```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from google.colab import files
```

```
data = pd.read_csv('Books_data - Sheet1.csv')
# Check the first few rows of the dataset
print(data.head())
```

```
   bid                               title            author       category  \
0    1                         Steve Jobs   Walter Issacson      Biography
1    2                  Discovery of India  Jawaharlal Nehru        History
2    3          My Experiments with Truth    Mahatma Gandhi  Autobiography
3    4  Object Oriented Programming with C++    E Balagurusamy      Education
4    5                  Thinking with type      Ellen Lupton           Arts

   status
0  issued
1  issued
2  issued
3  issued
4  issued
```

```
# Get a summary of the dataset
print(data.info())
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 61 entries, 0 to 60
Data columns (total 5 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   bid       61 non-null     int64
 1   title     61 non-null     object
 2   author    61 non-null     object
 3   category  61 non-null     object
 4   status    61 non-null     object
dtypes: int64(1), object(4)
memory usage: 2.5+ KB
None
```

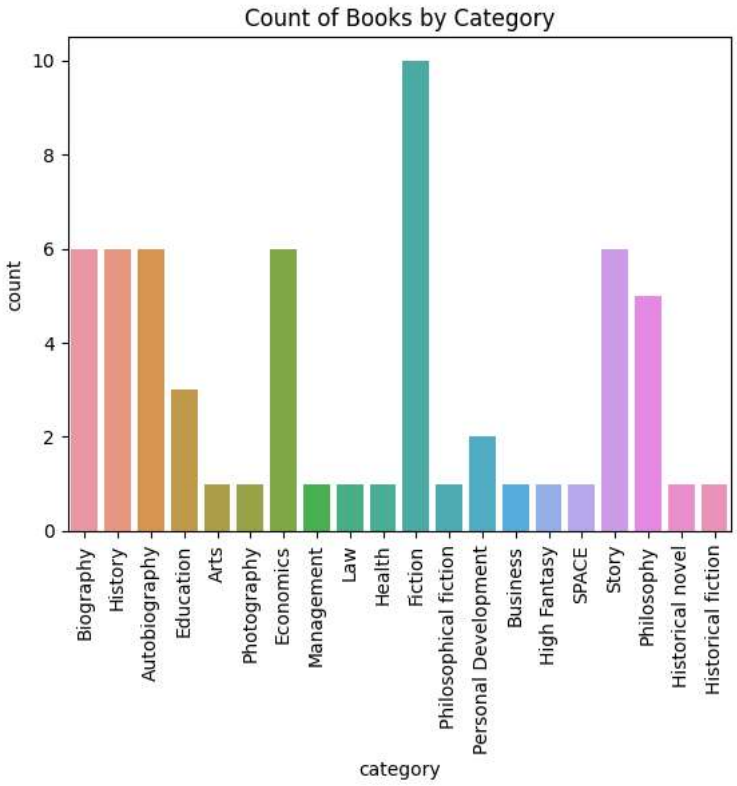
```
print(data.tail())
```

```
    bid                title              author            category  \
56   57    Alexander Hamilton         Ron Chernow           Biography
57   58            Barracoon  Zora Neale Hurston           Biography
58   59    CHURCHILL : A Life      MArtin Glibert           Biography
59   60  A Tale of Two Cities     Charles Dickens    Historical novel
60   61        Les Misérables         Victor Hugo  Historical fiction

      status
56  available
57  available
58  available
59     issued
60  available
```

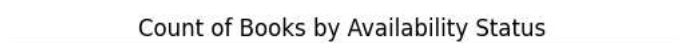
```
# Compute summary statistics
print(data.describe())
```

```
              bid
count  61.000000
mean   31.000000
std    17.752934
min     1.000000
25%    16.000000
50%    31.000000
75%    46.000000
max    61.000000
```

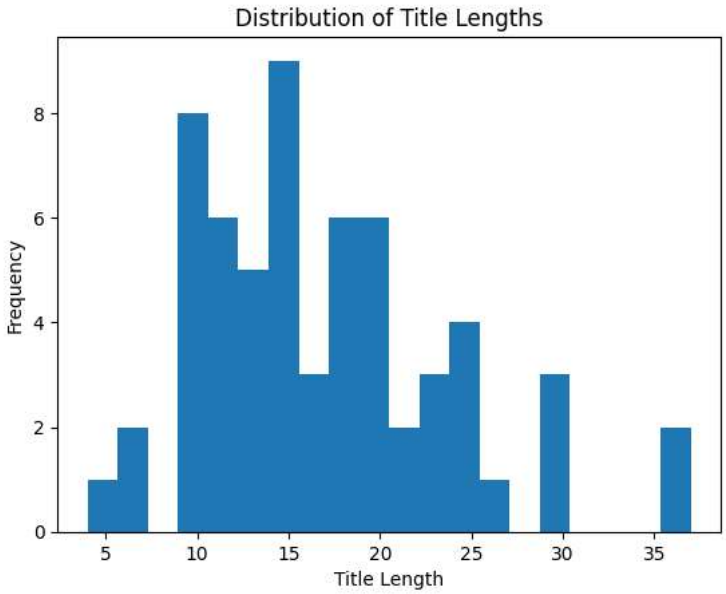
```
# Handle missing values
data.dropna(inplace=True)

# Convert 'status' column to categorical
data['status'] = data['status'].astype('category')
# Count plot for book categories
sns.countplot(x='category', data=data)
plt.title('Count of Books by Category')
plt.xticks(rotation=90)
plt.show()
```




```
# Bar plot for book availability status
sns.countplot(x='status', data=data)
plt.title('Count of Books by Availability Status')
plt.show()
```




```
# Histogram of book titles lengths
data['title_length'] = data['title'].apply(len)
plt.hist(data['title_length'], bins=20)
plt.title('Distribution of Title Lengths')
plt.xlabel('Title Length')
plt.ylabel('Frequency')
plt.show()
```

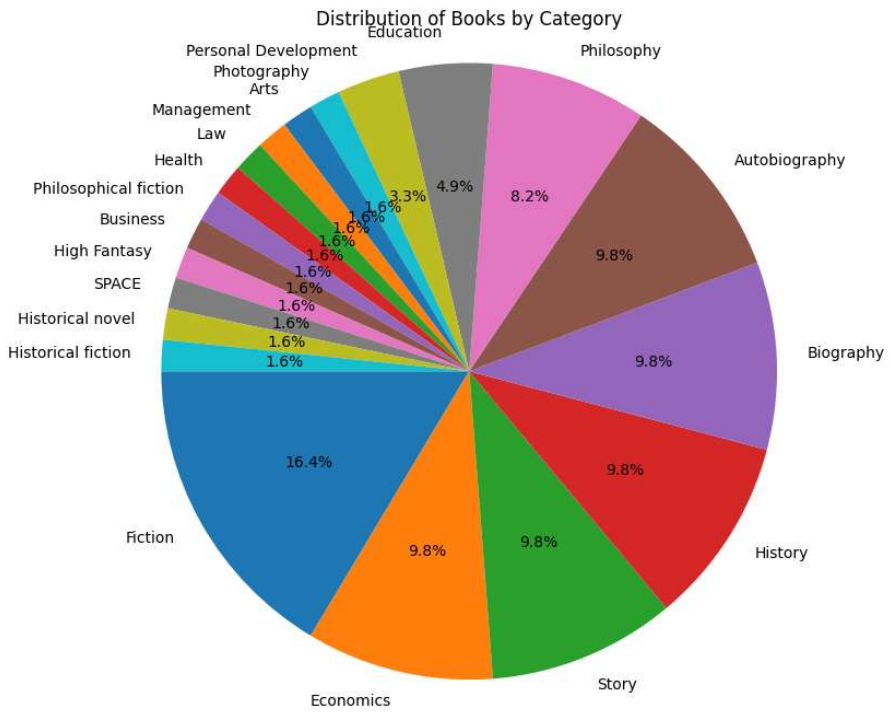



```
# Average title length by category
avg_title_length = data.groupby('category')['title_length'].mean()
print(avg_title_length)

# Proportion of books by availability status
status_proportion = data['status'].value_counts(normalize=True)
print(status_proportion)
```

```
category
Arts                    18.000000
Autobiography           16.833333
Biography               13.666667
Business                17.000000
Economics               16.833333
Education               29.000000
Fiction                 14.100000
Health                  11.000000
High Fantasy            19.000000
Historical fiction      14.000000
Historical novel        20.000000
History                 20.666667
Law                     19.000000
Management              30.000000
Personal Development    18.000000
Philosophical fiction   13.000000
Philosophy              13.200000
Photography             18.000000
SPACE                   23.000000
Story                   14.833333
Name: title_length, dtype: float64
available    0.639344
issued       0.360656
Name: status, dtype: float64
```

```
category_counts = data['category'].value_counts()

# Create the pie chart
plt.figure(figsize=(8, 8))
plt.pie(category_counts, labels=category_counts.index, autopct='%1.1f%%', startangle=180)
plt.title('Distribution of Books by Category')
```

```
plt.axis('equal') # Equal aspect ratio ensures the pie chart is circular
plt.show()
```




```
# Count Plot - Book Availability by Category
plt.figure(figsize=(10, 6))
sns.countplot(x='category', hue='status', data=data, palette='Set1')
plt.xticks(rotation=90)
plt.title('Availability of Books in Each Category')
plt.xlabel('Category')
plt.ylabel('Count')
plt.show()
```