

ID : 2102100

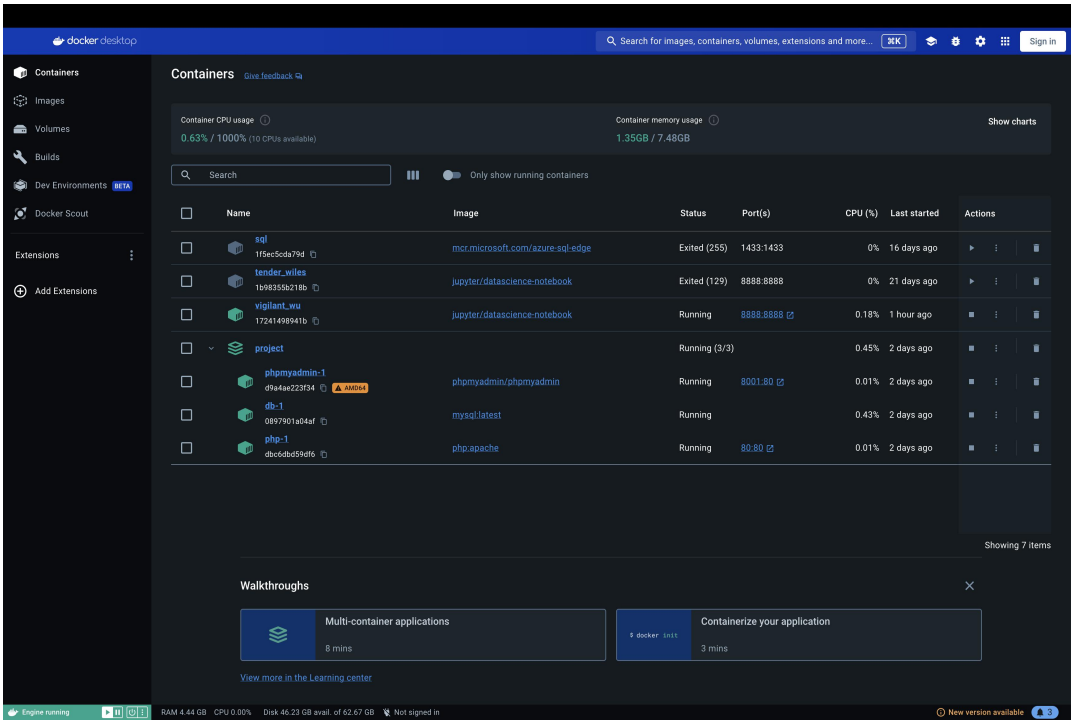
ID : 2102100

## ID : 2102100

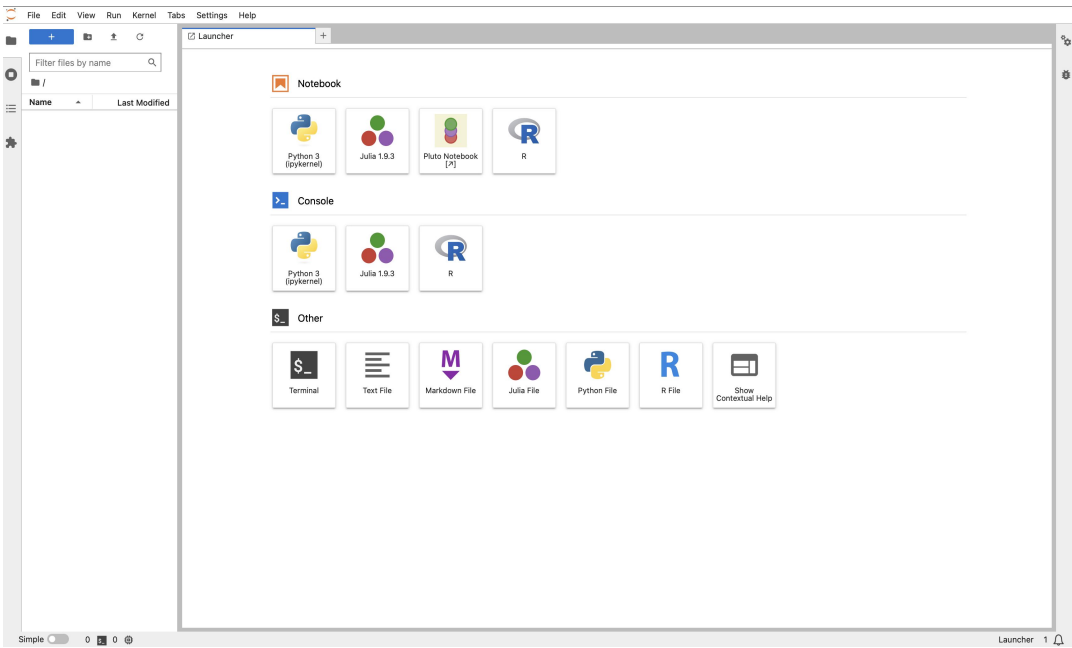
ID : 2102100ID : 2102100



2- Docker Container



3- Jupyter Notebook





## 4- Analysis

Apply data cleaning and preprocessing on popular books dataset then apply analysis on Harry Potter series to discover most selling books and average rating as shown in Harry Potter.ipynb)

Analysis shows that the average rating is 4.497 and the most selling series is Harry Potter and the Sorcerer's Stone

```
[27]: print(hp)
```

	authors	original_publication_year	\
1	J.K. Rowling, Mary GrandPré	1997.0	
6	J.K. Rowling, Mary GrandPré, Rufus Beck	1999.0	
8	J.K. Rowling, Mary GrandPré	2003.0	
9	J.K. Rowling, Mary GrandPré	1998.0	
10	J.K. Rowling, Mary GrandPré	2000.0	
11	J.K. Rowling, Mary GrandPré	2007.0	
12	J.K. Rowling, Mary GrandPré	2005.0	

	title	average_rating	\
1	Harry Potter and the Sorcerer's Stone (Harry P...	4.44	
6	Harry Potter and the Prisoner of Azkaban (Harr...	4.53	
8	Harry Potter and the Order of the Phoenix (Har...	4.46	
9	Harry Potter and the Chamber of Secrets (Harry...	4.37	
10	Harry Potter and the Goblet of Fire (Harry Pot...	4.53	
11	Harry Potter and the Deathly Hallows (Harry Po...	4.61	
12	Harry Potter and the Half-Blood Prince (Harry ...	4.54	

	ratings_count
1	4602479
6	1832823
8	1735368
9	1779331
10	1753043
11	1746574
12	1678823

```
• [28]: #calculate average rating of Harry Potter books  
hp.loc[:, 'average_rating'].mean()
```

```
[28]: 4.497142857142857
```

```
• [29]: #find most selling books within the Harry Potter series  
hp['ratings_count'].max()
```

```
[29]: 4602479
```