DS Tools Final Project

Team name: علي الله ان شاء الله

Members:

سيف الدين سامي حامد   23012082

محمد احمد رمضان   23011452

● Phase 1 Report:

1. Website to Scrape

For our project, we've chosen to scrape data from the [All products | Books to Scrape - Sandbox](#) website. This site provides a variety of books, including their ratings, prices, and categories, making it ideal for our analysis.

2. What Data Will Be Collected

From the Books to Scrape site, we will collect the following details for each book:

Book Title
Price
Star Rating
Availability (In stock or not)
Category/Genre
Product Description
UPC (Product Code)
Number of Reviews
Book URL (optional for reference)

This dataset will allow us to explore various aspects of the books' market trends, ratings, and pricing.

3. Project Objective

The goal of this project is to analyze the books on the website to discover:
The relationship between price and star rating for books.
The distribution of books by category/genre.
Which categories have the highest-rated or most-reviewed books.
Identify the books with the highest reviews or prices (top N books).
These insights will be valuable for book lovers looking to discover trends in the market.

4. Data Cleaning & Processing

Once the raw data is extracted, we will clean and organize it using Pandas.
The data will be converted into a structured DataFrame, and saved as a JSON file for easy future use.
We will also apply Regular Expressions (Regex) to handle text cleaning tasks like:

Removing unwanted characters or formatting.
Extracting and cleaning text fields (like descriptions).
Standardizing date formats (if applicable).
The result will be a clean and ready-to-analyze dataset.

5. Data Analysis & Visualization

Using Seaborn and Matplotlib, we will perform both statistical and visual analysis to uncover meaningful trends. Our focus will be on answering questions like:

What is the relationship between price and star rating for the books?
What is the distribution of books by category/genre?
Which categories have the highest-rated or most-reviewed books?
Which books have the highest reviews or prices?

Planned Visuals:

Bar Charts for top genres, book categories, and rating distributions.

Scatter Plots to show relationships between price and star ratings.

Pie Charts to show the distribution of books across categories.

Word Clouds to highlight frequent keywords from book descriptions or categories.

Box Plots to examine price distributions across categories.

6. Data Storage

After processing, we'll save the final dataset as a JSON file.
This format is easy to work with, portable, and can be reused in
future projects, web apps, or databases.

7. (Optional) Web App Deployment with Streamlit

If time allows, we'd like to build a simple Streamlit app
to showcase our findings interactively.
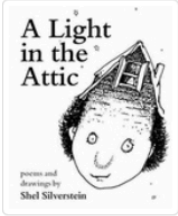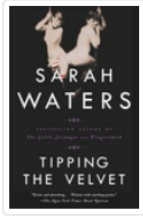Users could:
Browse visual summaries and trends.
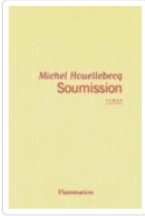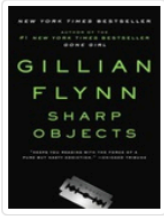Filter results by genre, price range, or rating.
Explore the relationships between different book
attributes in a dynamic way.
This would be a bonus step to improve the presentation and
make the project more engaging.

● Data Extraction:

Using [All products | Books to Scrape - Sandbox](#) website, we scraped the data of the books that looks like this on the website:

A Light in the ...	Tipping the Velvet	Soumission	Sharp Objects
★★★☆☆	★☆☆☆☆	★☆☆☆☆	★★★★☆
£51.77	£53.74	£50.10	£47.82
✔ In stock	✔ In stock	✔ In stock	✔ In stock
Add to basket	Add to basket	Add to basket	Add to basket

Home / Books / Poetry / A Light in the Attic
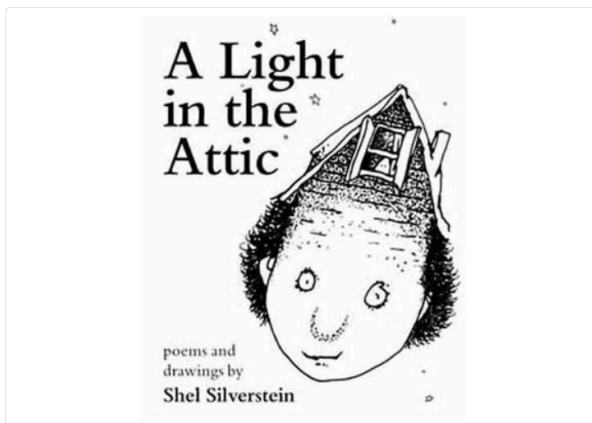
A Light in the Attic

£51.77

✔ In stock (22 available)

★★★☆☆

> **Warning!** This is a demo website for web scraping purposes. Prices and ratings here were randomly assigned and have no real meaning.

Product Description

It's hard to imagine a world without A Light in the Attic. This now-classic collection of poetry and drawings from Shel Silverstein celebrates its 20th anniversary with this special edition. Silverstein's humorous and creative verse can amuse the dowdiest of readers. Lemon-faced adults and fidgety kids sit still and read these rhythmic words and laugh and smile and love th It's hard to imagine a world without A Light in the Attic. This now-classic collection of poetry and drawings from Shel Silverstein celebrates its 20th anniversary with this special edition. Silverstein's humorous and creative verse can amuse the dowdiest of readers. Lemon-faced adults and fidgety kids sit still and read these rhythmic words and laugh and smile and love that Silverstein. Need proof of his genius? RockabyeRockabye baby, in the treetopDon't you know a treetopIs no safe place to rock?And who put you up there,And your cradle, too?Baby, I think someone down here'sGot it in for you. Shel, you never sounded so good. ...more

Using libraries like requests, and BeautifulSoup for scraping:

```python
booksData = []

for page in range(1, 51):

    print(f'Scraping page-{page}...')

    if page == 1:
        url = base_url
    else:
        url = page_url.format(page)


    res = requests.get(url)
    soup = BeautifulSoup(res.text, 'html.parser')

    for book in soup.select('article.product_pod'):
        title = book.h3.a['title']
        price = book.select_one('.price_color').text
        availability = book.select_one('.availability').text.strip()

        rating_class = book.select_one('p.star-rating')['class']
        star_word = [cls for cls in rating_class if cls != 'star-rating'][0]
        stars = rating_map.get(star_word, 0)

        relative_link = book.h3.a['href']
        if 'catalogue/' not in relative_link:
            relative_link = 'catalogue/' + relative_link
        book_url = urljoin(base_url, relative_link)

        detail_res = requests.get(book_url)
        detail_soup = BeautifulSoup(detail_res.text, 'html.parser')

        desc_tag = detail_soup.select_one('#product_description ~ p')
        description = desc_tag.text.strip() if desc_tag else None

        table = detail_soup.select_one('table.table.table-striped')
        info = {row.th.text: row.td.text for row in table.select('tr')} if table else None
        upc = info.get('UPC', '') if info else None


        genre_tags = detail_soup.select('ul.breadcrumb li > a')

        genre_tag = genre_tags[2].text.strip() if len(genre_tags) > 2 else None


        booksData.append({
            'title': title,
            'price': price,
            'availability': availability,
            'stars': stars,
            'url': book_url,
            'description': description,
            'upc': upc,
            'genre': genre_tag,
        })
```

Then saving the data scraped as csv for next steps:

```
title,price,availability,stars,url,description,upc,genre
A Light in the Attic,Â£51.77,In stock,3,https://books.toscrape.com/catalogue/a-light-in-the-attic_1000/ind
Tipping the Velvet,Â£53.74,In stock,1,https://books.toscrape.com/catalogue/tipping-the-velvet_999/index.ht
Soumission,Â£50.10,In stock,1,https://books.toscrape.com/catalogue/soumission_998/index.html,"Dans une Fra
Sharp Objects,Â£47.82,In stock,4,https://books.toscrape.com/catalogue/sharp-objects_997/index.html,"WICKED
Sapiens: A Brief History of Humankind,Â£54.23,In stock,5,https://books.toscrape.com/catalogue/sapiens-a-br
The Requiem Red,Â£22.65,In stock,1,https://books.toscrape.com/catalogue/the-requiem-red_995/index.html,"Pa
The Dirty Little Secrets of Getting Your Dream Job,Â£33.34,In stock,4,https://books.toscrape.com/catalogue
"The Coming Woman: A Novel Based on the Life of the Infamous Feminist, Victoria Woodhull",Â£17.93,In stock
The Boys in the Boat: Nine Americans and Their Epic Quest for Gold at the 1936 Berlin Olympics,Â£22.60,In
The Black Maria,Â£52.15,In stock,1,https://books.toscrape.com/catalogue/the-black-maria_991/index.html,"Pr
"Starving Hearts (Triangular Trade Trilogy, #1)",Â£13.99,In stock,2,https://books.toscrape.com/catalogue/s
Shakespeare's Sonnets,Â£20.66,In stock,4,https://books.toscrape.com/catalogue/shakespeares-sonnets_989/ind
Set Me Free,Â£17.46,In stock,5,https://books.toscrape.com/catalogue/set-me-free_988/index.html,"Aaron Ledb
Scott Pilgrim's Precious Little Life (Scott Pilgrim #1),Â£52.29,In stock,5,https://books.toscrape.com/cata
Rip it Up and Start Again,Â£35.02,In stock,5,https://books.toscrape.com/catalogue/rip-it-up-and-start-aga
"Our Band Could Be Your Life: Scenes from the American Indie Underground, 1981-1991",Â£57.25,In stock,3,ht
Olio,Â£23.88,In stock,1,https://books.toscrape.com/catalogue/olio_984/index.html,"Part fact, part fiction,
Mesaerion: The Best Science Fiction Stories 1800-1849,Â£37.59,In stock,1,https://books.toscrape.com/catalo
Libertarianism for Beginners,Â£51.33,In stock,2,https://books.toscrape.com/catalogue/libertarianism-for-be
It's Only the Himalayas,Â£45.17,In stock,2,https://books.toscrape.com/catalogue/its-only-the-himalayas_981
In Her Wake,Â£12.84,In stock,1,https://books.toscrape.com/catalogue/in-her-wake_980/index.html,"A perfect
How Music Works,Â£37.32,In stock,2,https://books.toscrape.com/catalogue/how-music-works_979/index.html,"Hc
"Foolproof Preserving: A Guide to Small Batch Jams, Jellies, Pickles, Condiments, and More: A Foolproof Gu
```

● Data Cleaning & Preprocessing:

Check data info with info() function

```
[7]: df.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1000 entries, 0 to 999
Data columns (total 8 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   title         1000 non-null   object
 1   price         1000 non-null   object
 2   availability  1000 non-null   object
 3   stars         1000 non-null   int64
 4   url           1000 non-null   object
 5   description   998 non-null    object
 6   upc           1000 non-null   object
 7   genre         1000 non-null   object
dtypes: int64(1), object(7)
memory usage: 62.6+ KB
```

There are 2 missing values in the description column.

```
[8]: df.dropna(inplace=True)
     df.info()

     <class 'pandas.core.frame.DataFrame'>
     Index: 998 entries, 0 to 999
     Data columns (total 8 columns):
      #   Column        Non-Null Count  Dtype
     ---  ------        --------------  -----
      0   title         998 non-null    object
      1   price         998 non-null    object
      2   availability  998 non-null    object
      3   stars         998 non-null    int64
      4   url           998 non-null    object
      5   description   998 non-null    object
      6   upc           998 non-null    object
      7   genre         998 non-null    object
     dtypes: int64(1), object(7)
     memory usage: 70.2+ KB
```

Now no missing values in data
Next we need to check for duplicated rows :

```
[9]: df.duplicated().sum()

[9]: 0
```

	title	price	availability	stars	url	description	upc	genre
919	The Selfish Gene	£29.45	In stock	1	https://books.toscrape.com/catalogue/the-selfi...	Richard Dawkins' brilliant reformulation of th...	6f2c0a9adbb2bca9	Science
62	The Most Perfect Thing: Inside (and Outside) a...	£42.96	In stock	4	https://books.toscrape.com/catalogue/the-most-...	Renowned ornithologist Tim Birkhead opens this...	aadee1c326d286e3	Science
307	Every Last Word	£46.47	In stock	3	https://books.toscrape.com/catalogue/every-las...	If you could read my mind, you wouldn't be smi...	a2bd4ca82d908442	Default
530	How to Be a Domestic Goddess: Baking and the A...	£28.25	In stock	2	https://books.toscrape.com/catalogue/how-to-be...	Nigella Lawson's How to Be a Domestic Goddess ...	eff5710404e2ff78	Food and Drink
210	Anonymous	£46.82	In stock	4	https://books.toscrape.com/catalogue/anonymous...	Logan Matthews is a father, architect, and wid...	d24b829be2ca81c4	Default
110	Don't Be a Jerk: And Other Practical Advice fr...	£37.97	In stock	2	https://books.toscrape.com/catalogue/dont-be-a...	A Radical but Reverent Paraphrasing of Dogenâ...	fce19f9b4a943f22	Religion
617	Last One Home (New Beginnings #1)	£59.98	In stock	3	https://books.toscrape.com/catalogue/last-one-...	Growing up, Cassie Carter and her sisters, Kar...	07e6810fd3236bda	Fiction
650	The Jazz of Physics: The Secret Link Between M...	£38.71	In stock	3	https://books.toscrape.com/catalogue/the-jazz-...	More than fifty years ago, John Coltrane drew ...	6a57286a8890d8c5	Nonfiction
809	Lust & Wonder	£11.87	In stock	2	https://books.toscrape.com/catalogue/lust-wond...	First came Running with Scissors. Then came Dr...	0dcadb06206abd3c	Autobiography
826	If I Run (If I Run #1)	£49.97	In stock	4	https://books.toscrape.com/catalogue/if-i-run-...	Casey knows the truth.But it wonâ t set her f...	61569d12d4ae7161	Christian Fiction

We have to fix price column data type:

```
[10]: df['price'] = df['price'].apply(lambda x: re.sub(r'Â£','', x))
      df.head()
```

	title	price	availability	stars	url	description	upc	genre
0	A Light in the Attic	51.77	In stock	3	https://books.toscrape.com/catalogue/a-light-i...	It's hard to imagine a world without A Light i...	a897fe39b1053632	Poetry
1	Tipping the Velvet	53.74	In stock	1	https://books.toscrape.com/catalogue/tipping-t...	"Erotic and absorbing...Written with starling ...	90fa61229261140a	Historical Fiction
2	Soumission	50.10	In stock	1	https://books.toscrape.com/catalogue/soumissio...	Dans une France assez proche de la nÂ´tre, un ...	6957f44c3847a760	Fiction
3	Sharp Objects	47.82	In stock	4	https://books.toscrape.com/catalogue/sharp-obj...	WICKED above her hipbone, GIRL across her hear...	e00eb4fd7b871a48	Mystery
4	Sapiens: A Brief History of Humankind	54.23	In stock	5	https://books.toscrape.com/catalogue/sapiens-a...	From a renowned historian comes a groundbreaki...	4165285e1663650f	History

```
[11]: #fix dtype for price col
      df['price']=df['price'].astype('float')
      df.dtypes
```

```
[11]: title           object
      price          float64
      availability    object
      stars            int64
      url             object
      description     object
      upc             object
      genre           object
      dtype: object
```

We used regex library to Remove unwanted characters such as "Â£" then transform price column to float dtype

```
[12]: #We are making sure the URL is correct using regex
      matches=df.url.apply(lambda x: bool(re.match(r'https://books.toscrape.com/catalogue/[\w-]+/index.html',x)))
      df=df[matches]
      df.head()
```

	title	price	availability	stars	url	description	upc	genre
0	A Light in the Attic	51.77	In stock	3	https://books.toscrape.com/catalogue/a-light-i...	It's hard to imagine a world without A Light i...	a897fe39b1053632	Poetry
1	Tipping the Velvet	53.74	In stock	1	https://books.toscrape.com/catalogue/tipping-t...	"Erotic and absorbing...Written with starling ...	90fa61229261140a	Historical Fiction
2	Soumission	50.10	In stock	1	https://books.toscrape.com/catalogue/soumissio...	Dans une France assez proche de la nÂ´tre, un ...	6957f44c3847a760	Fiction
3	Sharp Objects	47.82	In stock	4	https://books.toscrape.com/catalogue/sharp-obj...	WICKED above her hipbone, GIRL across her hear...	e00eb4fd7b871a48	Mystery
4	Sapiens: A Brief History of Humankind	54.23	In stock	5	https://books.toscrape.com/catalogue/sapiens-a...	From a renowned historian comes a groundbreaki...	4165285e1663650f	History

We are making sure the url for each book is correct by using regex

● Data Analysis & Visualization:

```
[14]: data=df.groupby('genre')['stars'].mean().round()
      df.genre.unique()
```

```
[14]: array(['Poetry', 'Historical Fiction', 'Fiction', 'Mystery', 'History',
             'Young Adult', 'Business', 'Default', 'Sequential Art', 'Music',
             'Science Fiction', 'Politics', 'Travel', 'Thriller',
             'Food and Drink', 'Romance', 'Childrens', 'Nonfiction', 'Art',
             'Spirituality', 'Philosophy', 'New Adult', 'Contemporary',
             'Fantasy', 'Add a comment', 'Science', 'Health', 'Horror',
             'Self Help', 'Religion', 'Christian', 'Crime', 'Autobiography',
             'Christian Fiction', 'Biography', 'Womens Fiction', 'Erotica',
             'Cultural', 'Psychology', 'Humor', 'Historical', 'Novels',
             'Short Stories', 'Suspense', 'Classics', 'Academic',
             'Sports and Games', 'Adult Fiction', 'Parenting', 'Paranormal'],
            dtype=object)
```

We grouped data by genre and find average star rating for each genre and now plot it

```
[15]:  plt.figure(figsize=(10,10))
       sns.barplot(y=data.index,x=data.values,palette='viridis')
       plt.ylabel('Book Category')
       plt.xlabel('Averge Ratings for each Category')
       plt.title('Book categories and average ratings')
       plt.show()
```

C:\Users\Mohamed\AppData\Local\Temp\ipykernel_6180\3383276283.py:2: FutureWarning:

Passing `palette` without assigning `hue` is deprecated and will be removed in v0.14.0. Assign the `y` variable to `hue` and set `legend=False` for the same effect.

 sns.barplot(y=data.index,x=data.values,palette='viridis')
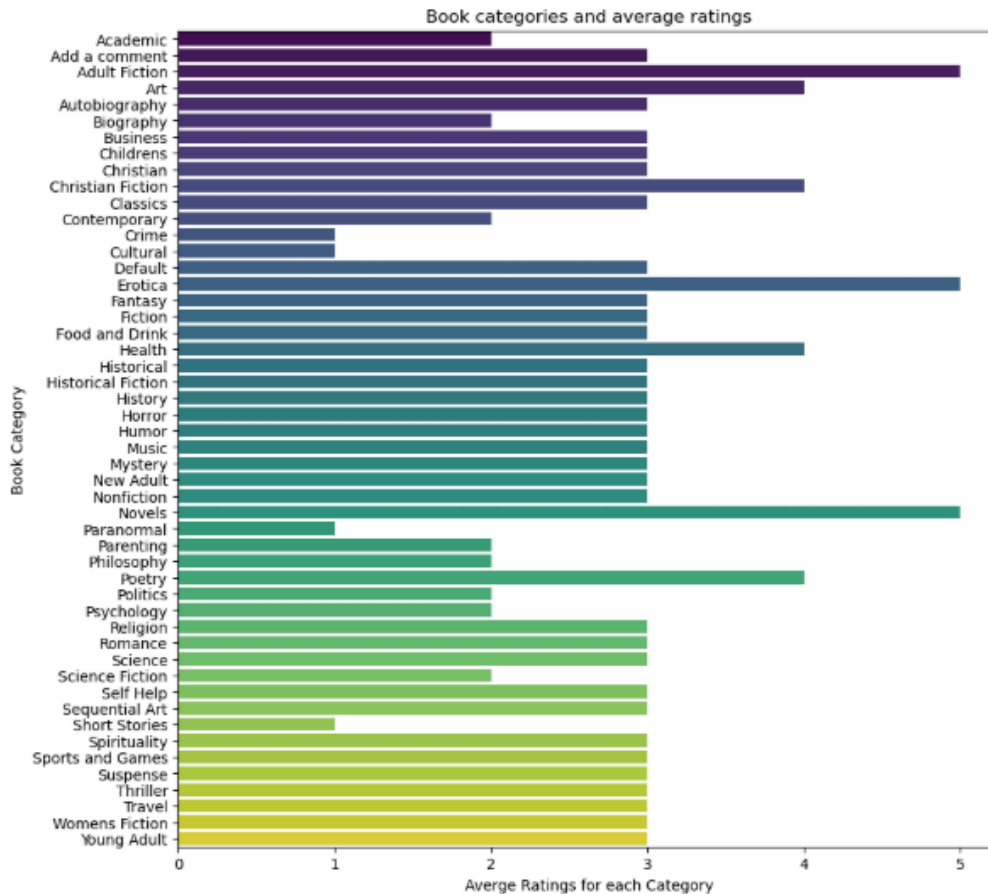
We used bar chart to make it easy to see which genre has high average star rating

```
[95]:  grouped_data_avg=df.groupby('genre')['price'].mean().reset_index()
        df2=grouped_data_avg
        df2.head()
```

[95]:

	genre	price
0	Academic	13.120000
1	Add a comment	35.796418
2	Adult Fiction	15.360000
3	Art	38.520000
4	Autobiography	37.053333

```
[99]:  plt.figure(figsize=(15,5))
        plt.plot(df2.genre,df2.price)
        plt.xticks(rotation=90)
        plt.xlabel('Book Category')
        plt.ylabel('avg price')
        plt.title('Comparison between genre of books and avg price for it')
        plt.show()
```
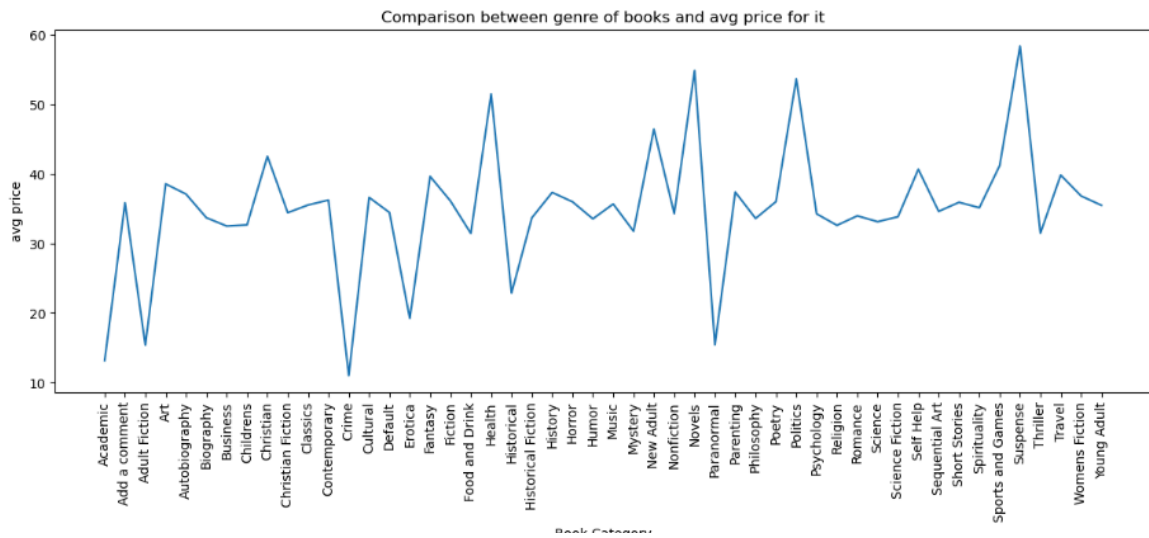

This chart shows us the average price per book for each category.
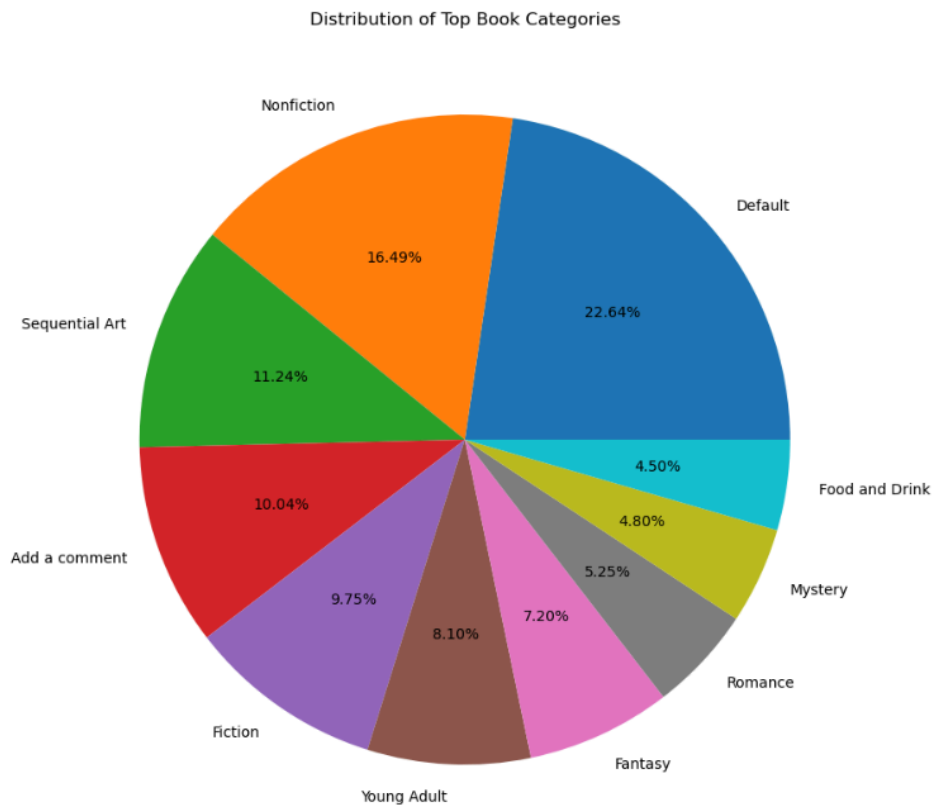
```
[16]:  plt.scatter(df.price,df.stars)
        plt.title('Relationship between price and Star Ratings')
        plt.xlabel('Price')
        plt.ylabel('Star Rating')
        plt.show()
```
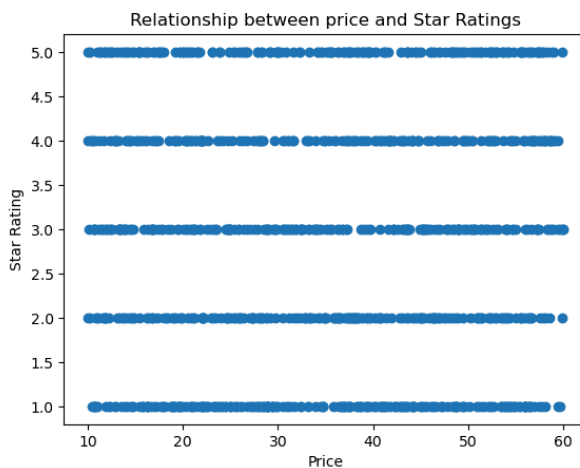

A plot showing no relationship between book price and ratings.

We used a scatter plot to see whether book ratings affect their prices, but it turned out there is no relationship between them.

```
[19]: plt.figure(figsize=(10,10))
      plt.pie(categories['count'],labels=categories['genre'],autopct='%0.2f%%')
      plt.title('Distribution of Top Book Categories')
      plt.show()
```


Distribution of Top Book Categories

We used a pie chart to identify the most common book category.

```
[24]: word=WordCloud(width=800,height=800).generate(art_description_text)
       plt.figure(figsize=(10,15))
       plt.imshow(word)
       plt.axis('off')
       plt.show()
```


We used a word cloud to find the most frequently occurring words in the descriptions of books from the Sequential Art category.

● Saving Data using MongoDB:

Saving data in mongoDB

```
[26]: #import lib which deal with mongoDB through python
      from pymongo import MongoClient
```

```
[27]: #connect to mongoDB
      client=MongoClient('mongodb://localhost:27017')
      #create data base('My_Data_Base')
      DB=client['My_Data_Base']
      #create a collection in ('My_Data_Base') it's like table in SQL but here it saves data as a documents
      Documents=DB['book_doc']
```

The first step is import pymongo to make a connection to the server, and create the database where we will store the data.

```
[29]: list_of_dict=df.to_dict('records')
      list_of_dict
```

```
[29]: [{'title': 'A Light in the Attic',
        'price': 51.77,
        'availability': 'In stock',
        'stars': 3,
        'url': 'https://books.toscrape.com/catalogue/a-light-in-the-attic_1000/index.html',
        'description': "It's hard to imagine a world without A Light in the Attic. This now-classic collection of poetry and drawings from Shel Silverstein c
     elebrates its 20th anniversary with this special edition. Silverstein's humorous and creative verse can amuse the dowdiest of readers. Lemon-faced adul
     ts and fidgety kids sit still and read these rhythmic words and laugh and smile and love th It's hard to imagine a world without A Light in the Attic.
     This now-classic collection of poetry and drawings from Shel Silverstein celebrates its 20th anniversary with this special edition. Silverstein's humor
     ous and creative verse can amuse the dowdiest of readers. Lemon-faced adults and fidgety kids sit still and read these rhythmic words and laugh and smi
     le and love that Silverstein. Need proof of his genius? RockabyeRockabye baby, in the treetopDon't you know a treetopIs no safe place to rock?And who p
     ut you up there,And your cradle, too?Baby, I think someone down here'sGot it in for you. Shel, you never sounded so good. ...more",
        'upc': 'a897fe39b1053632',
        'genre': 'Poetry'},
       {'title': 'Tipping the Velvet',
        'price': 53.74,
        'availability': 'In stock',
        'stars': 1
```

We converted the DataFrame into a list of dictionaries because the data is stored as documents

```
[30]: #insert data in mongoDB
      Documents.insert_many(list_of_dict)
      #it autmatically assigns an id for each document
```

```
[30]: InsertManyResult([ObjectId('6806ace8d45f8b2c7b609dea'), ObjectId('6806ace8d45f8b2c7b609deb'), ObjectId('6806ace8d45f8b2c7b609dec'), ObjectId('6806ace8d
      45f8b2c7b609ded'), ObjectId('6806ace8d45f8b2c7b609dee'), ObjectId('6806ace8d45f8b2c7b609def'), ObjectId('6806ace8d45f8b2c7b609df0'), ObjectId('6806ace8
      d45f8b2c7b609df1'), ObjectId('6806ace8d45f8b2c7b609df2'), ObjectId('6806ace8d45f8b2c7b609df3'), ObjectId('6806ace8d45f8b2c7b609df4'), ObjectId('6806ace
      8d45f8b2c7b609df5'), ObjectId('6806ace8d45f8b2c7b609df6'), ObjectId('6806ace8d45f8b2c7b609df7'), ObjectId('6806ace8d45f8b2c7b609df8'), ObjectId('6806ac
      e8d45f8b2c7b609df9'), ObjectId('6806ace8d45f8b2c7b609dfa'), ObjectId('6806ace8d45f8b2c7b609dfb'), ObjectId('6806ace8d45f8b2c7b609dfc'), ObjectId('6806a
      ce8d45f8b2c7b609dfd'), ObjectId('6806ace8d45f8b2c7b609dfe'), ObjectId('6806ace8d45f8b2c7b609dff'), ObjectId('6806ace8d45f8b2c7b609e00'), ObjectId('6806
      ace8d45f8b2c7b609e01'), ObjectId('6806ace8d45f8b2c7b609e02'), ObjectId('6806ace8d45f8b2c7b609e03'), ObjectId('6806ace8d45f8b2c7b609e04'), ObjectId('680
      6ace8d45f8b2c7b609e05'), ObjectId('6806ace8d45f8b2c7b609e06'), ObjectId('6806ace8d45f8b2c7b609e07'), ObjectId('6806ace8d45f8b2c7b609e08'), ObjectId('68
      06ace8d45f8b2c7b609e09'), ObjectId('6806ace8d45f8b2c7b609e0a'), ObjectId('6806ace8d45f8b2c7b609e0b'), ObjectId('6806ace8d45f8b2c7b609e0c'), ObjectId('6
      806ace8d45f8b2c7b609e0d'), ObjectId('6806ace8d45f8b2c7b609e0e'), ObjectId('6806ace8d45f8b2c7b609e0f'), ObjectId('6806ace8d45f8b2c7b609e10'), ObjectId
      ('6806ace8d45f8b2c7b609e11'), ObjectId('6806ace8d45f8b2c7b609e12'), ObjectId('6806ace8d45f8b2c7b609e13'), ObjectId('6806ace8d45f8b2c7b609e14'), ObjectI
      d('6806ace8d45f8b2c7b609e15'), ObjectId('6806ace8d45f8b2c7b609e16'), ObjectId('6806ace8d45f8b2c7b609e17'), ObjectId('6806ace8d45f8b2c7b609e18'), Object
      Id('6806ace8d45f8b2c7b609e19'), ObjectId('6806ace8d45f8b2c7b609e1a'), ObjectId('6806ace8d45f8b2c7b609e1b'), ObjectId('6806ace8d45f8b2c7b609e1c'), Objec
      tId('6806ace8d45f8b2c7b609e1d'), ObjectId('6806ace8d45f8b2c7b609e1e'), ObjectId('6806ace8d45f8b2c7b609e1f'), ObjectId('6806ace8d45f8b2c7b609e20'), Obje
      ctId('6806ace8d45f8b2c7b609e21'), ObjectId('6806ace8d45f8b2c7b609e22'), ObjectId('6806ace8d45f8b2c7b609e23'), ObjectId('6806ace8d45f8b2c7b609e24'), Obj
      ectId('6806ace8d45f8b2c7b609e25'), ObjectId('6806ace8d45f8b2c7b609e26'), ObjectId('6806ace8d45f8b2c7b609e27'), ObjectId('6806ace8d45f8b2c7b609e28'), Ob
      jectId('6806ace8d45f8b2c7b609e29'), ObjectId('6806ace8d45f8b2c7b609e2a'), ObjectId('6806ace8d45f8b2c7b609e2b'), ObjectId('6806ace8d45f8b2c7b609e2c'), O
      bjectId('6806ace8d45f8b2c7b609e2d'), ObjectId('6806ace8d45f8b2c7b609e2e'), ObjectId('6806ace8d45f8b2c7b609e2f'), ObjectId('6806ace8d45f8b2c7b609e30'),
```

insert_many() function used for insert data in the Database.

```
[31]: #for retrive data we use find funcion
      retrevied_data=list(Documents.find())
```

```
[32]: retrevied_data=pd.DataFrame(retrevied_data)
      retrevied_data.head()
```

[32]:

	_id	title	price	availability	stars	url	description	upc	genre
0	68027b4ead852950191ccff0	A Light in the Attic	51.77	In stock	3	https://books.toscrape.com/catalogue/a-light-i...	It's hard to imagine a world without A Light i...	a897fe39b1053632	Poetry
1	68027b4ead852950191ccff1	Tipping the Velvet	53.74	In stock	1	https://books.toscrape.com/catalogue/tipping-t...	"Erotic and absorbing...Written with starling ...	90fa61229261140a	Historical Fiction
2	68027b4ead852950191ccff2	Soumission	50.10	In stock	1	https://books.toscrape.com/catalogue/soumissio...	Dans une France assez proche de la nÃ´tre, un ...	6957f44c3847a760	Fiction
3	68027b4ead852950191ccff3	Sharp Objects	47.82	In stock	4	https://books.toscrape.com/catalogue/sharp-obj...	WICKED above her hipbone, GIRL across her hear...	e00eb4fd7b871a48	Mystery
4	68027b4ead852950191ccff4	Sapiens: A Brief History of Humankind	54.23	In stock	5	https://books.toscrape.com/catalogue/sapiens-a...	From a renowned historian comes a groundbreaki...	4165285e1663650f	History

find() function is used to retrieve data from the database .

● Streamlit app:

Using Streamlit, we turned the visualization interactive:
For Example:

```python
df2 = df.groupby('genre')['price'].mean().reset_index()

data_df = pd.DataFrame({
    'Genre': df2['genre'],
    'Average Price': df2['price'],
})

st.markdown("### Comparison between genre of books and avg price for it:")

data_df = data_df.set_index('Genre')

st.line_chart(data_df)

st.divider()
```
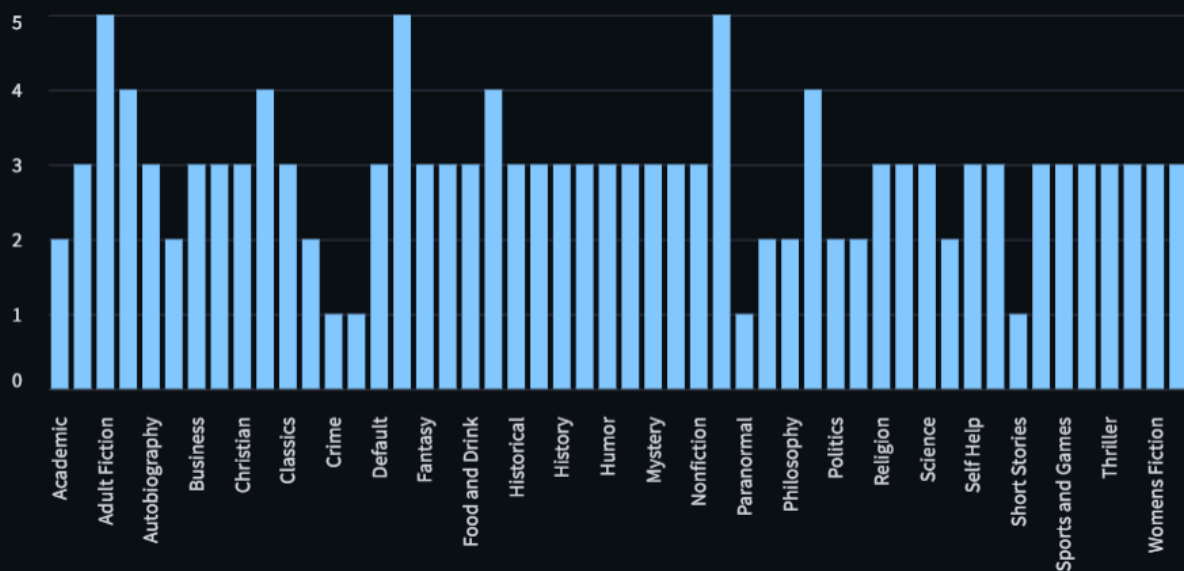
To turn the Line Chart to be like this:

Book categories and average ratings:

Also using plotly.express library to cover the other forms of visualization not covered in streamlit itself Like:

```python
genre_counts = df['genre'].value_counts().reset_index()
genre_counts.columns = ['Genre', 'Count']

categories = genre_counts.sort_values(by='Count', ascending=False).head(10)

st.markdown("### Distribution of Top Book Categories:")

fig = px.pie(
    categories,
    names='Genre',
    values='Count',
    color_discrete_sequence=px.colors.sequential.Viridis
)

st.plotly_chart(fig, use_container_width=True)

st.divider()
```
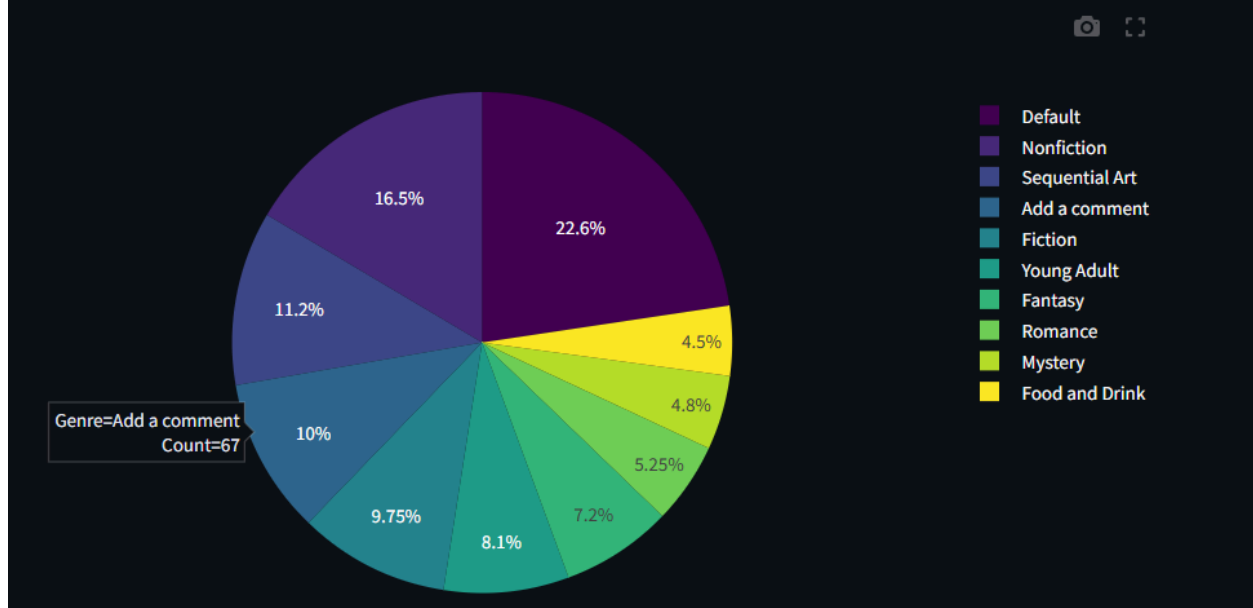
Distribution of Top Book Categories:

Then deployed the app on Streamlit with its link being:
https://sam50x-tools-project-appapp-y3toyv.streamlit.app/