Assignment2

الاسم: يارا عمادالدين متولي أحمد

ID: 2022595822

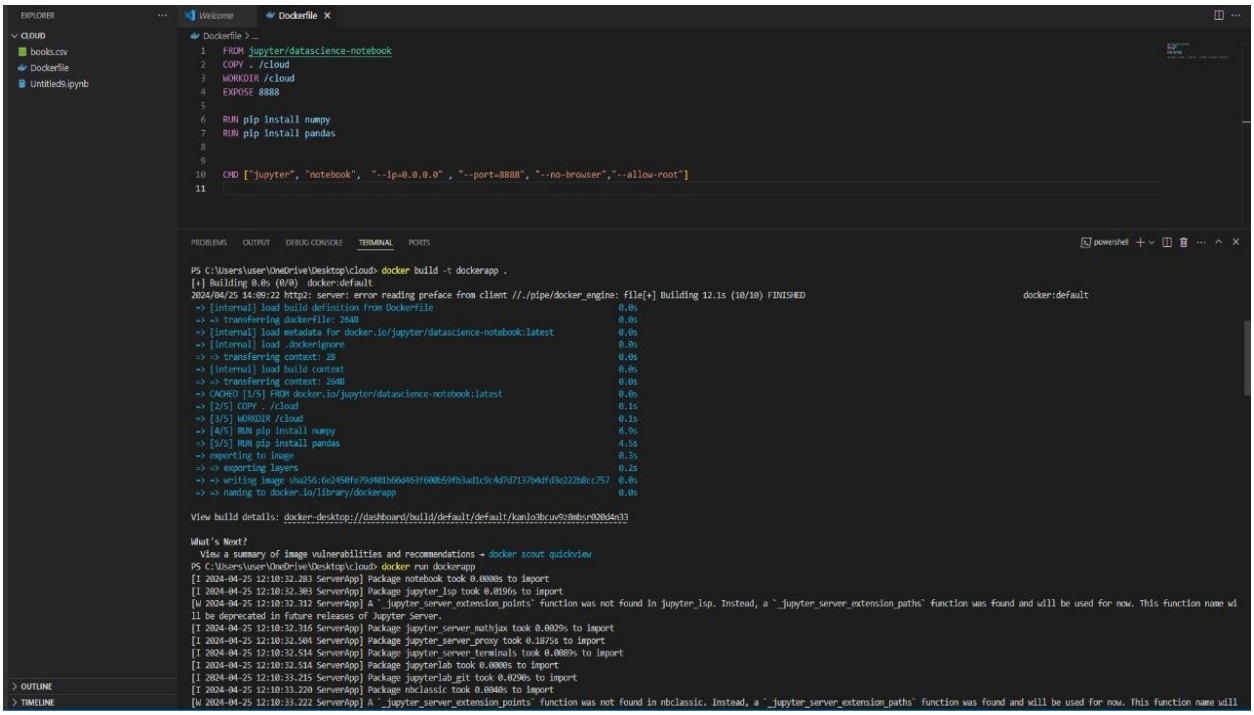
القسم: تحليلات الوسائط الإعلامية

First, we should run this in terminal.

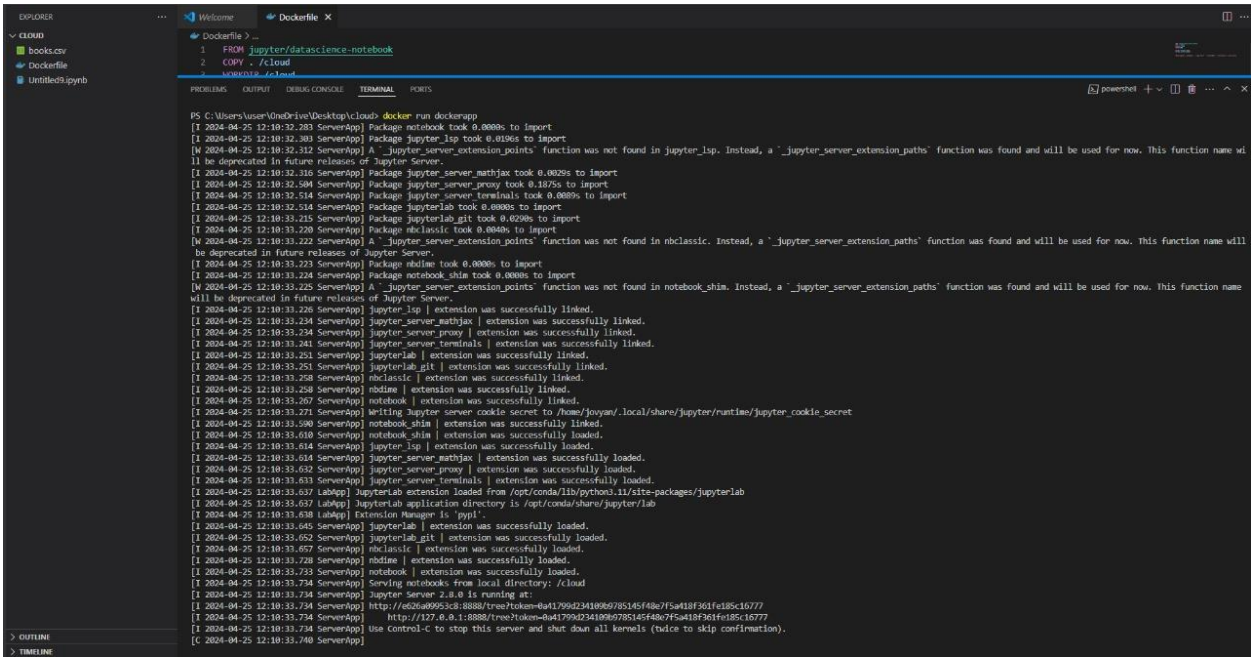
```
C:\Users\user>docker pull jupyter/datascience-notebook
Using default tag: latest
latest: Pulling from jupyter/datascience-notebook
aece8493d397: Pull complete
fd92c719666c: Pull complete
088f11eb1e74: Pull complete
4f4fb700ef54: Pull complete
ef8373d600b0: Pull complete
77e45ee945dc: Pull complete
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.

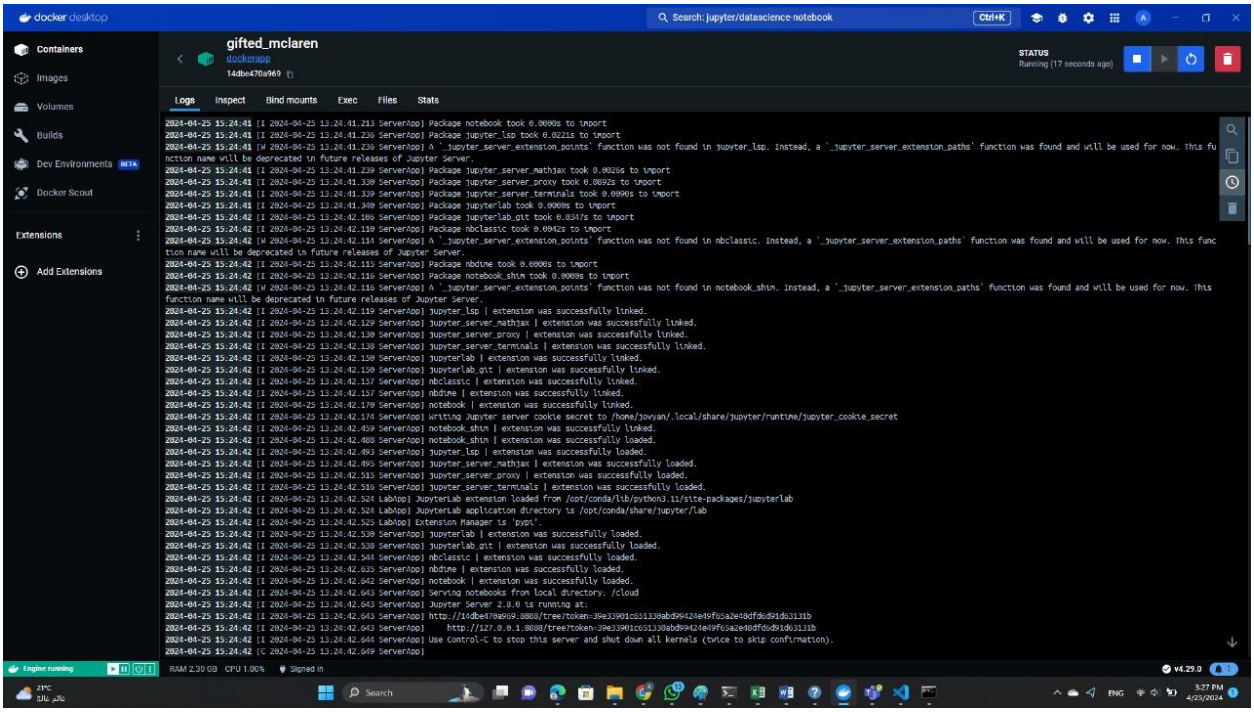
In the terminal we write *docker build -t (name of project docker name)* and then we write *docker run (name of project docker name)*.



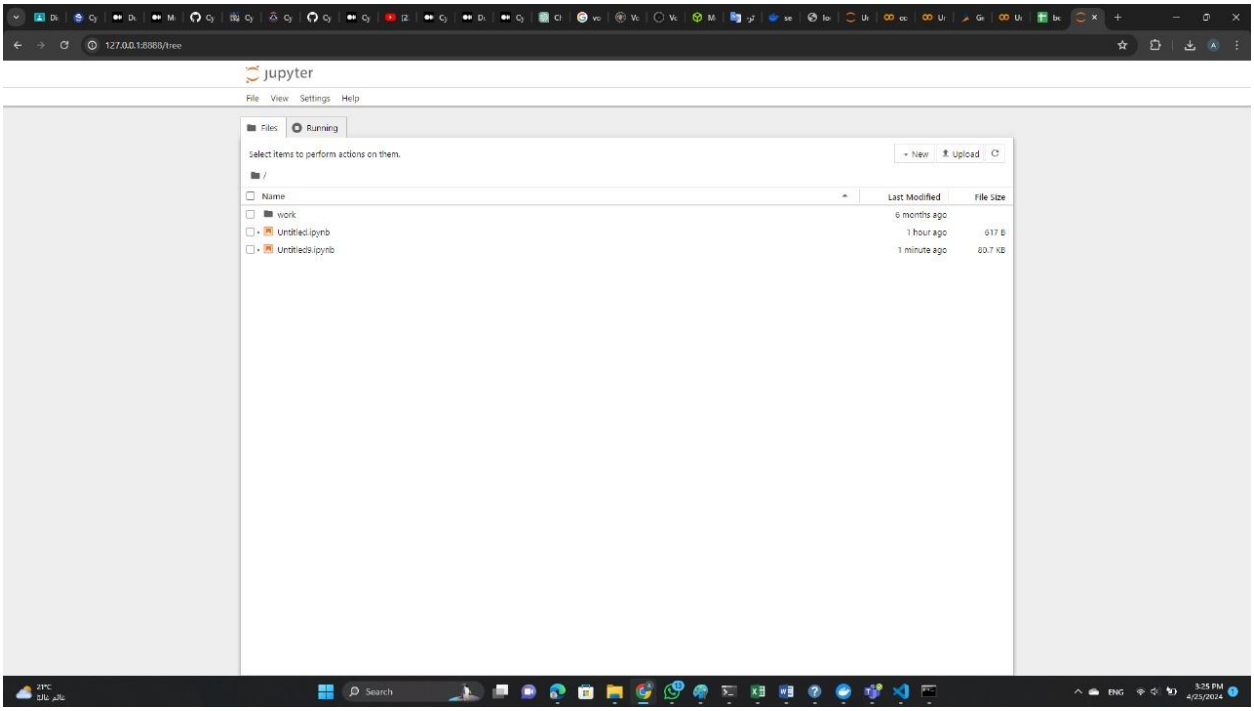
And this is the result of run.



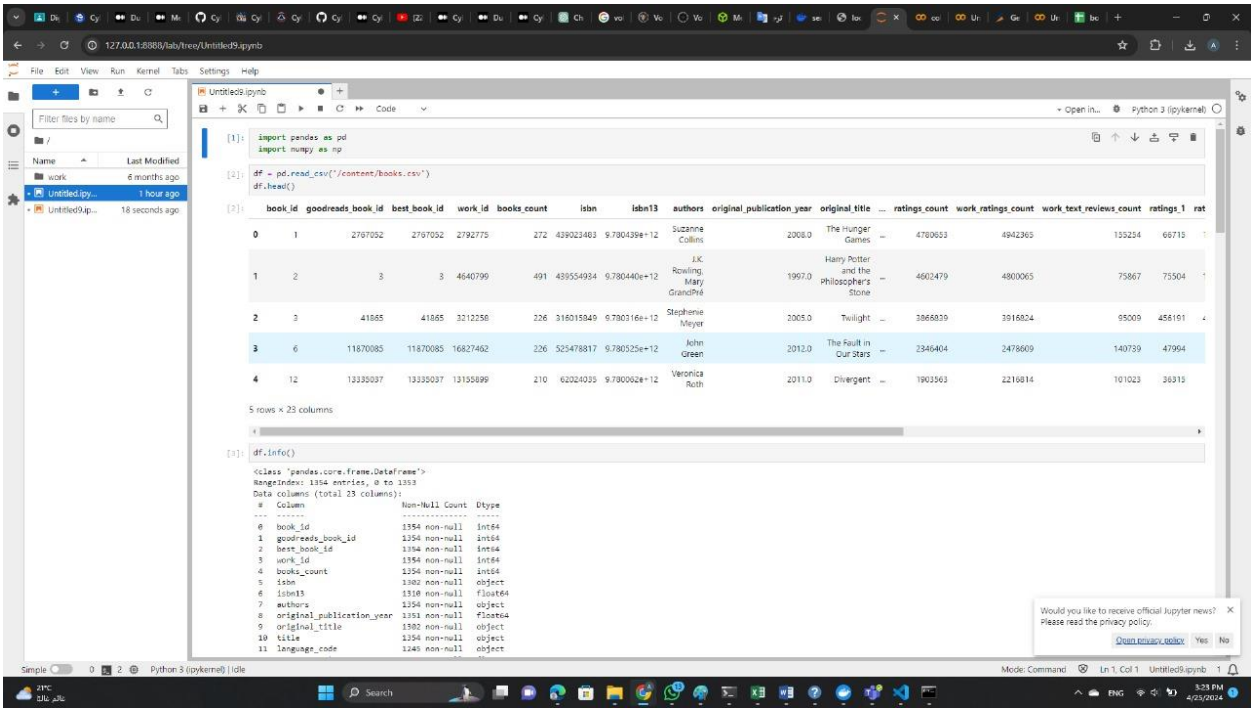
Open Docker to run the code and get the notebook link to write the code to clean the data preprocessing.

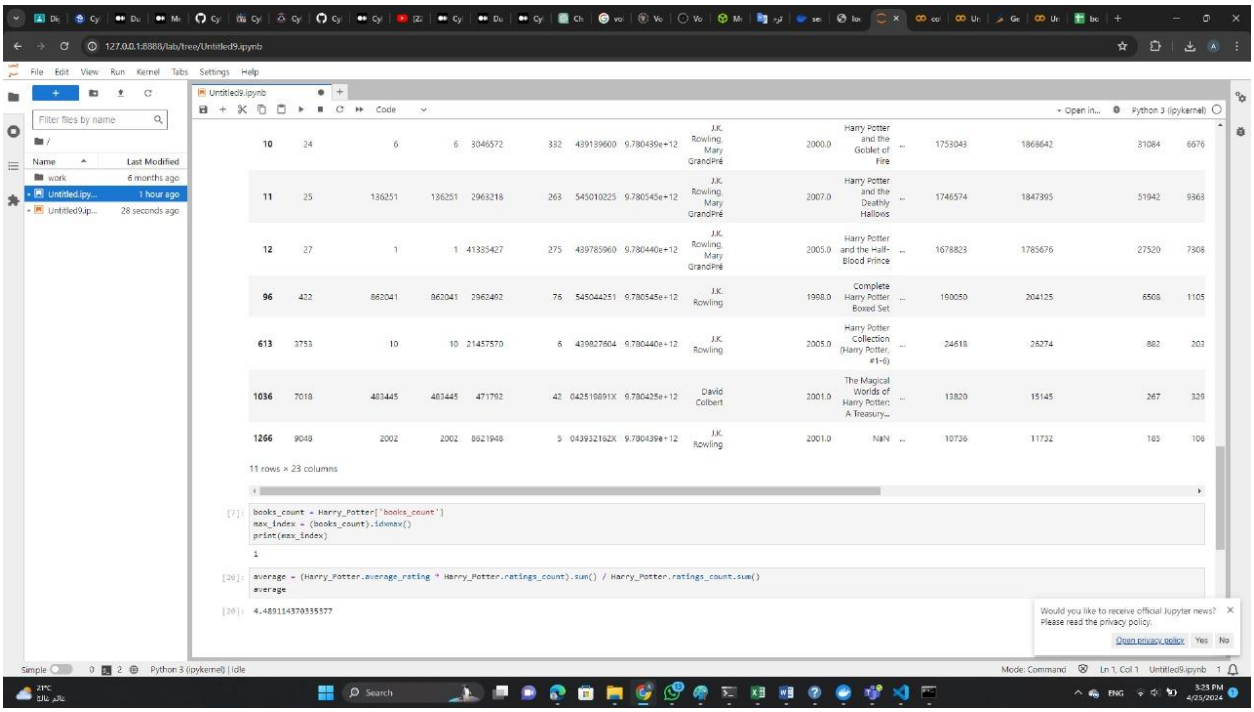


When we click here in link that window will open.



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 Tabs Settings Help
Untitled9.ipynb
Code
Open in...
Python 3 (ipykernel)
Filter files by name
Name
Last Modified
work
6 months ago
Untitled.ipy...
1 hour ago
Untitled9.ip...
18 seconds ago
[1]:
import pandas as pd
import numpy as np
[2]:
df = pd.read_csv('/content/books.csv')
df.head()
[2]:
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 2767052 2767052 2792775 272 439023483 9.780439e+12 Suzanne Collins 2008.0 The Hunger Games ... 4780653 4942365 155254 66715
1 2 3 3 4640799 491 439554934 9.780440e+12 J.K. Rowling, Mary GrandPré 1997.0 Harry Potter and the Philosopher's Stone ... 4602479 4800065 75867 75504
2 3 41865 41865 3212258 226 316015849 9.780316e+12 Stephenie Meyer 2005.0 Twilight ... 3866839 3916824 95009 456191
3 6 11870085 11870085 16827462 226 525478817 9.780525e+12 John Green 2012.0 The Fault in Our Stars ... 2346404 2478609 140739 47994
4 12 13335037 13335037 13155899 210 62024035 9.780062e+12 Veronica Roth 2011.0 Divergent ... 1903563 2216814 101023 38315
5 rows × 23 columns
[3]:
df.info()
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 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 1316 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 1240 non-null object
Would you like to receive official Jupyter news?
Please read the privacy policy.
Open privacy policy
Yes
No
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
Untitled9.ipynb
Search
ENG
4/25/2024


127.0.0.1:8888/lab/tree/Untitled9.ipynb
File Edit View Run Kernel Tabs Settings Help
Untitled9.ipynb
Code
Open in...
Python 3 (ipykernel)
Filter files by name
Name
Last Modified
work
6 months ago
Untitled.ipy...
1 hour ago
Untitled9.ip...
28 seconds ago
10 24 6 6 3046572 332 439139600 9.780439e+12 J.K. Rowling, Mary GrandPré 2000.0 Harry Potter and the Goblet of Fire ... 1753043 1868642 31084 6676
11 25 136251 136251 2963218 263 545010225 9.780545e+12 J.K. Rowling, Mary GrandPré 2007.0 Harry Potter and the Deathly Hallows ... 1746574 1847395 51942 9363
12 27 1 1 41335427 275 439785960 9.780440e+12 J.K. Rowling, Mary GrandPré 2005.0 Harry Potter and the Half-Blood Prince ... 1678823 1785676 27520 7308
96 422 862041 862041 2963462 76 545044251 9.780545e+12 J.K. Rowling 1998.0 Complete Harry Potter Boxed Set ... 190050 204125 6508 1105
613 3753 10 10 21457570 6 439827604 9.780440e+12 J.K. Rowling 2005.0 Harry Potter Collection (Harry Potter, #1-6) ... 24618 26274 862 203
1036 7018 483445 483445 471792 42 042519891X 9.780425e+12 David Colbert 2001.0 The Magical Worlds of Harry Potter: A Treasury... ... 13820 15145 267 329
1266 9048 2002 2002 8621948 5 043932162X 9.780439e+12 J.K. Rowling 2001.0 NaN ... 10736 11732 165 106
11 rows × 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
[20]:
4.489114370335377
Would you like to receive official Jupyter news?
Please read the privacy policy.
Open privacy policy
Yes
No
Simple
Python 3 (ipykernel) | Idle
Mode: Command
Ln 1, Col 1
Untitled9.ipynb
Search
ENG
4/25/2024