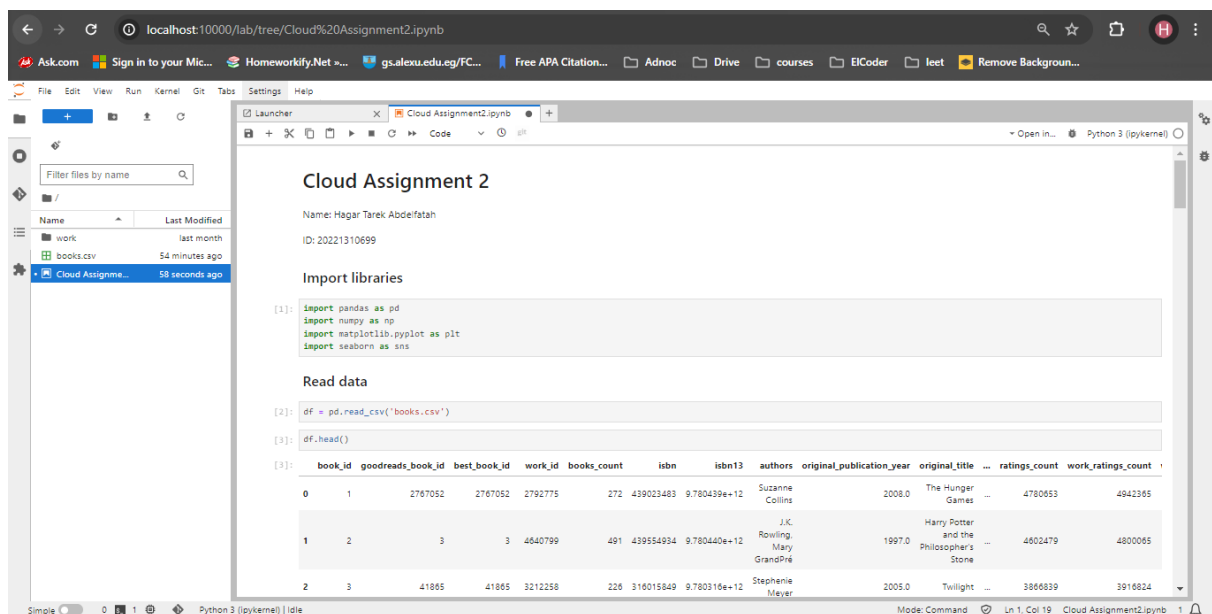


Cloud Computing Assignment 2: Dockerfile and Data Analysis with Popular Books Dataset

Name: Hagar Tarek Abdelfatah

ID: 20221310699

Analysis Results



The screenshot displays a Jupyter Notebook environment with the following content:

Cloud Assignment 2

Name: Hagar Tarek Abdelfatah
ID: 20221310699

Import libraries

```
[1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

Read data

```
[2]: df = pd.read_csv('books.csv')
[3]: df.head()
```

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	...	ratings_count	work_ratings_count
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	...	4780653	4942365
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	...	4602479	4800065
2	3	41865	41865	3212258	226	310015849	9.780316e+12	Stephenie Meyer	2005.0	Twilight	...	3866839	3916824

localhost:10000/lab/tree/Cloud%20Assignment2.ipynb

Ask.com Sign in to your Mic... Homeworkify.Net ... gs.alexu.edu.eg/FC... Free APA Citation... Adnoc Drive courses ElCoder leet Remove Background...

File Edit View Run Kernel Git Tabs Settings Help

Launcher Cloud Assignment2.ipynb

Filter files by name

Name Last Modified

work last month

books.csv 56 minutes ago

Cloud Assignme... 20 seconds ago

2	3	41855	41855	3212258	226	316015849	9.780315e+12	Stephanie Meyer	2005.0	Twilight	...	3866839	3916824
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	...	2346404	2478609
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	...	1903563	2216814

5 rows x 23 columns

Clean and Preprocess the data

```
[4]: # Check for null values
print(df.isnull().sum())

book_id          0
goodreads_book_id 0
best_book_id     0
work_id          0
books_count      0
isbn            52
isbn13          44
authors          0
original_publication_year 0
original_title   52
title           0
language_code    109
average_rating   0
ratings_count    0
work_ratings_count 0
work_text_reviews_count 0
ratings_1        0
ratings_2        0
ratings_3        0
ratings_4        0
```

Simple Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 19 Cloud Assignment2.ipynb 1

localhost:10000/lab/tree/Cloud%20Assignment2.ipynb

Ask.com Sign in to your Mic... Homeworkify.Net ... gs.alexu.edu.eg/FC... Free APA Citation... Adnoc Drive courses ElCoder leet Remove Background...

File Edit View Run Kernel Git Tabs Settings Help

Launcher Cloud Assignment2.ipynb

Filter files by name

Name Last Modified

work last month

books.csv 56 minutes ago

Cloud Assignme... 50 seconds ago

```
[5]: # Remove null values
df.dropna(inplace=True)

[6]: print(df.isnull().sum())

book_id          0
goodreads_book_id 0
best_book_id     0
work_id          0
books_count      0
isbn            0
isbn13          0
authors          0
original_publication_year 0
original_title   0
title           0
language_code    0
average_rating   0
ratings_count    0
work_ratings_count 0
work_text_reviews_count 0
ratings_1        0
ratings_2        0
ratings_3        0
ratings_4        0
ratings_5        0
image_url       0
small_image_url 0
dtype: int64

[7]: # Check for duplicates
duplicates = df[df.duplicated()]

# Display the duplicates
print(duplicates)
```

Simple Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 19 Cloud Assignment2.ipynb 1

localhost:10000/lab/tree/Cloud%20Assignment2.ipynb

Ask.com Sign in to your Mic... Homeworkify.Net... gs.alexu.edu.eg/FC... Free APA Citation... Adnoc Drive courses EICoder leet Remove Background...

File Edit View Run Kernel Git Tabs Settings Help

Launcher Cloud Assignment2.ipynb

Filter files by name

Name Last Modified

work last month

books.csv 57 minutes ago

Cloud Assignme... 1 minute ago

```
[7]: # Check for duplicates
duplicates = df[df.duplicated()]

# Display the duplicates
print(duplicates)

Empty DataFrame
Columns: [book_id, goodreads_book_id, best_book_id, work_id, books_count, isbn, isbn1, 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]
Index: []

[0 rows x 23 columns]

[8]: # Drop duplicates
df.drop_duplicates(inplace=True)
df.shape

[8]: (1153, 23)
```

Harry Potter Books Analysis

```
[11]: # Extract the data for Harry Potter books
authors = df.authors.unique()
jkr_books = df[df.authors.str.contains('J.K. Rowling')]
hp_books = jkr_books[jkr_books.title.str.contains('Harry Potter and ')]
hp_books = hp_books.sort_values(by='original_publication_year', ascending=True)
hp_books
```

al_title	ratings_count	work_ratings_count	work_text_reviews_count	ratings_1	ratings_2	ratings_3	ratings_4	ratings_5	image_url	small_image_url
J Potter and the opher's Stone	4602479	4800065	75867	75504	101676	455024	1156318	3011543	https://images-gr-assets.com/books/1474154022m...	https://images-gr-assets.com/books/1474154022s...

Simple Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 19 Cloud Assignment2.ipynb 1

localhost:10000/lab/tree/Cloud%20Assignment2.ipynb

Ask.com Sign in to your Mic... Homeworkify.Net... gs.alexu.edu.eg/FC... Free APA Citation... Adnoc Drive courses EICoder leet Remove Background...

File Edit View Run Kernel Git Tabs Settings Help

Launcher Cloud Assignment2.ipynb

Filter files by name

Name Last Modified

work last month

books.csv 1 hour ago

Cloud Assignme... 4 minutes ago

Harry Potter Books Analysis

```
[11]: # Extract the data for Harry Potter books
authors = df.authors.unique()
jkr_books = df[df.authors.str.contains('J.K. Rowling')]
hp_books = jkr_books[jkr_books.title.str.contains('Harry Potter and ')]
hp_books = hp_books.sort_values(by='original_publication_year', ascending=True)
hp_books
```

al_title	ratings_count	work_ratings_count	work_text_reviews_count	ratings_1	ratings_2	ratings_3	ratings_4	ratings_5	image_url	small_image_url
J Potter and the opher's Stone	4602479	4800065	75867	75504	101676	455024	1156318	3011543	https://images-gr-assets.com/books/1474154022m...	https://images-gr-assets.com/books/1474154022s...
J Potter and the riber of Secrets	1779331	1906199	34172	8253	42251	242345	548266	1065084	https://images-gr-assets.com/books/1474169725m...	https://images-gr-assets.com/books/1474169725s...
J Potter and the oner of zikaban	1832823	1969375	36099	6716	20413	166129	509447	1266670	https://images-gr-assets.com/books/1499277281m...	https://images-gr-assets.com/books/1499277281s...
J Potter and the oblet of Fine	1753043	1868642	31084	6676	20210	151785	494926	1195045	https://images-gr-assets.com/books/1381482611m...	https://images-gr-assets.com/books/1381482611s...
J Potter and the r of the Phoenix	1735368	1840548	28685	9528	31577	180210	494427	1124806	https://images-gr-assets.com/books/1387141547m...	https://images-gr-assets.com/books/1387141547s...

Simple Python 3 (ipykernel) | Idle Mode: Command Ln 1, Col 19 Cloud Assignment2.ipynb 1

