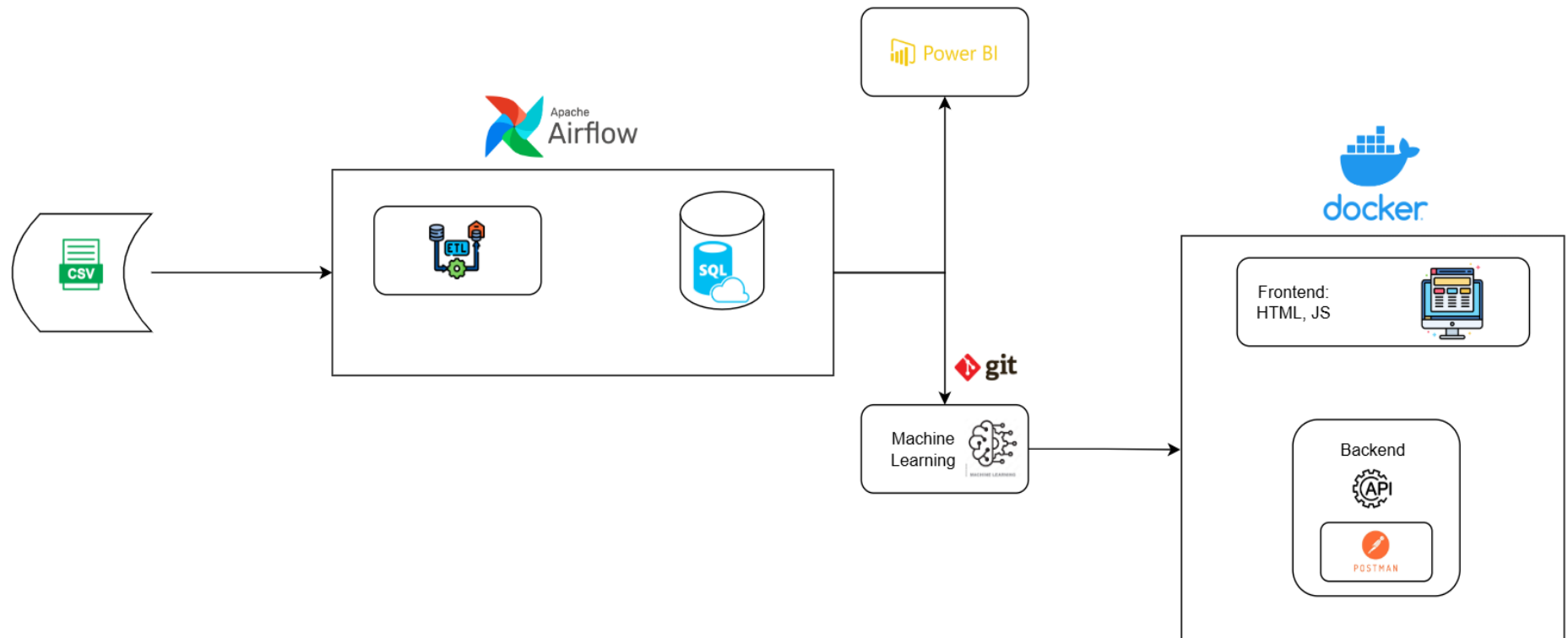
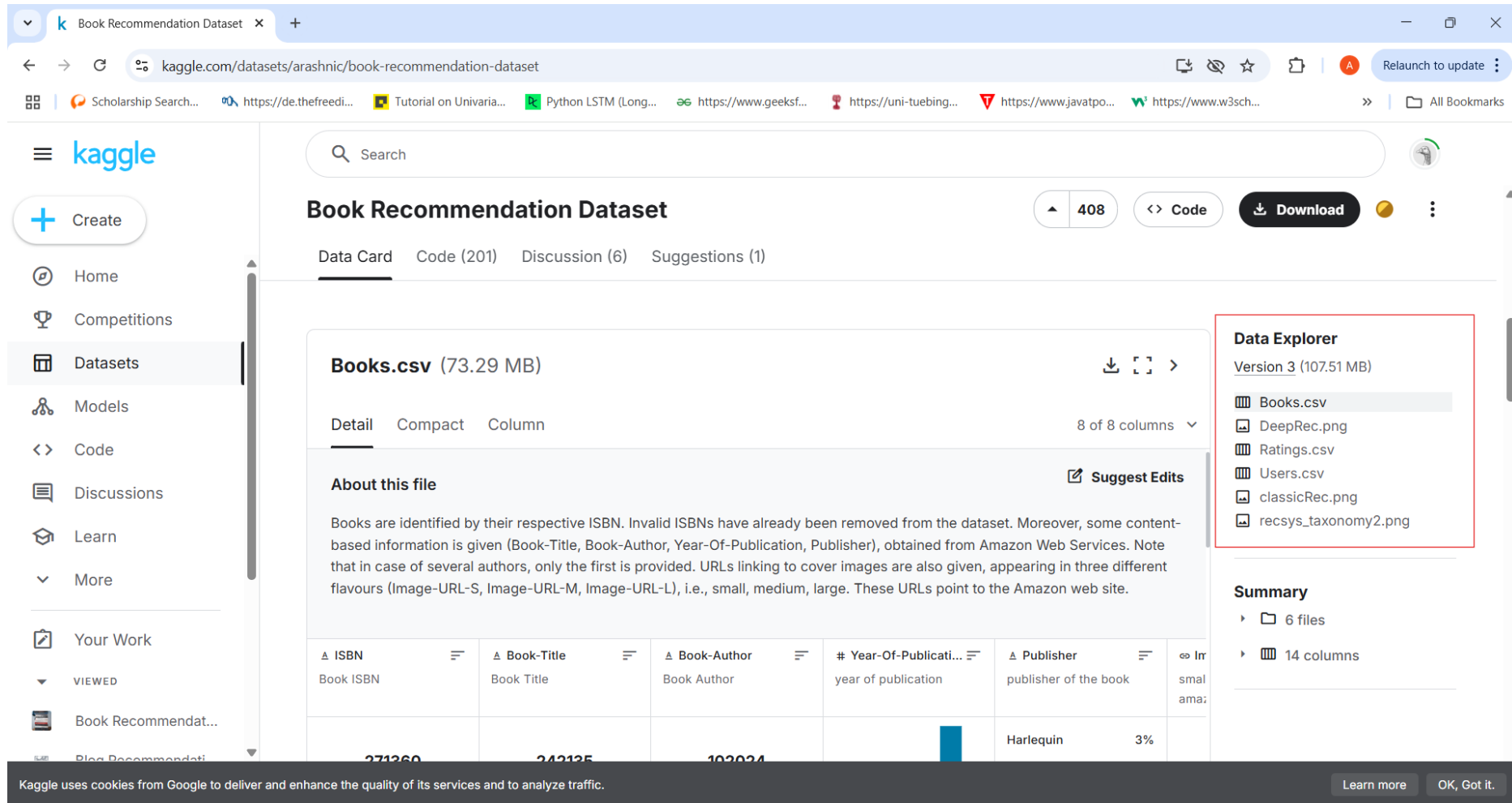


Data Engineering and Data Science Full End-to-End Project Structure



1) Getting data source



The screenshot shows the Kaggle interface for the 'Book Recommendation Dataset'. The main content area displays the 'Books.csv' file (73.29 MB) with a 'Detail' tab selected. The 'About this file' section provides information about the dataset, including ISBNs, book titles, authors, and publication years. The 'Data Explorer' sidebar on the right lists the files in the dataset: Books.csv, DeepRec.png, Ratings.csv, Users.csv, classicRec.png, and recsys_taxonomy2.png. The 'Summary' section shows 6 files and 14 columns.

Books.csv (73.29 MB)

Detail Compact Column 8 of 8 columns

About this file [Suggest Edits](#)

Books are identified by their respective ISBN. Invalid ISBNs have already been removed from the dataset. Moreover, some content-based information is given (Book-Title, Book-Author, Year-Of-Publication, Publisher), obtained from Amazon Web Services. Note that in case of several authors, only the first is provided. URLs linking to cover images are also given, appearing in three different flavours (Image-URL-S, Image-URL-M, Image-URL-L), i.e., small, medium, large. These URLs point to the Amazon web site.

ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S	Image-URL-M	Image-URL-L
031260	242125	102024		Harlequin	3%		

Data Explorer
Version 3 (107.51 MB)

- Books.csv
- DeepRec.png
- Ratings.csv
- Users.csv
- classicRec.png
- recsys_taxonomy2.png

Summary

- 6 files
- 14 columns

Kaggle uses cookies from Google to deliver and enhance the quality of its services and to analyze traffic. [Learn more](#) [OK, Got it.](#)

2) Dataset location in our project

The screenshot displays the Visual Studio Code (VS Code) interface. On the left, the Explorer sidebar shows a project named 'BOOK_RECOMMENDATION'. The project structure includes a 'dataset' folder, a 'model' folder, and a 'server' folder. The 'dataset' folder is expanded, showing three files: 'Books.csv', 'Ratings.csv', and 'Users.csv'. These three files are highlighted with a red rectangular box. Below the Explorer sidebar, the 'OUTLINE' and 'TIMELINE' panels are visible. The main editor area shows the 'Books.csv' file, which contains a list of books with their ISBNs, titles, authors, and publication years. The file is open in the 'book_recommendation' workspace. The bottom status bar shows the current file is 'main*' and the editor is in 'UTF-8' encoding.

```
model > airflow > dags > dataset > Books.csv
1  ISBN,Book-Title,Book-Author,Year-Of-Publication,Publisher,Image-URL-S,Image-URL-M,Image-URL-L
2  0195153448,Classical Mythology,Mark P. O. Morford,2002,Oxford University Press,http://images.amazon.com/images/P/
3  0002005018,Clara Callan,Richard Bruce Wright,2001,HarperFlamingo Canada,http://images.amazon.com/images/P/0002005
4  0060973129,Decision in Normandy,Carlo D'Este,1991,HarperPerennial,http://images.amazon.com/images/P/0060973129.01
5  0374157065,Flu: The Story of the Great Influenza Pandemic of 1918 and the Search for the Virus That Caused It,Gin
6  0393045218,The Mummies of Urumchi,E. J. W. Barber,1999,W. W. Norton & Company,http://images.amazon.com/images
7  0399135782,The Kitchen God's Wife,Amy Tan,1991,Putnam Pub Group,http://images.amazon.com/images/P/0399135782.01.T
8  0425176428,What If?: The World's Foremost Military Historians Imagine What Might Have Been,Robert Cowley,2000,Ber
9  0671870432,PLEADING GUILTY,Scott Turow,1993,Audioworks,http://images.amazon.com/images/P/0671870432.01.THUMBZZZ.j
10 0679425608,Under the Black Flag: The Romance and the Reality of Life Among the Pirates,David Cordingly,1996,Rando
11 074322678X,Where You'll Find Me: And Other Stories,Ann Beattie,2002,Scribner,http://images.amazon.com/images/P/07
12 0771074670,Nights Below Station Street,David Adams Richards,1988,Emblem Editions,http://images.amazon.com/images/
13 080652121X,Hitler's Secret Bankers: The Myth of Swiss Neutrality During the Holocaust,Adam Lebor,2000,Citadel Pre
14 0887841740,The Middle Stories,Sheila Heti,2004,House of Anansi Press,http://images.amazon.com/images/P/0887841740
15 1552041778,Jane Doe,R. J. Kaiser,1999,Mira Books,http://images.amazon.com/images/P/1552041778.01.THUMBZZZ.jpg,htt
16 1558746218,A Second Chicken Soup for the Woman's Soul (Chicken Soup for the Soul Series),Jack Canfield,1998,Healt
17 1567407781,The Witchfinder (Amos Walker Mystery Series),Loren D. Estleman,1998,Brilliance Audio - Trade,http://im
18 1575663937,More Cunning Than Man: A Social History of Rats and Man,Robert Hendrickson,1999,Kensington Publishing
19 1881320189,Goodbye to the Buttermilk Sky,Julia Oliver,1994,River City Pub,http://images.amazon.com/images/P/18813
20 0440234743,The Testament,John Grisham,1999,Dell,http://images.amazon.com/images/P/0440234743.01.THUMBZZZ.jpg,http
21 0452264464,Beloved (Plume Contemporary Fiction),Toni Morrison,1994,Plume,http://images.amazon.com/images/P/045226
22 0609804618,Our Dumb Century: The Onion Presents 100 Years of Headlines from America's Finest News Source,The Onio
23 1841721522,New Vegetarian: Bold and Beautiful Recipes for Every Occasion,Celia Brooks Brown,2001,Ryland Peters &
24 1879384493,If I'd Known Then What I Know Now: Why Not Learn from the Mistakes of Others? : You Can't Afford to Ma
25 0061076034,Mary Kate & Ashley's Catching Santa (Mary Kate and Ashley's Christmas)
```

3) Initiating Apache Airflow locally

The screenshot displays the Visual Studio Code interface with a project named 'book_recommendation'. The Explorer sidebar on the left shows the project structure, including a 'model' directory with an 'airflow' subdirectory. The 'airflow' directory contains files like 'config', ' dags', ' __pycache__', ' dataset', ' Books.csv', ' Ratings.csv', ' Users.csv', ' create_sql_tables.py', ' etl_books.py', ' etl_ratings.py', ' etl_users.py', ' test.py', ' users_data_cleaning.py', ' logs', and ' plugins'. The 'docker-compose.yml' file is selected and open in the editor.

The 'docker-compose.yml' file content is as follows:

```
model > airflow > docker-compose.yml
47 x-airflow-common:
54   environment:
75   volumes:
76     - ${AIRFLOW_PROJ_DIR:-.}/dags:/opt/airflow/dags
77     - ${AIRFLOW_PROJ_DIR:-.}/logs:/opt/airflow/logs
78     - ${AIRFLOW_PROJ_DIR:-.}/config:/opt/airflow/config
79     - ${AIRFLOW_PROJ_DIR:-.}/plugins:/opt/airflow/plugins
80   user: "${AIRFLOW_UID:-50000}:0"
81   depends_on:
82     &airflow-common-depends-on
83     redis:
84       condition: service_healthy
85     postgres:
86       condition: service_healthy
87
88 services:|
89   postgres:
90     image: postgres:13
91     environment:
92       POSTGRES_USER: airflow
93       POSTGRES_PASSWORD: airflow
94       POSTGRES_DB: airflow
95     volumes:
96       - postgres-db-volume:/var/lib/postgresql/data
97   ports:
```

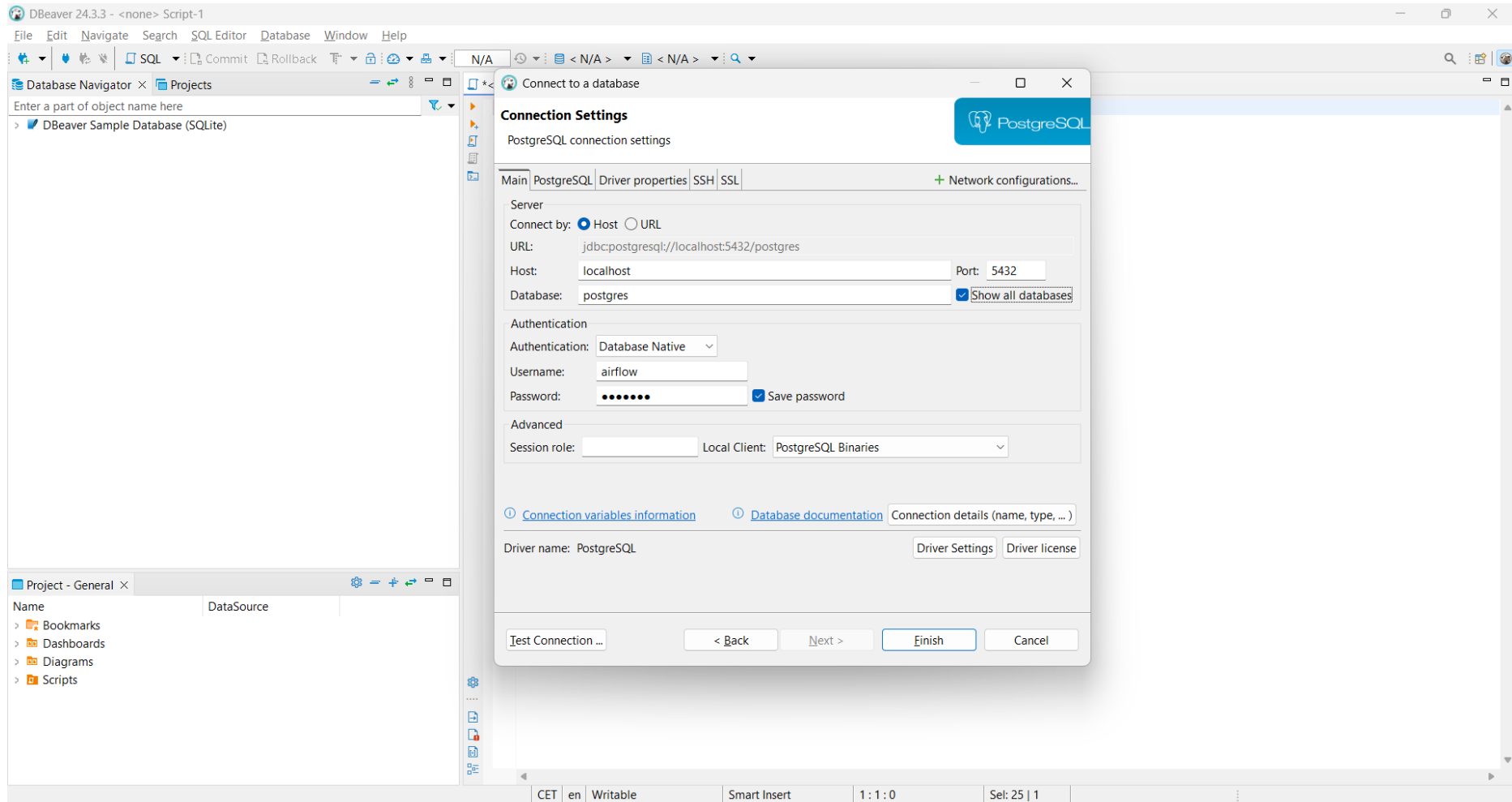
The terminal window at the bottom shows the command 'docker compose up' being executed. The output indicates that the containers were created successfully, with some warnings about the 'AIRFLOW_UID' variable not being set. The output is as follows:

```
C:\_Local Disk D\My_projects\recommendation_books_final\book_recommendation\model\airflow>docker compose up
time="2025-03-31T01:24:26+02:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
time="2025-03-31T01:24:26+02:00" level=warning msg="The \"AIRFLOW_UID\" variable is not set. Defaulting to a blank string."
[+] Running 7/0
✔ Container airflow-redis-1 Created 0.0s
✔ Container airflow-postgres-1 Created 0.0s
✔ Container airflow-airflow-init-1 Created 0.0s
✔ Container airflow-airflow-scheduler-1 Created 0.0s
✔ Container airflow-airflow-triggerer-1 Created 0.0s
✔ Container airflow-airflow-worker-1 Created 0.0s
```

4) Apache Airflow Interface

Version: v2.10.5
Git Version: .release:b93c3db6b1641b0840bd15ac7d05bc58ff2cccbf

5) Setting up SQL Server



6) Connecting Airflow to SQL Server

Book Recommendation Dataset x Add Connection - Airflow x +

localhost:8080/connection/add

Scholarship Search... https://de.thefreedi... Tutorial on Univaria... Python LSTM (Long... https://www.geeksf... https://uni-tuebing... https://www.javatpo... https://www.w3sch...

Airflow DAGs Cluster Activity Datasets Security Browse Admin Docs

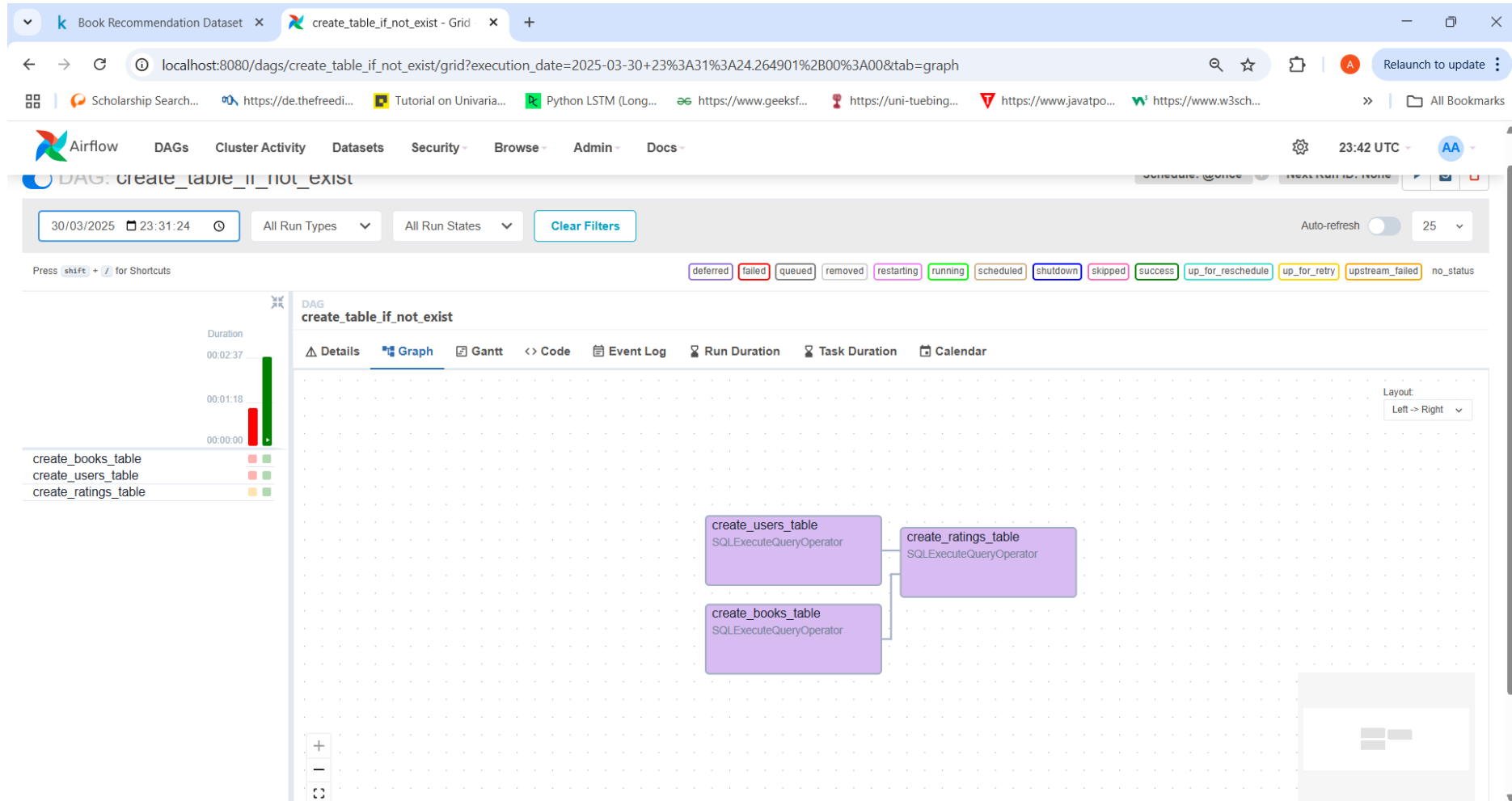
23:33 UTC AA

Warning: Fields that are currently populated can be modified but cannot be deleted. To delete data from a field, delete the Connection object and create a new one.

Add Connection

Connection Id *	postgres_localhost
Connection Type *	Postgres <small>Connection Type missing? Make sure you've installed the corresponding Airflow Provider Package.</small>
Description	
Host	host.docker.internal
Database	postgres
Login	airflow
Password	*****
Port	5432

7) Creating Tables on SQL using Apache Airflow DAGs



8) ETL Process

Book Recommendation Dataset x DAGs - Airflow x +

localhost:8080/home

Scholarship Search... https://de.thefreedi... Tutorial on Univaria... Python LSTM (Long... https://www.geeksf... https://uni-tuebing... https://www.javatpo... https://www.w3sch...

Airflow DAGs Cluster Activity Datasets Security Browse Admin Docs

23:42 UTC AA

DAGs

All 4 Active 4 Paused 0 Running 0 Failed 1 Filter DAGs by tag Search DAGs Auto-refresh

DAG	Owner	Runs	Schedule	Last Run	Next Run	Recent Tasks	Actions	Links
create_table_if_not_exist	azizjon	1 1	@once	2025-03-30, 23:31:24		3		...
import_books_to_database	azizjon	1	0 0 * * *	2025-03-30, 23:34:30	2025-03-31, 00:00:00	1		...
import_ratings_to_database	azizjon	1	0 0 * * *	2025-03-30, 23:41:22	2025-03-31, 00:00:00	1		...
import_users_to_database	azizjon	1	0 0 * * *	2025-03-30, 23:35:10	2025-03-31, 00:00:00	1		...

« < 1 > »

Showing 1-4 of 4 DAGs

Version: v2.10.5
Git Version: .release:b93c3db6b1641b0840bd15ac7d05bc58ff2cccbf

9) Running SQL commands to check for

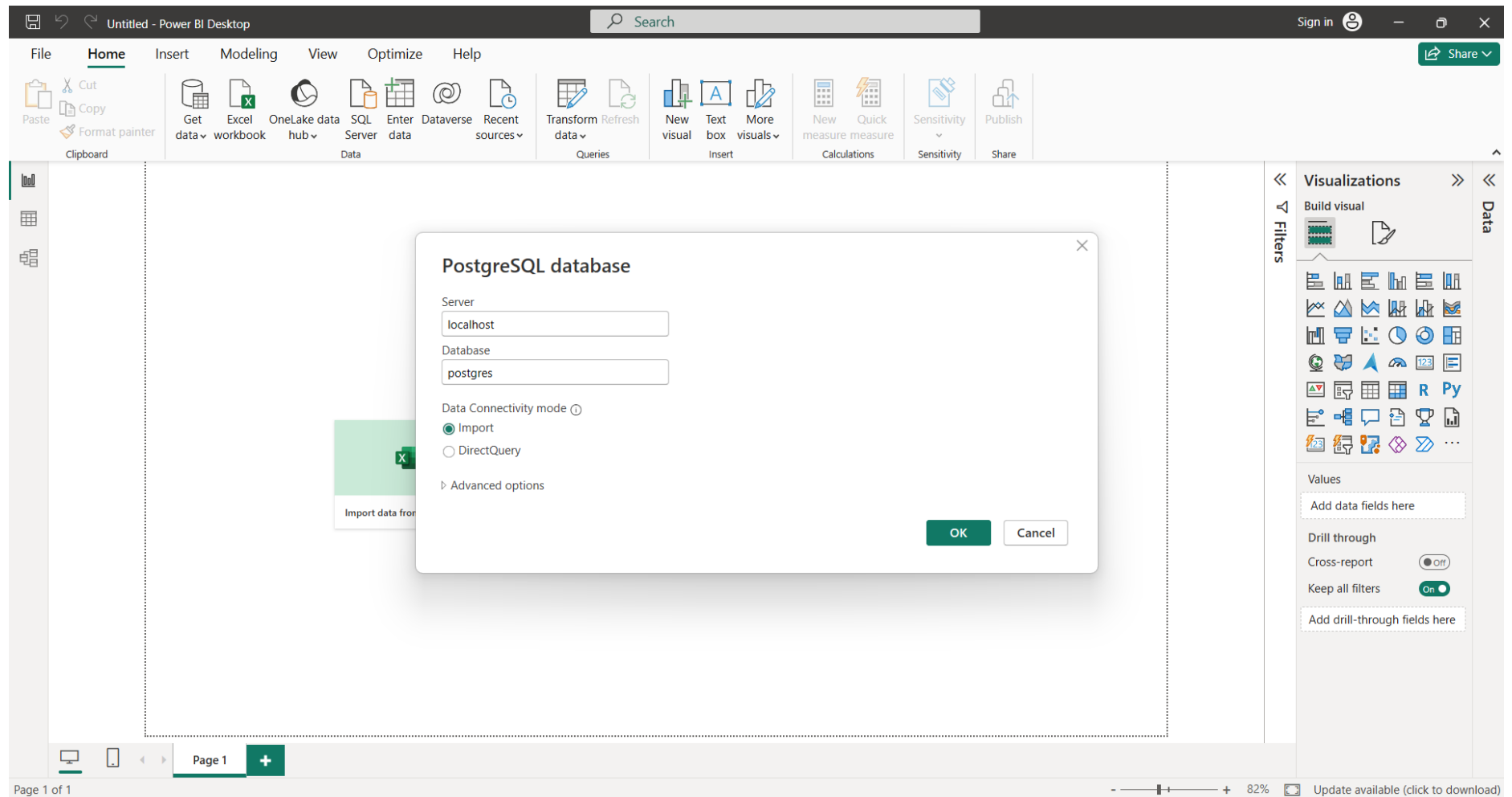
The screenshot displays the DBeaver 24.3.3 interface with a PostgreSQL database connection. The left sidebar shows the database structure, including the 'public' schema and the 'books' table. The main editor window contains a SQL query that selects book titles, authors, age groups, and average ratings, grouped by book title and author, and ordered by age group and average rating.

```
SELECT
    b.book_title,
    b.book_author,
    CASE
        WHEN u.age < 20 THEN 'Teens'
        WHEN u.age < 30 THEN '20s'
        WHEN u.age < 40 THEN '30s'
        WHEN u.age < 50 THEN '40s'
        ELSE '50+'
    END AS age_group,
    COUNT(*) AS rating_count,
    ROUND(AVG(r.book_rating), 2) AS avg_rating
FROM ratings r
JOIN books b ON r.ISBN = b.ISBN
JOIN users u ON r.user_id = u.user_id
GROUP BY b.book_title, b.book_author, age_group
HAVING COUNT(*) >= 50
ORDER BY age_group, avg_rating DESC;
```

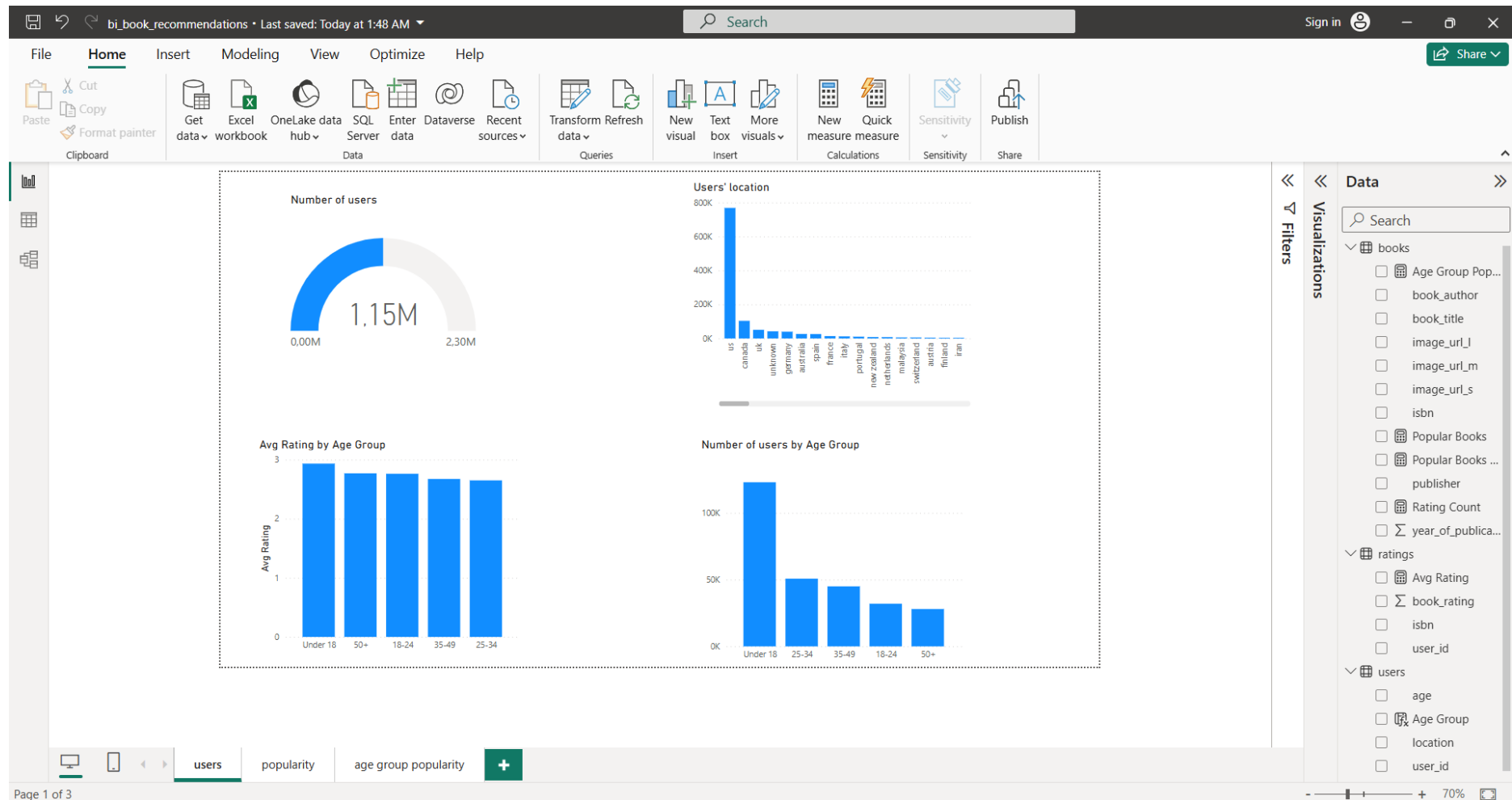
The results are displayed in a table with 13 rows and 6 columns. The first row is highlighted, showing the book 'Harry Potter and the Sorcerer's Stone' by J. K. Rowling, categorized as '20s', with a rating count of 123 and an average rating of 7.06.

	book_title	book_author	age_group	rating_count	avg_rating
1	Harry Potter and the Sorcerer's Stone	J. K. Rowling	20s	123	7.06
2	Ender's Game (Ender Wiggins Saga (Pa	Orson Scott Card	20s	54	6.69
3	Harry Potter and the Goblet of Fire (Bo	J. K. Rowling	20s	100	6.44
4	Harry Potter and the Prisoner of Azkaba	J. K. Rowling	20s	106	6.33
5	Harry Potter and the Order of the Phoe	J. K. Rowling	20s	90	6.31
6	The Two Towers (The Lord of the Rings	J.R.R. TOLKIEN	20s	55	5.89
7	The Five People You Meet in Heaven	Mitch Albom	20s	71	5.46
8	Harry Potter and the Chamber of Secre	J. K. Rowling	20s	136	5.35
9	The Fellowship of the Ring (The Lord o	J.R.R. TOLKIEN	20s	73	5.33
10	Harry Potter and the Sorcerer's Stone (J. K. Rowling	20s	143	5.23
11	The Catcher in the Rye	J.D. Salinger	20s	125	5.14
12	The Gunslinger (The Dark Tower, Book	Stephen King	20s	62	5.13
13	To Kill a Mockingbird	Harper Lee	20s	111	4.94

10) Connecting PowerBI to SQL Server



11) SQL Database usage on PowerBI



12) ML Model Backend – Frontend: recommendation.ipynb file must be executed to get server artifacts

The screenshot shows a VS Code workspace for a project named 'book_recommendation'. The Explorer sidebar on the left shows the project structure, with the 'airflow' folder and its contents highlighted. The main editor displays the 'Recommendation system' notebook, which contains a cell with the code `data.head()`. The output of this cell is a table of book data. The bottom terminal panel shows the status of various containers, all marked as 'Stopped'.

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0... http://images.amazon.com/images:
2	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0... http://images.amazon.com/images:
3	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0... http://images.amazon.com/images:

Terminal output:

```
[+] Stopping 7/7
✓ Container airflow-airflow-scheduler-1 Stopped 17.0s
✓ Container airflow-airflow-webserver-1 Stopped 17.9s
✓ Container airflow-airflow-triggerer-1 Stopped 17.2s
✓ Container airflow-airflow-worker-1 Stopped 17.4s
✓ Container airflow-airflow-init-1 Stopped 0.0s
✓ Container airflow-redis-1 Stopped 1.7s
✓ Container airflow-postgres-1 Stopped 2.4s
canceled
```

13) Deploying ML Model using Docker

The screenshot displays the Visual Studio Code interface with a Jupyter Notebook titled "Recommendation system". The Explorer sidebar on the left shows a project structure for "BOOK_RECOMMENDATION" with folders for dataset, model, airflow, server, and temp. The airflow folder is highlighted, containing a file named "recommendation.ipynb".

The Jupyter Notebook content shows a code cell with the command `data.head()` executed, resulting in a table of book data:

	ISBN	Book-Title	Book-Author	Year-Of-Publication	Publisher	Image-URL-S
1	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0... http://images.amazon.com/images:
2	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0... http://images.amazon.com/images:
3	0002005018	Clara Callan	Richard Bruce Wright	2001	HarperFlamingo Canada	http://images.amazon.com/images/P/0002005018.0... http://images.amazon.com/images:

The Terminal at the bottom shows the output of the `airflow` command, indicating that the containers are being stopped:

```
[+] Stopping 7/7
✓ Container airflow-airflow-scheduler-1 Stopped 17.0s
✓ Container airflow-airflow-webserver-1 Stopped 17.9s
✓ Container airflow-airflow-triggerer-1 Stopped 17.2s
✓ Container airflow-airflow-worker-1 Stopped 17.4s
✓ Container airflow-airflow-init-1 Stopped 0.0s
✓ Container airflow-redis-1 Stopped 1.7s
✓ Container airflow-postgres-1 Stopped 2.4s
canceled
```

14) Postman to check for API backend

The screenshot displays the Postman web interface. At the top, there's a navigation bar with 'Home', 'Workspaces', and 'Explore' links, along with a search bar and 'Sign In'/'Create Account' buttons. A blue banner below the navigation bar states: 'You are using the Lightweight API Client, sign in or create an account to work with collections, environments and unlock all free features in Postman.'

The main interface is divided into three sections:

- History:** A list of recent requests on the left. It includes several POST requests to `http://localhost:8000/recommend` and `http://localhost:8000/high-rated`, and two GET requests to `http://localhost:8000/existing-users` and `http://localhost:8000/`.
- Request Editor:** The central area where a new POST request is configured. The URL is `http://localhost:8000/recommend`. The 'Body' tab is selected, showing a JSON payload:

```
1 {
2   "num_recommendations": 5,
3   "user_id": 8
4 }
5
```
- Response View:** The bottom section shows the response from the server. The status is '200 OK' with a time of '25 ms' and a size of '1.13 KB'. The response is displayed in 'Pretty' JSON format:

```
1 {
2   "Book-Title": "The Progress of Love (King Penguin)",
3   "Book-Author": "Alice Munro",
4   "Year-Of-Publication": 1987,
5   "Publisher": "Penguin Books",
6   "Image-URL-L": "http://images.amazon.com/images/P/0140098798.01.LZZZZZZZ.jpg"
7 },
8 {
9   "Book-Title": "To Kill a Mockingbird",
10  "Book-Author": "Harper Lee",
11  "Year-Of-Publication": 1988,
12  "Publisher": "Little Brown & Company",
13 }
```

At the bottom left, there's a 'Create collections in Postman' prompt with a 'Create a Collection' button.

15) Simple Web UI

Book Recommendation System x FastAPI - Swagger UI x +

127.0.0.1:8000/docs/#

Scholarship Search... https://de.thefreedi... Tutorial on Univaria... Python LSTM (Long... https://www.geeksf... https://uni-tuebing... https://www.javatpo... https://www.w3sch...

Alle Lesezeichen

FastAPI 0.1.0 OAS 3.1

/openapi.json

default

- GET** / Home
- POST** /high-rated High Rated
- POST** /low-rated Low Rated
- GET** /existing-users Existing Users
- POST** /recommend Recommend Books

Schemas

16) Webpage using HTML and JS

Book Recommendation System

←

→

↻

📄

Date

c%3A/Users/AZTURAEV/main_folder/selbst/recommendation_books/server/index.html

☆

📁

📄

⌵

⋮

Scholarship Search...https://de.thefreedi...Tutorial on Univaria...Python LSTM (Long...https://www.geeksf...https://uni-tuebing...https://www.javatpo...https://www.w3sch...

»Alle Lesezeichen

Book Recommendation System

Low-Rated Books

High-Rated Books

Existing Users

Get Recommendations

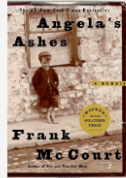
Number of Books:

57

User ID:

99

Submit

Book Title	Author	Year	Publisher	Image
ANGELA'S ASHES	Frank McCourt	1996	Scribner	
The Fall of Hyperion	Dan Simmons	1991	Bantam	