### Анализ и выводы по графикам

#### 1. Распределение рейтингов книг

На гистограмме видно, что большинство книг имеют рейтинг от 3,5 до 4,5. Это говорит о том, что книги в выборке в основном имеют положительные оценки, а негативные или низкие рейтинги встречаются реже. Это может свидетельствовать о популярности книг в данной базе данных или о том, что пользователи чаще оставляют высокие оценки.

\*\*Рекомендации:\*\*

- Учитывая, что распределение сконцентрировано в области высоких оценок, можно изучить, какие факторы способствуют получению высоких рейтингов (жанр, авторы, и т.д.).

#### 2. Категоризация цен на книги

Текущий график отображает цены в виде столбчатой диаграммы. Однако из-за длинных подписей на оси X и их расположения под углом, график воспринимается неинтуитивно.

\*\*Рекомендации и изменения:\*\*

- Заменить столбчатую диаграмму на круговую диаграмму, чтобы лучше визуализировать долю каждой категории цен.

- Использовать сокращенные названия для категорий, например: "До 5", "5-10", "10-20", "20-50", "50+", чтобы уменьшить количество текста и улучшить читаемость.

\*\*Мини-вывод:\*\*

Большинство книг имеют стоимость до 10 евро, что говорит о доступности большинства книг для широкой аудитории.

#### 3. Топ-10 популярных жанров

На графике показаны самые популярные жанры книг. Однако, длинные подписи жанров срезаются, что затрудняет восприятие.

\*\*Рекомендации:\*\*

- Увеличить размер поля графика или уменьшить шрифт подписей, чтобы все жанры отображались корректно.

- Добавить пояснения к некоторым жанрам, чтобы улучшить понимание (например, если "Fiction" и "Nonfiction" слишком обширные категории, можно дополнительно разбить их на подкатегории).

\*\*Мини-вывод:\*\*

Самыми популярными жанрами являются Fiction и Nonfiction, что указывает на большой интерес к этим типам литературы. На третьем месте Fantasy, что говорит о популярности вымышленных миров.

#### 4. Корреляционная матрица

На графике корелляционной матрицы показано, как различные показатели взаимосвязаны между собой. Видно, что сильная корреляция наблюдается между рейтингом и количеством оценок, а также между рейтингом и процентом прочтения.

\*\*Рекомендации:\*\*

- Увеличить поля вокруг графика или уменьшить шрифт подписей, чтобы текст на осях не обрезался.

- Рассмотреть дополнительные корреляции между параметрами, чтобы выявить возможные закономерности.

\*\*Мини-вывод:\*\*

Высокая корреляция между рейтингом и количеством прочтений указывает на то, что более популярные книги чаще получают высокие оценки.

#### 5. Боксплот цен на книги

Боксплот цен имеет много выбросов, что указывает на значительные отклонения от среднего значения. Вертикальная черта вместо прямоугольника может указывать на проблемы в распределении данных.

\*\*Рекомендации:\*\*

- Изучить выбросы и определить, действительно ли эти книги существуют с такой высокой ценой или это ошибки в данных.

- Рассмотреть применение логарифмической шкалы для лучшей визуализации.

\*\*Мини-вывод:\*\*

Большинство книг имеют цену до 50 евро, однако существуют экземпляры с ценами выше 800 евро, что может требовать дополнительной проверки.

#### 6. Боксплот количества страниц

Аналогичная ситуация с боксплотом количества страниц: есть большое количество выбросов. Книги с более чем 1000 страниц встречаются редко и требуют проверки.

\*\*Рекомендации:\*\*

- Анализировать выбросы и исключить аномальные значения.

- Рассмотреть распределение страниц по различным жанрам, чтобы понять, в каких жанрах обычно встречаются длинные книги.

\*\*Мини-вывод:\*\*

Большинство книг имеют от 100 до 500 страниц, что соответствует среднему объему книг, но есть редкие случаи с более чем 14000 страниц, которые требуют анализа.

#### 7. Зависимость рейтинга от цены

На графике показана слабая положительная зависимость между рейтингом и ценой, но большинство точек сконцентрированы в нижней части графика.

\*\*Рекомендации:\*\*

- Удалить или заменить красную линию на более релевантную линию регрессии, которая учитывает разброс данных.

- Рассмотреть возможность удаления крайних значений (выбросов) для лучшего понимания тренда.

\*\*Мини-вывод:\*\*

Цена книги не оказывает значительного влияния на её рейтинг, что говорит о том, что пользователи оценивают содержание, а не стоимость.

#### 8. Зависимость рейтинга от количества страниц

Слабая положительная зависимость указывает на то, что более длинные книги могут иметь немного более высокий рейтинг.

\*\*Рекомендации:\*\*

- Удалить красную линию или заменить её на более точное отображение данных.

- Проанализировать выбросы, чтобы понять, почему книги с маленьким количеством страниц имеют высокие оценки.

\*\*Мини-вывод:\*\*

Рейтинг книг не сильно зависит от количества страниц, однако более длинные книги могут восприниматься как более информативные или захватывающие.

#### 9. Боксплот годов публикаций

Большое количество выбросов до 1900 года и после 2020 указывают на ошибки или редкие случаи.

\*\*Рекомендации:\*\*

- Проверить значения и исправить ошибки, если они есть (например, книги с датой 2050).

- Изучить влияние года публикации на рейтинг и количество прочтений.

\*\*Мини-вывод:\*\*

Большинство книг из базы данных опубликованы в последние десятилетия, что отражает текущие тенденции в чтении и публикации книг.

1. **Gender Distribution of Top-100 Book Authors**

На круговой диаграмме видно, что среди топ-100 книг большинство авторов составляют мужчины (53.8%), на втором месте находятся женщины (27.7%), и около 18.5% авторов имеют неопределенный пол.

Эта диаграмма демонстрирует, что мужские авторы доминируют в топ-100 книг. Однако, женщины также составляют значительную часть, что говорит о растущем присутствии женских авторов в литературе.

2. **Rating vs Price by Gender**

На графике видна зависимость рейтинга от цены, где синий цвет представляет мужских авторов, красный — женских, а серый — неизвестный пол.

Основная масса данных сосредоточена в нижней части графика, что указывает на то, что книги с более низкой ценой имеют более широкое распространение среди авторов обоих полов.

Визуально можно заметить, что для книг с очень высокой ценой большинство авторов — мужчины, а также наблюдаются аномалии, где цена существенно выше среднего.

Возможно, стоит рассмотреть добавление линий среднего значения для каждого пола или добавление точек-медиан для более наглядного представления.

Вывод к графику **"Price Distribution by Gender"**:

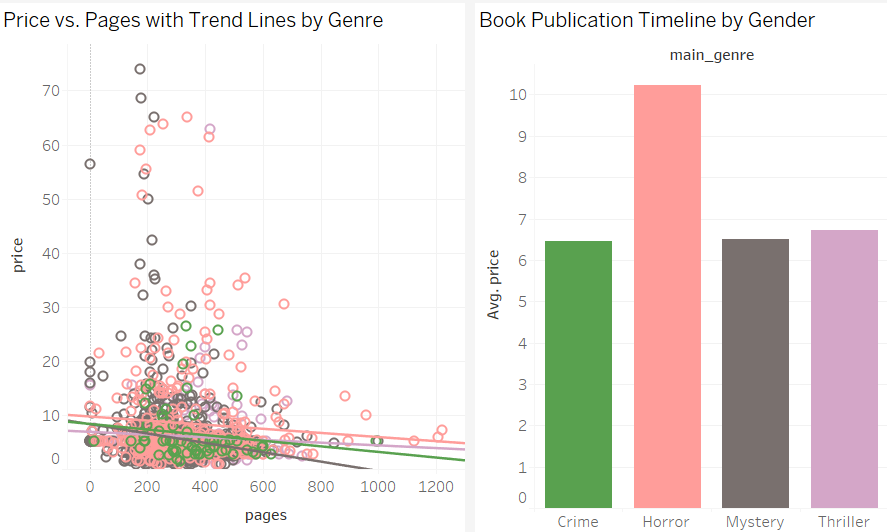
На графике представлено распределение цен на книги по гендеру авторов.

1. **Общее распределение:**
   * В обеих категориях (мужские и женские авторы) большая часть цен сосредоточена в нижней части графика, что указывает на то, что большинство книг имеют относительно низкую стоимость.
2. **Средние значения и межквартильный размах:**
   * Средние цены на книги авторов обоих полов примерно одинаковы и находятся в нижнем ценовом диапазоне.
   * Межквартильный размах также примерно одинаков, что указывает на схожее распределение цен на книги среди мужчин и женщин.
3. **Аномальные значения:**
   * Есть несколько выбросов (outliers), где цены на книги значительно выше средней стоимости. Это может свидетельствовать о специальных изданиях или редких коллекционных экземплярах.
   * У женщин наблюдается больше выбросов в диапазоне 100-500, что может указывать на наличие отдельных дорогостоящих книг.
4. **Рекомендации:**
   * Можно дополнительно исследовать книги с высокими ценами для более детального анализа причин таких выбросов.
   * Для улучшения визуализации можно добавить медианные значения и дополнительные пояснения для выбросов.

График демонстрирует, что книги с высоким ценовым диапазоном встречаются у авторов обоих полов, но в основном остаются в нижней части ценовой шкалы.

* ***Interactive Tableau Dashboard.***

***The dashboard allows for easy comparison between genres, showing how different genres perform in terms of ratings, prices, and number of books published.***



* ***The dashboard includes various visualizations such as bar charts, line graphs, and scatter plots that offer detailed insights into the book market.***

***For instance, you can see which genres have the highest average ratings or which genres are dominated by male or female authors***