Name: Marina George Zarif

ID:20221443430

Department: Intelligent System

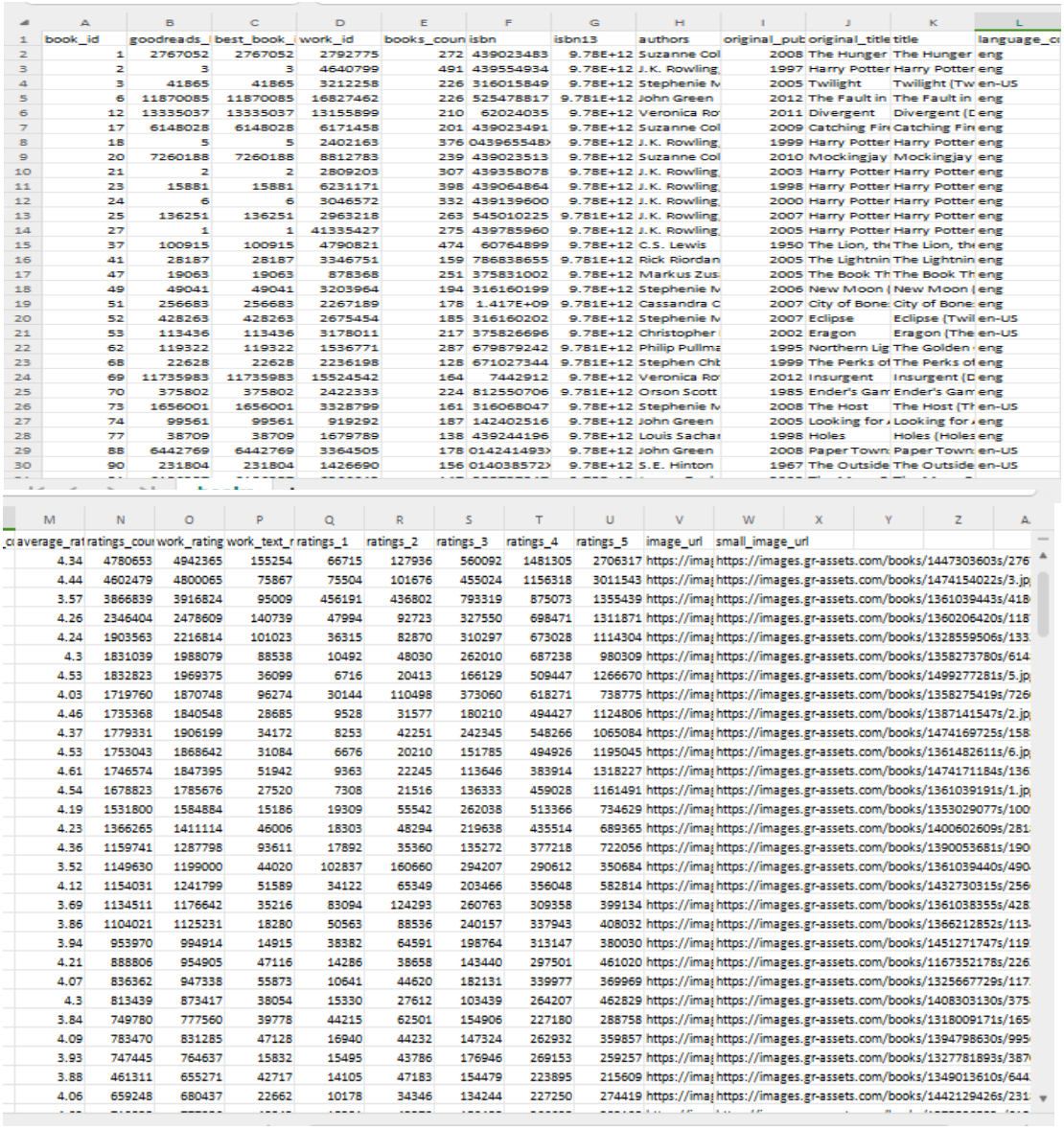
Assignment Cloud Computing

Eng\Sarah

1) First I download the dataset from Kaggle

"Popular Books Dataset"

This file contains information about popular books. Each row represents a book and includes details such as the book ID, Goodreads book ID, best book ID, work ID, number of books in the series, ISBN and ISBN13 numbers, authors, original publication year, original title, title, language code, average rating, number of ratings, number of work ratings, number of text reviews, rating counts for different rating levels (1-5 stars), and URLs for the book's image and small image.



book_id	goodreads_	best_book_	work_id	books_coun	isbn	isbn13	authors	original_pub	original_title	title	language_c
1	2767052	2767052	2792775	272	439023483	9.78E+12	Suzanne Col	2008	The Hunger	The Hunger	eng
2	3	3	4640799	491	439554934	9.78E+12	J.K. Rowling	1997	Harry Potter	Harry Potter	eng
3	41865	41865	3212258	226	316015849	9.78E+12	Stephenie M	2005	Twilight	Twilight (Tw	en-US
6	11870085	11870085	16827462	226	525478817	9.78E+12	John Green	2012	The Fault in	The Fault in	eng
12	13335037	13335037	13155899	210	62024035	9.78E+12	Veronica Ro	2011	Divergent	Divergent (D	eng
17	6148028	6148028	6171458	201	439023491	9.78E+12	Suzanne Col	2009	Catching Fir	Catching Fir	eng
18	5	5	2402163	376	043965548X	9.78E+12	J.K. Rowling	1999	Harry Potter	Harry Potter	eng
20	7260188	7260188	8812783	239	439023513	9.78E+12	Suzanne Col	2010	Mockingjay	Mockingjay	eng
21	2	2	2809203	307	439358078	9.78E+12	J.K. Rowling	2003	Harry Potter	Harry Potter	eng
23	15881	15881	6231171	398	439064864	9.78E+12	J.K. Rowling	1998	Harry Potter	Harry Potter	eng
24	6	6	3046572	332	439139600	9.78E+12	J.K. Rowling	2000	Harry Potter	Harry Potter	eng
25	136251	136251	2963218	263	545010225	9.78E+12	J.K. Rowling	2007	Harry Potter	Harry Potter	eng
27	1	1	41335427	275	439785960	9.78E+12	J.K. Rowling	2005	Harry Potter	Harry Potter	eng
37	100915	100915	4790821	474	60764899	9.78E+12	C.S. Lewis	1950	The Lion, th	The Lion, th	eng
41	28187	28187	3346751	159	786838655	9.78E+12	Rick Riordan	2005	The Lightnin	The Lightnin	eng
47	19063	19063	878368	251	375831002	9.78E+12	Markus Zus	2005	The Book Th	The Book Th	eng
49	49041	49041	3203964	194	316160199	9.78E+12	Stephenie M	2006	New Moon (	New Moon (	eng
51	256683	256683	2267189	178	1.417E+09	9.78E+12	Cassandra C	2007	City of Bone	City of Bone	eng
52	428263	428263	2675454	185	316160202	9.78E+12	Stephenie M	2007	Eclipse	Eclipse (Twil	en-US
53	113436	113436	3178011	217	375826696	9.78E+12	Christopher	2002	Eragon	Eragon (The	en-US
62	119322	119322	1536771	287	679879242	9.78E+12	Philip Pullma	1995	Northern Lig	The Golden	eng
68	22628	22628	2236198	128	671027344	9.78E+12	Stephen Chb	1999	The Perks of	The Perks of	eng
69	11735983	11735983	15524542	164	7442912	9.78E+12	Veronica Ro	2012	Insurgent	Insurgent (D	eng
70	375802	375802	2422333	224	812550706	9.78E+12	Orson Scott	1985	Ender's Gam	Ender's Gam	eng
73	1656001	1656001	3328799	161	316068047	9.78E+12	Stephenie M	2008	The Host	The Host (TI	en-US
74	99561	99561	919292	187	142402516	9.78E+12	John Green	2005	Looking for	Looking for	eng
77	38709	38709	1679789	138	439244196	9.78E+12	Louis Sachar	1998	Holes	Holes (Holes	eng
88	6442769	6442769	3364505	178	014241493X	9.78E+12	John Green	2008	Paper Town	Paper Town	en-US
90	231804	231804	1426690	156	014038572X	9.78E+12	S.E. Hinton	1967	The Outside	The Outside	en-US

average_rat	ratings_cou	work_rating	work_text_r	ratings_1	ratings_2	ratings_3	ratings_4	ratings_5	image_url	small_image_url
4.34	4780653	4942365	155254	66715	127936	560092	1481305	2706317	https://ima	https://images.gr-assets.com/books/1447303603s/276
4.44	4602479	4800065	75867	75504	101676	455024	1156318	3011543	https://ima	https://images.gr-assets.com/books/1474154022s/3.jp
3.57	3866839	3916824	95009	456191	436802	793319	875073	1355439	https://ima	https://images.gr-assets.com/books/1361039443s/418
4.26	2346404	2478609	140739	47994	92723	327550	698471	1311871	https://ima	https://images.gr-assets.com/books/1360206420s/118
4.24	1903563	2216814	101023	36315	82870	310297	673028	1114304	https://ima	https://images.gr-assets.com/books/1328559506s/133
4.3	1831039	1988079	88538	10492	48030	262010	687238	980309	https://ima	https://images.gr-assets.com/books/1358273780s/614
4.53	1832823	1969375	36099	6716	20413	166129	509447	1266670	https://ima	https://images.gr-assets.com/books/1499277281s/5.jp
4.03	1719760	1870748	96274	30144	110498	373060	618271	738775	https://ima	https://images.gr-assets.com/books/1358275419s/726
4.46	1735368	1840548	28685	9528	31577	180210	494427	1124806	https://ima	https://images.gr-assets.com/books/1387141547s/2.jp
4.37	1779331	1906199	34172	8253	42251	242345	548266	1065084	https://ima	https://images.gr-assets.com/books/1474169725s/158
4.53	1753043	1868642	31084	6676	20210	151785	494926	1195045	https://ima	https://images.gr-assets.com/books/1361482611s/6.jp
4.61	1746574	1847395	51942	9363	22245	113646	383914	1318227	https://ima	https://images.gr-assets.com/books/1474171184s/136
4.54	1678823	1785676	27520	7308	21516	136333	459028	1161491	https://ima	https://images.gr-assets.com/books/1361039191s/1.jp
4.19	1531800	1584884	15186	19309	55542	262038	513366	734629	https://ima	https://images.gr-assets.com/books/1353029077s/100
4.23	1366265	1411114	46006	18303	48294	219638	435514	689365	https://ima	https://images.gr-assets.com/books/1400602609s/281
4.36	1159741	1287798	93611	17892	35360	135272	377218	722056	https://ima	https://images.gr-assets.com/books/1390053681s/190
3.52	1149630	1199000	44020	102837	160660	294207	290612	350684	https://ima	https://images.gr-assets.com/books/1361039440s/490
4.12	1154031	1241799	51589	34122	65349	203466	356048	582814	https://ima	https://images.gr-assets.com/books/1432730315s/256
3.69	1134511	1176642	35216	83094	124293	260763	309358	399134	https://ima	https://images.gr-assets.com/books/1361038355s/428
3.86	1104021	1125231	18280	50563	88536	240157	337943	408032	https://ima	https://images.gr-assets.com/books/1366212852s/113
3.94	953970	994914	14915	38382	64591	198764	313147	380030	https://ima	https://images.gr-assets.com/books/1451271747s/119
4.21	888806	954905	47116	14286	38658	143440	297501	461020	https://ima	https://images.gr-assets.com/books/1167352178s/226
4.07	836362	947338	55873	10641	44620	182131	339977	369969	https://ima	https://images.gr-assets.com/books/1325667729s/117
4.3	813439	873417	38054	15330	27612	103439	264207	462829	https://ima	https://images.gr-assets.com/books/1408303130s/375
3.84	749780	777560	39778	44215	62501	154906	227180	288758	https://ima	https://images.gr-assets.com/books/1318009171s/165
4.09	783470	831285	47128	16940	44232	147324	262932	359857	https://ima	https://images.gr-assets.com/books/1394798630s/995
3.93	747445	764637	15832	15495	43786	176946	269153	259257	https://ima	https://images.gr-assets.com/books/1327781893s/387
3.88	461311	655271	42717	14105	47183	154479	223895	215609	https://ima	https://images.gr-assets.com/books/1349013610s/644
4.06	659248	680437	22662	10178	34346	134244	227250	274419	https://ima	https://images.gr-assets.com/books/1442129426s/231

2) Second I made the jupyter notebook that contains the data cleaning and preprocessing of the data:

Dockerfile app.ipynb

python_image > app.ipynb > ...

+ Code + Markdown | Run All | Clear All Outputs | Outline ··· Python 3.12.2

```
import pandas as pd
```
[1] Python

```
df = pd.read_csv("books.csv")
df.head()
```
[20] Python

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...

5 rows × 23 columns

```
df.shape
```
[21]

```
(1354, 23)
```

```
df.columns
```
[24]

```
Index(['book_id', 'goodreads_book_id', 'best_book_id', 'work_id',
       'books_count', 'isbn', 'isbn13', 'authors', 'original_publication_year',
       'original_title', 'title', 'language_code', 'average_rating',
       'ratings_count', 'work_ratings_count', 'work_text_reviews_count',
       'ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5',
       'image_url', 'small_image_url'],
      dtype='object')
```

```
columns_to_drop = ['work_id', 'isbn', 'isbn13', 'image_url', 'small_image_url']
df = df.drop(columns=columns_to_drop)
```
[13]

```
rows_with_null = df[df.isnull().any(axis=1)]
print("\nRows with Null Values:")
print(rows_with_null)
```
[14]

```
[15] df.dropna(inplace=True)
```
Python

```
[16] df.shape
```
Python

```
(1197, 18)
```

```
[18] duplicate_rows = df[df.duplicated()]
     print("Duplicate Rows:")
     print(duplicate_rows)
```
Python

```
Duplicate Rows:
Empty DataFrame
Columns: [book_id, goodreads_book_id, best_book_id, books_count, authors, original_publication_year, original_title, title, language_code, a
Index: []
```

```
[30] harry_potter_df = df[df['title'].str.contains('Harry Potter')]
```
Python

Dockerfile | app.ipynb

python_image > app.ipynb > ...

+ Code | + Markdown | Run All | Clear All Outputs | Outline | ... | Python 3.12

```
[31] harry_potter_df.shape
```
Python

```
(11, 23)
```

```
[32] harry_potter_df.drop_duplicates(inplace=True)
```
Python

```
<ipython-input-32-3be7da392e4d>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  harry_potter_df.drop_duplicates(inplace=True)
```

```
[33] harry_potter_df.shape
```
Python

```
(11, 23)
```

```
[41] harry_potter_df['total_ratings'] = harry_potter_df['ratings_1'] + harry_potter_df['ratings_2'] + harry_potter_df['ratings_3'] + harry_po
```
Python

```
<ipython-input-41-086740a25b8f>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame
```

```
most_selling_books = harry_potter_df.groupby('title')['total_ratings'].sum()
most_selling_books
```

[43] Python

```
title
Harry Potter Boxset (Harry Potter, #1-7)                                                                  204125
Harry Potter Collection (Harry Potter, #1-6)                                                               26274
Harry Potter Schoolbooks Box Set: Two Classic Books from the Library of Hogwarts School of Witchcraft and Wizardry    11732
Harry Potter and the Chamber of Secrets (Harry Potter, #2)                                               1906199
Harry Potter and the Deathly Hallows (Harry Potter, #7)                                                  1847395
Harry Potter and the Goblet of Fire (Harry Potter, #4)                                                   1868642
Harry Potter and the Half-Blood Prince (Harry Potter, #6)                                                1785676
Harry Potter and the Order of the Phoenix (Harry Potter, #5)                                             1840548
Harry Potter and the Prisoner of Azkaban (Harry Potter, #3)                                              1969375
Harry Potter and the Sorcerer's Stone (Harry Potter, #1)                                                 4800065
The Magical Worlds of Harry Potter: A Treasury of Myths, Legends, and Fascinating Facts                    15145
Name: total_ratings, dtype: int64
```

```
harry_potter_df['average_rating'] = harry_potter_df['average_rating'].astype(float)
```

[38] Python

```
<ipython-input-38-39a35a381676>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
```

Dockerfile | app.ipynb

python_image > app.ipynb > ...

+ Code + Markdown | Run All | Clear All Outputs | Outline ... Python 3.12.

[38] Python

```
<ipython-input-38-39a35a381676>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  harry_potter_df['average_rating'] = harry_potter_df['average_rating'].astype(float)
```

```
harry_potter_df['average_rating'] = harry_potter_df[['ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5']].mean(axis=1)
```

[45] Python

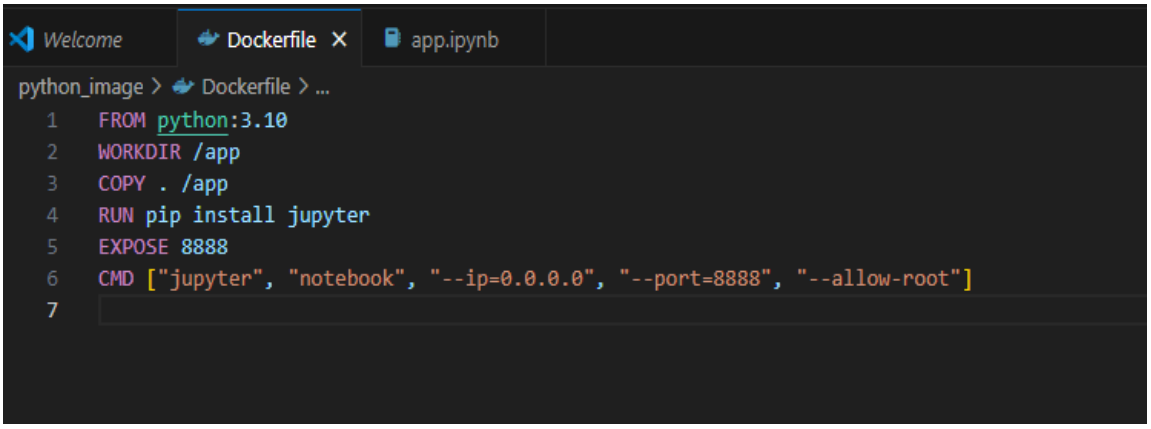
```
<ipython-input-45-7fb65d8fc19a>:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
  harry_potter_df['average_rating'] = harry_potter_df[['ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5']].mean(axis=1)
```

```
average_rating = harry_potter_df['average_rating'].mean()
average_rating
```

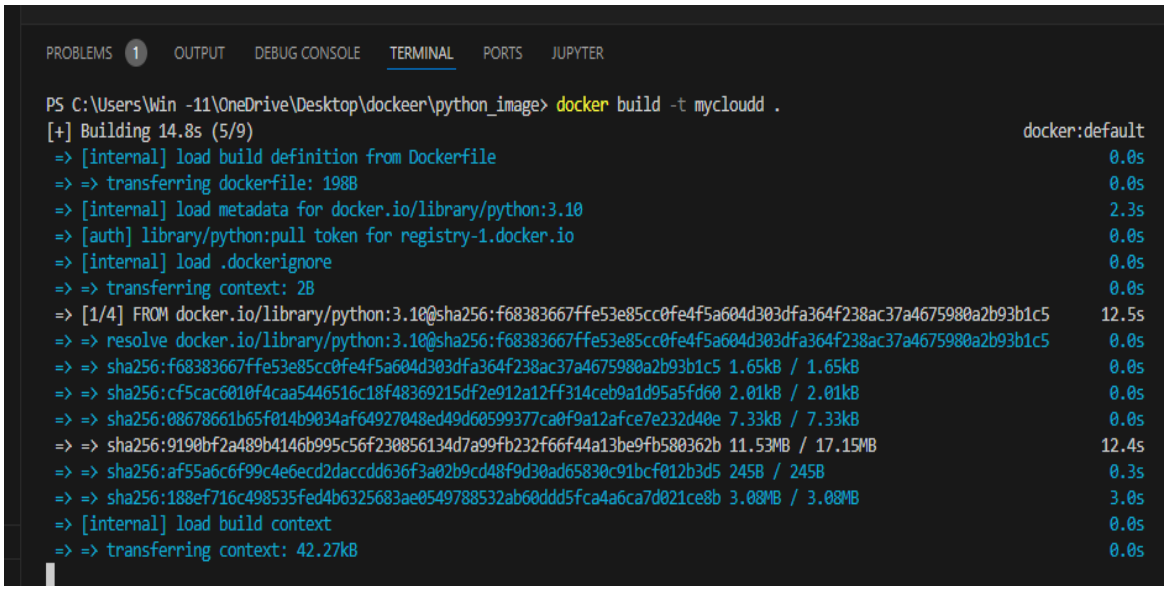
[46] Python

```
295912.2909090909
```

3) After that I made a docker file that connects to python and jupyter notebook

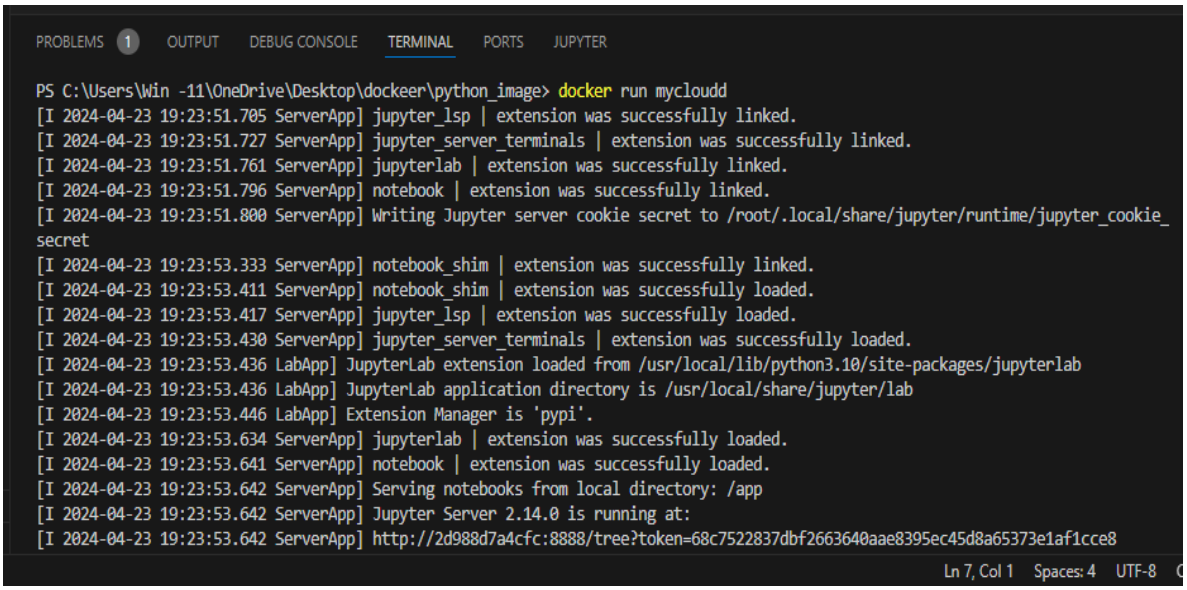


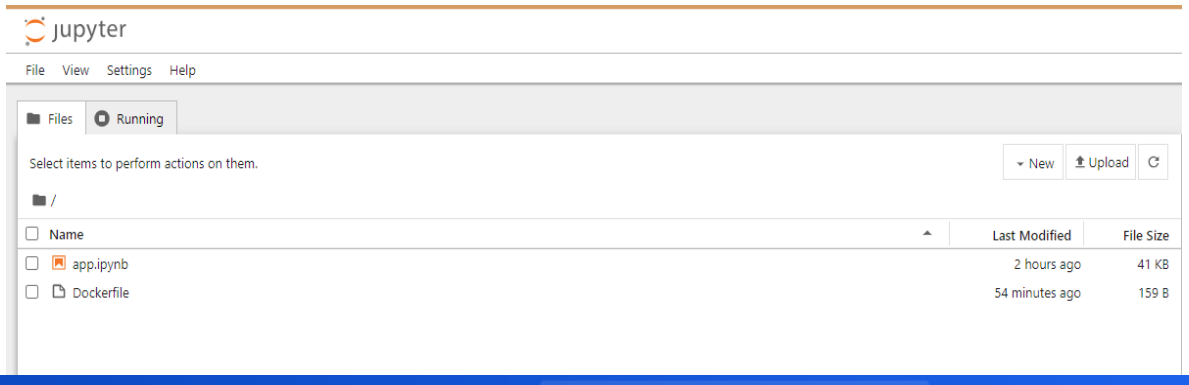
- Build docker file that have the name mucloudd

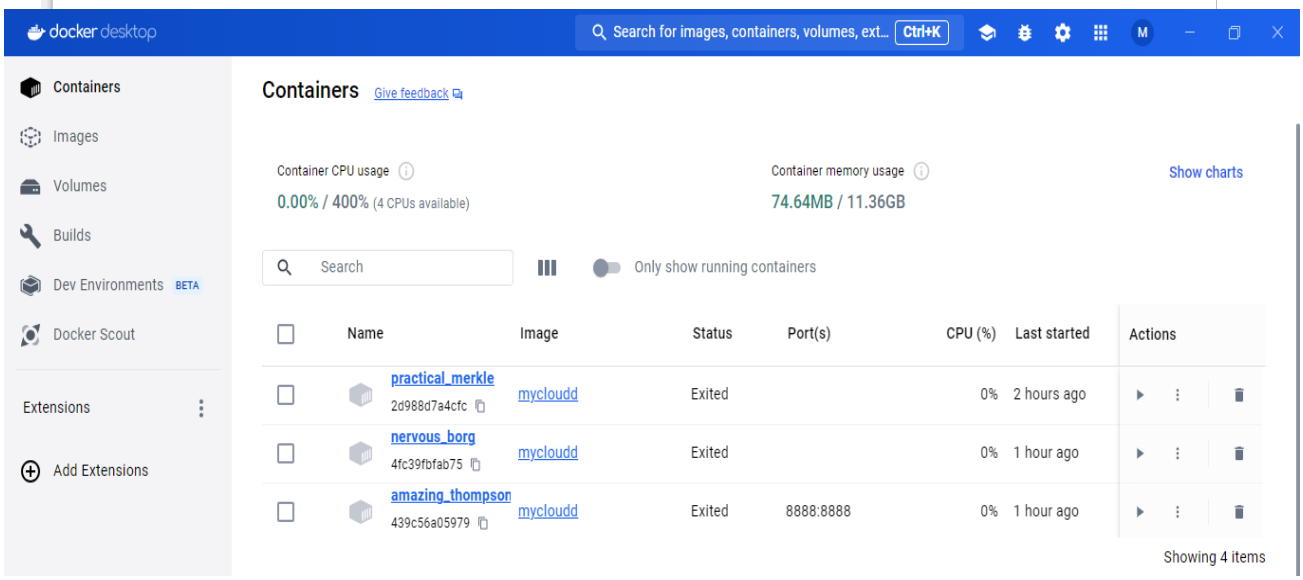


-

Run docker


PROBLEMS 1
OUTPUT
DEBUG CONSOLE
TERMINAL
PORTS
JUPYTER
PS C:\Users\Win -11\OneDrive\Desktop\dockeer\python_image> docker run mycloudd
[I 2024-04-23 19:23:51.705 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-04-23 19:23:51.727 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-04-23 19:23:51.761 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-04-23 19:23:51.796 ServerApp] notebook | extension was successfully linked.
[I 2024-04-23 19:23:51.800 ServerApp] Writing Jupyter server cookie secret to /root/.local/share/jupyter/runtime/jupyter_cookie_secret
[I 2024-04-23 19:23:53.333 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-04-23 19:23:53.411 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-04-23 19:23:53.417 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-04-23 19:23:53.430 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-04-23 19:23:53.436 LabApp] JupyterLab extension loaded from /usr/local/lib/python3.10/site-packages/jupyterlab
[I 2024-04-23 19:23:53.436 LabApp] JupyterLab application directory is /usr/local/share/jupyter/lab
[I 2024-04-23 19:23:53.446 LabApp] Extension Manager is 'pypi'.
[I 2024-04-23 19:23:53.634 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-04-23 19:23:53.641 ServerApp] notebook | extension was successfully loaded.
[I 2024-04-23 19:23:53.642 ServerApp] Serving notebooks from local directory: /app
[I 2024-04-23 19:23:53.642 ServerApp] Jupyter Server 2.14.0 is running at:
[I 2024-04-23 19:23:53.642 ServerApp] http://2d988d7a4cfc:8888/tree?token=68c7522837dbf2663640aae8395ec45d8a65373e1af1cce8
Ln 7, Col 1
Spaces: 4
UTF-8


jupyter
File
View
Settings
Help
Files
Running
Select items to perform actions on them.
New
Upload
/
Name
Last Modified
File Size
app.ipynb
2 hours ago
41 KB
Dockerfile
54 minutes ago
159 B


docker desktop
Search for images, containers, volumes, ext...
Ctrl+K
Containers
Images
Volumes
Builds
Dev Environments
BETA
Docker Scout
Extensions
Add Extensions
Containers
Give feedback
Container CPU usage
0.00% / 400% (4 CPUs available)
Container memory usage
74.64MB / 11.36GB
Show charts
Search
Only show running containers
Name
Image
Status
Port(s)
CPU (%)
Last started
Actions
practical_merkle
2d988d7a4cfc
mycloudd
Exited
0%
2 hours ago
nervous_borg
4fc39fbfab75
mycloudd
Exited
0%
1 hour ago
amazing_thompson
439c56a05979
mycloudd
Exited
8888:8888
0%
1 hour ago
Showing 4 items

- **Find docker images**