

# **Project Documentation**

## **For**

# **BookScape Explorer**

Version 1.0

**Prepared by**

Name: Mohamed Niyas

Batch No: MAE1

## BookScape Explorer Project Documentation

---

### Release Notice

This is Version 1.0 of BookScape Explorer Project Documentation

| Name          | Batch No | Organization |
|---------------|----------|--------------|
| Prepared By   |          |              |
| Mohamed Niyas | MAE1     | Guvi         |

### Version History:

| Version No. | Date Issued | Remarks         |
|-------------|-------------|-----------------|
| 1.0         | 23 Dec 2024 | Initial Release |

---

## Table of Contents

|           |   |           |
|-----------|---|-----------|
| <b>1</b>  | <b>INTRODUCTION.....</b>                          | <b>4</b>  |
| 1.1.1     | <i>Purpose Of the Document .....</i>              | <i>4</i>  |
| 1.1.2     | <i>Scope .....</i>                                | <i>4</i>  |
| 1.1.3     | <i>Business Use cases .....</i>                   | <i>5</i>  |
| <b>2</b>  | <b>BOOKSCAPE ARCHITECTURE DIAGRAM.....</b>        | <b>6</b>  |
| 2.1.1     | <i>Overview.....</i>                              | <i>6</i>  |
| 2.1.2     | <i>Data Extraction: .....</i>                     | <i>6</i>  |
| <b>3</b>  | <b>INSTALLATION .....</b>                         | <b>7</b>  |
| 3.1.1     | <i>Prerequisites.....</i>                         | <i>7</i>  |
| <b>4</b>  | <b>CLONE THE REPOSITORY.....</b>                  | <b>8</b>  |
| <b>5</b>  | <b>INSTALL DEPENDENCIES .....</b>                 | <b>9</b>  |
| 5.1.1     | <i>requirements.txt file.....</i>                 | <i>9</i>  |
| 5.1.2     | <i>install the dependencies:.....</i>             | <i>9</i>  |
| <b>6</b>  | <b>MYSQL DATABASE SETUP .....</b>                 | <b>10</b> |
| 6.1.1     | <i>Create a MySQL Database: .....</i>             | <i>10</i> |
| 6.1.2     | <i>Create the books Table:.....</i>               | <i>10</i> |
| <b>7</b>  | <b>CONFIGURATION .....</b>                        | <b>11</b> |
| 7.1.1     | <i>Update Database Connection .....</i>           | <i>11</i> |
| 7.1.2     | <i>Google Books API Key.....</i>                  | <i>11</i> |
| <b>8</b>  | <b>RUN THE APPLICATION.....</b>                   | <b>12</b> |
| 8.1.1     | <i>Access the Application.....</i>                | <i>12</i> |
| <b>9</b>  | <b>FEATURES .....</b>                             | <b>13</b> |
| 9.1.1     | <i>Home Page .....</i>                            | <i>13</i> |
| 9.1.2     | <i>Home Page Wireframe.....</i>                   | <i>13</i> |
| 9.1.3     | <i>Search Page .....</i>                          | <i>13</i> |
| 9.1.4     | <i>Search Page Wireframe.....</i>                 | <i>14</i> |
| 9.1.5     | <i>Data Analysis Page .....</i>                   | <i>14</i> |
| 9.1.6     | <i>Data Analysis Options and Wireframes .....</i> | <i>14</i> |
| <b>10</b> | <b>INSIGHTS AND CHALLENGES.....</b>               | <b>23</b> |
| 10.1.1    | <i>Insights .....</i>                             | <i>23</i> |
| 10.1.2    | <i>Challenges .....</i>                           | <i>23</i> |
| <b>11</b> | <b>TROUBLESHOOTING.....</b>                       | <b>24</b> |
| 11.1.1    | <i>Common Issues.....</i>                         | <i>24</i> |

# 1 Introduction

Welcome to **Bookscape Explorer**, a comprehensive tool designed to enhance your book discovery and data analysis experience. This project leverages the power of the Google Books API to fetch detailed information about books and stores this data in a MySQL database. With Bookscape Explorer, users can search for books, view detailed information, and perform various data analyses to gain insights into book trends and statistics.

Bookscape Explorer is built using Streamlit, a powerful framework for creating interactive web applications in Python. The application integrates several key technologies, including Pandas for data manipulation, SQLAlchemy for database interactions, and Plotly for data visualization.

### 1.1.1 Purpose Of the Document

This documentation will guide you through the following sections to help you set up, configure, and use Bookscape Explorer effectively:

- **Installation:** Step-by-step instructions to install all necessary dependencies and set up the project on your local machine.
- **Configuration:** Detailed guidance on configuring the application, including setting up the MySQL database and integrating the Google Books API.
- **Usage:** Instructions on how to run the application and navigate through its features.
- **Features:** An overview of the main features of Bookscape Explorer, including book search, data storage, and data analysis.
- **Data Analysis Options:** A list of the various data analysis options available within the application.
- **Wireframe:** Visual representations of the application's user interface to help you understand its layout and functionality.
- **Insights and Challenges:** Insights gained during the development process and challenges faced, along with solutions implemented.
- **Troubleshooting:** Common issues and their solutions to help you resolve any problems you might encounter.

### 1.1.2 Scope

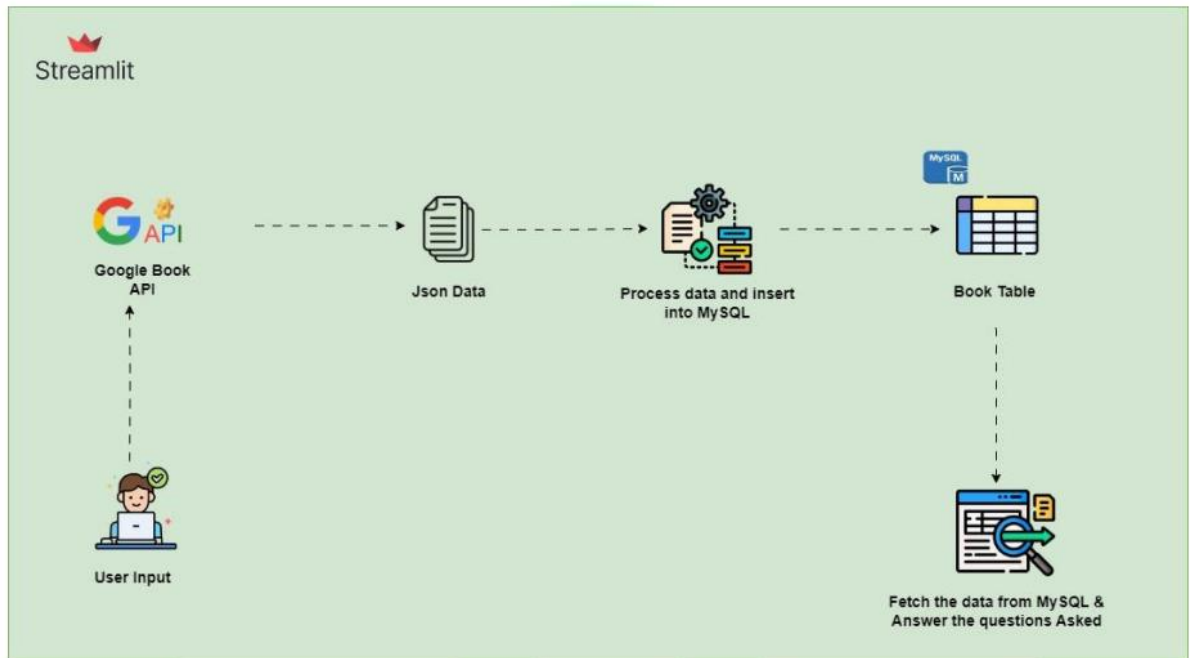
The BookScape Explorer project aims to facilitate users in discovering and analysing book data through a web application. The application will extract data from online book APIs, store this information in a SQL database, and enable data analysis using SQL queries. The project will provide insights into book trends, user preferences, and reviews, helping users make informed reading choices while also offering a platform for community engagement. This initiative targets avid readers, researchers, and book enthusiasts.

### **1.1.3 Business Use cases**

- Search Optimization: Filter books based on genre, author, or publication year.
- Trend Analysis: Identify trending genres or authors over time.
- Data Insights: Perform analysis on user reviews and ratings to identify popular books.
- Decision Support: Provide insights for libraries or bookstores to stock trending books.

## 2 BookScape Architecture Diagram

### 2.1.1 Overview



### 2.1.2 Data Extraction:

Utilize the Google Books API to gather comprehensive data on a variety of books. Extract information such as book titles, authors, publication dates, genres, descriptions, user reviews etc

## **3 Installation**

### **3.1.1 Prerequisites**

- Python 3.7 or higher
- MySQL Server
- MySQL Workbench (optional, for database management)

## **4 Clone the Repository**

git clone <https://github.com/niyasatg/BookScape-Explorer>



## 5 Install Dependencies

### 5.1.1 requirements.txt file

Create a requirements.txt file with the following content:

```
pandas==2.2.3  
requests==2.32.3  
streamlit==1.41.1  
plotly==5.24.1  
SQLAlchemy==2.0.36  
PyMySQL==1.1.1  
streamlit-option-menu==0.4.0
```

### 5.1.2 install the dependencies:

```
pip install -r requirements.txt
```

## 6 MySQL Database Setup

### 6.1.1 Create a MySQL Database:

```
CREATE DATABASE bookscope;
```

### 6.1.2 Create the books Table:

```
USE bookscope;
```

```
CREATE TABLE books (  
    book_id VARCHAR(36) PRIMARY KEY,  
    search_key VARCHAR(255),  
    book_title VARCHAR(255),  
    book_subtitle VARCHAR(255),  
    book_authors TEXT,  
    book_description TEXT,  
    industryIdentifiers TEXT,  
    text_readingModes INT,  
    image_readingModes INT,  
    pageCount INT,  
    languages VARCHAR(50),  
    imageLinks TEXT,  
    ratingsCount INT,  
    averageRating FLOAT,  
    country VARCHAR(50),  
    saleability VARCHAR(50),  
    isEbook BOOLEAN,  
    amount_listPrice FLOAT,  
    currencyCode_listPrice VARCHAR(10),  
    amount_retailPrice FLOAT,  
    currencyCode_retailPrice VARCHAR(10),  
    buyLink TEXT,  
    Publishedyear VARCHAR(10),  
    categories TEXT,  
    publisher VARCHAR(255)  
);
```

## **7 Configuration**

### **7.1.1 Update Database Connection**

Update the database connection string in your code:

```
engine = create_engine('mysql+pymysql://root:yourpassword@localhost:3306/bookscape')
```

### **7.1.2 Google Books API Key**

Replace "Your api key" with your actual Google Books API key in the code:

```
api_key = "Your api key"
```

## **8 Run the Application**

`streamlit run app.py`

### **8.1.1 Access the Application**

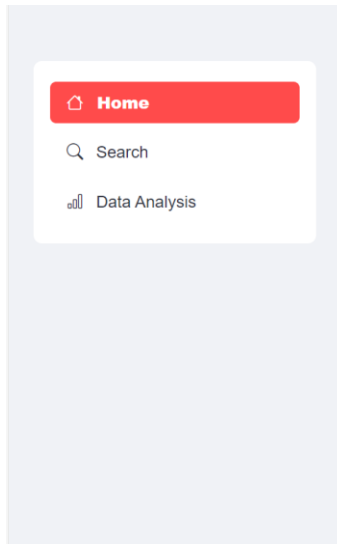
Open your web browser and go to <http://localhost:8501>.

## 9 Features

### 9.1.1 Home Page

- Introduction to Bookscape Explorer.

### 9.1.2 Home Page Wireframe



Bookscape explorer offers you over a million titles across genres like **Fiction, Non-fiction, Crime, Thriller, Comics** and more, and across categories like **Children's Books, Young Adults, Academic Books and Textbooks, Graphic Novels, Manga, Indian Writing, Classics** and more.

If reading is your passion or pastime, head over to **Bookscape** to browse through a massive selection of titles in multiple languages as well. From new releases, bestsellers to curated editor's picks, there is something for everyone.

**Students** can take their pick from exam prep books, university textbooks and references, school textbooks and guides. **Young teenagers** will find fiction and fantasy books from the Young Adults section interesting. Lovers of true crime, suspense and horror stories would find the Crime, Mystery & Thrillers section of the site an absolute delight.

### 9.1.3 Search Page

- Search for books based on genre, authors, etc.
- Fetch book details from the Google Books API.
- Store book details in the MySQL database.

9.1.4 Search Page Wireframe

Home

Search

Data Analysis

Deploy

BookScape

Search books

Non-fiction

Search

| Title                                     | Authors                 | Page Count | Published Date | Categories            | Publisher            |
|---|-------------------------|------------|----------------|-----------------------|----------------------|
| I Have No Mouth & I Must Scream           | Harlan Ellison          | 167        | 2014-04-29     | [[Fiction],]          | Open Road Media      |
| What Is Non-fiction Cinema?               | Trevor Ponech           | 267        | 2021-11-28     | [[Social Science],]   | Routledge            |
| Guiding Readers Through Non-fiction       | Dave Leochko, Tom Rossi | 186        | 2007           | [[Guided reading],]   | Portage & Main Press |
| Mastering Nonfiction Writing with ChatGPT | Maximus Wilson          | 71         | 2023-03-19     | [[Computers],]        | Adidas Wilson        |
| Nonfiction Matters                        | Stephanie Harvey        | 249        | 1998           | [[English language],] | Stenhouse Publishers |

Books data extracted and Data inserted successfully.

9.1.5 Data Analysis Page

- Perform various data analyses on the stored book data.

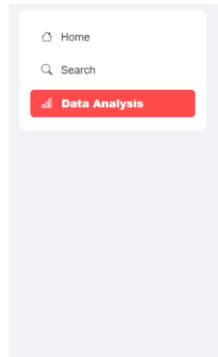
9.1.6 Data Analysis Options and Wireframes

- Check Availability of eBooks vs Physical Books
- Wireframe:

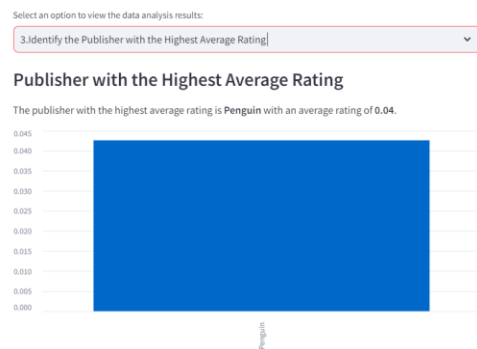
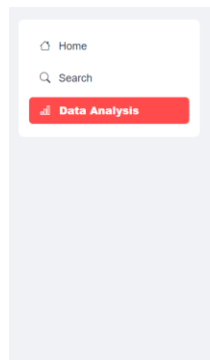


## BookScape Explorer Project Documentation

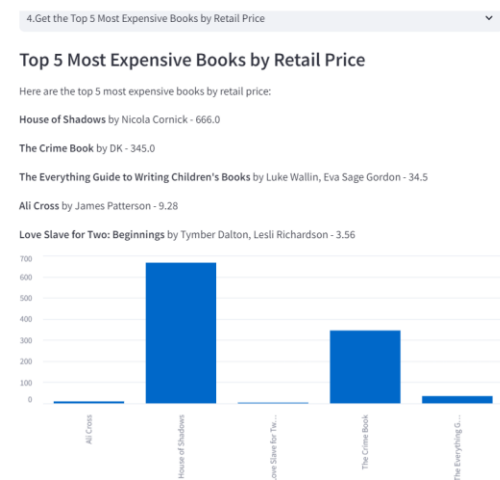
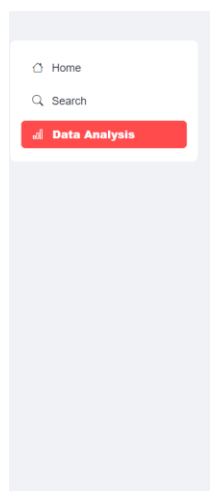
- Find the Publisher with the Most Books Published
- Wireframe:



- Identify the Publisher with the Highest Average Rating
- Wireframe:



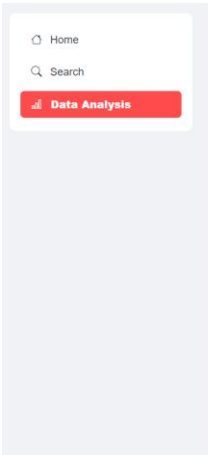
- Get the Top 5 Most Expensive Books by Retail Price
- Wireframe:



- Find Books Published After 2010 with at Least 500 Pages

BookScape Explorer Project Documentation

- Wireframe:



Select an option to view the data analysis results:

5. Find Books Published After 2010 with at Least 500 Pages

**Books Published After 2010 with at Least 500 Pages**

Here are the books published after 2010 with at least 500 pages:

**The Life of Jonathan Swift, Dean of St. Patrick's, Dublin** by Henry Craik - 606 pages, published in 2024-05-01

**Crime and Criminal Justice** by Stacy L. Mallicoat, Denise Paquette Boots - 738 pages, published in 2023-12-13

**Handbook of Mathematical Models and Algorithms in Computer Vision and Imaging** by Ke Chen, Carola-Bibiane Schönlieb, Xue-Cheng Tai, Laurent Younes - 1981 pages, published in 2023-02-24

**Albert Winsemius And Singapore: Here It Is Going To Happen** by Euston Quah, Luke Nursultan Iuldashov, Zach Lee - 1049 pages, published in 2022-05-18

**America's Book** by Mark A. Noll - 865 pages, published in 2022

**Uncertainty Quantification in Variational Inequalities** by Joachim Gwinner, Baasansuren Jadamba, Akhtar A. Khan, Fabio Raciti - 594 pages, published in 2021-12-21

**Anson's Navy** by Brian Lavery - 675 pages, published in 2021-11-30

**Applied Computer Sciences in Engineering** by Juan Carlos Figueroa-García, Fabián Steven Garay-Rairán, Germán Jairo Hernández-Pérez, Yesid Díaz-Gutiérrez - 536 pages, published in 2020-10-07

- List Books with Discounts Greater than 20%
- Wireframe:



Select an option to view the data analysis results:

6. List Books with Discounts Greater than 20%

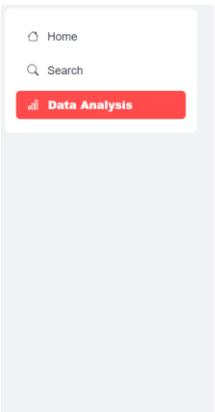
**Books with Discounts Greater than 20%**

|   | book_title | book_authors    | amount_listPrice | amount_retailPrice | discount_percentage |
|---|------------|-----------------|------------------|--------------------|---------------------|
| 0 | All Cross  | James Patterson | 13.65            | 9.28               | 32.0147             |

Here are the books with discounts greater than 20%:

**All Cross** by James Patterson - List Price: 13.65, Retail Price: 9.28, Discount: 32.01%

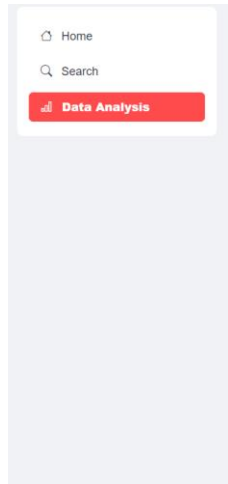
- Find the Average Page Count for eBooks vs Physical Books
- Wireframe:



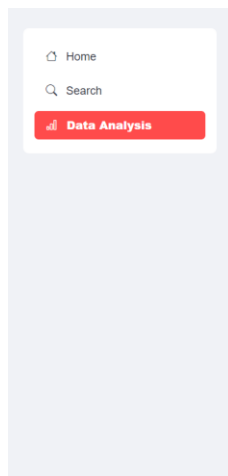


## BookScape Explorer Project Documentation

- Find the Top 3 Authors with the Most Books
- Wireframe:



- List Publishers with More than 10 Books
- Wireframe:



- Find the Average Page Count for Each Category
- Wireframe:

BookScape Explorer Project Documentation

Home

Search

Data Analysis

BookScape

Select an option to view the data analysis results:

10.Find the Average Page Count for Each Category

Average Page Count for Each Category

| Category                         | Average Page Count |
|----------------------------------|--------------------|
| 0 ARCHITECTURE                   | 312                |
| 1 Art                            | 227.8              |
| 2 Best books                     | 960                |
| 3 Biography & Autobiography      | 247.2              |
| 4 Bodybuilders                   | 324                |
| 5 Business & Economics           | 306.2              |
| 6 Children's literature, English | 116                |
| 7 Comics & Graphic Novels        | 365.5              |
| 8 Computers                      | 110                |

- Retrieve Books with More than 3 Authors
- Wireframe:

Home

Search

Data Analysis

Select an option to view the data analysis results:

11.Retrieve Books with More than 3 Authors

Books with More than 3 Authors

| book_title   | book_authors                                     |
|--|--|
| 0 Applied Computer Sciences in Engineering               | Juan Carlos Figueroa-García, Fabián Steven Gara  |
| 1 Captain America  | Mark Waid, James Felder, Brian K. Vaughan, Roge  |
| 2 Captain America  | Mark Waid, James Felder, Brian K. Vaughan, Roge  |
| 3 Confronting Peace                                      | Susan H. Allen, Landon E. Hancock, Christopher I |
| 4 Handbook of Mathematical Models and Algorithms in Comp | Ke Chen, Carola-Bibiane Schönlieb, Xue-Cheng T.  |
| 5 Hub-and-Spoke Cartels                                  | Luke Garrod, Joseph E. Harrington, Jr., Matthew  |
| 6 Laterality in Sports                                   | Florian Löffing, Norbert Hagemann, Bernd Strau   |
| 7 Teaching and Research in Contemporary Higher Education | Jung Cheol Shin, Akira Arimoto, William K. Cum   |
| 8 The Changing Academic Profession in Japan              | Akira Arimoto, William K. Cummings, Futao Huan   |
| 9 The Most Important Comic Book on Earth                 | Cara Delevingne, Ricky Gervais, Jane Goodall, Sc |

Here are the books with more than 3 authors:

Applied Computer Sciences in Engineering by Juan Carlos Figueroa-García, Fabián Steven Garay-Rairán, Germán Jairo Hernández-Pérez, Yesid Díaz-Gutiérrez - 536 pages, published in 2020-10-07

- Books with Ratings Count Greater Than the Average
- Wireframe:

Home

Search

Data Analysis

BookScape

Select an option to view the data analysis results:

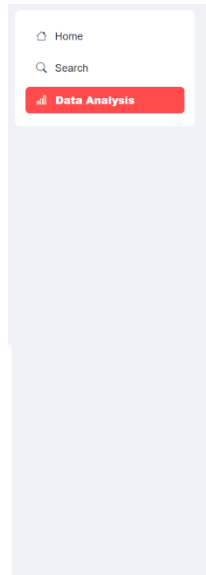
12.Books with Ratings Count Greater Than the Average

Books with Ratings Count Greater Than the Average

| book_title           | book_authors   |
|----------------------|----------------|
| 0 America (the Book) | Richard Moreno |
| 1 Candida            | Bernard Shaw   |

## BookScape Explorer Project Documentation

- Books with the Same Author Published in the Same Year
- Wireframe:



Select an option to view the data analysis results:

13. Books with the Same Author Published in the Same Year

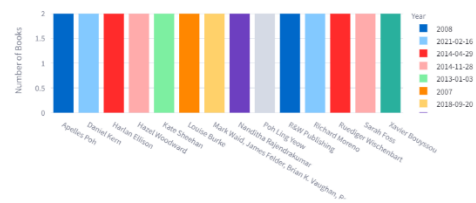
### Books with the Same Author Published in the Same Year

Here are the books with the same author published in the same year:

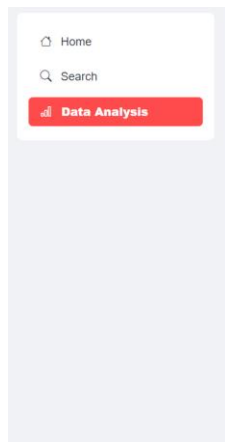
Apelles Poh (2008): Live Well, Love Much, Laugh Often, Live Well, Love Much, Laugh Often

Daniel Kern (2021-02-16): How to Write an eBook Bestseller, How to Write an eBook Bestseller

Number of Books by the Same Author Published in the Same Year



- Books with a Specific Keyword in the Title
- Wireframe:



Select an option to view the data analysis results:

14. Books with a Specific Keyword in the Title

### Books with a Specific Keyword in the Title

Enter a keyword to search in book titles:

Drama

|   | book_title | book_authors              | pageCount | Publishedyear |
|---|------------|---------------------------|-----------|---------------|
| 0 | Drama      | David Rockwell, Bruce Mau | 312       | 2021          |

Here are the books with the keyword 'Drama' in the title:

Drama by David Rockwell, Bruce Mau - 312 pages, published in 2021

- Year with the Highest Average Book Price
- Wireframe:

BookScape Explorer Project Documentation

Home

Search

Data Analysis

Select an option to view the data analysis results:

15.Year with the Highest Average Book Price

|   | Publishedyear | average_price |
|---|---------------|---------------|
| 0 | 2022-04-18    | 600           |
| 1 | 2017-04-03    | 250           |
| 2 | 2011-01-18    | 25            |
| 3 | 2019-12-12    | 13.65         |
| 4 | 2024-09-17    | 4.29          |
| 5 | 2024-03-16    | 1.92          |
| 6 | 2022          | 0             |
| 7 | 2018-11-10    | 0             |
| 8 | 2024          | 0             |
| 9 | 2013          | 0             |

The year with the highest average book price is 2022-04-18 with an average price of 600.00.

- Count Authors Who Published 3 Consecutive Years
- Wireframe:

Home

Search

Data Analysis

Select an option to view the data analysis results:

16.Count Authors Who Published 3 Consecutive Years

Count of Authors Who Published in 3 Consecutive Years

|   | book_authors   | Publishedyear | next_year | year_after_next | author_count |
|---|----------------|---------------|-----------|-----------------|--------------|
| 0 | Richard Moreno | 1972          | 1971      | 1970            | 1            |
| 1 | Richard Moreno | 1972          | 1971      | 1970            | 1            |

The number of authors who published books in three consecutive years is: 1

- Find Authors Who Published Books in the Same Year but Under Different Publishers
- Wireframe:

Home

Search

Data Analysis

Select an option to view the data analysis results:

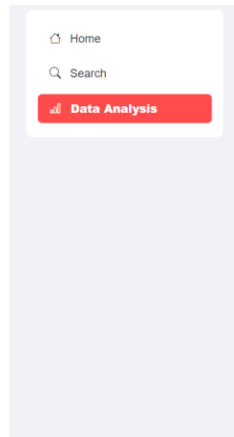
17.Write a SQL query to find authors who have published books in the same year but under diff...

Authors Who Published Books in the Same Year with Different Publishers

|   | book_authors     | Publishedyear | publishers                        | book_count |
|---|------------------|---------------|-----------------------------------|------------|
| 0 | Daniel Kern      | 2021-02-16    | BoD – Books on Demand, Penguin    | 2          |
| 1 | Hazel Woodward   | 2014-11-28    | Penguin, Ubiquity Press           | 2          |
| 2 | Kate Sheehan     | 2013-01-03    | Bloomsbury Publishing USA, Pengu  | 2          |
| 3 | Louise Burke     | 2007          | Human Kinetics, National Geograph | 2          |
| 4 | R&N Publishing   | 2006-12       | Andrew J Ruzicho, Penguin         | 2          |
| 5 | Ruediger Wischer | 2013-02-11    | "O'Reilly Media, Inc.", Penguin   | 2          |

## BookScape Explorer Project Documentation

- Write a SQL query to find authors who have published books in the same year but under different publishers. Return the authors, year, and the COUNT of books they published in that year.
- Wireframe:



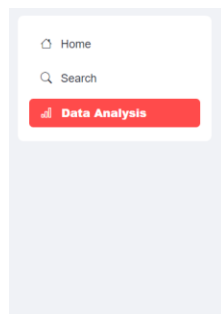
Select an option to view the data analysis results:

17. Write a SQL query to find authors who have published books in the same year but under diff... ▼

### Authors Who Published Books in the Same Year with Different Publishers

|   | book_authors     | Publishedyear | publishers                          | book_count |
|---|------------------|---------------|-------------------------------------|------------|
| 0 | Daniel Kern      | 2021-02-16    | BoD – Books on Demand, Penguin      | 2          |
| 1 | Hazel Woodward   | 2014-11-28    | Penguin, Ubiquity Press             | 2          |
| 2 | Kate Sheehan     | 2013-01-03    | Bloomsbury Publishing USA, Penguin  | 2          |
| 3 | Louise Burke     | 2007          | Human Kinetics, National Geographic | 2          |
| 4 | R&W Publishing   | 2006-12       | Andrew J Ruzicho, Penguin           | 2          |
| 5 | Ruediger Wischer | 2013-02-11    | "O'Reilly Media, Inc.", Penguin     | 2          |

- Create a query to find the average amount `retailPrice` of eBooks and physical books. Return a single result set with columns for `avg_ebook_price` and `avg_physical_price`. Ensure to handle cases where either category may have no entries
- Wireframe:



Select an option to view the data analysis results:

18. Create a query to find the average amount `retailPrice` of eBooks and physical books. Return ... ▼

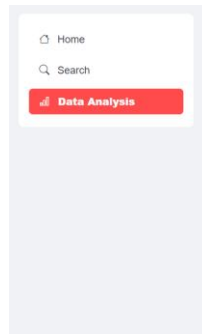
### Average Retail Price of eBooks and Physical Books

Average eBook Price: 11.50

Average Physical Book Price: 2.08

## BookScape Explorer Project Documentation

- To identify books that have an averageRating that is more than two standard deviations away from the average rating of all books. Return the title, averageRating, and ratingsCount for these outliers.
- Wireframe:



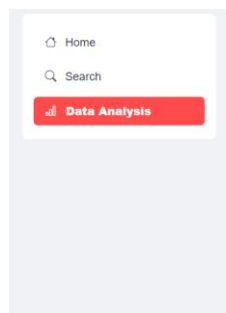
Select an option to view the data analysis results:

19.To identify books that have an averageRating that is more than two standard deviations awa...

### Books with Ratings More Than Two Standard Deviations Away from the Average

|   | book_title         | averageRating | ratingsCount |
|---|--------------------|---------------|--------------|
| 0 | Candida            | 4             | 224          |
| 1 | America (the Book) | 4             | 226          |

- Determines which publisher has the highest average rating among its books, but only for publishers that have published more than 10 books. Return the publisher, average\_rating, and the number of books published.
- Wireframe



Select an option to view the data analysis results:

20.Determines which publisher has the highest average rating among its books, but only for pu...

### Publisher with the Highest Average Rating

|   | publisher | average_rating | book_count |
|---|-----------|----------------|------------|
| 0 | Penguin   | 0.0426         | 188        |

## 10 Insights and Challenges

### 10.1.1 Insights

- **API Integration:** Integrating with the Google Books API was straightforward, but handling rate limits required implementing retry logic.
- **Data Storage:** Using SQLAlchemy for database interactions provided a robust and flexible way to handle database operations.
- **Data Visualization:** Plotly and Streamlit's built-in charting capabilities made it easy to visualize data.

### 10.1.2 Challenges

- **Design issue:** I faced some design issues while populating data from data frame to HTML.
- **Connection error:** I experienced Connection error issues when connecting to my database. "Connection error Is Streamlit still running?". This issue was resolved after installing sqlalchemy
- **Rate Limiting:** The Google Books API has rate limits, which required implementing retry logic to handle HTTP 429 errors.
- **Data Consistency:** Ensuring data consistency when inserting into the database was crucial, especially with varying data formats from the API.
- **Error Handling:** Comprehensive error handling was necessary to manage different types of exceptions, including IntegrityError and DataError.

## 11 Troubleshooting

### 11.1.1 Common Issues

- **Database Connection Error:** Ensure your MySQL server is running and the connection string is correct.
- **API Rate Limit Exceeded:** If you encounter a rate limit error, wait for a while before making new requests.