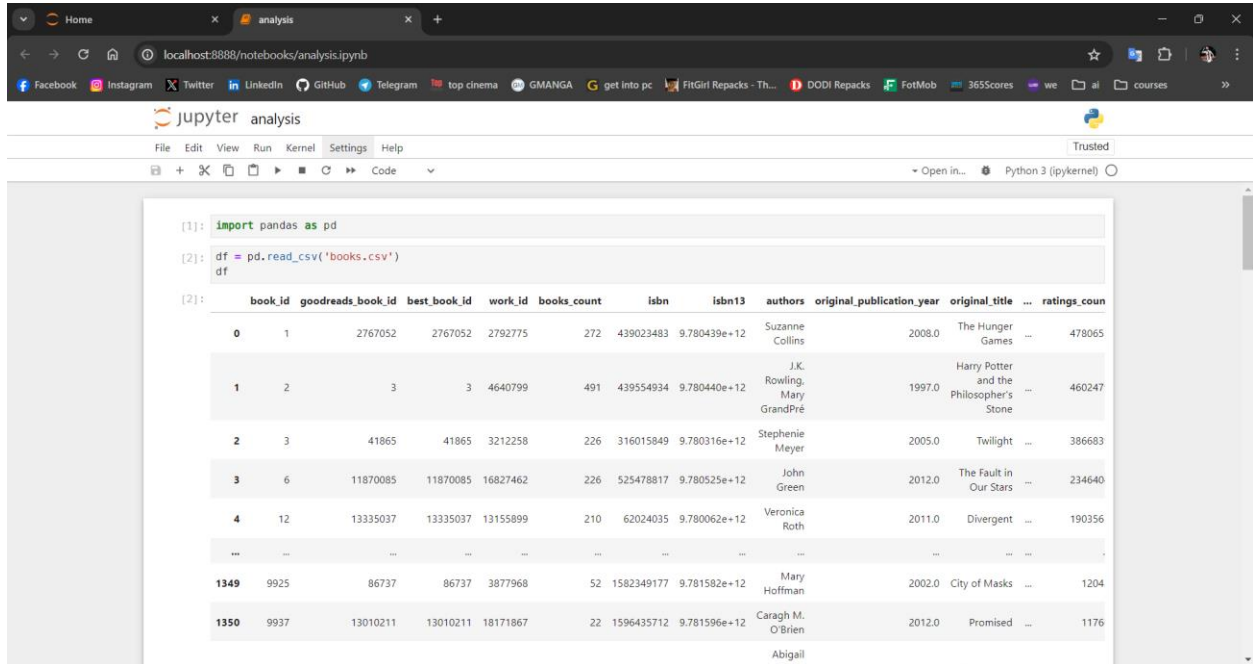


# Assignment 2

Name: Mina Nader ID: 22010451

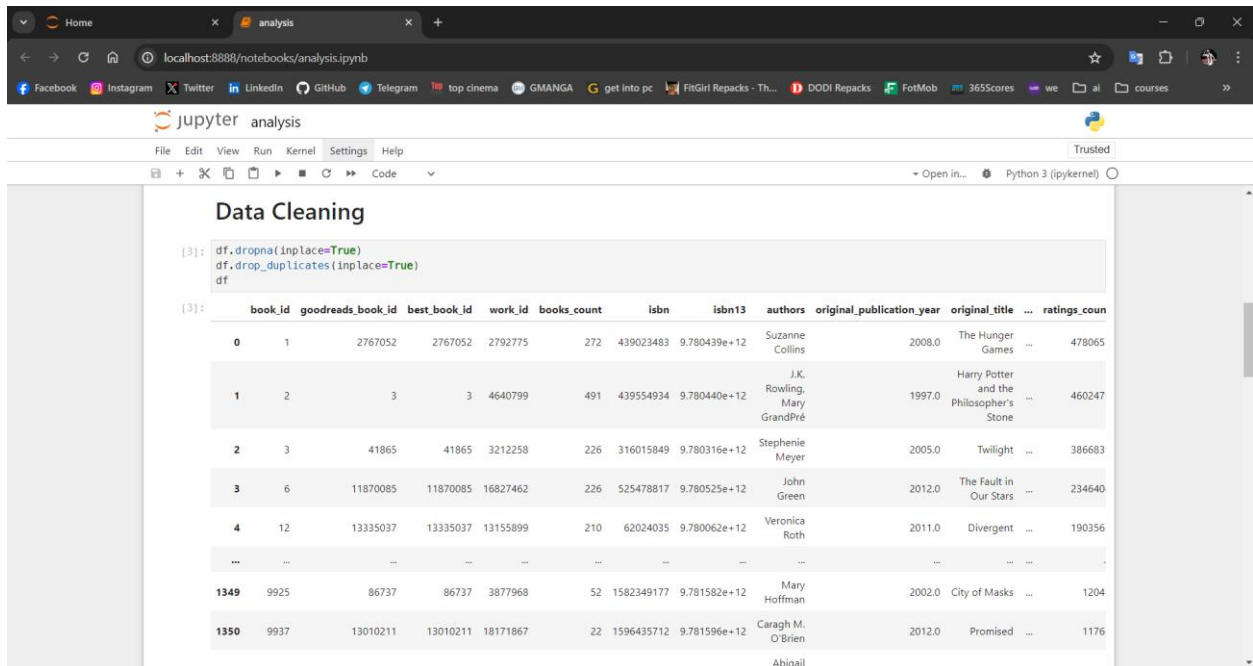


The screenshot shows a Jupyter Notebook interface with the following code and output:

```
[1]: import pandas as pd
[2]: df = pd.read_csv('books.csv')
df
```

The output is a DataFrame with the following columns: book\_id, goodreads\_book\_id, best\_book\_id, work\_id, books\_count, isbn, isbn13, authors, original\_publication\_year, original\_title, ratings\_count. The data includes books like 'The Hunger Games', 'Harry Potter and the Philosopher's Stone', 'Twilight', 'The Fault in Our Stars', 'Divergent', 'City of Masks', and 'Promised'.

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	ratings_count
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	478065
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	460247
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephanie Meyer	2005.0	Twilight	386683
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	234640
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	190356
...	...	...	...	...	...	...	...	...	...	...	...
1349	9925	86737	86737	3877968	52	1582349177	9.781582e+12	Mary Hoffman	2002.0	City of Masks	1204
1350	9937	13010211	13010211	18171867	22	1596435712	9.781596e+12	Caragh M. O'Brien	2012.0	Promised	1176



The screenshot shows a Jupyter Notebook interface with the following code and output:

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

The output is the same DataFrame as above, but with the first column (index) removed, starting from 0.

	book_id	goodreads_book_id	best_book_id	work_id	books_count	isbn	isbn13	authors	original_publication_year	original_title	ratings_count
0	1	2767052	2767052	2792775	272	439023483	9.780439e+12	Suzanne Collins	2008.0	The Hunger Games	478065
1	2	3	3	4640799	491	439554934	9.780440e+12	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	460247
2	3	41865	41865	3212258	226	316015849	9.780316e+12	Stephanie Meyer	2005.0	Twilight	386683
3	6	11870085	11870085	16827462	226	525478817	9.780525e+12	John Green	2012.0	The Fault in Our Stars	234640
4	12	13335037	13335037	13155899	210	62024035	9.780062e+12	Veronica Roth	2011.0	Divergent	190356
...	...	...	...	...	...	...	...	...	...	...	...
1349	9925	86737	86737	3877968	52	1582349177	9.781582e+12	Mary Hoffman	2002.0	City of Masks	1204
1350	9937	13010211	13010211	18171867	22	1596435712	9.781596e+12	Caragh M. O'Brien	2012.0	Promised	1176

Home analysis

localhost:8888/notebooks/analysis.ipynb

Facebook Instagram Twitter LinkedIn GitHub Telegram top cinema GMANGA get into pc FITGirl Repacks - Th... DODI Repacks FotMob 365Scores we al courses

jupyter analysis

File Edit View Run Kernel Settings Help

Trusted

Open in... Python 3 (ipykernel)

## Data Preprocessing

```
[4]: df = df[df['original_title'].str.contains('Harry Potter', case=False)]
columns_to_drop = [
    'book_id', 'goodreads_book_id', 'best_book_id', 'work_id',
    'books_count', 'isbn', 'isbn13', 'work_ratings_count',
    'work_text_reviews_count', 'image_url', 'small_image_url', 'language_code'
]

df.drop(columns_to_drop, axis=1, inplace=True)
df
```

/tmp/ipykernel\_301/4203864638.py:8: 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)  
df.drop(columns\_to\_drop, axis=1, inplace=True)

```
[4]:
```

	authors	original_publication_year	original_title	title	average_rating	ratings_count	ratings_1	ratings_2	ratings_3	ratings_4	ratings_5
1	J.K. Rowling, Mary GrandPré	1997.0	Harry Potter and the Philosopher's Stone	Harry Potter and the Sorcerer's Stone (Harry P...	4.44	4602479	75504	101676	455024	1156318	3011543
6	J.K. Rowling, Mary GrandPré, Rufus Beck	1999.0	Harry Potter and the Prisoner of Azkaban	Harry Potter and the Prisoner of Azkaban (Harr...	4.53	1832823	6716	20413	166129	509447	1266670
8	J.K. Rowling, Mary GrandPré	2003.0	Harry Potter and the Order of the Phoenix	Harry Potter and the Order of the Phoenix (Har...	4.46	1735368	9528	31577	180210	494427	1124806

Home analysis

localhost:8888/notebooks/analysis.ipynb

Facebook Instagram Twitter LinkedIn GitHub Telegram top cinema GMANGA get into pc FITGirl Repacks - Th... DODI Repacks FotMob 365Scores we al courses

jupyter analysis

File Edit View Run Kernel Settings Help

Trusted

Open in... Python 3 (ipykernel)

Potter: A Treasury... Treasury...

## the most selling books within the Harry Potter series

```
[5]: most_selling_books = df.sort_values(by="ratings_count", ascending=False)
most_selling_books[["original_title", "authors", "ratings_count"]]
```

```
[5]:
```

	original_title	authors	ratings_count
1	Harry Potter and the Philosopher's Stone	J.K. Rowling, Mary GrandPré	4602479
6	Harry Potter and the Prisoner of Azkaban	J.K. Rowling, Mary GrandPré, Rufus Beck	1832823
9	Harry Potter and the Chamber of Secrets	J.K. Rowling, Mary GrandPré	1779331
10	Harry Potter and the Goblet of Fire	J.K. Rowling, Mary GrandPré	1753043
11	Harry Potter and the Deathly Hallows	J.K. Rowling, Mary GrandPré	1746574
8	Harry Potter and the Order of the Phoenix	J.K. Rowling, Mary GrandPré	1735368
12	Harry Potter and the Half-Blood Prince	J.K. Rowling, Mary GrandPré	1678823
96	Complete Harry Potter Boxed Set	J.K. Rowling	190050
613	Harry Potter Collection (Harry Potter, #1-6)	J.K. Rowling	24618
1036	The Magical Worlds of Harry Potter: A Treasury...	David Colbert	13820

Type Markdown and LaTeX:  $\alpha^2$

Homeanalysis

localhost:8888/notebooks/analysis.ipynb

FacebookInstagramTwitterLinkedInGitHubTelegramtop cinemagMANGAget into pcFIRGirl Repacks - Th...DODI RepacksFotMob365Scoreswealcourses

jupyteranalysis

FileEditViewRunKernelSettingsHelp

Trusted

Python 3 (ipykernel)

11	Harry Potter and the Deathly Hallows	J.K. Rowling, Mary GrandPré	1746574
8	Harry Potter and the Order of the Phoenix	J.K. Rowling, Mary GrandPré	1735368
12	Harry Potter and the Half-Blood Prince	J.K. Rowling, Mary GrandPré	1678823
96	Complete Harry Potter Boxed Set	J.K. Rowling	190050
613	Harry Potter Collection (Harry Potter, #1-6)	J.K. Rowling	24618
1036	The Magical Worlds of Harry Potter: A Treasury...	David Colbert	13820

Type Markdown and LaTeX:  $\alpha^2$

### the average rating of the Harry Potter books

```
[6]: average_rating = df['average_rating'].mean()
average_rating

[6]: 4.4910000000000005
```