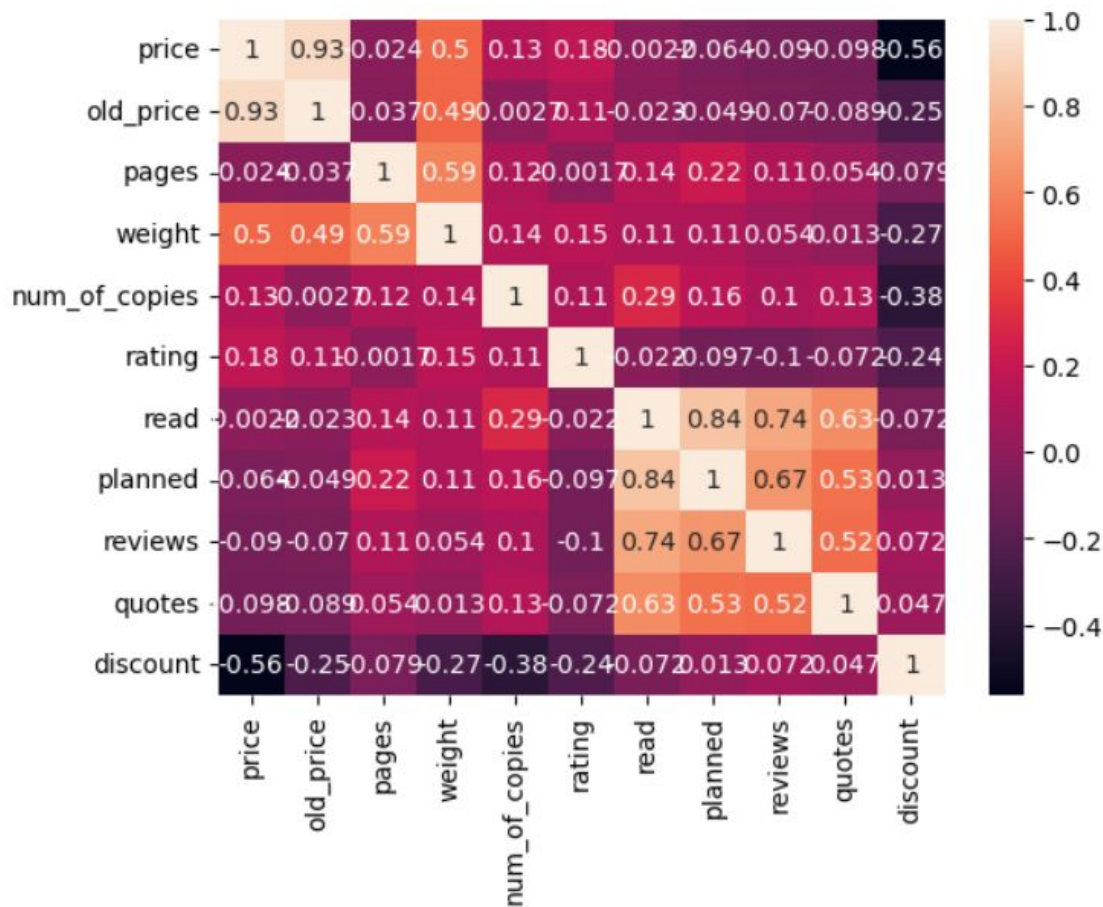
## 10. Более качественные (читаемые и высоко оцененные) книги стоят дороже

Коэффициент корреляции близок к нулю, только у рейтинга есть подозрение на зависимость от стоимости (или наоборот), поэтому гипотеза неверна.

```
rating    0.176852
read      -0.002163
planned   -0.063926
reviews   -0.089957
quotes    -0.098284
dtype: float64
```

Выведем сводную статистику по всем количественным данным датасета

	price	old_price	pages	weight	num_of_copies	rating	read	planned	reviews	quotes	discount
count	271.000000	271.000000	271.000000	271.000000	263.000000	271.000000	271.000000	271.000000	271.000000	271.000000	271.000000
mean	613.542435	757.107011	349.859779	394.221402	5217.490494	4.178598	120.073801	490.033210	9.228782	13.889299	19.599963
std	218.344336	205.291030	119.491526	128.525206	6763.056229	0.399239	220.191537	633.083151	15.472774	27.487956	9.599275
min	220.000000	220.000000	49.000000	129.000000	1000.000000	3.000000	2.000000	31.000000	0.000000	0.000000	0.000000
25%	485.000000	645.000000	272.000000	319.000000	2500.000000	3.900000	27.000000	139.500000	2.000000	0.000000	18.360000
50%	568.000000	725.000000	352.000000	380.000000	3000.000000	4.200000	54.000000	300.000000	4.000000	4.000000	18.660000
75%	684.000000	845.000000	416.000000	451.000000	5000.000000	4.500000	108.500000	570.000000	8.000000	16.000000	21.885000
max	1974.000000	1974.000000	768.000000	1110.000000	55000.000000	5.000000	2069.000000	4743.000000	112.000000	270.000000	40.000000



Составим тепловую карту корреляции между числовыми данными датасета. Все возможные корреляции между величинами были учтены в наших гипотезах

В данном исследовании были собраны данные о популярных книгах, а затем выявлены зависимости между этими данными. Были подтверждены 8 гипотез из 10, что говорит о том, что почти все мои предположения были верны.

Были выявлены наиболее популярные жанры, зависимость веса книги от количества страниц, стоимости от обложки, наличия скидки от издательства и процента скидки от первоначальной стоимости

