Cloud Computing Assignment

Dockerfile and Data Analysis with Popular Books Dataset

Name: Amal Khaled Abd El-moniem Elwan

ID:210475

Department: Media Analysis

- Download Jupyter docker image from a registry
- Use this command in cmd docker pull jupyter/datascience-notebook

```
C:\Users\user>docker pull jupyter/datascience-notebook
Using default tag: latest
latest: Pulling from jupyter/datascience-notebook
aece8493d397: Pull complete
fd92c719666c: Pull complete
088f11eb1e74: Pull complete
4f4fb700ef54: Pull complete
ef8373d600b0: Pull complete
77e45ee945dc: Pull complete
a30f89a0af6c: Pull complete
dc42adc7eb73: Pull complete
abaa8376a650: Pull complete
aa099bb9e49a: Pull complete
822c4cbcf6a6: Pull complete
d25166dcdc7b: Pull complete
964fc3e4ff9f: Pull complete
2c4c69587ee4: Pull complete
de2cdd875fa8: Pull complete
75d33599f5f2: Pull complete
31973ea82470: Pull complete
96ee7e4439c7: Pull complete
1f9ad23c07ac: Pull complete
d19266e0cb17: Pull complete
9a165b6e9dc7: Pull complete
5689442fd4e1: Pull complete
9a6a202f62a6: Pull complete
734ea0c3d94e: Pull complete
a21a167f7127: Pull complete
02c2173301db: Pull complete
e488194bf535: Pull complete
f5302bfd25be: Pull complete
5201d3116fb6: Pull complete
Digest: sha256:476c6e673e7d5d8b5059f8680b1c6a988942a79263da651bf302dc696ab311f2
Status: Downloaded newer image 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
```

- make dockerfile
- Create a directory & go to the directory
- use this command mkdir cloudAssignment

```
C:\Users\user>mkdir cloudAssignment

C:\Users\user>cd cloudAssignment

C:\Users\user\cloudAssignment>
```

- Create a Dockerfile without extension
    - In Dockerfile

```dockerfile
FROM  jupyter/datascience-notebook
COPY . /cloudAssignment
WORKDIR /cloudAssignment
EXPOSE 8888

# Install any needed packages specified in requirements.txt
RUN pip install -r requirements.txt
# Run Jupyter Notebook when the container launches
CMD ["jupyter", "notebook",  "--ip=0.0.0.0" , "--port=8888", "--no-browser","--allow-root"]

```

- In requirements.txt

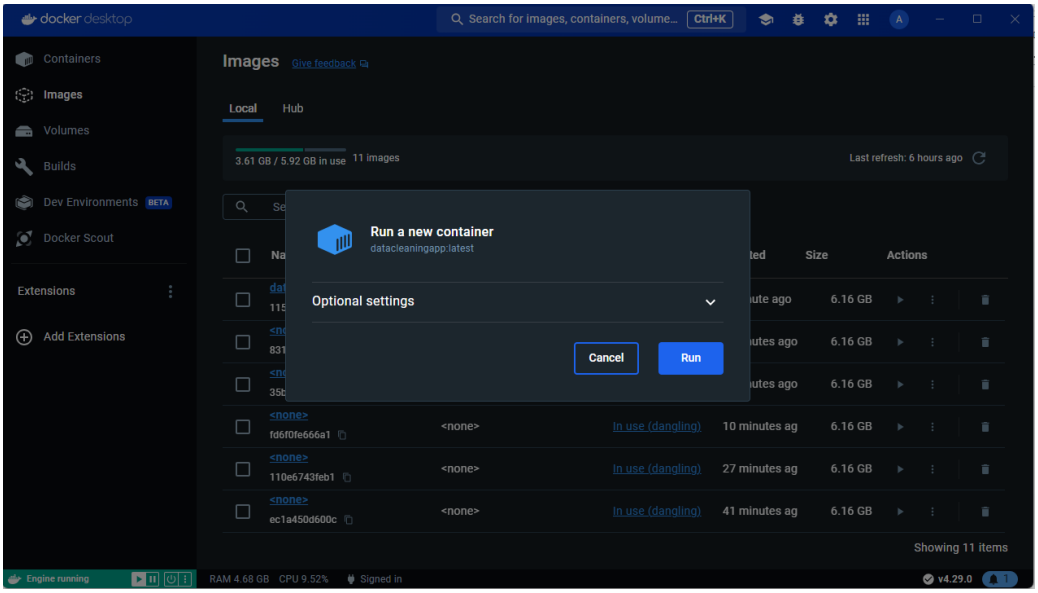
```
#  lists the Python libraries
pandas== 2.2.2
numpy==1.26.4
matplotlib==3.8.4
seaborn==0.13.2
```

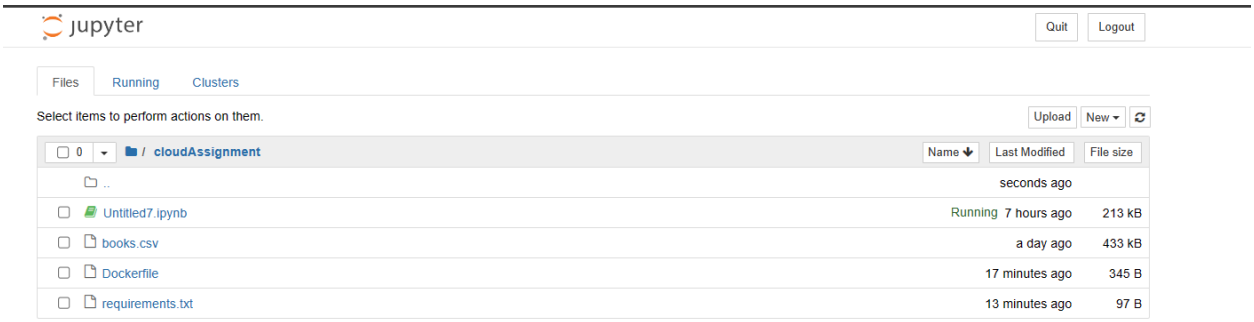
- In Dockerfile terminal we will write this command docker build -t datacleaningapp . to build Docker image

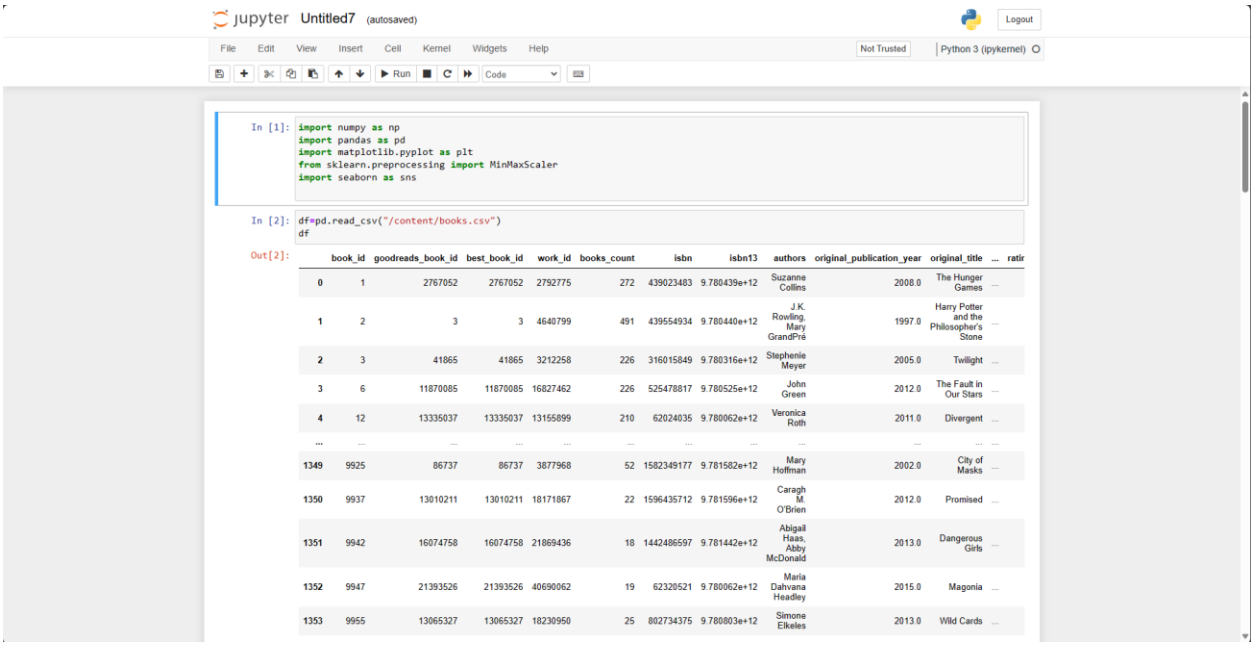
```
PS C:\Users\user\cloudAssignment> docker build -t datacleaningapp .
2024/04/23 22:41:54 http2: server: error reading preface from client //./pipe/docker_engine: file has already been closed
[+] Building 0.0s (0/0)  docker:default
[+] Building 43.5s (9/9) FINISHED
 => [internal] load build definition from Dockerfile
 => => transferring dockerfile: 271B
 => [internal] load metadata for docker.io/jupyter/datascience-notebook:latest
 => [internal] load .dockerignore
 => => transferring context: 2B
 => [internal] load build context
 => => transferring context: 505B
 => CACHED [1/4] FROM docker.io/jupyter/datascience-notebook:latest
 => [2/4] COPY . /cloudAssignment
 => [3/4] WORKDIR /cloudAssignment
 => [4/4] RUN pip install -r requirements.txt
 => exporting to image
 => => exporting layers
 => => writing image sha256:115aed894da230f4f99e9a51bc5210e4d01ed551d32a0e3bb0c819a1d470692f
 => => naming to docker.io/library/datacleaningapp

View build details: docker-desktop://dashboard/build/default/default/ugsh4eqgzatjr0jky6csikylu

What's Next?
  View a summary of image vulnerabilities and recommendations → docker scout quickview
```

- Run Docker Container , use this command docker run datacleaningapp

```
PS C:\Users\user\cloudAssignment> docker run  datacleaningapp
[I 2024-04-23 20:42:42.092 ServerApp] Package notebook took 0.0000s to import
[I 2024-04-23 20:42:42.133 ServerApp] Package jupyter_lsp took 0.0410s to import
[W 2024-04-23 20:42:42.134 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 wi
ll be deprecated in future releases of Jupyter Server.
[I 2024-04-23 20:42:42.137 ServerApp] Package jupyter_server_mathjax took 0.0028s to import
[I 2024-04-23 20:42:42.239 ServerApp] Package jupyter_server_proxy took 0.1007s to import
[I 2024-04-23 20:42:42.251 ServerApp] Package jupyter_server_terminals took 0.0119s to import
[I 2024-04-23 20:42:42.252 ServerApp] Package jupyterlab took 0.0000s to import
[I 2024-04-23 20:42:43.205 ServerApp] Package jupyterlab_git took 0.0694s to import
[I 2024-04-23 20:42:43.211 ServerApp] Package nbclassic took 0.0048s to import
[W 2024-04-23 20:42:43.215 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 2024-04-23 20:42:43.215 ServerApp] Package nbdime took 0.0000s to import
[I 2024-04-23 20:42:43.216 ServerApp] Package notebook_shim took 0.0000s to import
[W 2024-04-23 20:42:43.216 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 2024-04-23 20:42:43.218 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-04-23 20:42:43.226 ServerApp] jupyter_server_mathjax | extension was successfully linked.
[I 2024-04-23 20:42:43.226 ServerApp] jupyter_server_proxy | extension was successfully linked.
[I 2024-04-23 20:42:43.233 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-04-23 20:42:43.244 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-04-23 20:42:43.244 ServerApp] jupyterlab_git | extension was successfully linked.
[I 2024-04-23 20:42:43.251 ServerApp] nbclassic | extension was successfully linked.
[I 2024-04-23 20:42:43.251 ServerApp] nbdime | extension was successfully linked.
[I 2024-04-23 20:42:43.262 ServerApp] notebook | extension was successfully linked.
[I 2024-04-23 20:42:43.264 ServerApp] Writing Jupyter server cookie secret to /home/jovyan/.local/share/jupyter/runtime/jupyter_cookie_secret
[I 2024-04-23 20:42:43.739 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-04-23 20:42:43.763 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-04-23 20:42:43.766 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-04-23 20:42:43.768 ServerApp] jupyter_server_mathjax | extension was successfully loaded.
[I 2024-04-23 20:42:43.786 ServerApp] jupyter_server_proxy | extension was successfully loaded.
[I 2024-04-23 20:42:43.790 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-04-23 20:42:43.796 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.11/site-packages/jupyterlab
[I 2024-04-23 20:42:43.796 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 2024-04-23 20:42:43.796 LabApp] Extension Manager is "pypi".
[I 2024-04-23 20:42:43.803 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-04-23 20:42:43.818 ServerApp] jupyterlab_git | extension was successfully loaded.
[I 2024-04-23 20:42:43.818 ServerApp] nbclassic | extension was successfully loaded.
[I 2024-04-23 20:42:43.919 ServerApp] nbdime | extension was successfully loaded.
[I 2024-04-23 20:42:43.924 ServerApp] notebook | extension was successfully loaded.
[I 2024-04-23 20:42:43.925 ServerApp] Serving notebooks from local directory: /cloudAssignment
[I 2024-04-23 20:42:43.925 ServerApp] Jupyter Server 2.8.0 is running at:
[I 2024-04-23 20:42:43.925 ServerApp] http://0143187d05a:8888/tree?token=7eea0a860c48b3987f1f7543eb32c097d74e60e95d91fce8
[I 2024-04-23 20:42:43.925 ServerApp]     http://127.0.0.1:8888/tree?token=7eea0a860c48b3987f1f7543eb32c097d74e60e95d91fce8
[I 2024-04-23 20:42:43.925 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2024-04-23 20:42:43.930 ServerApp]
```







- most selling books within the Harry Potter series. & the average rating of the Harry Potter books

```
#index maximum value
max_index = Harry_Potter_books['books_count'].idxmax()
#get value of maximum index
max_value = df.loc[max_index, 'books_count']
title_value = df.loc[max_index, 'title']
print("most selling books :", title_value ," books_count: " ,max_value)
```

```
most selling books : Harry Potter and the Sorcerer's Stone (Harry Potter, #1)  books_count:  491
```

```
average = (Harry_Potter_books.average_rating * Harry_Potter_books.ratings_count).sum() / Harry_Potter_books.ratings_count.sum()
average
```

```
4.489114370335377
```