BookScape Explorer - Project Documentation

1. Project Overview

BookScape Explorer is a data-driven web application that fetches book data from the Google Books API, stores it in a MySQL database, and allows users to analyze and visualize the data through an interactive Streamlit web interface.

Objectives:

- Provide an intuitive way to search and explore books.

- Store book details in a structured database for analysis.

- Enable users to visualize key insights using charts and tables.

- Implement pagination to fetch more than 40 books per query.

---

2. Technical Implementation

2.1 Technology Stack

- Frontend: Streamlit (Python-based UI)

- Backend: Python (Flask for API calls, MySQL for data storage)

- Database: MySQL (Relational database to store book data)

- External API: Google Books API (Fetch book information)

2.2 API Integration

- Fetches book data dynamically using the Google Books API.

- Implements pagination to retrieve more than 40 books per request.

- Handles API errors, timeout issues, and quota limits.

2.3 Database Design

Table: `books`

Column Name	Data Type	Description
book_id	VARCHAR(255)	Unique identifier
search_key	VARCHAR(255)	Search term used
book_title	TEXT	Title of the book
book_authors	TEXT	List of authors
book_description	TEXT	Book summary
categories	TEXT	Book categories
language	VARCHAR(10)	Language code
averageRating	DECIMAL(3,2)	Avg rating
ratingsCount	INT	No. of reviews
pageCount	INT	Page count
publisher	TEXT	Publisher name
year	TEXT	Publication year

---

3. Challenges & Solutions

Challenge 1: API Request Limitations

- Issue: Google Books API limits results to 40 books per request.

- Solution: Implemented pagination using `startIndex` and multiple API calls.

Challenge 2: Handling Duplicate Entries

- Issue: Repeated searches could insert duplicate books.

- Solution: Used `ON DUPLICATE KEY UPDATE` in MySQL queries.

Challenge 3: Data Inconsistency

- Issue: Some books lacked key details like authors or prices.

- Solution: Used default values (e.g., "Unknown") for missing fields.

---

4. Insights & Future Enhancements

Current Insights:

- Majority of books are eBooks over physical copies.

- Certain publishers dominate highly rated books.

- Most books fall within specific price ranges.

Future Enhancements:

- Implement user authentication for saved searches.

- Allow custom filtering options (e.g., price range, genre).

- Deploy on AWS/GCP for better scalability.

- Enable export to CSV for analytics.

---

5. README for GitHub

Project Name: BookScape Explorer

Installation & Setup

```bash
# Clone the repository

git clone https://github.com/yourusername/bookscape-explorer.git

cd bookscape-explorer

# Create and activate virtual environment

python -m venv venv

source venv/bin/activate  # On Windows: venv\Scripts ctivate

# Install dependencies

pip install -r requirements.txt
```

```
CREATE DATABASE bookscape;

USE bookscape;
```

```
streamlit run app.py
```