

Name: Youmna Ihab EL-Awady (2203166)

## Assignment 2 | Dockerfile and Data Analysis with Popular Books Dataset

### Requirements:

#### 1. Docker Setup:

Download and pull the jupyter docker image.

```
Microsoft Windows [Version 10.0.19045.4291]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Dell>docker images

REPOSITORY              TAG         IMAGE ID      CREATED        SIZE
popular-books-assignment latest      e1ade3a0ae1a  14 hours ago  5.92GB
jupyter/datascience-notebook latest      f78a42f3bc9a  6 months ago  5.92GB
hello-world             latest      d2c94e258dcb  11 months ago 13.3kB
```

Name	Tag	Status	Created	Size	Actions
<a href="#">popular-books-assignment</a>	latest	<a href="#">In use</a>	3 minutes ago	5.92 GB	<a href="#">▶</a> <a href="#">⋮</a> <a href="#">🗑</a>
<a href="#">jupyter/datascience-notebook</a>	latest	<a href="#">In use</a>	6 months ago	5.92 GB	<a href="#">▶</a> <a href="#">⋮</a> <a href="#">🗑</a>
<a href="#">hello-world</a>	latest	<a href="#">In use</a>	1 year ago	13.25 KB	<a href="#">▶</a> <a href="#">⋮</a> <a href="#">🗑</a>

#### 2. Create dockerfile:

Dockerfile created and *image is build*

The screenshot shows the VS Code interface with the Dockerfile editor and the terminal output of the Docker build command. The Dockerfile content is as follows:

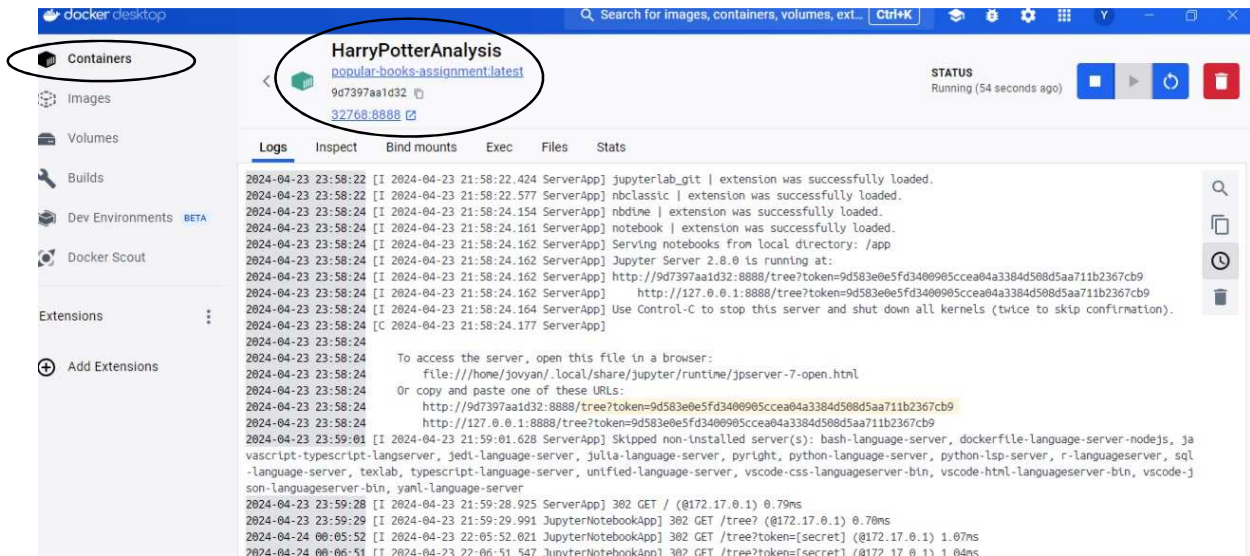
```
Dockerfile > ...
1 FROM jupyter/datascience-notebook
2 WORKDIR /app
3 COPY Books.csv /app
4 COPY Booksnotebook.ipynb /app
5 EXPOSE 8888
6 CMD ["jupyter", "notebook", "--ip='0.0.0.0'", "--port=8888", "--no-browser", "--allow-root"]
```

The terminal output shows the successful execution of the Docker build command:

```
PS E:\College Youmna\Term 4\Cloud Computing\Popular Books Assignment> docker build -t booksnotebook .
[+] Building 1.1s (9/9) FINISHED
=> [internal] load build definition from Dockerfile
=> => transferring dockerfile: 240B
=> [internal] load metadata for docker.io/jupyter/datascience-notebook:latest
=> [internal] load .dockerignore
=> => transferring context: 2B
=> [1/4] FROM docker.io/jupyter/datascience-notebook:latest
=> [internal] load build context
=> => transferring context: 71B
=> CACHED [2/4] WORKDIR /app
=> CACHED [3/4] COPY Books.csv /app
=> CACHED [4/4] COPY Booksnotebook.ipynb /app
=> exporting to image
=> => exporting layers
=> writing image sha256:e1ade3a0ae1aa6e653bc79d55e4f51d36b4b1a5b4317135fe8ba5a6f7332c99e
=> naming to docker.io/library/booksnotebook

View build details: docker-desktop://dashboard/build/default/default/rkm0b6oa39dhxyjd19filmnvi
What's Next?
```

## Container launched

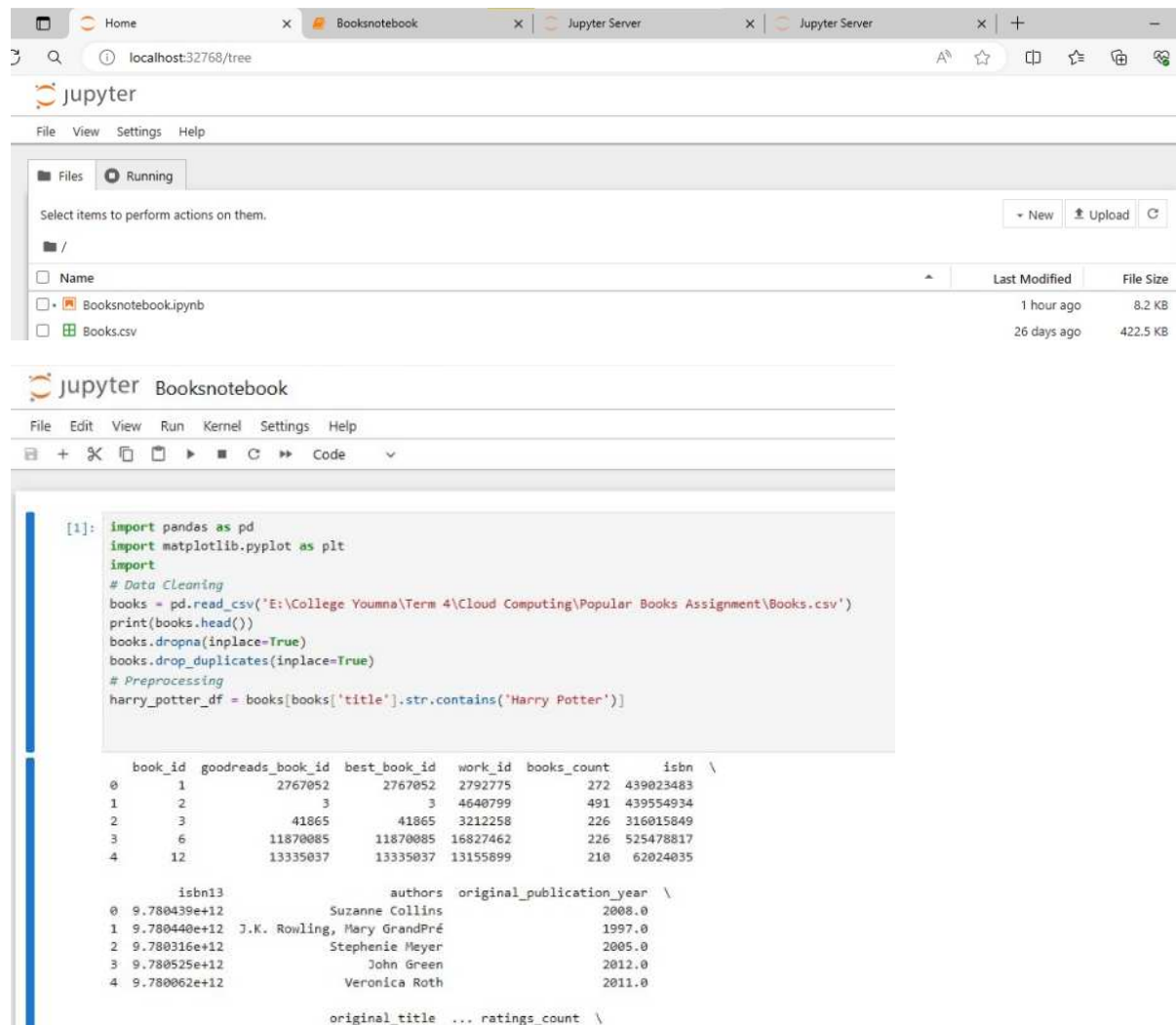


The screenshot shows the Docker Desktop interface. On the left sidebar, the 'Containers' tab is selected. The main area displays a container named 'HarryPotterAnalysis' with ID '9d7397aa1d32'. The container is running. The 'Logs' tab is active, showing the following log entries:

```
2024-04-23 23:58:22 [I 2024-04-23 21:58:22.424 ServerApp] jupyterlab_glt | extension was successfully loaded.
2024-04-23 23:58:22 [I 2024-04-23 21:58:22.577 ServerApp] nbclassic | extension was successfully loaded.
2024-04-23 23:58:24 [I 2024-04-23 21:58:24.154 ServerApp] nbdtline | extension was successfully loaded.
2024-04-23 23:58:24 [I 2024-04-23 21:58:24.161 ServerApp] notebook | extension was successfully loaded.
2024-04-23 23:58:24 [I 2024-04-23 21:58:24.162 ServerApp] Serving notebooks from local directory: /app
2024-04-23 23:58:24 [I 2024-04-23 21:58:24.162 ServerApp] Jupyter Server 2.8.0 is running at:
2024-04-23 23:58:24 [I 2024-04-23 21:58:24.162 ServerApp] http://9d7397aa1d32:8888/tree?token=9d583e0e5fd3400905ccea04a3384d508d5aa711b2367cb9
2024-04-23 23:58:24 [I 2024-04-23 21:58:24.162 ServerApp] http://127.0.0.1:8888/tree?token=9d583e0e5fd3400905ccea04a3384d508d5aa711b2367cb9
2024-04-23 23:58:24 [I 2024-04-23 21:58:24.164 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
2024-04-23 23:58:24 [C 2024-04-23 21:58:24.177 ServerApp]

To access the server, open this file in a browser:
file:///home/jovyan/.local/share/jupyter/runtime/jpserver-7-open.html
Or copy and paste one of these URLs:
http://9d7397aa1d32:8888/tree?token=9d583e0e5fd3400905ccea04a3384d508d5aa711b2367cb9
http://127.0.0.1:8888/tree?token=9d583e0e5fd3400905ccea04a3384d508d5aa711b2367cb9
2024-04-23 23:59:01 [I 2024-04-23 21:59:01.628 ServerApp] Skipped non-installed server(s): bash-language-server, dockerfile-language-server-nodejs, ja
vascript-typescript-languageserver, jedi-language-server, julia-language-server, pyright, python-language-server, python-lsp-server, r-languageserver, sql
-language-server, texlab, typescript-language-server, unified-language-server, vscode-css-languageserver-bin, vscode-html-languageserver-bin, vscode-j
son-languageserver-bin, yamll-language-server
2024-04-23 23:59:28 [I 2024-04-23 21:59:28.925 ServerApp] 302 GET / (@172.17.0.1) 0.79ms
2024-04-23 23:59:29 [I 2024-04-23 21:59:29.991 JupyterNotebookApp] 302 GET /tree? (@172.17.0.1) 0.70ms
2024-04-24 00:05:52 [I 2024-04-23 22:05:52.021 JupyterNotebookApp] 302 GET /tree?token=[secret] (@172.17.0.1) 1.07ms
2024-04-24 00:06:51 [I 2024-04-23 22:06:51.547 JupyterNotebookApp] 302 GET /tree?token=[secret] (@172.17.0.1) 1.04ms
```

## Configure the Dockerfile to run jupyter Notebook when the container launches



The screenshot shows a Jupyter Notebook interface. The file explorer on the left shows a file named 'Booksnotebook.ipynb'. The main area displays the following Python code:

```
[1]: import pandas as pd
import matplotlib.pyplot as plt
import
# Data Cleaning
books = pd.read_csv('E:\College Youmna\Term 4\Cloud Computing\Popular Books Assignment\Books.csv')
print(books.head())
books.dropna(inplace=True)
books.drop_duplicates(inplace=True)
# Preprocessing
harry_potter_df = books[books['title'].str.contains('Harry Potter')]
```

The output of the code is displayed below:

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn
0	1	2767052	2767052	2792775	272	439023483
1	2	3	4640799	491	439554934	
2	3	41865	41865	3212258	226	316015049
3	6	11870085	11870085	16827462	226	525478817
4	12	13335037	13335037	13155899	210	62024035

	isbn13	authors	original_publication_year
0	9.780439e+12	Suzanne Collins	2008.0
1	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0
2	9.780316e+12	Stephenie Meyer	2005.0
3	9.780525e+12	John Green	2012.0
4	9.780062e+12	Veronica Roth	2011.0

	original_title	ratings_count
--	----------------	---------------

### 3. Notebook Development:

```
import pandas as pd
import matplotlib.pyplot as plt
# Data Cleaning
books = pd.read_csv('E:\College Youmna\Term 4\Cloud Computing\Popular Books Assignment\Books.csv')
print(books.head())
books.dropna(inplace=True)
books.drop_duplicates(inplace=True)
# Preprocessing
harry_potter_df = books[books['title'].str.contains('Harry Potter')]
```

Python

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn \
0	1	2767052	2767052	2792775	272	439023483
1	2	3	3	4640799	491	439554934
2	3	41865	41865	3212258	226	316015849
3	6	11870085	11870085	16827462	226	525478817
4	12	13335037	13335037	13155899	210	62024035

	isbn13	authors	original_publication_year \
0	9.780439e+12	Suzanne Collins	2008.0
1	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0
2	9.780316e+12	Stephenie Meyer	2005.0
3	9.780525e+12	John Green	2012.0
4	9.780062e+12	Veronica Roth	2011.0

	original_title	... ratings_count \
0	The Hunger Games	4780653

Activa

```
# Analyzing
most_selling_books = harry_potter_df.sort_values(by='books_count', ascending=False).head()
# Calculate Average Rating
average_rating = harry_potter_df['average_rating'].mean()

# Display the results
print("Most Selling Harry Potter Books:", most_selling_books)
print("\nAverage Rating of Harry Potter Books:", average_rating)
```

Python

```
... Most Selling Harry Potter Books:
book_id  goodreads_book_id  best_book_id  work_id  books_count \
1         2                 3              3  4640799         491
9         23                15881         15881  6231171         398
6         18                 5              5  2402163         376
10        24                 6              6  3046572         332
8         21                 2              2  2809203         307
```

	isbn	isbn13	authors \
1	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré
9	439064864	9.780439e+12	J.K. Rowling, Mary GrandPré
6	043965548X	9.780440e+12	J.K. Rowling, Mary GrandPré, Rufus Beck
10	439139600	9.780439e+12	J.K. Rowling, Mary GrandPré
8	439358078	9.780439e+12	J.K. Rowling, Mary GrandPré

	original_publication_year	original_title	... \
1	1997.0	Harry Potter and the Philosopher's Stone	...
9	1998.0	Harry Potter and the Chamber of Secrets	...
6	1999.0	Harry Potter and the Prisoner of Azkaban	...
10	2000.0	Harry Potter and the Goblet of Fire	...
8	2003.0	Harry Potter and the Order of the Phoenix	...

Activate  
Go to Setti

```
8          2003.0 Harry Potter and the Order of the Phoenix ...
```

	ratings_count	work_ratings_count	work_text_reviews_count	ratings_1	\
1	4602479	4800065	75867	75504	
9	1779331	1906199	34172	8253	
6	1832823	1969375	36099	6716	

```
...
```

```
[5 rows x 23 columns]
```

Average Rating of Harry Potter Books: 4.4910000000000005

Output is truncated. View as a [scrollable element](#) or open in a [text editor](#). Adjust cell output [settings...](#)