**Docker Assignment**

**Harry Potter Books Analysis**

In this report I will be discussing the steps I went through to run my image in addition to Discussing the analysis about Harry Potter Books in jupyter Notebook.

1) Pull the Image

In this first step I pulled the image specified in the assignment

docker pull jupyter/datascience-notebook

so that it would be available for me to use it .

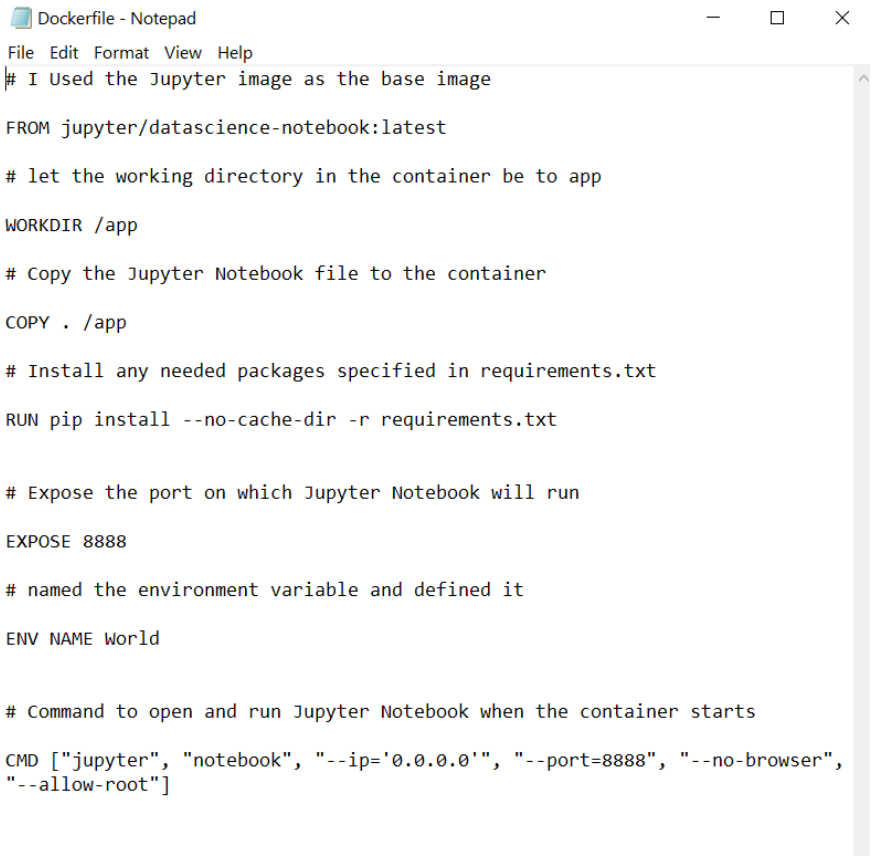
```
Microsoft Windows [Version 10.0.19045.4291]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Mega Store>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/datascien
ce-notebook
```

2) Dockerfile

In the second step I will prepare the Dockerfile. In my case I wrote the docker file in the notepad .




I used the Jupyter Docker image as a base image and exposed the port on which Jupyter Notebook will run, exposed port 8888, which is the default port used by Jupyter.

I called the file Dockerfile and changed its type from txt to file.

WORKDIR is a Dockerfile instruction used to set the working directory for instructions in the Dockerfile. It changes the working directory within the Docker container, like the cd command in Linux. The WORKDIR instruction only affects commands in the Dockerfile. It doesn't create a directory on my host machine it just sets the directory within the Docker container.

3) Copy path of Dockerfile

In the third step I copied the path to the Docker file I created in the previous step and used it in the command prompt to access this location.

```
C:\Users\Mega Store>cd "C:\Users\Mega Store\Documents\potterdocker"
```

4) Build the Image

After that I built the image, and named it dockerpotterimage.

docker build -t dockerpotterimage .

```
Microsoft Windows [Version 10.0.19045.4291]
(c) Microsoft Corporation. All rights reserved.

C:\Users\Mega Store>cd "C:\Users\Mega Store\Documents\potterdocker"

C:\Users\Mega Store\Documents\potterdocker>docker build -t dockerpotterimage .
[+] Building 0.3s (9/9) FINISHED                                                                    docker:default
 => [internal] load build definition from Dockerfile                                                         0.0s
 => => transferring dockerfile: 701B                                                                         0.0s
 => [internal] load metadata for docker.io/jupyter/datascience-notebook:latest                               0.0s
 => [internal] load .dockerignore                                                                            0.0s
 => => transferring context: 2B                                                                              0.0s
 => [1/4] FROM docker.io/jupyter/datascience-notebook:latest                                                 0.0s
 => [internal] load build context                                                                            0.0s
 => => transferring context: 143B                                                                            0.0s
 => CACHED [2/4] WORKDIR /app                                                                                0.0s
 => CACHED [3/4] COPY . /app                                                                                 0.0s
 => CACHED [4/4] RUN pip install --no-cache-dir -r requirements.txt                                          0.0s
 => exporting to image                                                                                       0.0s
 => => exporting layers                                                                                      0.0s
 => => writing image sha256:00d64ac435fddefe7db1db75c797835c1afc35dcbfe0ee671d97b5bc32e3657c                 0.0s
 => => naming to docker.io/library/dockerpotterimage                                                         0.0s

View build details: docker-desktop://dashboard/build/default/default/c6r3cdsuefwx6wwzd1u7gswm6
```

5) Run the image

Then I ran the Image that I created and defined the port

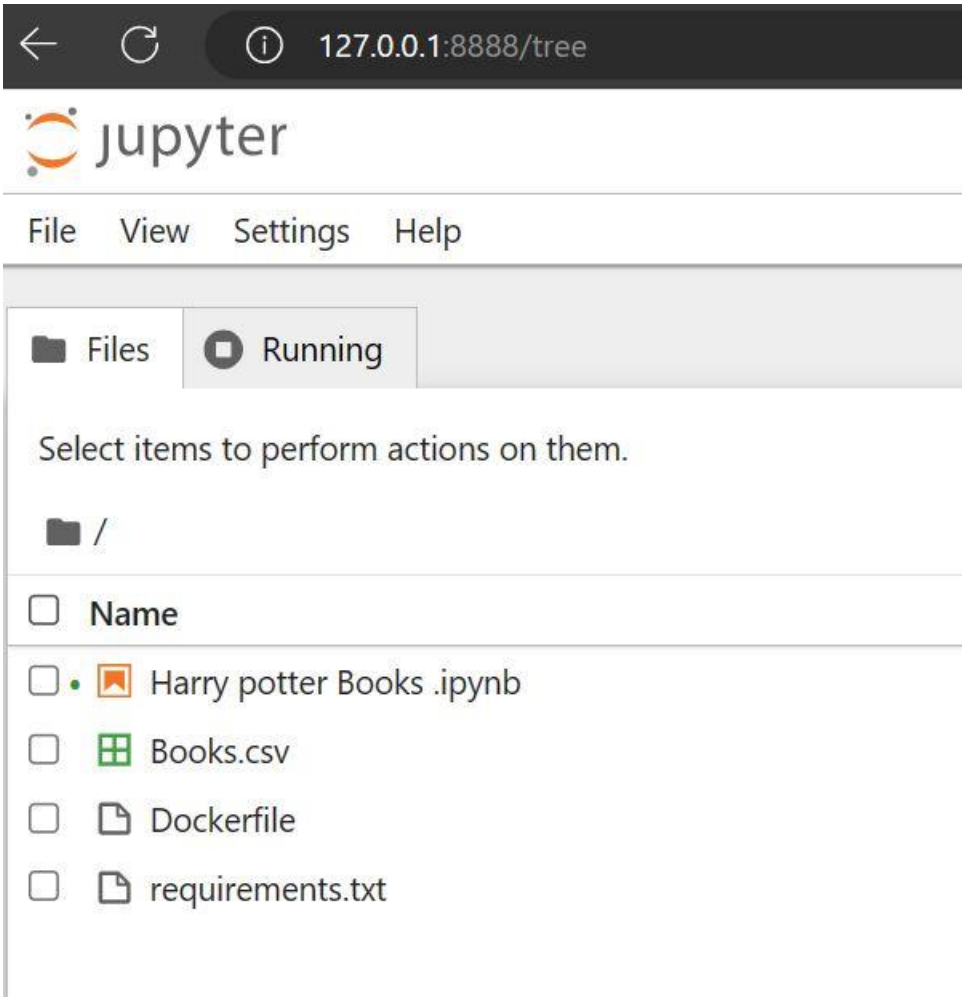
```
C:\Users\Mega Store\Documents\potterdocker>docker run -p 8888:8888 dockerpotterimage
[I 2024-04-23 19:38:47.409 ServerApp] Package notebook took 0.0000s to import
```

6) use URL

After running my image, the command prompt output a URL to be copied and used to open my Jupyter Notebook.

```
To access the server, open this file in a browser:
    file:///home/jovyan/.local/share/jupyter/runtime/jpserver-7-open.html
Or copy and paste one of these URLs:
    http://2e98d3e3446b:8888/tree?token=d162ebf484f675fadd95a2b2ea503d675224be60888ecb97
    http://127.0.0.1:8888/tree?token=d162ebf484f675fadd95a2b2ea503d675224be60888ecb97
```

After I pasted the URL in the browser, a Jupyter opened with my notebook including the analysis And the Books data set afterthat I ran the code.



# ***Harry Potter Books Analysis from Books Dataset***

*As for the python analysis on my jupyter notebook , detailed comments were included for each step.*

*And the answers for the questions regarding the ratings and most selling where:*

1)Find the most selling books within the Harry Potter series.

I arranged Harry Potter Books in order from most Book count to least Book count in order to Know the order of the most selling first going down to the least selling.

Finding the Most selling Harry Potter Books

```
[87]: bestsell=harrypotter.sort_values(by='books_count',ascending=False)
      bestsell
```

[87]:

| | book_id | goodreads_book_id | best_book_id | work_id | books_count | isbn | isbn13 | authors | original_publica |
|---|---|---|---|---|---|---|---|---|---|
| **1** | 2 | 3 | 3 | 4640799 | 491 | 439554934 | 9.780440e+12 | J.K. Rowling, Mary GrandPré | |
| **9** | 23 | 15881 | 15881 | 6231171 | 398 | 439064864 | 9.780439e+12 | J.K. Rowling, Mary GrandPré | |
| | | | | | | | | J.K. | |

Then I displayed the output with only the name of the book and its book count.

the most selling books for harry potter in order are:

| | title | books_count |
|---|---|---|
| 1 | Harry Potter and the Sorcerer's Stone (Harry P... | 491 |
| 9 | Harry Potter and the Chamber of Secrets (Harry... | 398 |
| 6 | Harry Potter and the Prisoner of Azkaban (Harr... | 376 |
| 10 | Harry Potter and the Goblet of Fire (Harry Pot... | 332 |
| 8 | Harry Potter and the Order of the Phoenix (Har... | 307 |
| 12 | Harry Potter and the Half-Blood Prince (Harry ... | 275 |
| 11 | Harry Potter and the Deathly Hallows (Harry Po... | 263 |

```
print("the most selling books for harry potter in order are:")
print(bestsell[['title','books_count']])
```

```
the most selling books for harry potter in order are:
                                                title  books_count
1    Harry Potter and the Sorcerer's Stone (Harry P...          491
9    Harry Potter and the Chamber of Secrets (Harry...          398
6    Harry Potter and the Prisoner of Azkaban (Harr...          376
10   Harry Potter and the Goblet of Fire (Harry Pot...          332
8    Harry Potter and the Order of the Phoenix (Har...          307
12   Harry Potter and the Half-Blood Prince (Harry ...          275
11   Harry Potter and the Deathly Hallows (Harry Po...          263
```

2)Calculate the average rating of the Harry Potter books.

As for the average rating, there were 2 ways to calculate it . and it was within 4.486117244334694 or 4.497142857142857

[84]:
```
avg_r=avg_ratings['average_rating'].sum()/avg_ratings['average_rating'].count()
avg_r
```

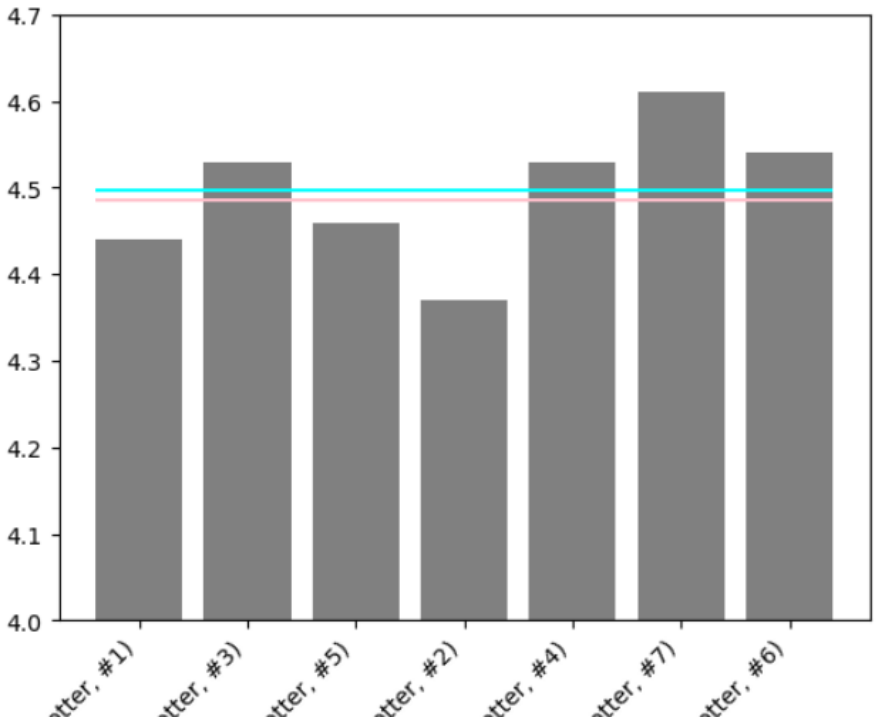
[84]:
```
4.497142857142857
```

[85]:
```
#another way to calculate the avg
weighted_avg = (harrypotter.average_rating * harrypotter.ratings_count).sum() / harrypotter.ratings_count.sum()
weighted_avg
```

[85]:
```
4.486117244334694
```

[86]:
```
#display the average of each book for visualization
plt.bar(avg_ratings.title,avg_ratings.average_rating,color='grey')
plt.xticks(rotation=45, ha='right')
# I added thes previous line because the name was too long to be displayed horizontaly on x axis
plt.ylim((4, 4.7))
plt.hlines(weighted_avg, xmin=-0.4, xmax=6.4, color='pink')
plt.hlines(avg_r, xmin=-0.4, xmax=6.4, color='cyan')
```

After I calculated the AVERAGE rating, I drew a line of the average compared to the rating of each Book to be able to visualize the difference.


4.7
4.6
4.5
4.4
4.3
4.2
4.1
4.0
tter, #1)
tter, #3)
tter, #5)
tter, #2)
tter, #4)
tter, #7)
tter, #6)