Cloud computing

Assignment 2

Name : Walaa Ahmed Hassan

ID : 2206192

Department : cyber security

1. downloading the dataset and creating directory named dockerimages and store csv file in it

2. i pulled the base image from docker hub using the following command :

docker pull jupyter/datascience-notebook

```
PS C:\Users\walaa> 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\walaa>
```

3. then I wrote Dockerfile using vs code

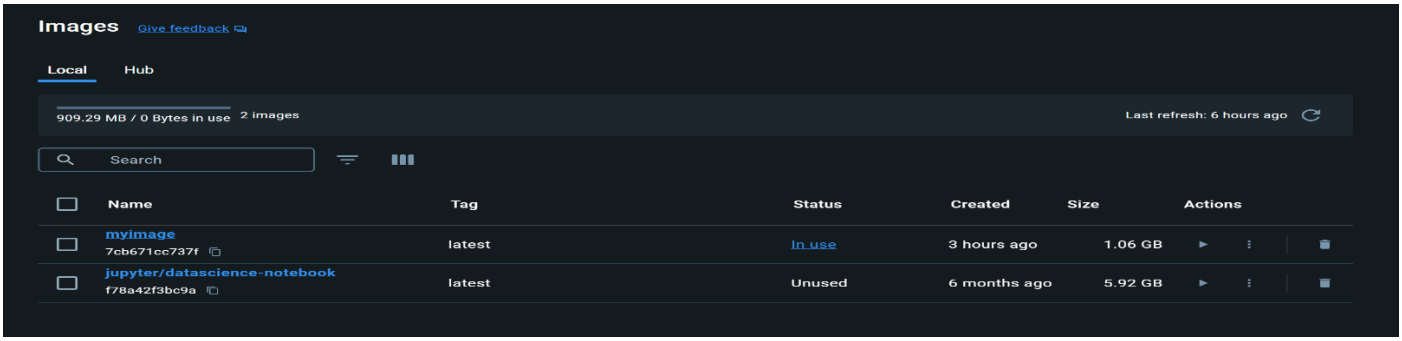
Dockerfile

```
#assigning base image
FROM jupyter/base-notebook:latest

# Set the working directory
WORKDIR /dockerimages

# Copy the notebook and dataset into the container
COPY books.csv /dockerimages/

# Expose the port
EXPOSE 8888

# the given Command to run Jupyter Notebook
CMD ["jupyter", "notebook", "--ip='0.0.0.0'", "--port=8888", "--no-browser", "--allow-root"]
```

4. build the image using the following command :

docker build -t myimage C:\dockerimages

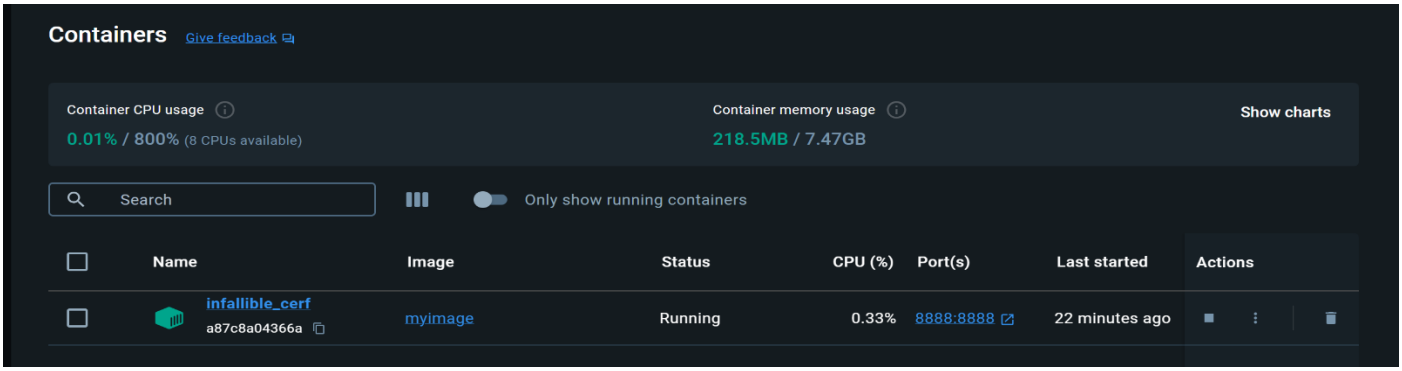
```
PS C:\Users\walaa> docker build -t myimage C:\dockerimages
[+] Building 1.6s (8/8) FINISHED                                                docker:default
 => [internal] load build definition from Dockerfile                                      0.0s
 => => transferring dockerfile: 413B                                                      0.0s
 => [internal] load metadata for docker.io/jupyter/base-notebook:latest                   1.5s
 => [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 sha256:7cb671cc737f190cbb28381f862c1bf8a0b2f9e3180cd475b7116f04d42e8eba  0.0s
 => => naming to docker.io/library/myimage                                                0.0s

View build details: docker-desktop://dashboard/build/default/default/t6zi1zqsgda7htxj9j5zu3jp7

What's Next?
  View a summary of image vulnerabilities and recommendations → docker scout quickview
PS C:\Users\walaa>
```



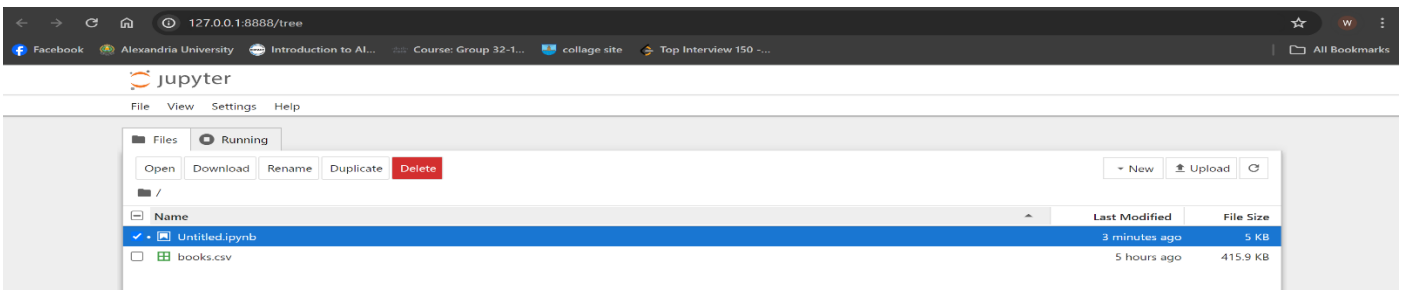
5. running the container and giving it the port using the following command :

docker run -p 8888:8888 myimage

```
PS C:\Users\walaa> docker run -p 8888:8888 harrybook
[I 2024-04-22 18:10:41.874 ServerApp] Package notebook took 0.0000s to import
[I 2024-04-22 18:10:41.883 ServerApp] Package jupyter_lsp took 0.0082s to import
[W 2024-04-22 18:10:41.883 ServerApp] A `_jupyter_server_extension_points` function was not found in jupyter_lsp. Instea
d, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be deprecate
d in future releases of Jupyter Server.
[I 2024-04-22 18:10:41.886 ServerApp] Package jupyter_server_terminals took 0.0035s to import
[I 2024-04-22 18:10:41.887 ServerApp] Package jupyterlab took 0.0000s to import
[I 2024-04-22 18:10:42.270 ServerApp] Package nbclassic took 0.0013s to import
[W 2024-04-22 18:10:42.271 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 18:10:42.272 ServerApp] Package notebook_shim took 0.0000s to import
[W 2024-04-22 18:10:42.272 ServerApp] A `_jupyter_server_extension_points` function was not found in notebook_shim. Inst
ead, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be depreca
ted in future releases of Jupyter Server.
[I 2024-04-22 18:10:42.272 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-04-22 18:10:42.276 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-04-22 18:10:42.280 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-04-22 18:10:42.283 ServerApp] nbclassic | extension was successfully linked.
[I 2024-04-22 18:10:42.286 ServerApp] notebook | extension was successfully linked.
[I 2024-04-22 18:10:42.288 ServerApp] Writing Jupyter server cookie secret to /home/jovyan/.local/share/jupyter/runtime/
jupyter_cookie_secret
[I 2024-04-22 18:10:42.438 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-04-22 18:10:42.454 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-04-22 18:10:42.456 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-04-22 18:10:42.458 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-04-22 18:10:42.460 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.11/site-packages/jupyterlab
[I 2024-04-22 18:10:42.460 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 2024-04-22 18:10:42.460 LabApp] Extension Manager is 'pypi'.
[I 2024-04-22 18:10:42.462 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-04-22 18:10:42.464 ServerApp] nbclassic | extension was successfully loaded.
[I 2024-04-22 18:10:42.466 ServerApp] notebook | extension was successfully loaded.
[I 2024-04-22 18:10:42.466 ServerApp] Serving notebooks from local directory: /dockerimages
[I 2024-04-22 18:10:42.466 ServerApp] Jupyter Server 2.8.0 is running at:
[I 2024-04-22 18:10:42.466 ServerApp] http://49e2ddbe14a7:8888/tree?token=ed601b8e8547e42057d8d663f5c387ea85be758f464891
95
```



6. open my jupyter notebook



7. write my own code

```
pip install pandas
```

```
Collecting pandas
  Downloading pandas-2.2.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (19 kB)
Collecting numpy>=1.23.2 (from pandas)
  Downloading numpy-1.26.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl.metadata (61 kB)
     ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 61.0/61.0 kB 843.0 kB/s eta 0:00:00 0:00:01
Requirement already satisfied: python-dateutil>=2.8.2 in /opt/conda/lib/python3.11/site-packages (from pandas) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/lib/python3.11/site-packages (from pandas) (2023.3.post1)
Collecting tzdata>=2022.7 (from pandas)
  Downloading tzdata-2024.1-py2.py3-none-any.whl.metadata (1.4 kB)
Requirement already satisfied: six>=1.5 in /opt/conda/lib/python3.11/site-packages (from python-dateutil>=2.8.2->pandas) (1.16.0)
Downloading pandas-2.2.2-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (13.0 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 13.0/13.0 MB 3.8 MB/s eta 0:00:0000:0100:01
Downloading numpy-1.26.4-cp311-cp311-manylinux_2_17_x86_64.manylinux2014_x86_64.whl (18.3 MB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 18.3/18.3 MB 3.4 MB/s eta 0:00:0000:0100:01
Downloading tzdata-2024.1-py2.py3-none-any.whl (345 kB)
   ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 345.4/345.4 kB 2.6 MB/s eta 0:00:00a 0:00:01m
Installing collected packages: tzdata, numpy, pandas
Successfully installed numpy-1.26.4 pandas-2.2.2 tzdata-2024.1
Note: you may need to restart the kernel to use updated packages.
```

```python
import pandas as pd

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

# Data cleaning and preprocessing
# Drop rows with null values
df.dropna(inplace=True)

# Filter rows related to Harry Potter series
harry_potter_books = df[df['title'].str.contains('Harry Potter', case=False)]

# Find the most selling books within the Harry Potter series
most_selling_books = harry_potter_books.sort_values(by='ratings_count', ascending=False).head(5)

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

print("Most selling Harry Potter books:")
print(most_selling_books[['title', 'ratings_count']])
print("\nAverage rating of Harry Potter books:", average_rating)
```

```
Most selling Harry Potter books:
                                          title  ratings_count
1   Harry Potter and the Sorcerer's Stone (Harry P...        4602479
6   Harry Potter and the Prisoner of Azkaban (Harr...        1832823
9   Harry Potter and the Chamber of Secrets (Harry...        1779331
10  Harry Potter and the Goblet of Fire (Harry Pot...        1753043
11  Harry Potter and the Deathly Hallows (Harry Po...        1746574

Average rating of Harry Potter books: 4.4910000000000005
```