

Assignment 2

Cloud computing

Name :Youstena Malak Tadros.



ID: 2206199.

Department of cyper security(level 2).

Steps:

- 1) I create a directory called dockerimages, then I download the data set and store it in this file.

his PC > Local Disk (C:) > dockerimages

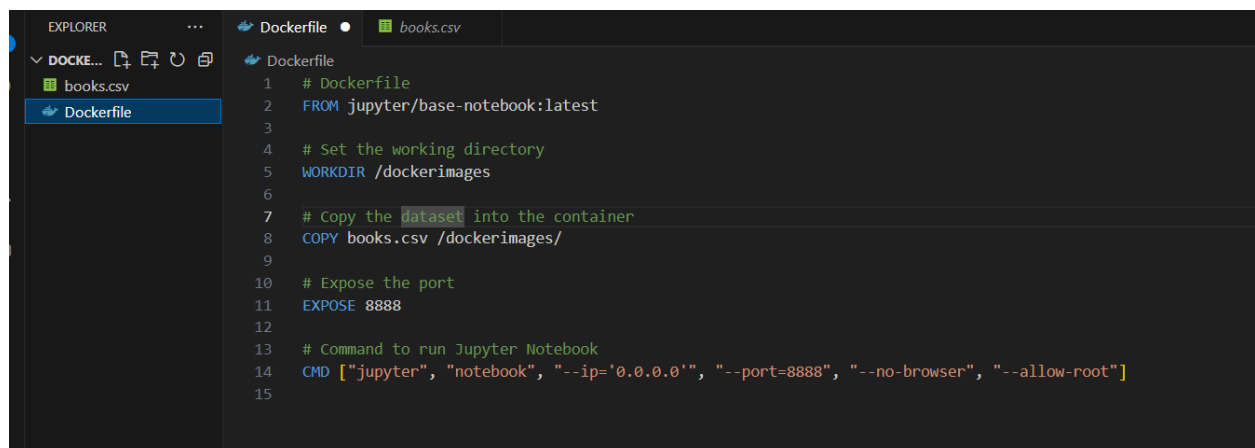
Name	Date modified	Type	Size
 books	4/22/2024 3:44 PM	Microsoft Excel Co...	423 KB
 Dockerfile	4/22/2024 8:07 PM	File	1 KB

- 2) Pull the base image from docker hub by this written command in the power shell

```
PS C:\Users\LENOVO> docker pull jupyter/datascience-notebook
Using default tag: latest
latest: Pulling from jupyter/datascience-notebook
Digest: sha256:476c6e673e7d5d8b5059f8680b1c6a988942a79263da651bf302dc696ab311f2
Status: Image is up to date for jupyter/datascience-notebook:latest
docker.io/jupyter/datascience-notebook:latest

What's Next?
View a summary of image vulnerabilities and recommendations → docker scout quickview jupyter/datascience-notebook
PS C:\Users\LENOVO> docker pull jupyter/datascience-notebook
```

- 3) Open vs code, then write these commands to make a Dockerfile (the explanation of each command is in the above comment)



The screenshot shows the VS Code interface with the Explorer sidebar on the left. The 'Dockerfile' is selected in the Explorer. The main editor area displays the content of the Dockerfile, which is as follows:

```
1 # Dockerfile
2 FROM jupyter/base-notebook:latest
3
4 # Set the working directory
5 WORKDIR /dockerimages
6
7 # Copy the dataset into the container
8 COPY books.csv /dockerimages/
9
10 # Expose the port
11 EXPOSE 8888
12
13 # Command to run Jupyter Notebook
14 CMD ["jupyter", "notebook", "--ip='0.0.0.0'", "--port=8888", "--no-browser", "--allow-root"]
15
```

4) In CMD, I wrote this command to build the image









```
C:\Users\LENOVO>docker build -t harrybook C:\dockerimages
2024/04/22 20:07:41 http2: server: error reading preface from client //./pipe/docker_engine: file has already been closed
[+] Building 2.7s (8/8) FINISHED                                docker:default
=> [internal] load build definition from Dockerfile              0.0s
=> => transferring dockerfile: 429B                             0.0s
=> [internal] load metadata for docker.io/jupyter/base-notebook:latest 2.5s
=> [internal] load .dockerignore                                0.0s
=> => transferring context: 2B                                   0.0s
=> [1/3] FROM docker.io/jupyter/base-notebook:latest@sha256:8c903974902b0e9d45d9823c2234411de0614c5c98c4bb782b3d 0.0s
=> [internal] load build context                                0.0s
=> => transferring context: 32B                                  0.0s
=> CACHED [2/3] WORKDIR /dockerimages                          0.0s
=> CACHED [3/3] COPY books.csv /dockerimages/                  0.0s
=> exporting to image                                           0.0s
=> => exporting layers                                           0.0s
=> => writing image sha256:45aaf0a585262a413ed4ec2b3264f26673e7db56fb14871c6749918bd80b0dd8 0.0s
=> => naming to docker.io/library/harrybook                     0.0s
```


5) Running container

```
C:\Users\LENOVO>docker run -p 8888:8888 harrybook
[I 2024-04-22 18:07:59.833 ServerApp] Package notebook took 0.0000s to import
[I 2024-04-22 18:07:59.851 ServerApp] Package jupyter_lsp took 0.0168s to import
[W 2024-04-22 18:07:59.851 ServerApp] A `jupyter_server_extension_points` function was not found in jupyter_lsp. Instead, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be deprecated in future releases of Jupyter Server.
[I 2024-04-22 18:07:59.858 ServerApp] Package jupyter_server_terminals took 0.0064s to import
[I 2024-04-22 18:07:59.859 ServerApp] Package jupyterlab took 0.0000s to import
[I 2024-04-22 18:08:00.389 ServerApp] Package nbclassic took 0.0022s to import
[W 2024-04-22 18:08:00.391 ServerApp] A `jupyter_server_extension_points` function was not found in nbclassic. Instead, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be deprecated in future releases of Jupyter Server.
[I 2024-04-22 18:08:00.392 ServerApp] Package notebook_shim took 0.0000s to import
[W 2024-04-22 18:08:00.392 ServerApp] A `jupyter_server_extension_points` function was not found in notebook_shim. Instead, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be deprecated in future releases of Jupyter Server.
[I 2024-04-22 18:08:00.392 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-04-22 18:08:00.398 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-04-22 18:08:00.404 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-04-22 18:08:00.410 ServerApp] nbclassic | extension was successfully linked.
[I 2024-04-22 18:08:00.415 ServerApp] notebook | extension was successfully linked.
[I 2024-04-22 18:08:00.417 ServerApp] Writing Jupyter server cookie secret to /home/jovyan/.local/share/jupyter/runtime/jupyter_cookie_secret
[I 2024-04-22 18:08:00.639 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-04-22 18:08:00.658 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-04-22 18:08:00.661 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-04-22 18:08:00.662 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-04-22 18:08:00.663 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.11/site-packages/jupyterlab
[I 2024-04-22 18:08:00.663 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 2024-04-22 18:08:00.664 LabApp] Extension Manager is 'pypi'.
[I 2024-04-22 18:08:00.667 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-04-22 18:08:00.670 ServerApp] nbclassic | extension was successfully loaded.
[I 2024-04-22 18:08:00.674 ServerApp] notebook | extension was successfully loaded.
[I 2024-04-22 18:08:00.675 ServerApp] Serving notebooks from local directory: /dockerimages
[I 2024-04-22 18:08:00.675 ServerApp] Jupyter Server 2.8.0 is running at:
[I 2024-04-22 18:08:00.675 ServerApp] http://3690d74149e5:8888/tree?token=4ab46fe26933fe763ec7f38a55fddce74ffedbbe00099
[I 2024-04-22 18:08:00.675 ServerApp] http://127.0.0.1:8888/tree?token=4ab46fe26933fe763ec7f38a55fddce74ffedbbe00099
[I 2024-04-22 18:08:00.675 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2024-04-22 18:08:00.679 ServerApp]

To access the server, open this file in a browser:
file:///home/jovyan/.local/share/jupyter/runtime/jpserver-8-open.html
Or copy and paste one of these URLs:
http://3690d74149e5:8888/tree?token=4ab46fe26933fe763ec7f38a55fddce74ffedbbe00099
http://127.0.0.1:8888/tree?token=4ab46fe26933fe763ec7f38a55fddce74ffedbbe00099
[I 2024-04-22 18:08:01.121 ServerApp] Skipped non-installed server(s): bash-language-server, dockerfile-language-server-nodejs, javascript-typescript-langserver, jedi-language-server, julia-language-server, pyri
ght, 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-json-languageserver-bin, yaml-language-server
0.00s - Debugger warning: It seems that frozen modules are being used, which may
0.00s - make the debugger miss breakpoints. Please pass -Xfrozen_modules=off
0.00s - to python to disable frozen modules.
0.00s - Note: Debugging will proceed. Set PYDEVD_DISABLE_FILE_VALIDATION=1 to disable this validation.
[I 2024-04-22 18:08:37.779 ServerApp] Creating new notebook in
```

6) Open docker to make sure the images and container are exist

<input type="checkbox"/>	Name	Tag	Status	Created	Size	Actions
<input type="checkbox"/>	harrybook	latest	In use	52 minutes ag	1.06 GB	  
<input type="checkbox"/>	45aaf0a58526 					
<input type="checkbox"/>	jupyter/datascience-notebook	latest	Unused	6 months ago	5.92 GB	  
<input type="checkbox"/>	f78a42f3bc9a 					

<input type="checkbox"/>	Name	Image	Status	CPU (%)	Port(s)	Last started	Actions
<input type="checkbox"/>	great_dewd 3690d74149e	 harrybook	Running	0%	8888:8888 🔗	2 days ago	<div> <div></div> <div></div> <div></div> </div>

7) Enter to this link

 127.0.0.1:8888/tree

8) Go to this notebook to write the code

Files

Running

Open

Download

Rename

Duplicate


Delete

New

Upload

🔄

/

Name	Last Modified	File Size
✓  Untitled.ipynb	1 minute ago	4.3 KB
 Untitled2.ipynb	2 minutes ago	3.2 KB
 books.csv	yesterday	422.5 KB
 harry_potter_best_selling_books.csv	1 minute ago	1.2 KB

9) Write the code

```
import pandas as pd

# Load the dataset
df = pd.read_csv("books.csv")

# Perform data cleaning and preprocessing
# Handle missing values
df.dropna(inplace=True)

# Remove unnecessary columns
df.drop(['book_id', 'goodreads_book_id', 'best_book_id', 'work_id', 'isbn', 'isbn13', 'image_url', 'small_image_url'], axis=1, inplace=True)

# Filter dataset for Harry Potter series
harry_potter_df = df[df['authors'].str.contains('J.K. Rowling') & df['title'].str.contains('Harry Potter')]

# Check the filtered dataset
harry_potter_df.head()

# Find the most selling books within the Harry Potter series
most_selling_books = harry_potter_df.sort_values(by='ratings_count', ascending=False).head()

# Display the most selling books
print(most_selling_books[['title', 'authors', 'ratings_count', 'average_rating']])

# Save the most selling books to a CSV file
most_selling_books.to_csv("harry_potter_best_selling_books.csv", index=False)

# Calculate the average rating of the Harry Potter books
average_rating_hp = harry_potter_df['average_rating'].mean()

# Display the average rating
print(f"Average rating of Harry Potter books: {average_rating_hp}")
```

10)Output

```

                                title \
1  Harry Potter and the Sorcerer's Stone (Harry P...
6  Harry Potter and the Prisoner of Azkaban (Harr...
9  Harry Potter and the Chamber of Secrets (Harry...
10 Harry Potter and the Goblet of Fire (Harry Pot...
11 Harry Potter and the Deathly Hallows (Harry Po...

                                authors ratings_count average_rating
1          J.K. Rowling, Mary GrandPré         4602479          4.44
6  J.K. Rowling, Mary GrandPré, Rufus Beck         1832823          4.53
9          J.K. Rowling, Mary GrandPré         1779331          4.37
10         J.K. Rowling, Mary GrandPré         1753043          4.53
11         J.K. Rowling, Mary GrandPré         1746574          4.61
Average rating of Harry Potter books: 4.550000000000001
```

- 11) Explain of this code: Importing pandas and read a CSV file, then Removing missing values and drop Unnecessary columns like 'book_id', 'goodreads_book_id', A new Data Frame is created by filtering rows where the 'authors' column contains 'J.K. Rowling' and the 'title' column contains 'Harry Potter'. After that Finding the most selling books to save the most selling books we create another csv file to store it then the code prints the details (title, authors, ratings count, and average rating) of the most selling Harry Potter books. Finally, Calculating average rating and print it.