([DataSet Presentation - Book Sales])

{Produced with Kaggle and Jupyter Notebook}

By Phillip Seggie

```
bk_df.dtypes
```

✓ 0.0s

```
index                  int64
Publishing Year      float64
Book Name             object
Author                object
language_code         object
Author_Rating         object
Book_average_rating  float64
Book_ratings_count     int64
genre                 object
gross sales          float64
publisher revenue    float64
sale price           float64
sales rank             int64
Publisher             object
units sold             int64
dtype: object
```

Book Sales and Ratings.

- This Dataset provides an insight into a multitude of different aspects of books such as:
- Book Names
- Publishing Year
- Name of Author
- Genre – Fiction, non-Fiction and Childrens
- Gross sales
- Sale price
- Units Sold

Read the Power of Data with Jupyter Notebook

In the world of data analysis, Jupyter Notebook stands out as an innovative tool that revolutionizes the way we explore, visualize, and interpret datasets. With its interactive environment, Jupyter Notebook empowers analysts, researchers, and data enthusiasts to perform real-time data manipulation and gain insights that drive strategic decisions.

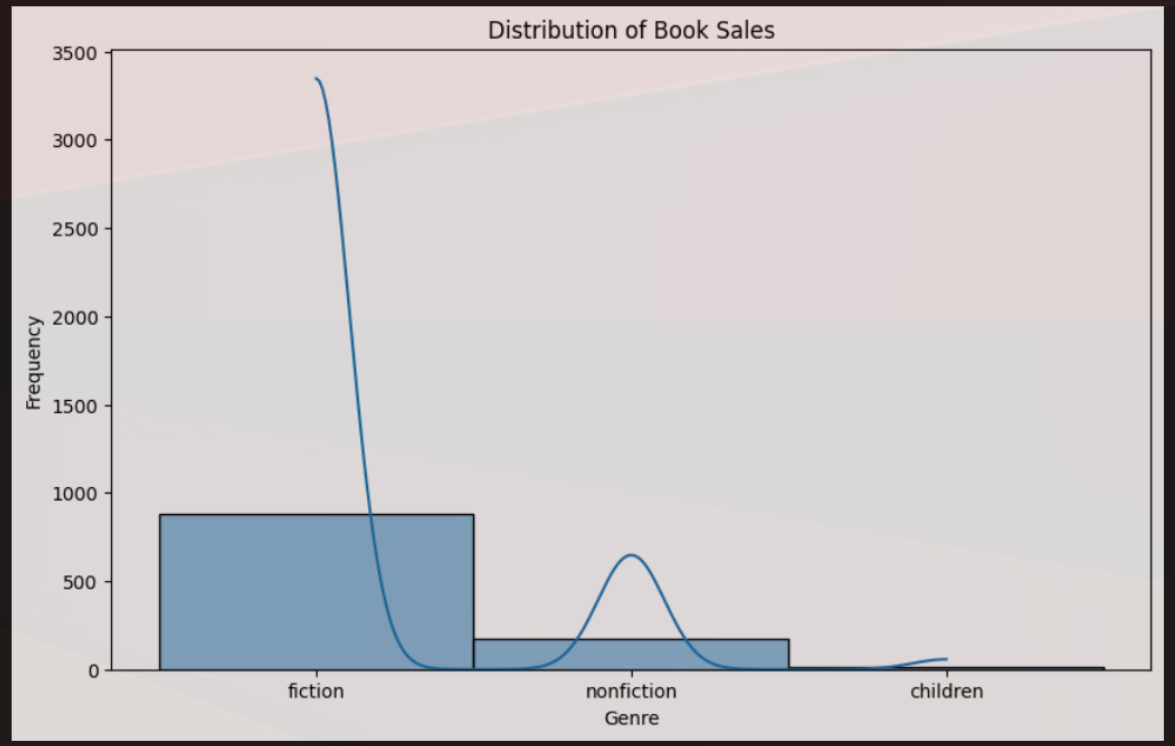
Wade through Jupyter Notebook and let the data tell its story.

Market Analysis

Distribution of Book Sales by Genre

```
plt.figure(figsize=(10, 6))
sns.histplot(bk['genre'], bins=20, kde=True)
plt.title('Distribution of Book Sales')
plt.xlabel('Genre')
plt.ylabel('Frequency')
plt.show()
```

✓ 0.1s

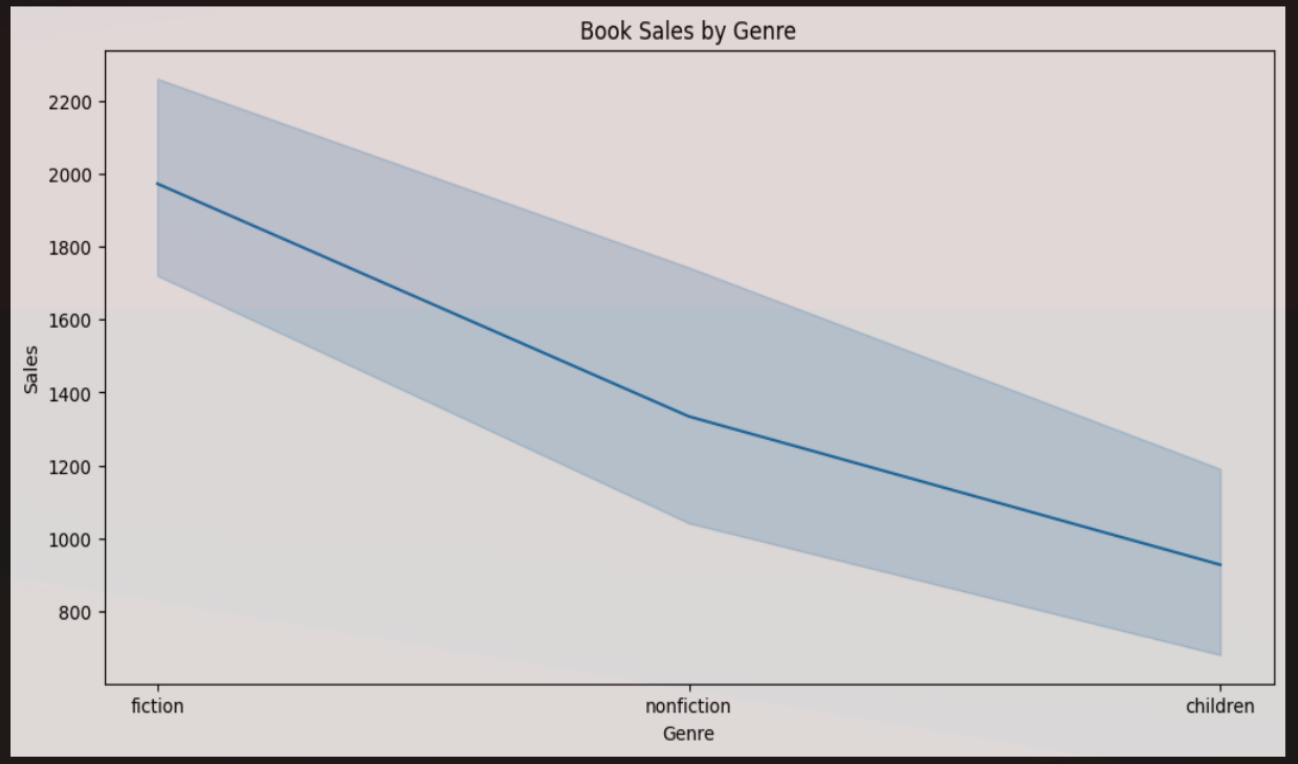


The graph illustrates the distribution of book sales across different genres. Each bar represents a genre, and the height of the bars indicates the frequency of book sales within that genre. This graph is valuable for discerning patterns in book sales, identifying popular genres, and gaining insights into the overall distribution of sales within the dataset.

Book Sales by Genre

```
plt.figure(figsize=(12, 6))
sns.lineplot(x='genre', y='gross sales', data=bk)
plt.title('Book Sales by Genre')
plt.xlabel('Genre')
plt.ylabel('Sales')
plt.show()
```

✓ 0.1s



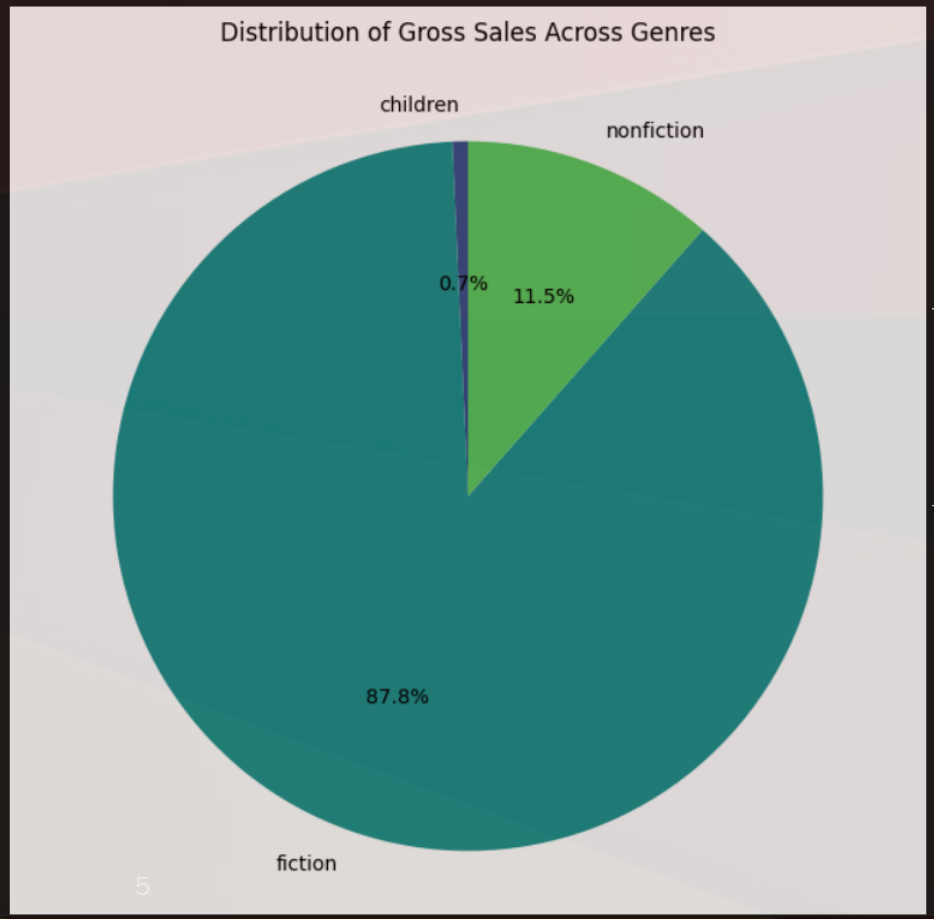
This visualization allows for a clear comparison of sales trends across genres, emphasizing any notable variations or patterns in the dataset. The line plot serves as an effective tool for tracking changes in sales over different genres and provides insights into the relative performance of each category.

Market Analysis

Distribution of Gross Sales across Genres

```
# pie chart for the overall distribution of gross sales across genres
plt.figure(figsize=(8, 8))
sales_by_genre = bk.groupby('genre')['gross sales'].sum()
plt.pie(sales_by_genre, labels=sales_by_genre.index, autopct='%1.1f%%', startangle=90, colors=sns.color_palette('viridis', len(sales_by_genre)))
plt.title('Distribution of Gross Sales Across Genres')
plt.show()
```

✓ 0.0s



The displayed pie chart offers a concise overview of the distribution of gross sales across different book genres.

We can start to see the relative importance of each genre within this dataset in terms of total sales, it would be used to aid understanding the market better and making any informed decisions as to where to allocate resource or marketing strategies.

Potential Market Strategies.

1. **Content Emphasis:** Publishers could prioritize acquiring and promoting fiction titles.

2. **Targeted Marketing:** Tailor marketing campaigns to highlight fiction books through various channels, such as social media, book clubs, and online platforms.

3. **Author Collaborations:** Partnering with successful fiction authors or emerging talents can enhance the appeal of a publisher's catalogue.

4. **Market Research:** Continuous monitoring of reader preferences and staying abreast of trends within the fiction genre can inform editorial decisions.

5. **Digital Presence:** Capitalize on the popularity of e-books and audiobooks within the fiction genre by optimizing digital distribution channels.

6. **Diversification Within Fiction:** While fiction dominates, consider exploring sub-genres or niche categories within fiction to cater to diverse reader interests.

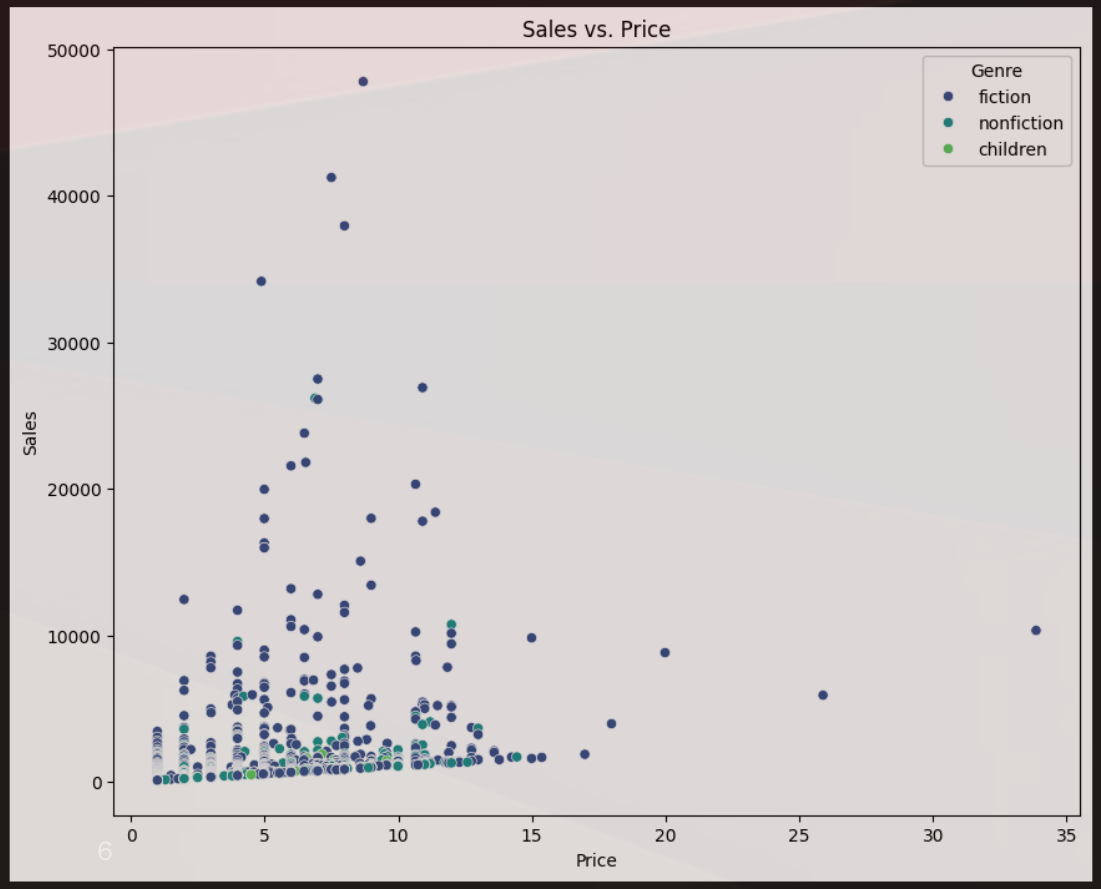
7. **Collaboration with Retailers:** Forge strategic partnerships with book retailers to ensure prominent placement and visibility for fiction titles in physical and online bookstores.

Market Analysis

Sales VS Price

```
plt.figure(figsize=(10, 8))
sns.scatterplot(x='sale price', y='gross sales', data=bk, hue='genre', palette='viridis')
plt.title('Sales vs. Price')
plt.xlabel('Price')
plt.ylabel('Sales')
plt.legend(title='Genre')
plt.show()
```

✓ 0.1s



Optimal Price Points: The plot may reveal clusters of points indicating price points where books within a particular genre consistently achieve higher sales.

Genre Preferences: Outliers or distinctive patterns for specific genres might suggest varying consumer behaviours and preferences.

Price Sensitivity: Analysis of the scatter plot may uncover genres where sales are particularly sensitive to price changes.

Identification of High-Performing Titles: Outliers with exceptionally high sales at specific price points may represent blockbuster titles.

Genre-Specific Trends: Certain genres may exhibit non-linear relationships between price and sales, indicating unique market dynamics.

Competitive Positioning: Understanding how different genres perform at various price levels can provide insights into the competitive landscape.

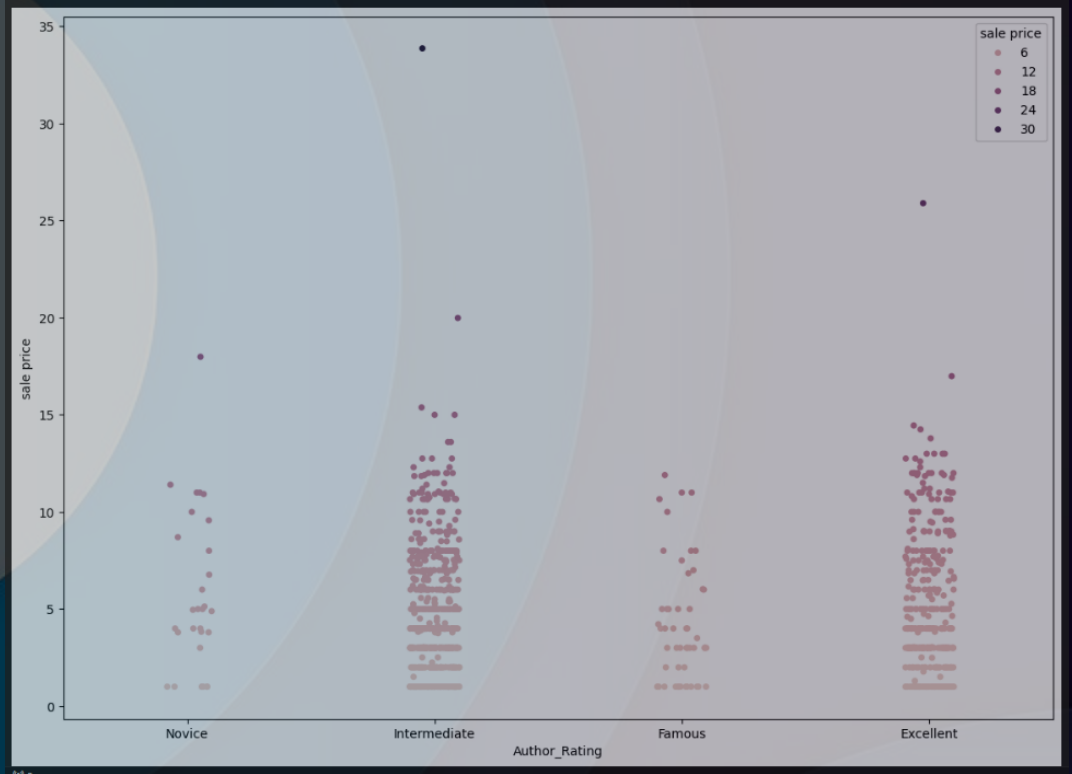
Consumer Behaviour Analysis: Examining clusters or outliers may uncover unexpected correlations between sales and price, shedding light on underlying consumer behaviours.

Sales Price vs. Author Rating

```
fig, axs = plt.subplots(nrows=1, ncols=1, figsize=(14,10))

sns.stripplot(data=bk,
              x="Author_Rating",
              y="sale price",
              hue='sale price',
              ax=axs)
```

✓ 0.1s

```
<Axes: xlabel='Author_Rating', ylabel='sale price'>
```



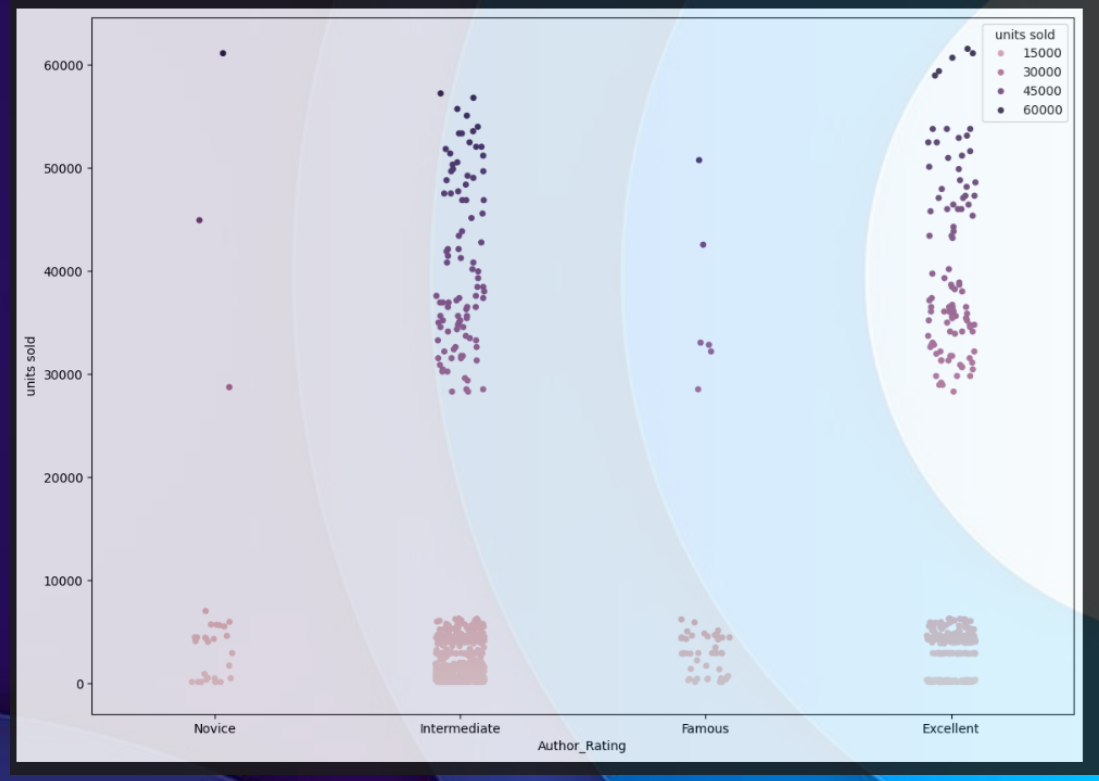
The Sales Price vs Author provides insights into pricing dynamics based on an author's reputation.

Units Sold vs. Author Rating

```
fig, axs = plt.subplots(nrows=1, ncols=1, figsize=(14,10))

sns.stripplot(data=bk,
              x="Author_Rating",
              y="units sold",
              hue='units sold',
              ax=axs)
```

✓ 0.2s

```
<Axes: xlabel='Author_Rating', ylabel='units sold'>
```



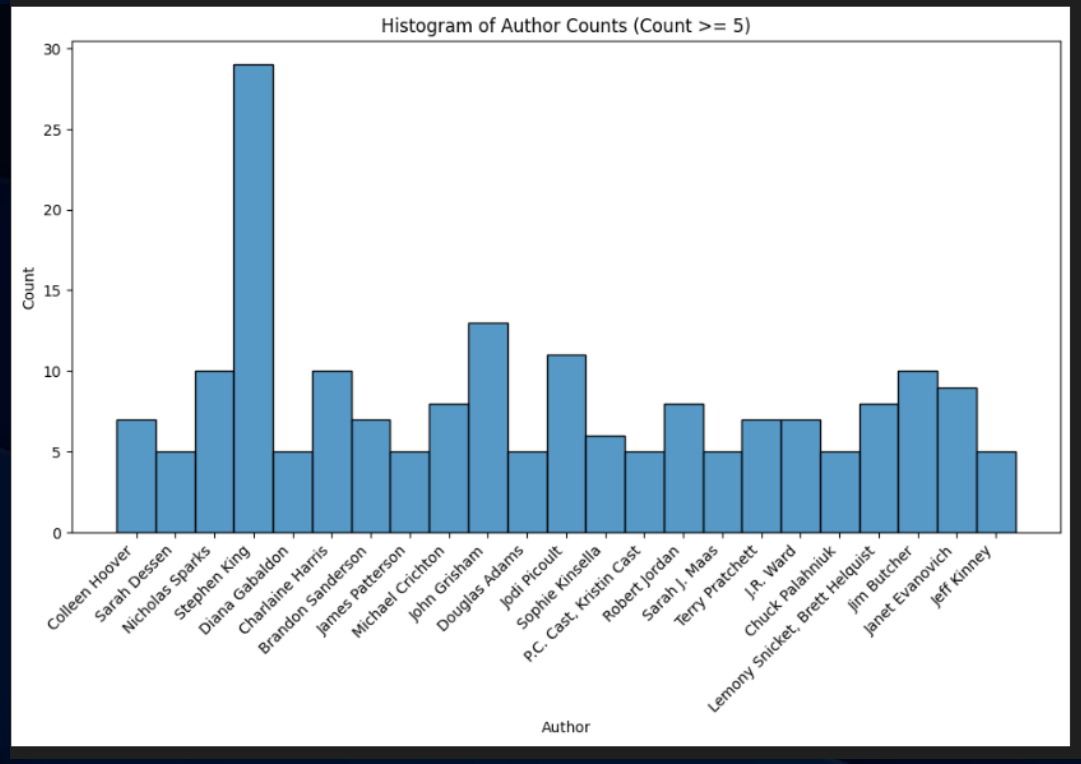
The Sold vs Author shows how the ratings correlate with actual sales volume.

Histogram of Authors Publications

Published 5 + Books

```
author_counts = bk['Author'].value_counts()

# Filter authors with counts greater than or equal to 5
selected_authors = author_counts[author_counts >= 5].index

# Filter the DataFrame to include only selected authors
filtered_data = bk[bk['Author'].isin(selected_authors)]

# Plot the histogram
plt.figure(figsize=(12, 6))
sns.histplot(data=filtered_data, x="Author", bins=len(selected_authors), multiple="stack")
plt.title('Histogram of Author Counts (Count >= 5)')
plt.xlabel('Author')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')  # Rotates x-axis labels for better visibility!
plt.show()
```

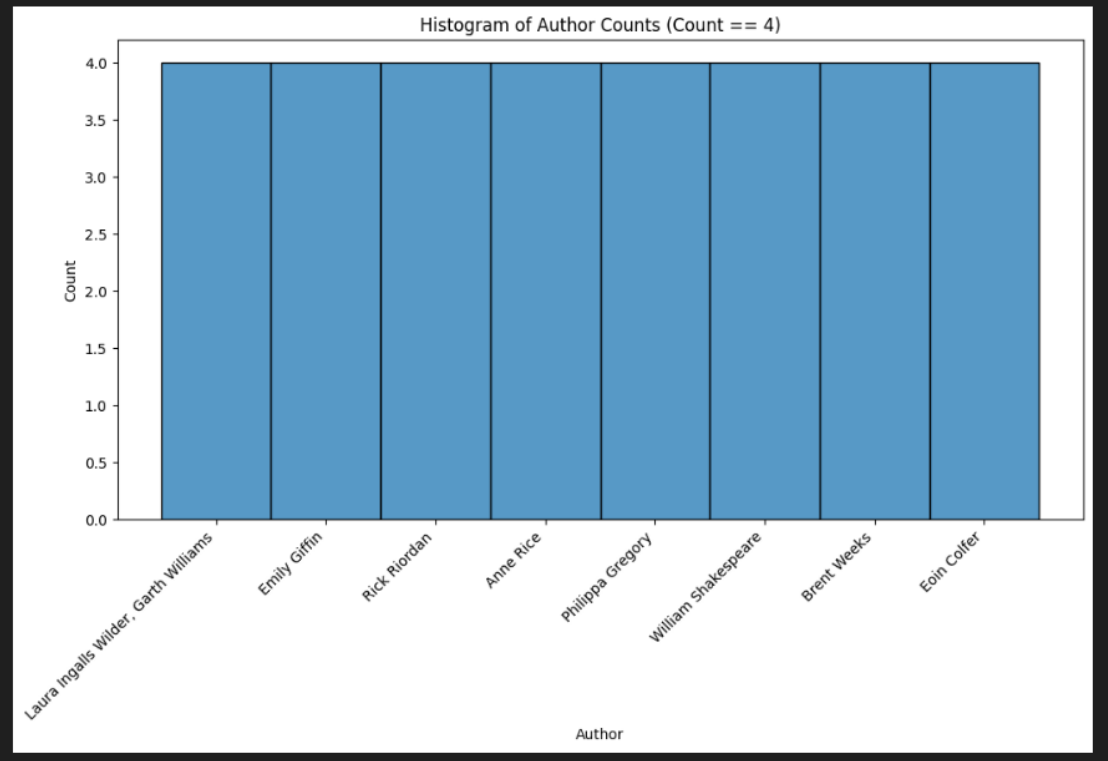
✓ 0.1s



Published 4 Books

```
author_counts = bk['Author'].value_counts()

# Filter authors with counts greater than or equal to 5
selected_authors = author_counts[author_counts == 4].index

# Filter the DataFrame to include only selected authors
filtered_data = bk[bk['Author'].isin(selected_authors)]

# Plot the histogram
plt.figure(figsize=(12, 6))
sns.histplot(data=filtered_data, x="Author", bins=len(selected_authors), multiple="stack")
plt.title('Histogram of Author Counts (Count == 4)')
plt.xlabel('Author')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```

✓ 0.1s

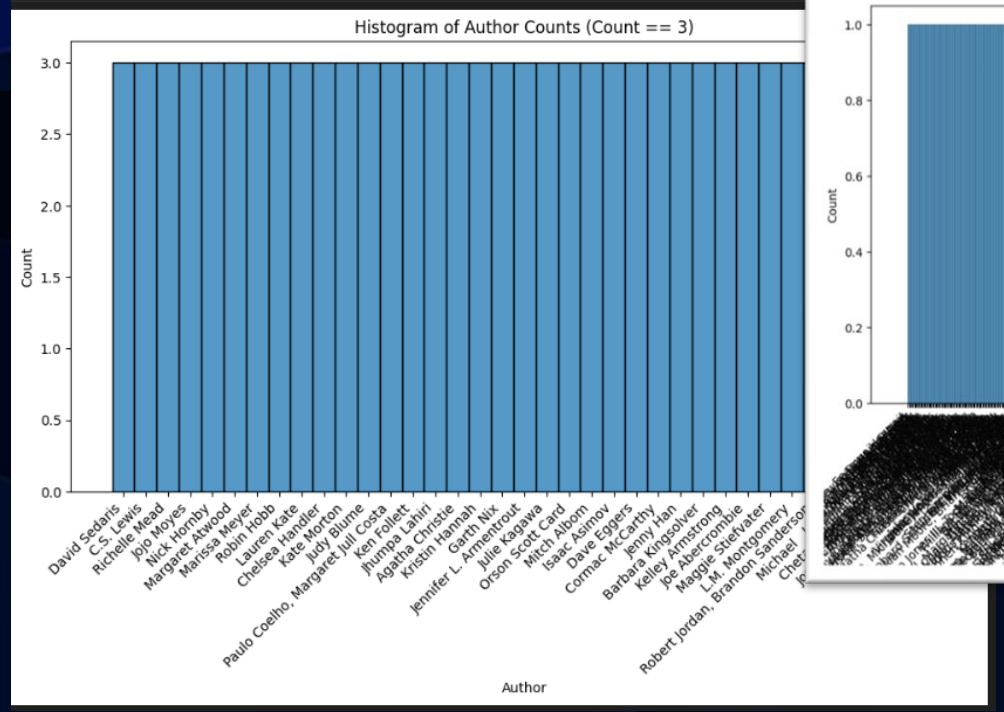


Histogram of Authors Publications

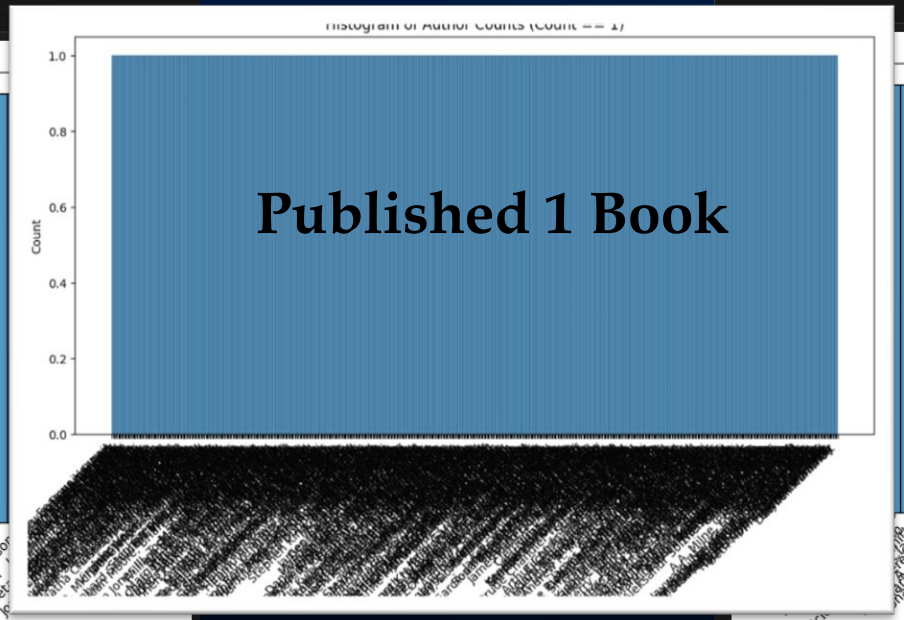
Published 3 Books

```
author_counts = bk['Author'].value_counts()

# Filter authors with counts greater than or equal to 5
selected_authors = author_counts[author_counts == 3].index

# Filter the DataFrame to include only selected authors
filtered_data = bk[bk['Author'].isin(selected_authors)]

# Plot the histogram
plt.figure(figsize=(12, 6))
sns.histplot(data=filtered_data, x="Author", bins=len(selected_authors), multiple="stack")
plt.title('Histogram of Author Counts (Count == 3)')
plt.xlabel('Author')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```

0.3s

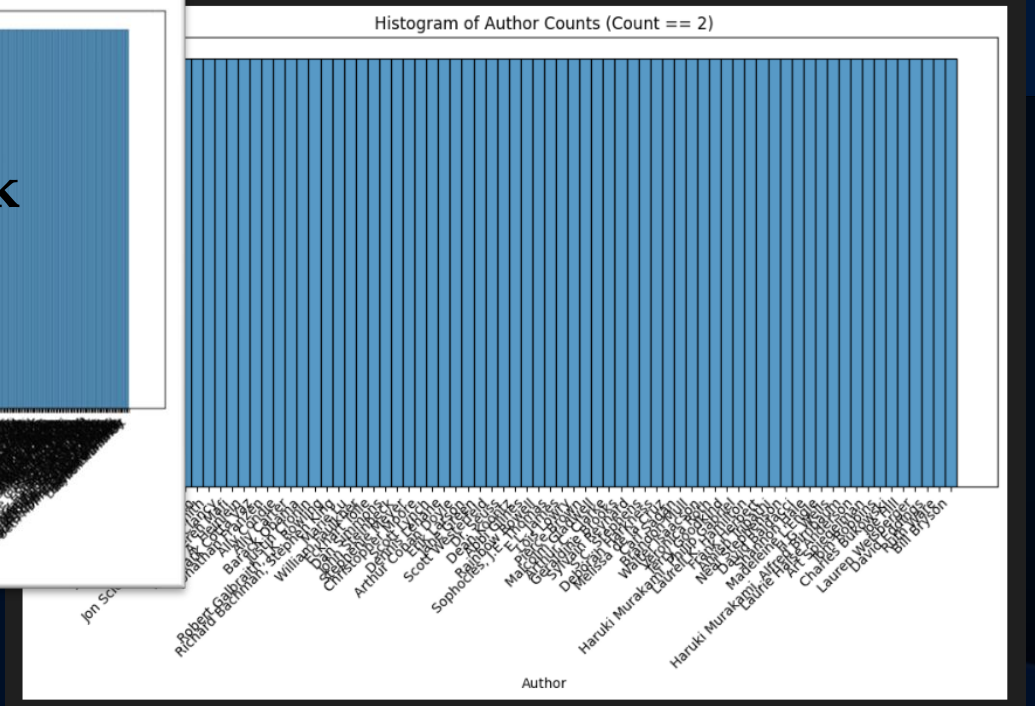


Published 1 Book



Published 2 Books

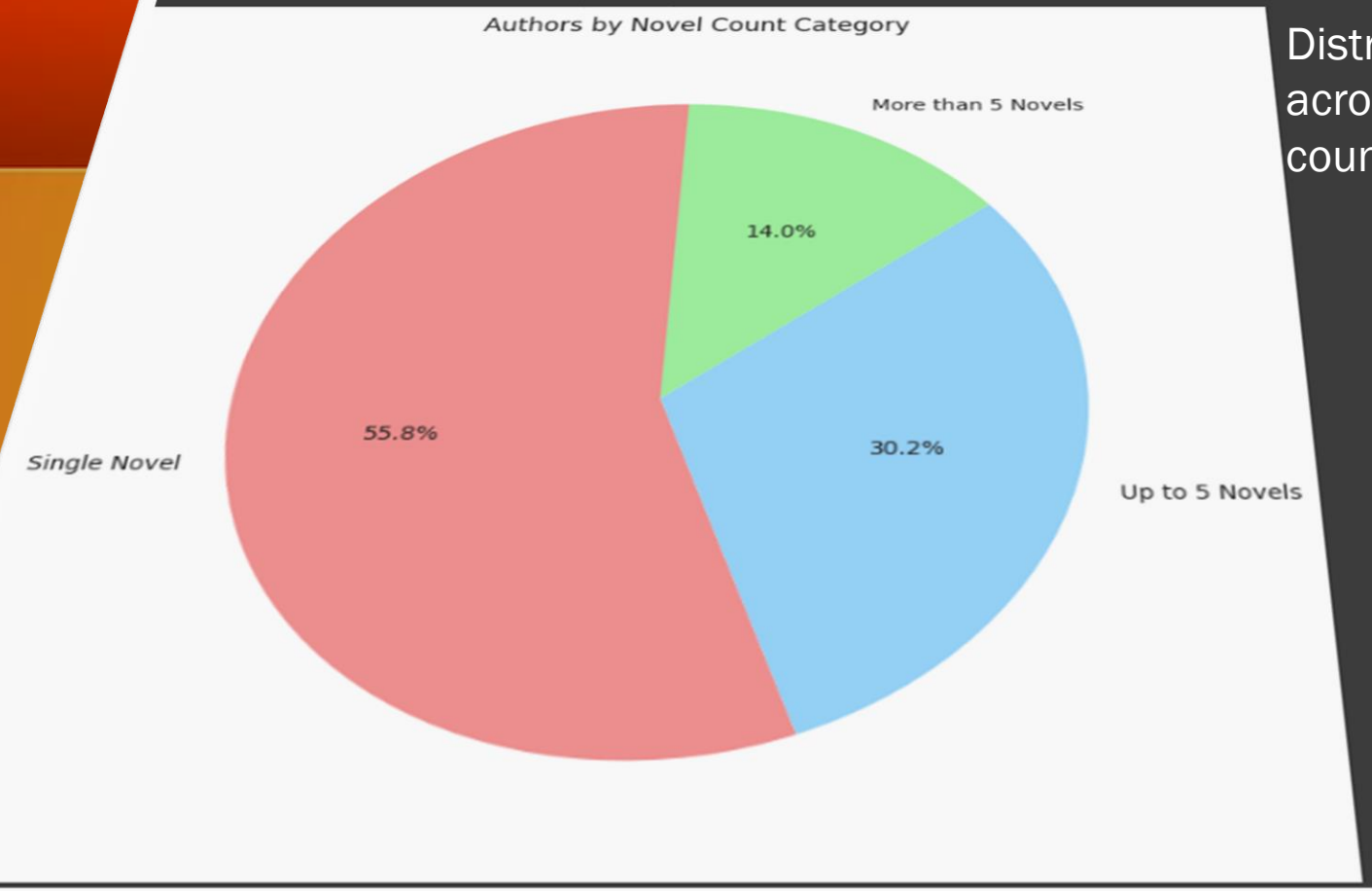
```
author_counts = bk['Author'].value_counts()

# Filter authors with counts greater than or equal to 5
selected_authors = author_counts[author_counts == 2].index

# Filter the DataFrame to include only selected authors
filtered_data = bk[bk['Author'].isin(selected_authors)]

# Plot the histogram
plt.figure(figsize=(12, 6))
sns.histplot(data=filtered_data, x="Author", bins=len(selected_authors), multiple="stack")
plt.title('Histogram of Author Counts (Count == 2)')
plt.xlabel('Author')
plt.ylabel('Count')
plt.xticks(rotation=45, ha='right')
plt.show()
```



Authors by Novel Count

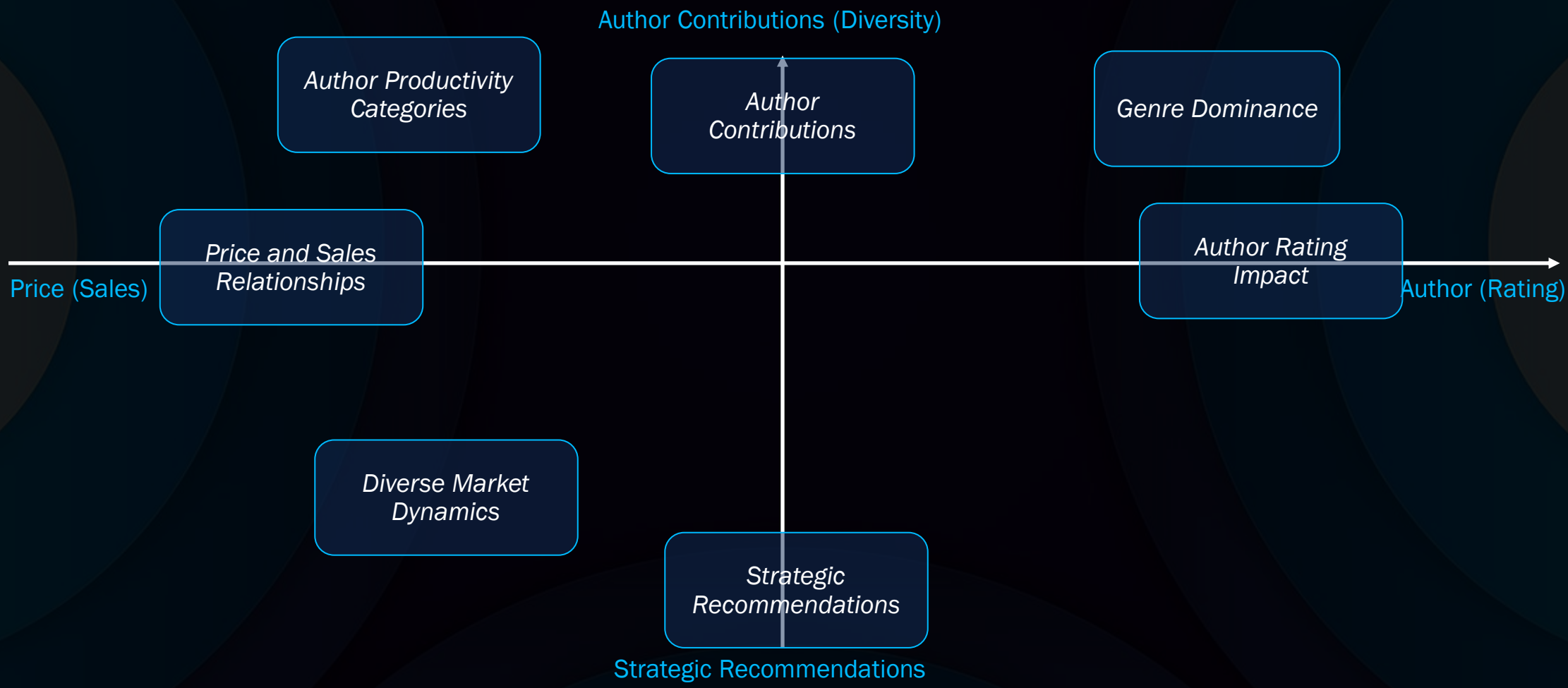
```
# New column 'NovelCount' with the number of novels each author wrote
bk['NovelCount'] = bk['Author'].map(author_counts)

# Categorize authors into 'Single Novel' or 'Up to 5 Novels'
bk['NovelCategory'] = pd.cut(bk['NovelCount'], bins=[-1, 1, 5, float('inf')], labels=['Single Novel', 'Up to 5 Novels', 'More than 5 Novels'])

# Count the number of authors in each category
novel_category_counts = bk['NovelCategory'].value_counts()

# Plot the pie chart
plt.figure(figsize=(8, 8))
plt.pie(novel_category_counts, labels=novel_category_counts.index, autopct='%1.1f%%', startangle=90, colors=['lightcoral', 'lightskyblue', 'lightgreen'])
plt.title('Authors by Novel Count Category')
plt.show()
```

0.0s



Distribution of authors across the three novel count categories.

Conclusion


Author Contributions (Diversity)
Author Productivity Categories
Author Contributions
Genre Dominance
Price and Sales Relationships
Author Rating Impact
Price (Sales)
Author (Rating)
Diverse Market Dynamics
Strategic Recommendations
Strategic Recommendations

Closing Statement

In summary, we looked at a multifaceted analysis of a dataset which pointed to the importance of understanding and leveraging diverse factors, including genre preferences, authors reputation, and pricing dynamics that could potentially inform strategic decisions in the book industry. These conclusions provide a foundation for optimising marketing efforts, pricing strategies and author engagement to enhance the overall success in the dynamic and varied landscape of book sales.

Thank you!

Phillip Seggie

Code Cumbria 2023