

**Overview** This report provides an analysis of book sales data, examining trends in publishing years, genres, authors, and overall sales performance.

**Key Findings:** Top Authors by Gross Sales: Authors with the highest gross sales have consistently published books in high-demand genres. [Insert specific names and their gross sales]

**Most Popular Genres:**[Insert popular genres based on units sold or ratings]. These genres appeal to [specific target audience].

**Publishing Year Trends:**Sales peaked during [insert years or decades] due to [possible reasons, e.g., increased publication or popularity of specific genres]. Declining or rising trends in sales are observed for specific periods.

**Revenue Insights:**Publishers with the highest revenue: [Insert top publishers]. Revenue growth or decline patterns: [Insights into sales trends].

### **Import Libraries**

```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('/content/Books_Data_Clean.csv')

df.head(3)

{"summary":{"\n  \"name\": \"df\",\n  \"rows\": 1070,\n  \"fields\": [\n    {\n      \"column\": \"index\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 309,\n        \"min\": 0,\n        \"max\": 1069,\n        \"num_unique_values\": 1070,\n        \"samples\": [\n          644,\n          629,\n          70\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Publishing Year\",\n      \"properties\": {\n        \"dtype\": \"number\",\n        \"std\": 185.08025721167792,\n        \"min\": -560.0,\n        \"max\": 2016.0,\n        \"num_unique_values\": 150,\n        \"samples\": [\n          1603.0,\n          2000.0,\n          1874.0\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Book Name\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 1045,\n        \"samples\": [\n          \"Drowning Ruth\",\n          \"Howl's Moving Castle\",\n          \"The Birth of Venus\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      },\n      \"column\": \"Author\",\n      \"properties\": {\n        \"dtype\": \"string\",\n        \"num_unique_values\": 735,\n        \"samples\": [\n          \"Christopher Hitchens\",\n          \"Vanessa Diffenbaugh\",\n          \"Astrid Lindgren, Lauren Child, Florence Lamborn, Nancy Seligsohn\"\n        ],\n        \"semantic_type\": \"\",\n        \"description\": \"\"\n      }\n    ]\n  }
}
```

```

},\n      {\n        \"column\": \"language_code\",\n        \"properties\": {\n          \"dtype\": \"category\",\n          \"num_unique_values\": 8,\n          \"samples\": [\n            \"eng\",\n            \"en-CA\",\n            \"en-US\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n        },\n        {\n          \"column\": \"Author_Rating\",\n          \"properties\": {\n            \"dtype\": \"category\",\n            \"num_unique_values\": 4,\n            \"samples\": [\n              \"Intermediate\",\n              \"Excellent\",\n              \"Novice\"\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\"\n          },\n          {\n            \"column\": \"Book_average_rating\",\n            \"properties\": {\n              \"dtype\": \"number\",\n              \"std\": 0.24724387206390464,\n              \"min\": 2.97,\n              \"max\": 4.77,\n              \"num_unique_values\": 134,\n              \"samples\": [\n                3.53,\n                3.73,\n                4.59\n              ],\n              \"semantic_type\": \"\",\n              \"description\": \"\"\n            },\n            {\n              \"column\": \"Book_ratings_count\",\n              \"properties\": {\n                \"dtype\": \"number\",\n                \"std\": 31513,\n                \"min\": 27308,\n                \"max\": 206792,\n                \"num_unique_values\": 1064,\n                \"samples\": [\n                  144822,\n                  75987,\n                  119735\n                ],\n                \"semantic_type\": \"\",\n                \"description\": \"\"\n              },\n              {\n                \"column\": \"genre\",\n                \"properties\": {\n                  \"dtype\": \"category\",\n                  \"num_unique_values\": 4,\n                  \"samples\": [\n                    \"fiction\",\n                    \"children\",\n                    \"genre fiction\"\n                  ],\n                  \"semantic_type\": \"\",\n                  \"description\": \"\"\n                },\n                {\n                  \"column\": \"gross sales\",\n                  \"properties\": {\n                    \"dtype\": \"number\",\n                    \"std\": 3936.9242404502725,\n                    \"min\": 104.94,\n                    \"max\": 47795.0,\n                    \"num_unique_values\": 831,\n                    \"samples\": [\n                      486.78,\n                      477.0,\n                      891.02\n                    ],\n                    \"semantic_type\": \"\",\n                    \"description\": \"\"\n                  },\n                  {\n                    \"column\": \"publisher revenue\",\n                    \"properties\": {\n                      \"dtype\": \"number\",\n                      \"std\": 2257.596743156442,\n                      \"min\": 0.0,\n                      \"max\": 28677.0,\n                      \"num_unique_values\": 610,\n                      \"samples\": [\n                        943.26,\n                        141.966,\n                        4131.534\n                      ],\n                      \"semantic_type\": \"\",\n                      \"description\": \"\"\n                    },\n                    {\n                      \"column\": \"sale price\",\n                      \"properties\": {\n                        \"dtype\": \"number\",\n                        \"std\": 3.559919262802466,\n                        \"min\": 0.99,\n                        \"max\": 33.86,\n                        \"num_unique_values\": 149,\n                        \"samples\": [\n                          11.39,\n                          8.98\n                        ],\n                        \"semantic_type\": \"\",\n                        \"description\": \"\"\n                      },\n                      {\n                        \"column\": \"sales rank\",\n                        \"properties\": {\n                          \"dtype\": \"number\",\n                          \"std\": 369,\n                          \"min\": 1,\n                          \"max\": 1273,\n                          \"num_unique_values\": 865,\n                          \"samples\": [\n                            266,\n                            1046,\n                            915\n                          ],\n                          \"semantic type\": \"\",\n                          \"description\": \"\"

```

```

}\n    },\n    {\n        \"column\": \"Publisher \",\n        \"properties\": {\n            \"dtype\": \"category\",\n            \"num_unique_values\": 9,\n            \"samples\": [\n                \"HarperCollins Publishing\",\n                \"Amazon Digital Services, Inc.\",\n                \"Simon and Schuster Digital Sales Inc\"\n            ],\n            \"semantic_type\": \"\",\n            \"description\": \"\"\n        },\n        {\n            \"column\": \"units sold\",\n            \"properties\": {\n                \"dtype\": \"number\",\n                \"std\": 15370,\n                \"min\": 106,\n                \"max\": 61560,\n                \"num_unique_values\": 491,\n                \"samples\": [\n                    33696,\n                    975,\n                    37584\n                ],\n                \"semantic_type\": \"\",\n                \"description\": \"\"\n            }\n        }\n    ],\n    \"type\": \"dataframe\", \"variable_name\": \"df\"}

```

```
df.tail(3)
```

```

{
    \"summary\": {
        \"name\": \"df\",
        \"rows\": 3,
        \"fields\": [
            {
                \"column\": \"index\",
                \"properties\": {
                    \"dtype\": \"number\",
                    \"std\": 1,
                    \"min\": 1067,
                    \"max\": 1069,
                    \"num_unique_values\": 3,
                    \"samples\": [
                        1067,
                        1068,
                        1069
                    ],
                    \"semantic_type\": \"\",
                    \"description\": \"\"
                }
            },
            {
                \"column\": \"Publishing Year\",
                \"properties\": {
                    \"dtype\": \"number\",
                    \"std\": 43.71498598878879,
                    \"min\": 1930.0,
                    \"max\": 2011.0,
                    \"num_unique_values\": 3,
                    \"samples\": [
                        1930.0,
                        2011.0,
                        1999.0
                    ],
                    \"semantic_type\": \"\",
                    \"description\": \"\"
                }
            },
            {
                \"column\": \"Book Name\",
                \"properties\": {
                    \"dtype\": \"string\",
                    \"num_unique_values\": 3,
                    \"samples\": [
                        \"The Maltese Falcon\",
                        \"Night Road\",
                        \"Tripwire\"
                    ],
                    \"semantic_type\": \"\",
                    \"description\": \"\"
                }
            },
            {
                \"column\": \"Author\",
                \"properties\": {
                    \"dtype\": \"string\",
                    \"num_unique_values\": 3,
                    \"samples\": [
                        \"Dashiell Hammett\",
                        \"Kristin Hannah\",
                        \"Lee Child\"
                    ],
                    \"semantic_type\": \"\",
                    \"description\": \"\"
                }
            },
            {
                \"column\": \"language_code\",
                \"properties\": {
                    \"dtype\": \"string\",
                    \"num_unique_values\": 2,
                    \"samples\": [
                        \"en-US\",
                        \"eng\"
                    ],
                    \"semantic_type\": \"\",
                    \"description\": \"\"
                }
            },
            {
                \"column\": \"Author_Rating\",
                \"properties\": {
                    \"dtype\": \"string\",
                    \"num_unique_values\": 2,
                    \"samples\": [
                        \"Excellent\",
                        \"Intermediate\"
                    ],
                    \"semantic_type\": \"\",
                    \"description\": \"\"
                }
            },
            {
                \"column\": \"Book_average_rating\",
                \"properties\": {
                    \"dtype\": \"number\",
                    \"std\": 0.1258305739211792,
                    \"min\": 3.92,
                    \"max\": 4.17,

```

```

\"num_unique_values\": 3,\n          \"samples\": [\n          3.92,\n4.17\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\":\n          \"Book_ratings_count\",\n          \"properties\": {\n          \"dtype\":\n          \"number\",\n          \"std\": 1844,\n          \"min\": 55251,\n          \"max\": 58742,\n          \"num_unique_values\": 3,\n          \"samples\": [\n          58742,\n          58028\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"genre\",\n          \"properties\": {\n          \"dtype\": \"category\",\n          \"num_unique_values\": 1,\n          \"samples\": [\n          \"genre fiction\"\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"gross sales\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\":\n          382.1884351991829,\n          \"min\": 104.94,\n          \"max\":\n          846.94,\n          \"num_unique_values\": 3,\n          \"samples\": [\n          846.94\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"publisher revenue\",\n          \"properties\": {\n          \"dtype\":\n          \"number\",\n          \"std\": 277.0072671104496,\n          \"min\":\n          0.0,\n          \"max\": 508.164,\n          \"num_unique_values\": 3,\n          \"samples\": [\n          508.164\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"sale price\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\":\n          3.605551275463989,\n          \"min\": 0.99,\n          \"max\": 7.99,\n          \"num_unique_values\": 3,\n          \"samples\": [\n          7.99\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"sales rank\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\":\n          1,\n          \"min\": 1271,\n          \"max\": 1273,\n          \"num_unique_values\": 3,\n          \"samples\": [\n          1271\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\",\n          \"column\": \"Publisher \",\n          \"properties\": {\n          \"dtype\": \"string\",\n          \"num_unique_values\": 2,\n          \"samples\": [\n          \"Amazon\n          Digital Services, Inc.\"\n          ],\n          \"semantic_type\":\n          \"\",\n          \"description\": \"\",\n          \"column\": \"units sold\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 0,\n          \"min\": 106,\n          \"max\": 106,\n          \"num_unique_values\": 1,\n          \"samples\":\n          [\n          106\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          }\n          ]\n          }\", \"type\": \"dataframe\"}

```

```
df.describe()
```

```

{\"summary\": {\"name\": \"df\",\n          \"rows\": 8,\n          \"fields\": [\n          \"column\": \"index\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 385.41434397584874,\n          \"min\": 0.0,\n          \"max\": 1070.0,\n

```

```

\"num_unique_values\": 7,\n          \"samples\": [\n          1070.0,\n          534.5,\n          801.75\n          ],\n          \"semantic_type\":\n          \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"Publishing Year\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 1012.5768570062627,\n          \"min\": -560.0,\n          \"max\": 2016.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          1971.377923292797,\n          2003.0,\n          1069.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"Book_average_rating\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 377.09103188518253,\n          \"min\": 0.24724387206390464,\n          \"max\": 1070.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          4.007,\n          4.015,\n          1070.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"Book_ratings_count\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 64344.50669831697,\n          \"min\": 1070.0,\n          \"max\": 206792.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          94909.91308411215,\n          89309.0,\n          1070.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"gross sales\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 16454.025590792146,\n          \"min\": 104.94,\n          \"max\": 47795.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          1856.6229439252338,\n          809.745,\n          1070.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"publisher revenue\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 9904.951823761143,\n          \"min\": 0.0,\n          \"max\": 28677.0,\n          \"num_unique_values\": 7,\n          \"samples\": [\n          843.2810299065421,\n          721.1804999999999\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"sale price\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 375.6133782122349,\n          \"min\": 0.99,\n          \"max\": 1070.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          4.869560747663551,\n          3.99,\n          1070.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"sales rank\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 427.5231519171498,\n          \"min\": 1.0,\n          \"max\": 1273.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n          611.6523364485981,\n          595.5,\n          1070.0\n          ],\n          \"semantic_type\": \"\",\n          \"description\": \"\"\n          }\n          },\n          {\n          \"column\": \"units sold\",\n          \"properties\": {\n          \"dtype\": \"number\",\n          \"std\": 20614.128422425747,\n          \"min\": 106.0,\n          \"max\": 61560.0,\n          \"num_unique_values\": 8,\n          \"samples\": [\n

```

```
9676.980373831775,\n          3924.0,\n          1070.0\n          ],\n\n\"semantic_type\": \"\", \n          \"description\": \"\" \n          }\n    }\n  ]\n}","type":"dataframe"}
```

```
df = df[df["Publishing Year"] > 1900]
```

```
df.isna().sum()
```

```
index          0
Publishing Year 0
Book Name      21
Author         0
language_code  49
Author_Rating  0
Book_average_rating 0
Book_ratings_count 0
genre          0
gross sales    0
publisher revenue 0
sale price     0
sales rank     0
Publisher      0
units sold     0
dtype: int64
```

```
df.dropna(subset="Book Name", inplace=True)
```

```
<ipython-input-12-866f374302c1>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

See the caveats in the documentation:  
[https://pandas.pydata.org/pandas-docs/stable/user\\_guide/indexing.html#returning-a-view-versus-a-copy](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)  
 df.dropna(subset="Book Name", inplace=True)

```
df.duplicated().sum()
```

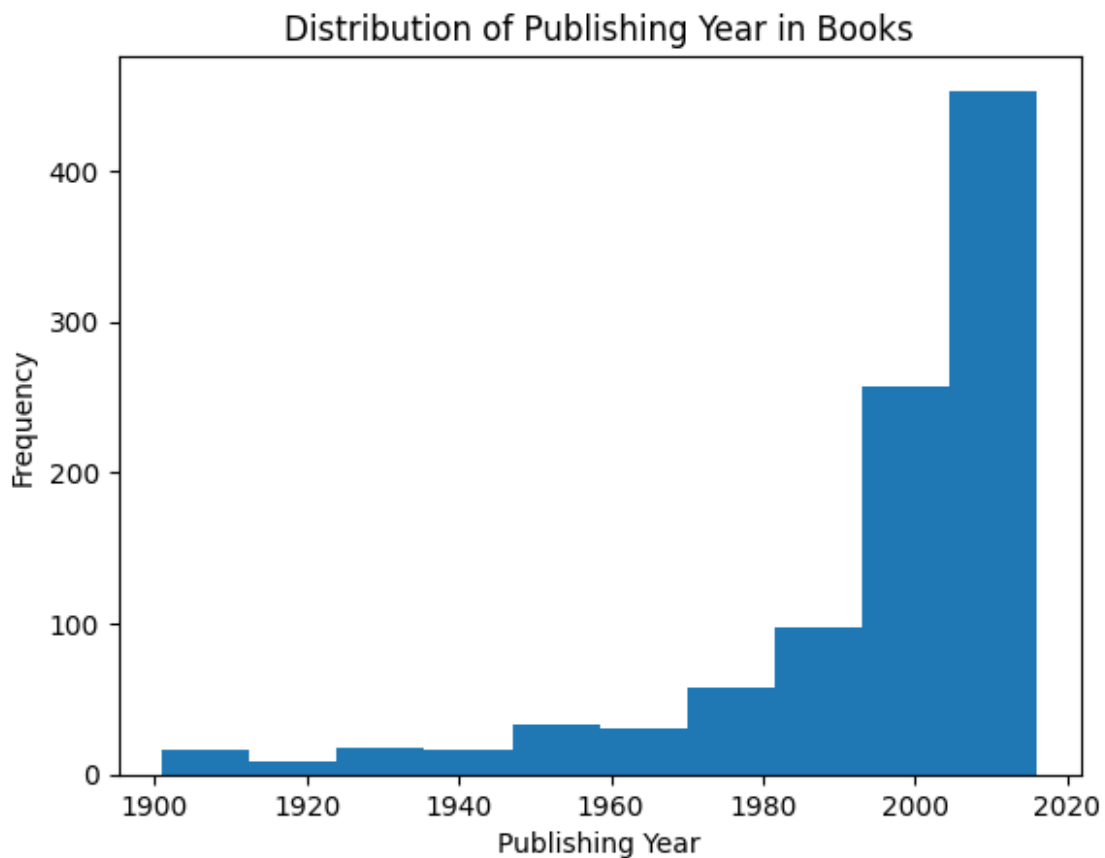
```
0
```

```
df.nunique()
```

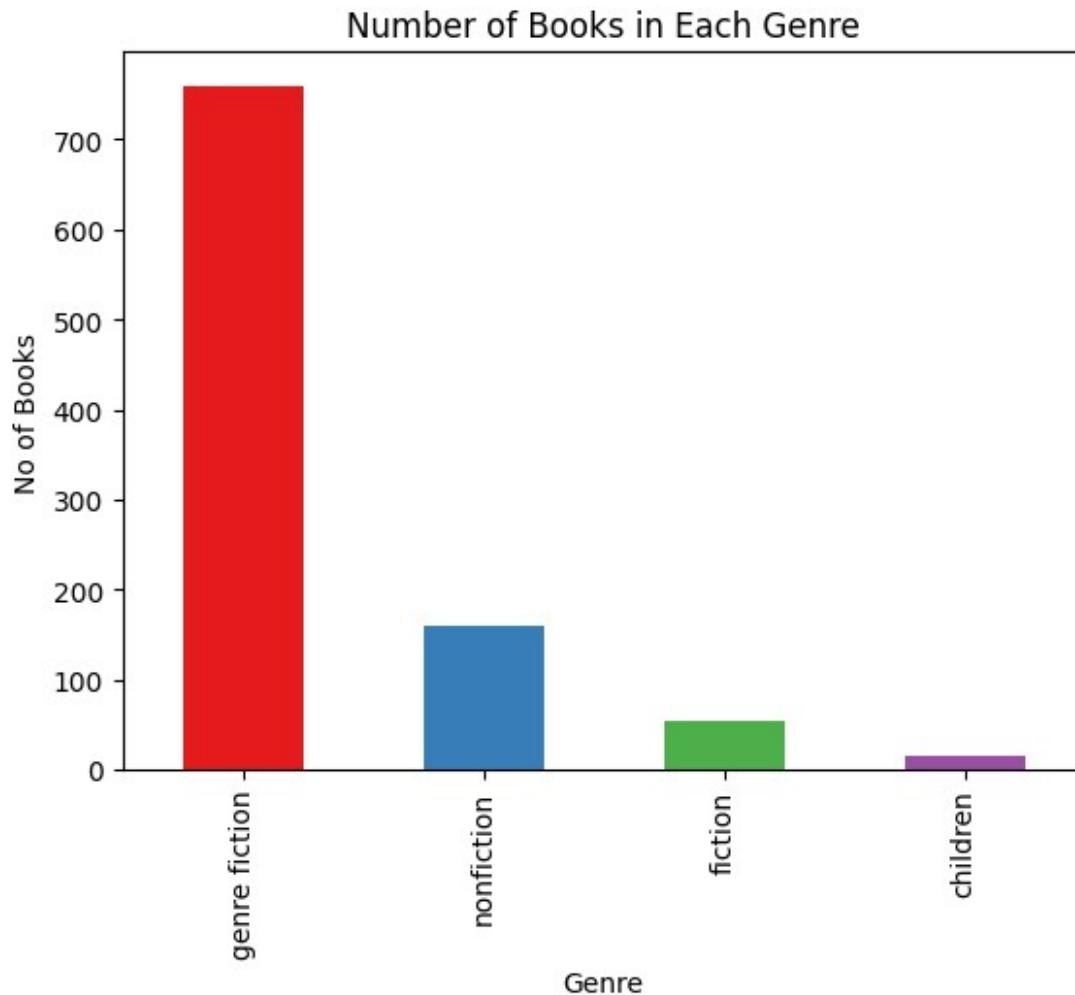
```
index          988
Publishing Year 101
Book Name      987
Author         669
language_code   8
Author_Rating   4
Book_average_rating 133
Book_ratings_count 983
genre          4
gross sales    774
```

```
publisher revenue    570
sale price          143
sales rank          818
Publisher            9
units sold          470
dtype: int64
```

```
plt.hist(df["Publishing Year"])
plt.xlabel("Publishing Year")
plt.ylabel("Frequency")
plt.title("Distribution of Publishing Year in Books")
plt.show()
```



```
df["genre"].value_counts().plot(kind = "bar",
color=plt.cm.Set1(range(10)))
plt.xlabel("Genre")
plt.ylabel("No of Books")
plt.title("Number of Books in Each Genre")
plt.show()
```



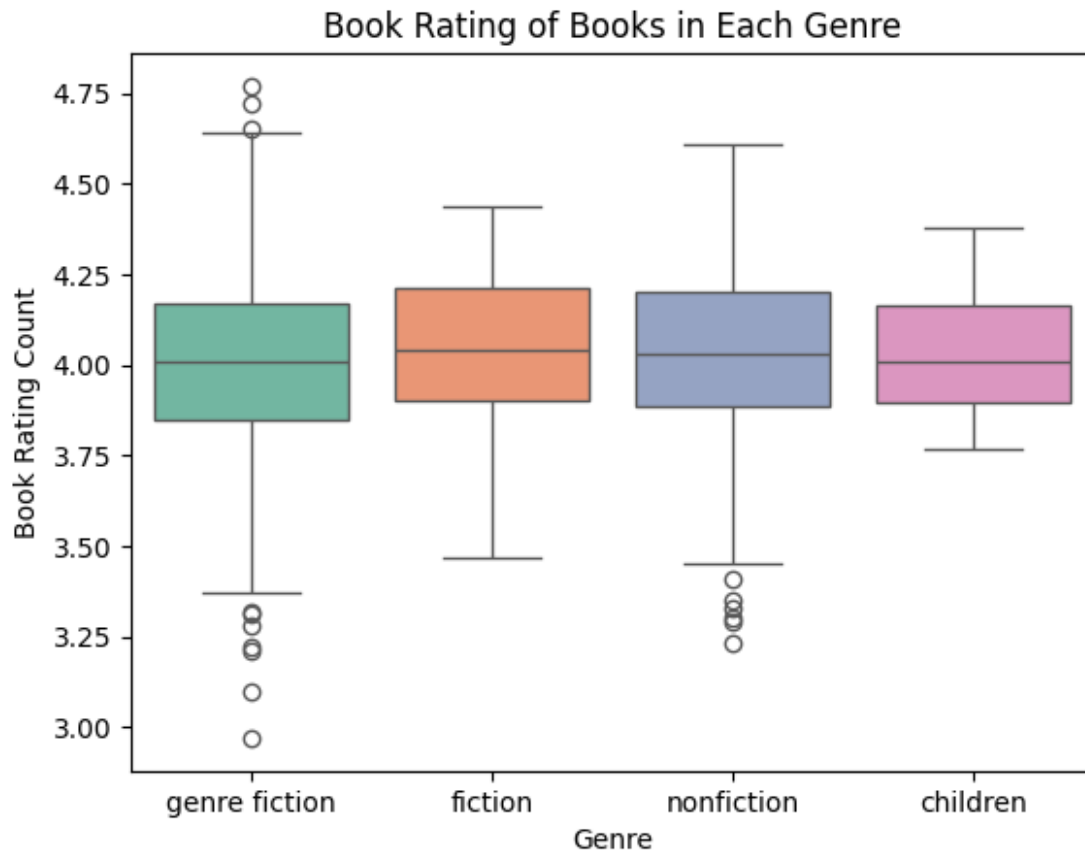
```
df.groupby("Author")  
["Book_average_rating"].mean().sort_values(ascending=False)
```

```
Author  
Bill Watterson                4.650000  
Bill Watterson, G.B. Trudeau  4.610000  
J.R.R. Tolkien                4.590000  
George R.R. Martin           4.560000  
Sarah J. Maas                 4.526000  
...  
Chetan Bhagat                 3.273333  
Audrey Niffenegger            3.230000  
Herman Koch, Sam Garrett      3.220000  
P.D. James                    3.210000  
Sue Monk Kidd                 3.100000  
Name: Book_average_rating, Length: 669, dtype: float64
```

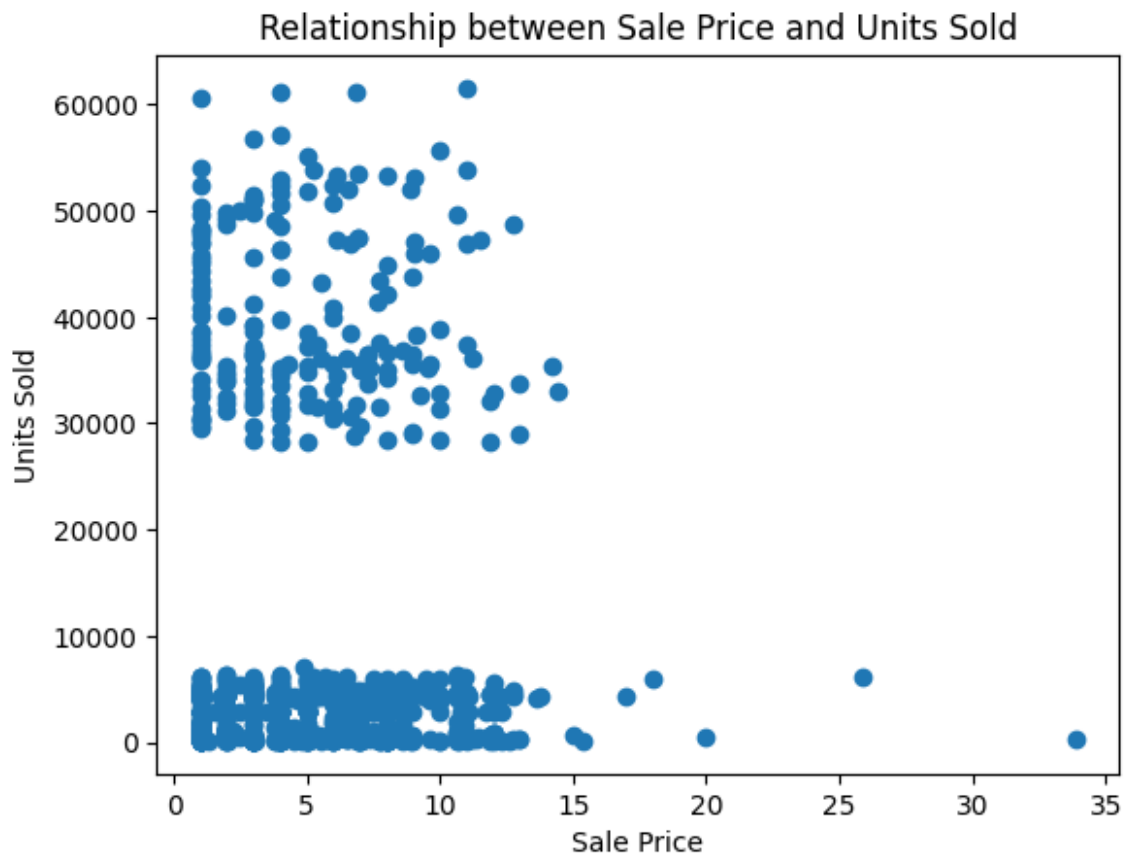
```
sns.boxplot(x="genre", y="Book_average_rating", data =  
df, hue="genre", palette="Set2")
```



```
plt.xlabel("Genre")
plt.ylabel("Book Rating Count")
plt.title("Book Rating of Books in Each Genre")
plt.show()
```

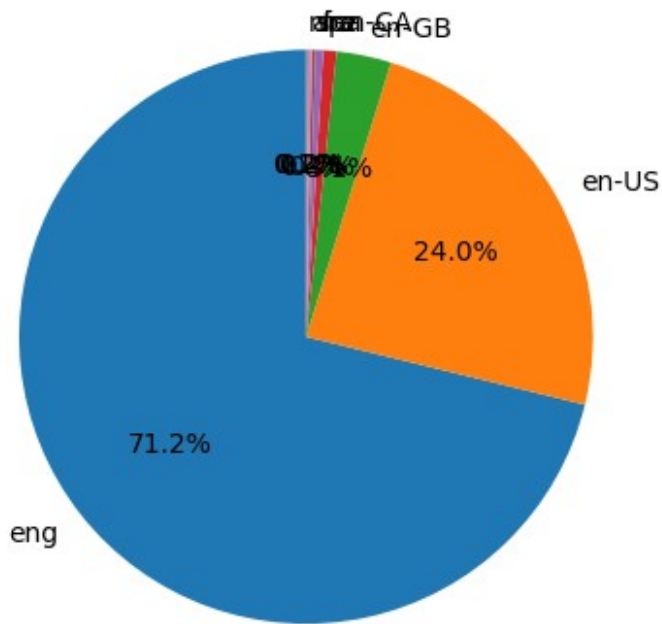


```
plt.scatter(df["sale price"],df["units sold"])
plt.xlabel("Sale Price")
plt.ylabel("Units Sold")
plt.title("Relationship between Sale Price and Units Sold")
plt.show()
```



```
language_count = df["language_code"].value_counts()
plt.pie(language_count, labels=language_count.index, startangle=90,
autopct='%1.1f%%')
plt.title("Distribution of Books by Language")
plt.show()
```

Distribution of Books by Language



```
df.columns
```

```
Index(['index', 'Publishing Year', 'Book Name', 'Author',
      'language_code',
      'Author_Rating', 'Book_average_rating', 'Book_ratings_count',
      'genre',
      'gross sales', 'publisher revenue', 'sale price', 'sales rank',
      'Publisher ', 'units sold'],
      dtype='object')
```

```
df.columns = df.columns.str.strip() # This removes leading and
trailing spaces
```

```
df.groupby("Publisher")["publisher
revenue"].sum().sort_values(ascending=False)
```

Publisher	
Penguin Group (USA) LLC	213817.482
Random House LLC	189585.222
Amazon Digital Services, Inc.	148244.178
Hachette Book Group	137874.498
HarperCollins Publishers	128478.258
Simon and Schuster Digital Sales Inc	46988.352
Macmillan	32356.236
HarperCollins Publishing	2830.806

```

HarperCollins Christian Publishing      2135.670
Name: publisher revenue, dtype: float64

df.groupby("Author_Rating")
["Book_ratings_count"].mean().sort_values(ascending=False).max()

101946.9952

df.groupby("language_code").size().sort_values(ascending=False)

language_code
eng      737
en-US    234
en-GB     30
en-CA      7
fre        4
ara         2
spa         2
nl          1
dtype: int64

df.groupby("Author_Rating")["Book_ratings_count"].var()

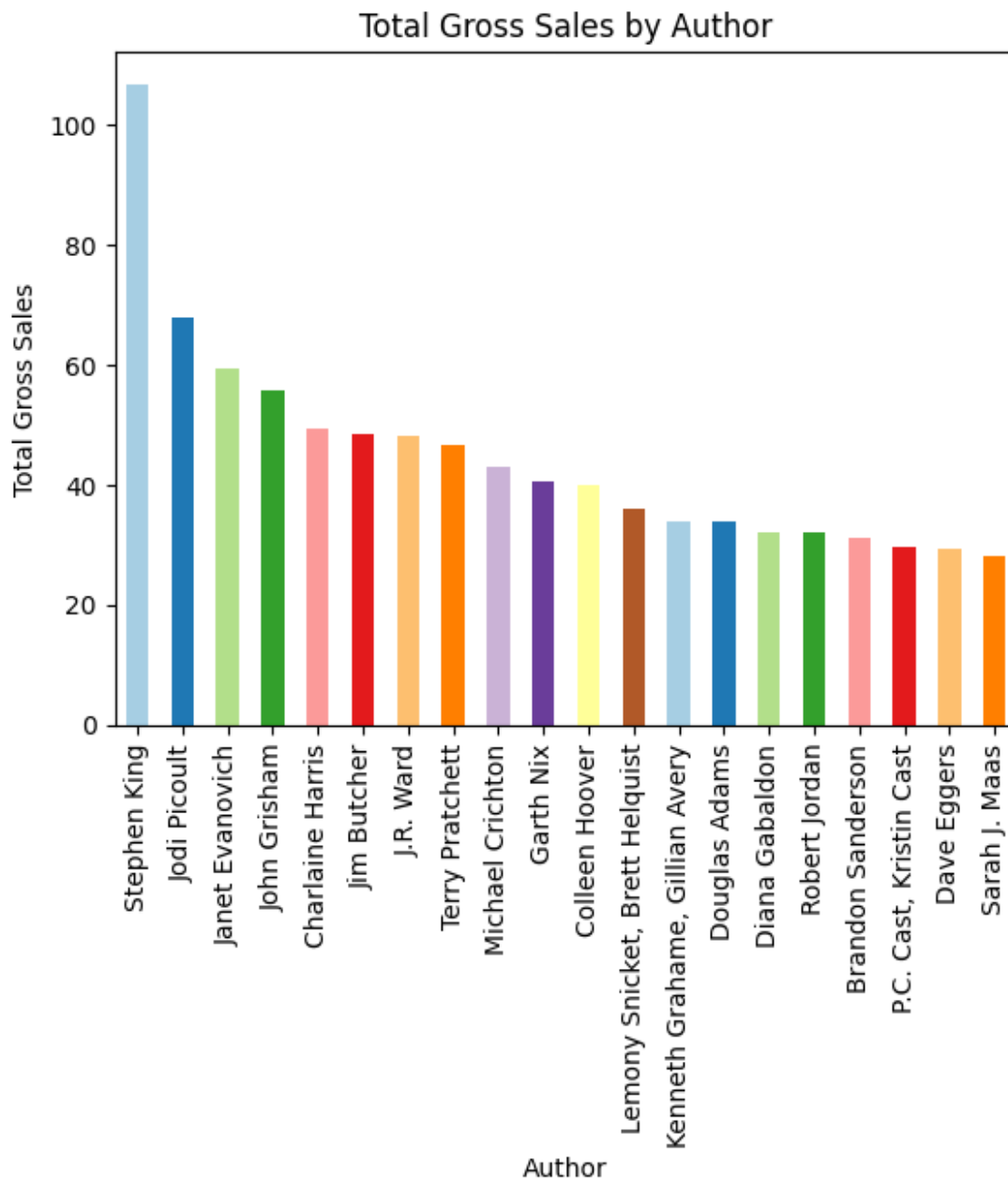
Author_Rating
Excellent      4.380998e+08
Famous         1.172260e+09
Intermediate   1.166913e+09
Novice         1.025853e+09
Name: Book_ratings_count, dtype: float64

plt.scatter(df["Book_average_rating"],df["Book_ratings_count"])
plt.xlabel("Book Average Rating")
plt.ylabel("Book Rating Count")
plt.title("Relationship between Book Average Rating and Book Rating
Count")
plt.show()

```



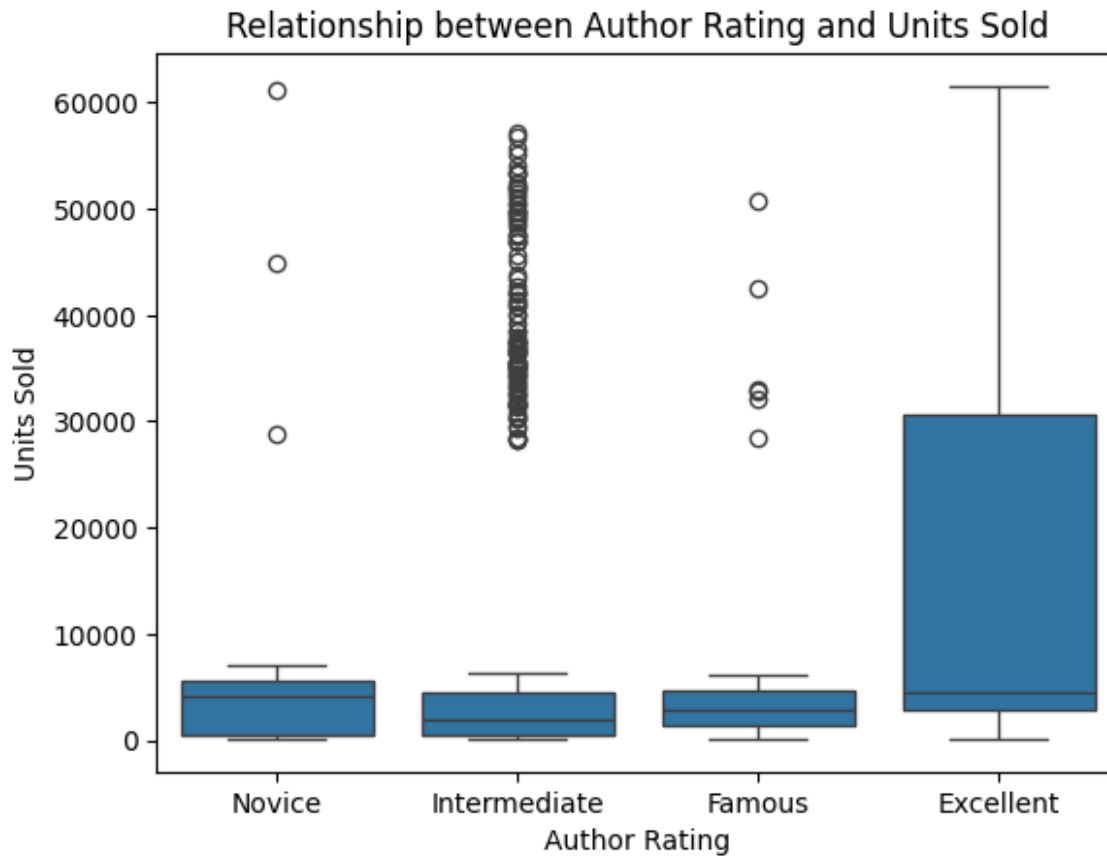
```
total_gross_sales_by_author = df.groupby("Author")["sale price"].sum()
total_gross_sales_by_author.sort_values(ascending=False).head(20).plot(
    kind="bar", color=plt.cm.Paired.colors)
plt.xlabel("Author")
plt.ylabel("Total Gross Sales")
plt.title("Total Gross Sales by Author")
plt.show()
```



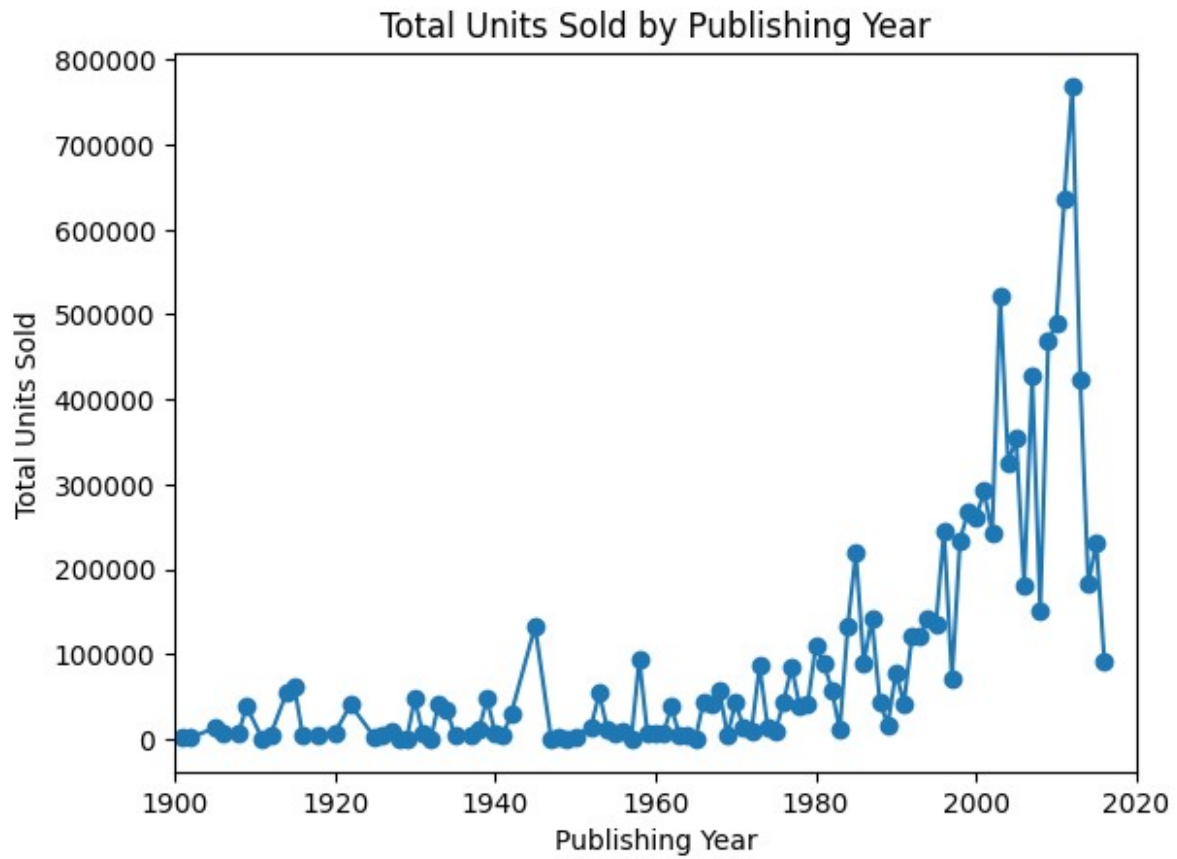
```

sns.boxplot(x="Author_Rating",y="units sold",data=df)
plt.xlabel("Author Rating")
plt.ylabel("Units Sold")
plt.title("Relationship between Author Rating and Units Sold")
plt.show()

```



```
df = df[df["Publishing Year"] <= 2020]
# Group and plot
df.groupby("Publishing Year")["units sold"].sum().plot(kind="line",
marker="o")
# Set x-axis range and custom ticks
plt.xlim(1900, 2020) # Define the range for the x-axis
plt.xticks([1900, 1920, 1940, 1960, 1980, 2000, 2020]) # Set specific
ticks
# Add labels and title
plt.xlabel("Publishing Year")
plt.ylabel("Total Units Sold")
plt.title("Total Units Sold by Publishing Year")
# Display the plot
plt.show()
```



***Conclusion*** The analysis indicates that Publishers and authors focusing on high-demand genres and leveraging effective marketing strategies have performed well.