

Name: Mickel Wassef Riad Wassef – ID : 22010449

Assignment 2

Dockerfile

Dockerfile > ...

```
1 FROM jupyter/datascience-notebook
2
3 WORKDIR /usr/src/app
4
5 COPY . /usr/src/app
6
7 RUN pip install -r requirements.txt
8
9 EXPOSE 8888
10
11 CMD ["jupyter", "notebook", "--ip='0.0.0.0'", "--port=8888", "--no-browser", "--allow-root"]
12
```

Windows PowerShell

```
PS C:\Users\LENOVO\Desktop\Assignment 2> docker build -t my-jupyter-notebook .
[+] Building 0.0s (0/0) docker:default
2024/04/24 12:04:57 http2: server: error reading preface from client //./pipe/docker_engine: file has already been close
[+] Building 3.5s (9/9) FINISHED docker:default
=> [internal] load build definition from Dockerfile 0.0s
=> => transferring dockerfile: 277B 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: 5.06kB 0.0s
=> CACHED [2/4] WORKDIR /usr/src/app 0.0s
=> [3/4] COPY . /usr/src/app 0.1s
=> [4/4] RUN pip install -r requirements.txt 2.9s
=> exporting to image 0.2s
=> => exporting layers 0.1s
=> => writing image sha256:2a5884b97e80191e1912487c2e7e0b1ef90eb0d1358ab51c618aa1a245a5464f 0.0s
=> => naming to docker.io/library/my-jupyter-notebook 0.0s
```

View build details: docker-desktop://dashboard/build/default/default/tc03ryf0pd9azecjta86ox3ed

What's Next?

View a summary of image vulnerabilities and recommendations → [docker scout quickview](#)

PS C:\Users\LENOVO\Desktop\Assignment 2> |

127.0.0.1:8888/tree

jupyter

File View Settings Help

Files Running

Select items to perform actions on them.

/

<input type="checkbox"/> Name	Last Modified	File Size
<input type="checkbox"/> data analysis.ipynb	2 minutes ago	4.6 KB
<input type="checkbox"/> books.csv	46 minutes ago	422.5 KB
<input type="checkbox"/> Dockerfile	1 minute ago	238 B
<input type="checkbox"/> requirements.txt	27 minutes ago	6 B

```
127.0.0.1:8888/notebooks/data%20analysis.ipynb

jupyter data analysis

File Edit View Run Kernel Settings Help Trusted

[18]: import pandas as pd

[19]: # Load the dataset
df = pd.read_csv('books.csv')

[20]: # Check for missing values in crucial columns
print(df[['title', 'average_rating']].isnull().sum())

title      0
average_rating  0
dtype: int64

[21]: # Calculate the mean of average_rating and fill missing values
average_rating_mean = df['average_rating'].mean()
df['average_rating'] = df['average_rating'].fillna(average_rating_mean)
print(df['average_rating'])

0      4.34
1      4.44
2      3.57
3      4.26
4      4.24
...
1349    3.90
1350    3.77
1351    4.14
1352    3.60
1353    3.95
Name: average_rating, Length: 1354, dtype: float64
```

```
127.0.0.1:8888/notebooks/data%20analysis.ipynb

jupyter data analysis

File Edit View Run Kernel Settings Help Trusted

[22]: # Drop rows with missing values
df.dropna(inplace=True)

[23]: # Focus your analysis on the Harry Potter book series.
harry_potter_books = df[df['title'].str.contains("Harry Potter", na=False)]

[24]: #Find the most selling books within the Harry Potter series.
most_selling_hp = harry_potter_books.loc[harry_potter_books['work_ratings_count'].idxmax()]
print(f"The most popular Harry Potter book is: {most_selling_hp['title']} with {most_selling_hp['work_ratings_count']} ratings.")

The most popular Harry Potter book is: Harry Potter and the Sorcerer's Stone (Harry Potter, #1) with 4800065 ratings.

[25]: # Calculate the average rating
average_rating = harry_potter_books['average_rating'].mean()
print(f"The average rating of the Harry Potter books is: {average_rating:.2f}")

The average rating of the Harry Potter books is: 4.49
```

Analysis :

- Find the most selling books within the Harry Potter series.

```
#Find the most selling books within the Harry Potter series.
most_selling_hp = harry_potter_books.loc[harry_potter_books['work_ratings_count'].idxmax()]
print(f"The most popular Harry Potter book is: {most_selling_hp['title']} with {most_selling_hp['work_ratings_count']} ratings.")

The most popular Harry Potter book is: Harry Potter and the Sorcerer's Stone (Harry Potter, #1) with 4800065 ratings.
```

Harry Potter and the Sorcerer's Stone (Harry Potter, #1) with 4800065 ratings.

- Calculate the average rating of the Harry Potter books.

```
# Calculate the average rating
average_rating = harry_potter_books['average_rating'].mean()
print(f"The average rating of the Harry Potter books is: {average_rating:.2f}")

The average rating of the Harry Potter books is: 4.49
```

4.49