

Assignment2

الاسم: يارا عمادالدين متولي أحمد

2022595822 :ID

القسم: تحليلات الوسائط الإعلامية

First, we should run this in terminal.

```
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
```

Then, download the Docker program, then create a file called it in any new folder on the computer, and then write these lines of code.

In the terminal we write *docker build -t (name of project docker name)* and then we write *docker run (name of project docker name)*.

The screenshot shows the VS Code interface with a Dockerfile being created in the Explorer. The Dockerfile content is as follows:

```
1 FROM jupyter/datascience-notebook
2 COPY . /cloud
3 WORKDIR /cloud
4 EXPOSE 8888
5
6 RUN pip install numpy
7 RUN pip install pandas
8
9
10 CMD ["jupyter", "notebook", "--ip=0.0.0.0", "--port=8888", "--no-browser", "--allow-root"]
11
```

The TERMINAL tab shows the build progress for the `docker:default` context:

```
PS C:\Users\User\OneDrive\Desktop\cloud> docker build -t dockerrapp .
[+] Building 0.0s (0/0) docker:default
2024/04/25 14:09:22 http: server: error reading preface from client //./pipe/docker_engine: file[s] building 12.1s (10/10) FINISHED
-> [internal] load build definition from Dockerfile
-> transferring dockerfile: 264B 0.0s
-> [internal] load metadata for docker.io/jupyter/datascience-notebook:latest 0.0s
-> [internal] load .dockerignore 0.0s
-> transferring context: 2B 0.0s
-> [internal] load build context 0.0s
-> transferring context: 264B 0.0s
-> CACHED [1/5] FROM docker.io/jupyter/datascience-notebook:latest 0.0s
-> [2/5] COPY . /cloud 0.1s
-> [3/5] WORKDIR /cloud 0.1s
-> [4/5] RUN pip install numpy 4.9s
-> [5/5] RUN pip install pandas 4.5s
-> exporting to image 0.3s
-> exporting layers 0.2s
-> writing image sha256:6c2450fc790401b06d463f000659fbbdd1c9e4f7d737944f43e22288cc757 0.0s
-> naming to docker.io/library/dockerrapp 0.0s
```

Below the build progress, there is a section titled "View build details: [docker-desktop://dashboard/build/default/default/kamls3cuae6b9c90646n3](#)".

Under "What's Next?", there is a summary of image vulnerabilities and recommendations from `docker scout quickview`:

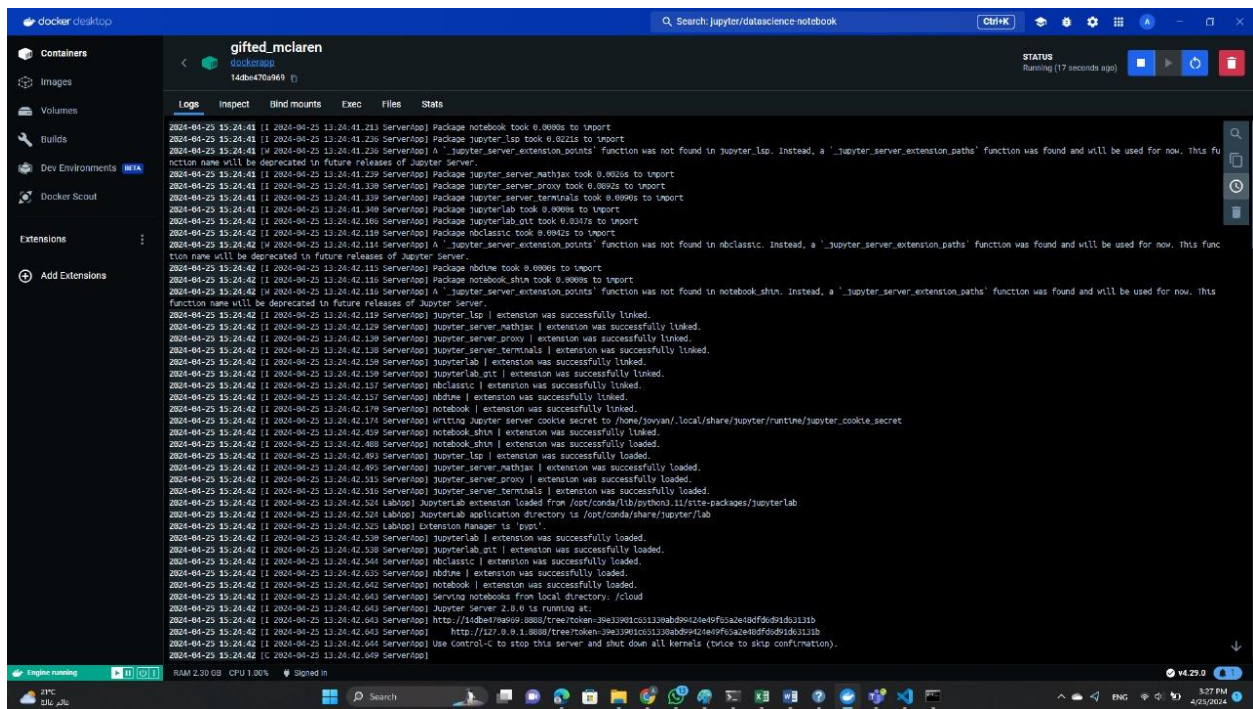
```
PS C:\Users\User\OneDrive\Desktop\cloud> docker run dockerrapp
[!] 2024-04-25 12:10:32.283 ServerApp) Package notebook took 0.0000s to import
[!] 2024-04-25 12:10:32.303 ServerApp) Package jupyter_lsp took 0.0196s to import
[W 2024-04-25 12:10:32.312 ServerApp) A "_jupyter_server_extension_points" 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 2024-04-25 12:10:32.316 ServerApp) Package jupyter_server_mathjax took 0.0020s to import
[I 2024-04-25 12:10:32.581 ServerApp) Package jupyter_server_proxy took 0.1875s to import
[I 2024-04-25 12:10:32.514 ServerApp) Package jupyter_server_terminals took 0.0080s to import
[I 2024-04-25 12:10:32.514 ServerApp) Package jupyterlab took 0.0000s to import
[I 2024-04-25 12:10:32.515 ServerApp) Package jupyterlab_git took 0.0209s to import
[I 2024-04-25 12:10:33.220 ServerApp) Package nbclassic took 0.0040s to import
[W 2024-04-25 12:10:33.222 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
```

And this is the result of run.

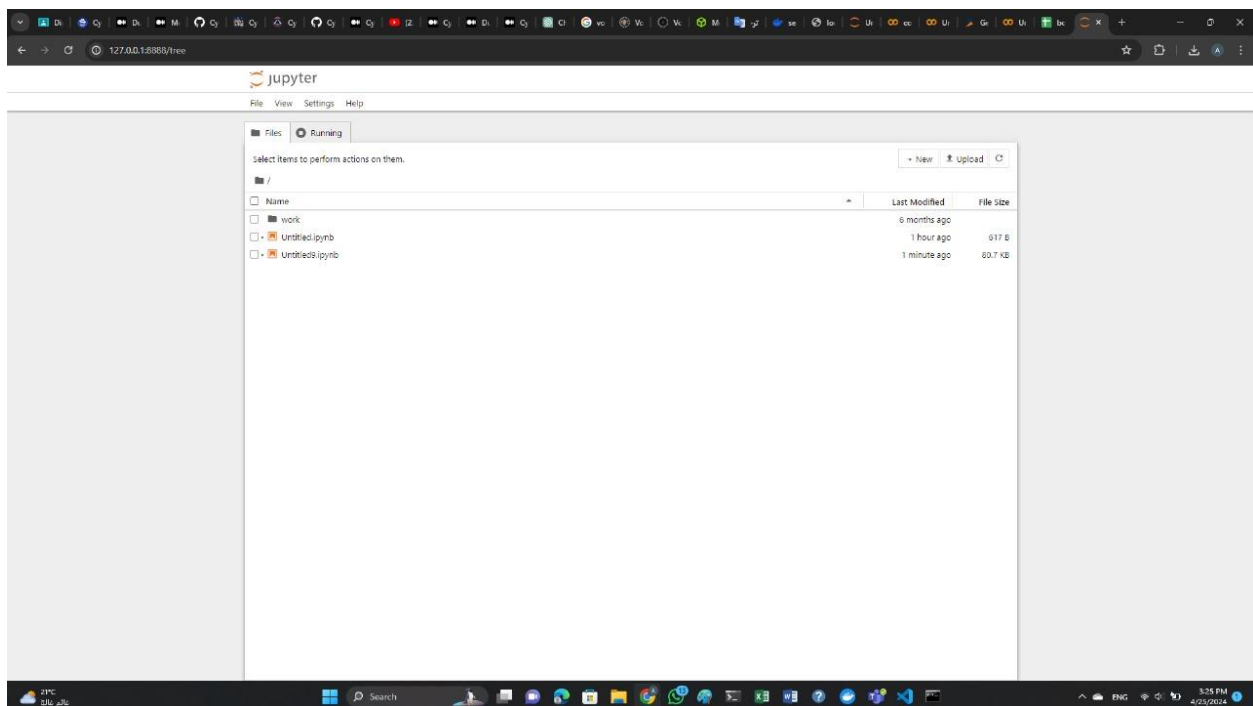
The screenshot shows the VS Code interface with the Dockerfile in the Explorer. The TERMINAL tab shows the output of running the `dockerrapp` container:

```
PS C:\Users\User\OneDrive\Desktop\cloud> docker run dockerrapp
[!] 2024-04-25 12:10:32.283 ServerApp) Package notebook took 0.0000s to import
[!] 2024-04-25 12:10:32.303 ServerApp) Package jupyter_lsp took 0.0196s to import
[W 2024-04-25 12:10:32.312 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 2024-04-25 12:10:32.316 ServerApp) Package jupyter_server_mathjax took 0.0020s to import
[I 2024-04-25 12:10:32.581 ServerApp) Package jupyter_server_proxy took 0.1875s to import
[I 2024-04-25 12:10:32.514 ServerApp) Package jupyter_server_terminals took 0.0080s to import
[I 2024-04-25 12:10:32.514 ServerApp) Package jupyterlab took 0.0000s to import
[I 2024-04-25 12:10:32.515 ServerApp) Package jupyterlab_git took 0.0209s to import
[I 2024-04-25 12:10:33.220 ServerApp) Package nbclassic took 0.0040s to import
[W 2024-04-25 12:10:33.222 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 2024-04-25 12:10:33.223 ServerApp) Package nbclassic took 0.0040s to import
[I 2024-04-25 12:10:33.224 ServerApp) Package notebook_shim took 0.0000s to import
[W 2024-04-25 12:10:33.225 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 2024-04-25 12:10:33.226 ServerApp) jupyter_lsp | extension was successfully linked.
[I 2024-04-25 12:10:33.234 ServerApp) jupyter_server_mathjax | extension was successfully linked.
[I 2024-04-25 12:10:33.234 ServerApp) jupyter_server_proxy | extension was successfully linked.
[I 2024-04-25 12:10:33.241 ServerApp) jupyter_server_terminals | extension was successfully linked.
[I 2024-04-25 12:10:33.251 ServerApp) jupyterlab | extension was successfully linked.
[I 2024-04-25 12:10:33.251 ServerApp) jupyterlab_git | extension was successfully linked.
[I 2024-04-25 12:10:33.258 ServerApp) nbclassic | extension was successfully linked.
[I 2024-04-25 12:10:33.258 ServerApp) nbclassic | extension was successfully linked.
[I 2024-04-25 12:10:33.267 ServerApp) notebook | extension was successfully linked.
[I 2024-04-25 12:10:33.271 ServerApp) Writing Jupyter server cookie secret to /home/jovyan/.local/share/jupyter/runtime/jupyter_cookie_secret
[I 2024-04-25 12:10:33.590 ServerApp) notebook_shim | extension was successfully linked.
[I 2024-04-25 12:10:33.610 ServerApp) notebook_shim | extension was successfully loaded.
[I 2024-04-25 12:10:33.614 ServerApp) jupyter_lsp | extension was successfully loaded.
[I 2024-04-25 12:10:33.614 ServerApp) jupyter_server_mathjax | extension was successfully loaded.
[I 2024-04-25 12:10:33.614 ServerApp) jupyter_server_proxy | extension was successfully loaded.
[I 2024-04-25 12:10:33.614 ServerApp) jupyter_server_terminals | extension was successfully loaded.
[I 2024-04-25 12:10:33.617 LabApp) JupyterLab extension loaded from /opt/conda/lib/python3.11/site-packages/jupyterlab
[I 2024-04-25 12:10:33.618 LabApp) JupyterLab application directory is /opt/conda/share/jupyterlab
[I 2024-04-25 12:10:33.618 LabApp) Extension Manager is 'jupyterlab'
[I 2024-04-25 12:10:33.645 ServerApp) jupyterlab | extension was successfully loaded.
[I 2024-04-25 12:10:33.652 ServerApp) jupyterlab_git | extension was successfully loaded.
[I 2024-04-25 12:10:33.652 ServerApp) nbclassic | extension was successfully loaded.
[I 2024-04-25 12:10:33.728 ServerApp) nbclassic | extension was successfully loaded.
[I 2024-04-25 12:10:33.733 ServerApp) notebook | extension was successfully loaded.
[I 2024-04-25 12:10:33.734 ServerApp) Saving notebooks from local directory: /cloud
[I 2024-04-25 12:10:33.734 ServerApp) Jupyter Server 2.14.0 is running at
[I 2024-04-25 12:10:33.734 ServerApp) http://127.0.0.1:8888/?token=ea41799d23418969785145f48e7f5a418f361fe185c1677
[I 2024-04-25 12:10:33.734 ServerApp) http://127.0.0.1:8888/?token=ea41799d23418969785145f48e7f5a418f361fe185c1677
[I 2024-04-25 12:10:33.734 ServerApp) Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2024-04-25 12:10:33.740 ServerApp)
```

Open Docker to run the code and get the notebook link to write the code to clean the data preprocessing.



When we click here in link that window will open.



Then in that notebook we can write code of data cleaning.

127.0.0.1:8888/lab/tree/Untitled9.ipynb

File Edit View Run Kernel Help

Filter files by name

Untitled9.ipynb 18 seconds ago

```
[1]: import pandas as pd
import numpy as np

[2]: df = pd.read_csv("../content/books.csv")
df.head()
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	ratings_count	work_ratings_count	work_text_reviews_count	ratings_1	rat
0	1	2787052	2787052	2792775	272	4390221403	97804398+12	Suzanne Collins	2008.0	The Hunger Games	4780653	4942365	130254	66713	1
1	2	3	3	4640799	491	439554934	9780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	4603479	4800065	75867	75504	1
2	3	41855	41855	3212256	326	316015949	6780316e+12	Stephenie Meyer	2005.0	Twilight	3866839	3916824	95009	456101	4
3	6	11870085	11870085	16827462	326	525478817	9780325e+12	John Green	2012.0	The Fault in Our Stars	2346404	2478609	140739	47964	1
4	12	13335037	13335037	13155899	210	62024025	9780002e+12	Veronica Roth	2011.0	Divergent	1903563	2216814	101023	26313	1

5 rows x 23 columns

```
[3]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 1354 entries, 0 to 1353
Data columns (total 23 columns):
 #   Column              Non-Null Count  Dtype
---  ---
 0   book_id             1354 non-null   int64
 1   goodreads_book_id   1354 non-null   int64
 2   best_book_id        1354 non-null   int64
 3   work_id             1354 non-null   int64
 4   books_count         1354 non-null   int64
 5   isbn                1302 non-null   object
 6   isbn13              1354 non-null   float64
 7   authors             1354 non-null   object
 8   original_publication_year  1351 non-null   float64
 9   original_title       1302 non-null   object
10   title               1354 non-null   object
11  language_code        1245 non-null   object
```

Would you like to receive official Jupyter news? Please read the privacy policy. [Open privacy policy](#) Yes No

Simple 0 2 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 Untitled9.ipynb 3:28 PM 4/23/2024

127.0.0.1:8888/lab/tree/Untitled9.ipynb

File Edit View Run Kernel Help

Untitled9.ipynb 28 seconds ago

10	24	6	6	3046572	332	439159609	9780439e+12	J.K. Rowling, Mary GrandPré	2000.0	Harry Potter and the Goblet of Fire	1753043	1868642	51084	6676	1
11	25	136251	136251	2963218	263	545010225	9780545e+12	J.K. Rowling, Mary GrandPré	2007.0	Harry Potter and the Deathly Hallows	1746574	1847395	51942	9363	1
12	27	1	1	41325427	275	439785960	9780440e+12	J.K. Rowling, Mary GrandPré	2005.0	Harry Potter and the Half-Blood Prince	1678823	1785676	27520	7308	1
96	422	862041	862041	2962462	75	545044251	9780545e+12	J.K. Rowling	1998.0	Complete Harry Potter Boxed Set	1500550	204125	8506	1105	1
613	3753	10	10	21457570	6	439827604	9780440e+12	J.K. Rowling	2005.0	Harry Potter Collection (Harry Potter, #1-6)	24618	26274	802	203	1
1036	7018	401445	401445	471792	42	042519891X	9780425e+12	David Colclert	2001.0	The Magical Worlds of Harry Potter: A Treasury...	13820	15145	267	329	1
1266	9049	2002	2002	8921949	5	043932162X	9780439e+12	J.K. Rowling	2001.0	N&N	10736	11732	183	108	1

11 rows x 23 columns

```
[7]: books_count = Harry_Potter['books_count']
max_index = (books_count).idxmax()
print(max_index)

1

[20]: average = (Harry_Potter.average_rating * Harry_Potter.ratings_count).sum() / Harry_Potter.ratings_count.sum()
average

[29]: 4.489114370355377
```

Would you like to receive official Jupyter news? Please read the privacy policy. [Open privacy policy](#) Yes No

Simple 0 2 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 1 Untitled9.ipynb 3:28 PM 4/23/2024

