

**22010133**

## 1) power shell commands

→ cd "D:\Dockerfile" (cd path of notepad)

→ **docker run -p 8888:8888 jupyter/datascience-notebook**

(<http://127.0.0.1:8888/lab?token=3e6e88ea94dd4a4ca907514f7b654aeb83f8bce0dedc250c>) to open jupyter in any browser

The screenshot shows a Windows PowerShell terminal window with the following content:

```
Windows PowerShell
Copyright (C) Microsoft Corporation. All rights reserved.

Install the latest PowerShell for new features and improvements! https://aka.ms/PSWindows

PS C:\Users\alvel> cd "D:\Dockerfile"
cd : cannot find path 'D:\Dockerfile' because it does not exist.
At line:1 char:1
+ cd "D:\Dockerfile"
+ ~~~~~
+ CategoryInfo          : ObjectNotFound: (D:\Dockerfile:String) [Set-Location], ItemNotFoundException
+ FullyQualifiedErrorId : PathNotFound, Microsoft.PowerShell.Commands.SetLocationCommand

PS C:\Users\alvel> Get-Childitem -Path D:\ -Filter Dockerfile -Recurse

Directory: D:\

Mode                LastWriteTime         Length Name
----                -
-a-----         4/26/2024  1:09 AM             888 Dockerfile

Directory: D:\anaconda last version\lib\site-packages\hnbclassic\static\components\codemirror\mode

Mode                LastWriteTime         Length Name
----                -
d-----         4/26/2024   4:11 AM             dockerfile

Directory: D:\anaconda last version\Library\kspecs\features\data\testserver

Mode                LastWriteTime         Length Name
----                -
-a-----        10/27/2023  10:02 AM             1212 Dockerfile

Directory: D:\anaconda last
version\pkgsrc\hnbclassic-4.2.8-py311h8a95532_0\lib\site-packages\hnbclassic\static\components\codemirror\mode

Mode                LastWriteTime         Length Name
----                -
d-----         4/26/2024   3:52 AM             dockerfile

Directory: D:\anaconda last version\pkgsrc\qt-main-5.15.2-h879a1e9_0\Library\kspecs\features\data\testserver

Mode                LastWriteTime         Length Name
----                -
-a-----        10/27/2023  10:02 AM             1212 Dockerfile
```

The terminal window is titled "Windows PowerShell" and shows the user's prompt as "PS C:\Users\alvel>". The user has navigated to the directory "D:\Dockerfile" and attempted to run the command "cd \"D:\Dockerfile\"", which resulted in an error: "cd : cannot find path 'D:\Dockerfile' because it does not exist." The user then ran "Get-Childitem -Path D:\ -Filter Dockerfile -Recurse", which displayed a list of files and directories containing the word "Dockerfile". The list includes "Dockerfile" in the root directory, "dockerfile" in the "anaconda last version\lib\site-packages\hnbclassic\static\components\codemirror\mode" directory, "Dockerfile" in the "anaconda last version\Library\kspecs\features\data\testserver" directory, and "Dockerfile" in the "anaconda last version\pkgsrc\hnbclassic-4.2.8-py311h8a95532\_0\lib\site-packages\hnbclassic\static\components\codemirror\mode" directory. The taskbar at the bottom shows the system clock as 3:02 AM on 4/26/2024.

## 2) notepad

FROM jupyter/base-notebook

WORKDIR /app

COPY . /app


RUN pip install --no-cache-dir -r requirements.txt

EXPOSE 8888

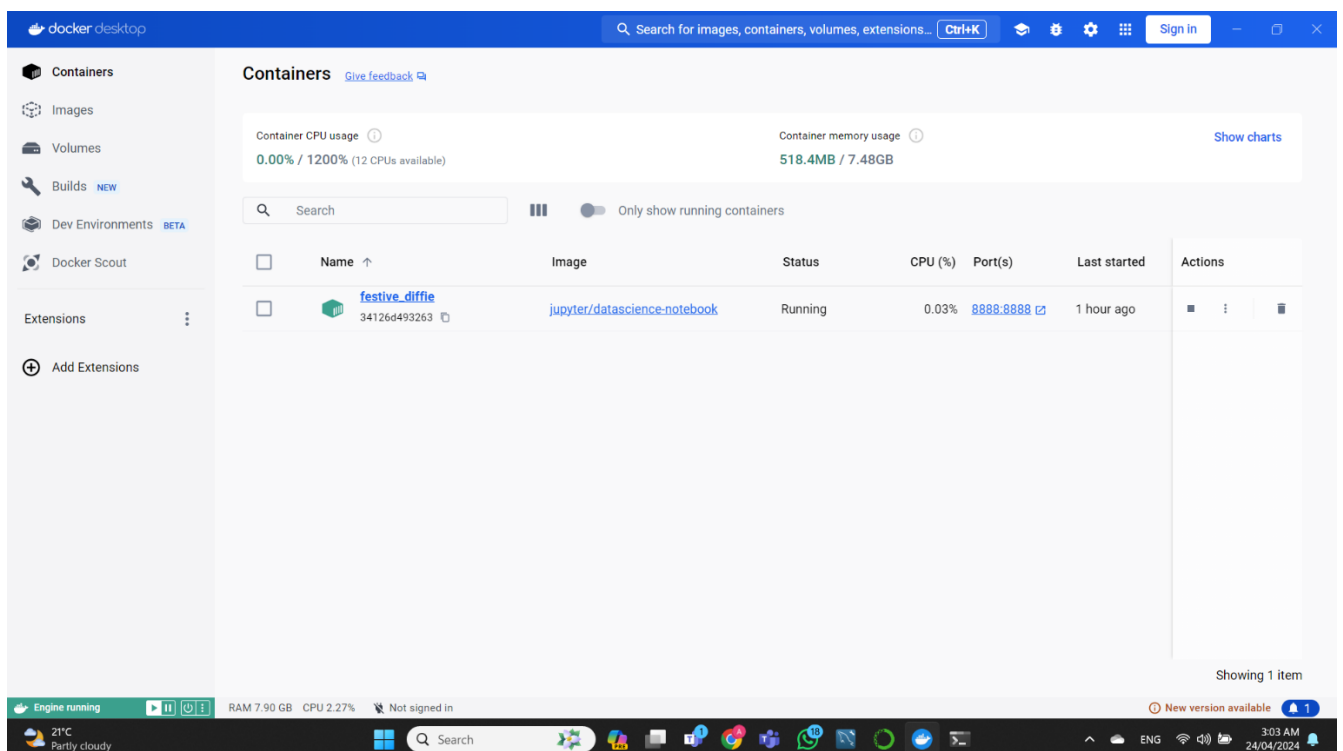
ENV NAME World

CMD ["jupyter", "notebook", "--ip='0.0.0.0'", "--port=8888", "--no-browser", "--allow-root"]

▼ Today

  Dockerfile	24/04/2024 1:09 AM	File	1 KB
--	--------------------	------	------

## 2) docker must be working



The screenshot shows the Docker Desktop application window. The left sidebar contains navigation options: Containers, Images, Volumes, Builds, Dev Environments, Docker Scout, Extensions, and Add Extensions. The main panel displays the 'Containers' view with a search bar and a toggle for 'Only show running containers'. A table lists the running containers:

Name	Image	Status	CPU (%)	Port(s)	Last started	Actions
festive_diffie 34126d493263	jupyter/datascience-notebook	Running	0.03%	8888:8888	1 hour ago	[Stop] [Refresh] [Delete]

At the bottom of the Docker Desktop window, a status bar shows 'Engine running', system resources (RAM 7.90 GB, CPU 2.27%), and a 'New version available' notification. The Windows taskbar at the very bottom shows the date and time as 3:03 AM on 24/04/2024.

### 3) jupyter (python compiler)

The screenshot shows the JupyterLab interface. On the left is a file browser with a search bar and a list of files: 'work' (6 months ago), 'books.csv' (59 minutes ago), and 'Untitled.ipynb' (22 minutes ago). The main area is a code editor for 'Untitled.ipynb' in Python 3 (ipykernel) mode. It contains two code cells. Cell [3] shows the output of `df.describe()`, which is a summary statistics table for the DataFrame. Cell [4] shows the output of `df.columns`, which is a list of column names.

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn13	original_publication_year	average_rating	ratings_count	work_ratings
count	1354.000000	1.354000e+03	1.354000e+03	1.354000e+03	1354.000000	1.310000e+03	1351.000000	1354.000000	1.354000e+03	1.3540
mean	4453.584195	5.951852e+06	6.120589e+06	8.707028e+06	50.330871	9.766700e+12	2003.422650	3.999357	9.160429e+04	9.9155
std	2894.277455	6.664595e+06	6.935008e+06	9.813696e+06	61.338867	3.572069e+11	16.779301	0.224263	2.871266e+05	3.0236
min	1.000000	1.000000e+00	1.000000e+00	1.150000e+02	1.000000	7.678361e+10	1868.000000	3.230000	6.221000e+03	8.8330
25%	1860.250000	1.537868e+05	1.537962e+05	1.375035e+06	22.000000	9.780152e+12	2003.000000	3.850000	1.759325e+04	1.9181
50%	4177.500000	3.305318e+06	3.422646e+06	4.005716e+06	37.000000	9.780440e+12	2008.000000	4.000000	2.943000e+04	3.2551
75%	6814.500000	9.917380e+06	1.019388e+07	1.435717e+07	58.000000	9.780805e+12	2011.000000	4.160000	6.073800e+04	6.6812
max	9955.000000	3.207567e+07	3.360215e+07	4.963819e+07	1314.000000	9.788424e+12	2017.000000	4.740000	4.780653e+06	4.9423

The screenshot shows the JupyterLab interface. On the left is a file browser with a search bar and a list of files: 'work' (6 months ago), 'books.csv' (58 minutes ago), and 'Untitled.ipynb' (21 minutes ago). The main area is a code editor for 'Untitled.ipynb' in Python 3 (ipykernel) mode. It contains two code cells. Cell [2] shows the output of `df.head()`, which is a table of book details. Cell [3] shows the output of `df.describe()`, which is a summary statistics table for the DataFrame.

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...	47
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	46
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...	38
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...	23
4	12	13335037	13335037	13155899	210	62024035	9.780602e+12	Veronica Roth	2011.0	Divergent	...	15

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/tree/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	22 minutes ago

```
[5]: 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              1310 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       1245 non-null  object  
12   average_rating      1354 non-null  float64
13   ratings_count       1354 non-null  int64  
14   work_ratings_count  1354 non-null  int64  
15   work_text_reviews_count 1354 non-null  int64  
16   ratings_1           1354 non-null  int64  
17   ratings_2           1354 non-null  int64  
18   ratings_3           1354 non-null  int64  
19   ratings_4           1354 non-null  int64  
20   ratings_5           1354 non-null  int64  
21   image_url           1354 non-null  object  
22   small_image_url     1354 non-null  object  
dtypes: float64(3), int64(13), object(7)
memory usage: 243.4+ KB
```

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1 3:03 AM 24/04/2024

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/tree/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	22 minutes ago

```
[6]: df.dropna(inplace=True)
df.drop_duplicates(inplace=True)
df
```

```
[6]:
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_co
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...	4780
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	4602
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...	3866
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...	2346
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...	1903
...	...	...	...	...	...	...	...	...	...	...	...	...
1349	9925	86737	86737	3877968	52	1582349177	9.781582e+12	Mary Hoffman	2002.0	City of Masks	...	12
1350	9937	13010211	13010211	18171867	22	1596435712	9.781596e+12	Caragh M. O'Brien	2012.0	Promised	...	11
1351	9942	16074758	16074758	21869436	18	1442486597	9.781442e+12	Abigail Haas, Abby	2013.0	Dangerous Girls	...	10

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1 3:03 AM 24/04/2024

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/tree/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	22 minutes ago

```
[11]: df = df.drop(columns=['goodreads_book_id'])
df['average_rating'] = df[['ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5']].mean(axis=1)
df.drop(columns=['ratings_1', 'ratings_2', 'ratings_3', 'ratings_4', 'ratings_5'], inplace=True)
df
```

```
[11]:
```

	book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	title	language_code	aver
0	1	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	The Hunger Games (The Hunger Games, #1)	eng	
1	2	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	Harry Potter and the Sorcerer's Stone (Harry P...	eng	
2	3	41865	3212258	226	316015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	Twilight (Twilight, #1)	en-US	
3	6	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	The Fault in Our Stars	eng	
4	12	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	Divergent (Divergent, #1)	eng	
...	...	...	...	...	...	...	...	...	...	...	...	...
1349	9925	86737	3877968	52	1582349177	9.781582e+12	Mary Hoffman	2002.0	City of Masks	City of Masks (Stravaganza, #1)	eng	

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1 3:03 AM 24/04/2024

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/tree/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	22 minutes ago

```
[12]: harry_potter_books = df[df['original_title'].str.contains('Harry Potter', case=False)]
harry_potter_books.dropna(inplace=True)
df
```

/tmp/ipykernel\_334/180751574.py:2: 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](https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy)  
harry\_potter\_books.dropna(inplace=True)

```
[12]:
```

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

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1 3:04 AM 24/04/2024

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	22 minutes ago

```
[15]: import numpy as np
import matplotlib.pyplot as plt

[16]: import seaborn as sns

[18]: harrypotter=df[df.authors=="J.K. Rowling, Mary GrandPré"]
harrypotter

[18]:
```

book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	title	language_code	average_rating
1	2	3	4640799	491 439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	Harry Potter and the Sorcerer's Stone (Harry P...	eng	96001
8	21	2	2809203	307 439358078	9.780439e+12	J.K. Rowling, Mary GrandPré	2003.0	Harry Potter and the Order of the Phoenix	Harry Potter and the Order of the Phoenix (Har...	eng	36810
9	23	15881	6231171	398 439064864	9.780439e+12	J.K. Rowling, Mary GrandPré	1998.0	Harry Potter and the Chamber of Secrets	Harry Potter and the Chamber of Secrets (Harry...	eng	38123

Simple 0 1 Python 3 (ipykernel) Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1

21°C Partly cloudy 3:04 AM 24/04/2024

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/tree/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	now

```
[19]: harrypotter.info()

<class 'pandas.core.frame.DataFrame'>
Index: 6 entries, 1 to 12
Data columns (total 17 columns):
 #   Column                                Non-Null Count  Dtype
---  -
 0   book_id                               6 non-null      int64
 1   best_book_id                         6 non-null      int64
 2   work_id                              6 non-null      int64
 3   books_count                          6 non-null      int64
 4   isbn                                  6 non-null      object
 5   isbn13                               6 non-null      float64
 6   authors                             6 non-null      object
 7   original_publication_year            6 non-null      float64
 8   original_title                       6 non-null      object
 9   title                               6 non-null      object
10   language_code                       6 non-null      object
11   average_rating                       6 non-null      float64
12   ratings_count                        6 non-null      int64
13   work_ratings_count                   6 non-null      int64
14   work_text_reviews_count              6 non-null      int64
15   image_url                           6 non-null      object
16   small_image_url                     6 non-null      object
dtypes: float64(3), int64(7), object(7)
memory usage: 864.0+ bytes

[20]: harrypotter.describe()

[20]:
```

	book_id	best_book_id	work_id	books_count	isbn13	original_publication_year	average_rating	ratings_count	work_ratings_count	work_text_rev
count	6.000000	6.000000	6.000000e+00	6.000000	6.000000e+00	6.000000	6.000000	6.000000e+00	6.000000e+00	

Simple 0 1 Python 3 (ipykernel) Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1

21°C Partly cloudy 3:04 AM 24/04/2024

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/tree/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	now

Untitled.ipynb

```
[20]: harrypotter.describe()
```

	book_id	best_book_id	work_id	books_count	isbn13	original_publication_year	average_rating	ratings_count	work_ratings_count	work_text_rev
count	6.000000	6.000000	6.000000e+00	6.000000	6.000000e+00	6.000000	6.000000	6.000000e+00	6.000000e+00	
mean	20.333333	25357.333333	1.017106e+07	344.333333	9.780457e+12	2001.666667	468284.166667	2.215936e+06	2.341421e+06	
std	9.201449	54696.572268	1.532470e+07	86.439960	4.312389e+07	3.983298	241025.416862	1.169634e+06	1.205127e+06	
min	2.000000	1.000000	2.809203e+06	263.000000	9.780439e+12	1997.000000	357135.200000	1.678823e+06	1.785676e+06	
25%	21.500000	2.250000	2.984056e+06	283.000000	9.780439e+12	1998.500000	368451.950000	1.738170e+06	1.842260e+06	
50%	23.500000	4.500000	3.843686e+06	319.500000	9.780439e+12	2001.500000	371603.700000	1.749808e+06	1.858018e+06	
75%	24.750000	11912.250000	5.833578e+06	381.500000	9.780440e+12	2004.500000	379361.950000	1.772759e+06	1.896810e+06	
max	27.000000	136251.000000	4.133543e+07	491.000000	9.780545e+12	2007.000000	960013.000000	4.602479e+06	4.800065e+06	

```
[24]: most_selling=harrypotter.sort_values(by='ratings_count',ascending = False)
most_selling_book = most_selling.iloc[0]
print("The most selling Harry Potter book based on book count is:")
print(most_selling_book[['original_title', 'ratings_count']])

The most selling Harry Potter book based on book count is:
original_title    Harry Potter and the Philosopher's Stone
ratings_count      4602479
Name: 1, dtype: object

[25]: print(most_selling[['original_title', 'ratings_count']])
```

original title ratings count

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1 21°C Partly cloudy 3:04 AM 24/04/2024

Untitled.ipynb - JupyterLab

127.0.0.1:8888/lab/tree/Untitled.ipynb

Gmail YouTube Maps Translate Netflix United Arab... Duolingo - The wor... Online Courses - Le... OmeTV Video Chat... Omegle: Talk to str... Telegram Web Messenger All Bookmarks

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Name	Last Modified
work	6 months ago
books.csv	59 minutes ago
Untitled.ipynb	20 seconds ago

Untitled.ipynb

```
[25]: print(most_selling[['original_title', 'ratings_count']])

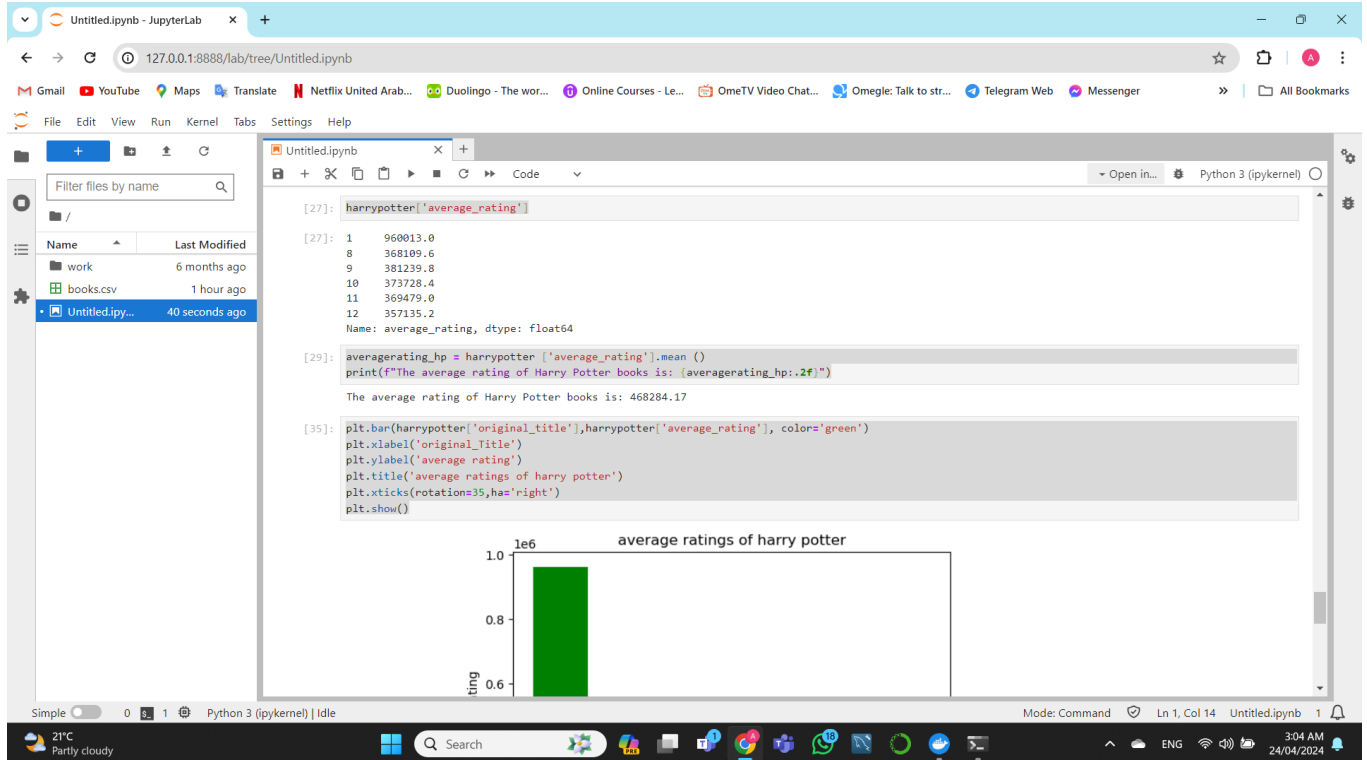
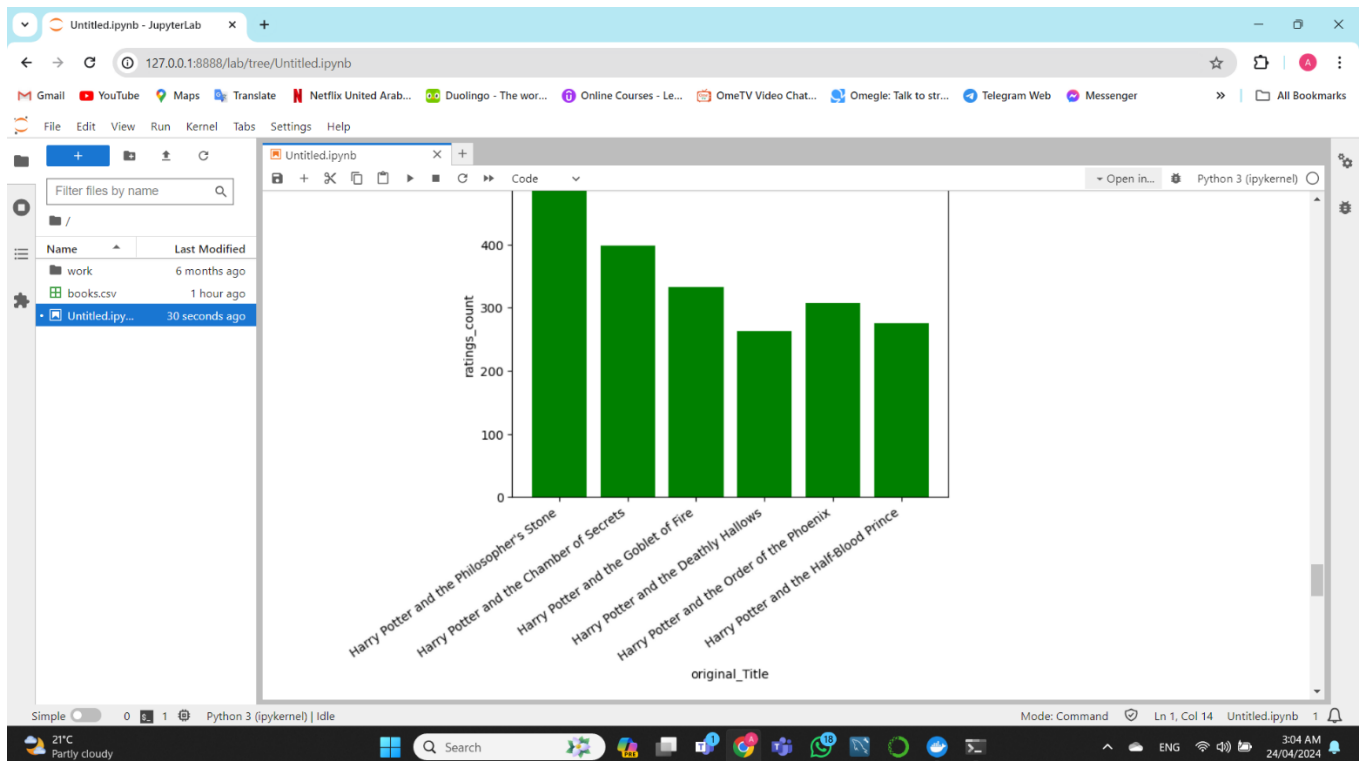
original_title ratings_count
1 Harry Potter and the Philosopher's Stone 4602479
9 Harry Potter and the Chamber of Secrets 1779331
10 Harry Potter and the Goblet of Fire 1753043
11 Harry Potter and the Deathly Hallows 1746574
8 Harry Potter and the Order of the Phoenix 1735368
12 Harry Potter and the Half-Blood Prince 1678823

[26]: plt.bar(most_selling['original_title'], most_selling['books_count'], color='green')
plt.xlabel('original_title')
plt.ylabel('ratings_count')
plt.title('Harry Potter Books by Books Count')
plt.xticks(rotation=35, ha='right')
plt.show()
```

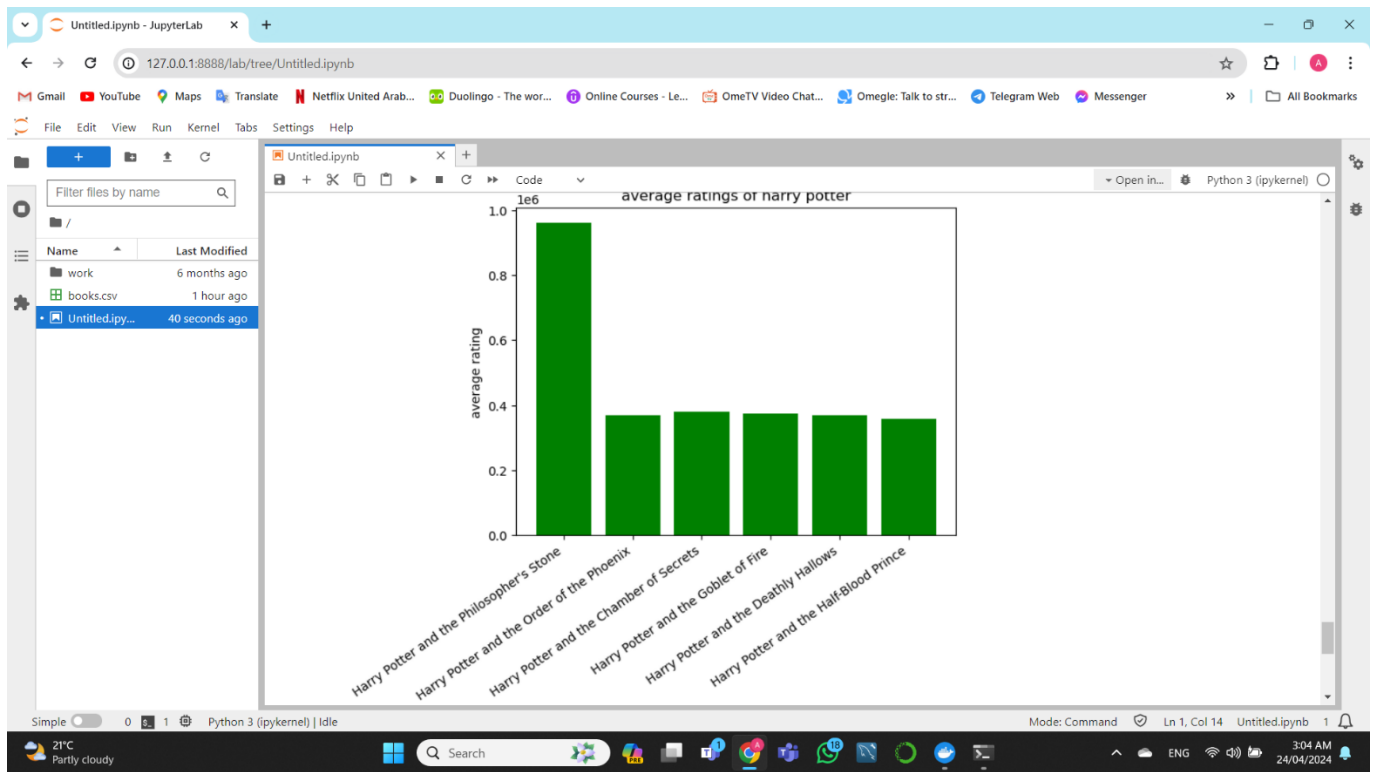
Harry Potter Books by Books Count

original_title	ratings_count
Harry Potter and the Philosopher's Stone	4602479
Harry Potter and the Chamber of Secrets	1779331
Harry Potter and the Goblet of Fire	1753043
Harry Potter and the Deathly Hallows	1746574
Harry Potter and the Order of the Phoenix	1735368
Harry Potter and the Half-Blood Prince	1678823

Simple 0 1 Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 14 Untitled.ipynb 1 21°C Partly cloudy 3:04 AM 24/04/2024







THANK YOUU