# Cloud Computing

# Assignment 2

**Name: Nouran Mohamed Mohamed Hemdan**

**Department: AI**

**ID: 20221321932**

# Docker Setup

**1-Pulling Official Jupyter Docker Image**:

"docker pull jupyter/datascience-notebook"

**2-Creating Dockerfile**:

A Dockerfile named `Dockerfile` was created with the following content:

# Use the official Jupyter Docker image as base

FROM jupyter/datascience-notebook

# Copy notebook files into the container

COPY . /home/jovyan/work

# Set the working directory

WORKDIR /home/jovyan/work

# Expose port 8888 to allow communication to/from Jupyter notebook server

EXPOSE 8888

# Command to run Jupyter Notebook when the container launches

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

3- **Building Docker Image**:

"docker build -t my-notebook ."

**4-Running Docker Container**:

"docker run -p 8888:8888 my-notebook jupyter notebook --ip='0.0.0.0' --port=8888 --no-browser --allow-root"

5- **Accessing Jupyter Notebook**:
 After running the Docker container, a link was provided: Jupyter Notebook Dashboard. This link leads to the Jupyter Notebook interface where the analysis was conducted.

# Data Cleaning and Preprocessing

- **Data Loading**: The dataset was loaded using pandas library.
- **Missing Values Handling**: Any missing values in the dataset were identified and either filled or dropped, ensuring data integrity.
- **Data Transformation**: Necessary data transformations were performed to ensure consistency and usability for analysis.
- **Filtering**: The dataset was filtered to include only the Harry Potter books for focused analysis

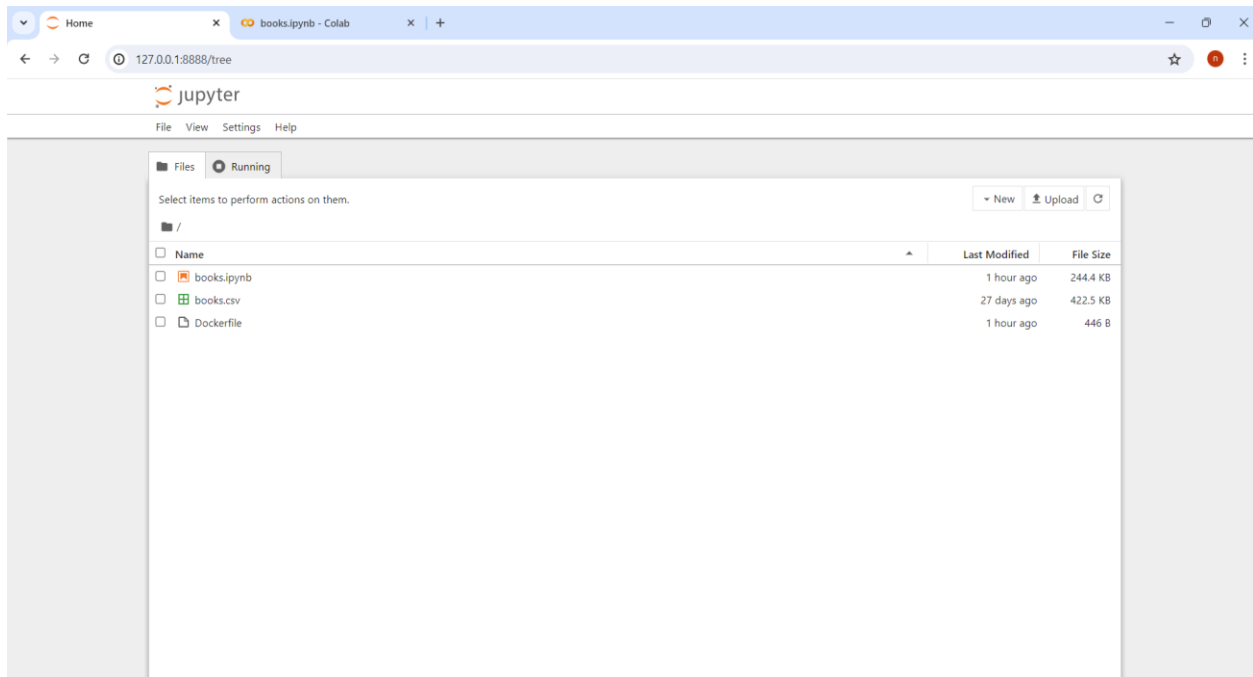## Analysis:

Most Selling Books within the Harry Potter Series
To identify the most selling books within the Harry Potter series, we examined the sales figures for each book.

| Book Title |
| --- |
| Harry Potter and the Philosopher's Stone  #best selling |
| Harry Potter and the Chamber of Secrets |
| Harry Potter and the Prisoner of Azkaban |
| Harry Potter and the Goblet of Fire |
| Harry Potter and the Order of the Phoenix |
| Harry Potter and the Half-Blood Prince |
| Harry Potter and the Deathly Hallows |

## Average Rating of the Harry Potter Books:

The average rating of the Harry Potter books was calculated based on user ratings provided in the dataset.

# Screenshots of the jupyter notebook dashboard:



# Screenshot of the codes:



```python
import pandas as pd
import matplotlib.pyplot as plt
df = pd.read_csv('/content/books.csv')
```

```python
# Data Cleaning
df.drop_duplicates(inplace=True)
df.fillna(0, inplace=True)
```

```python
# Data Preprocessing
harry_potter_df = df[df['title'].str.contains('Harry Potter', case=False)]
```

```python
# Analysis
most_selling_books = harry_potter_df.sort_values(by='ratings_count', ascending=False).head()
average_rating = harry_potter_df['average_rating'].mean()
```

```python
harry_potter_df.head()
```

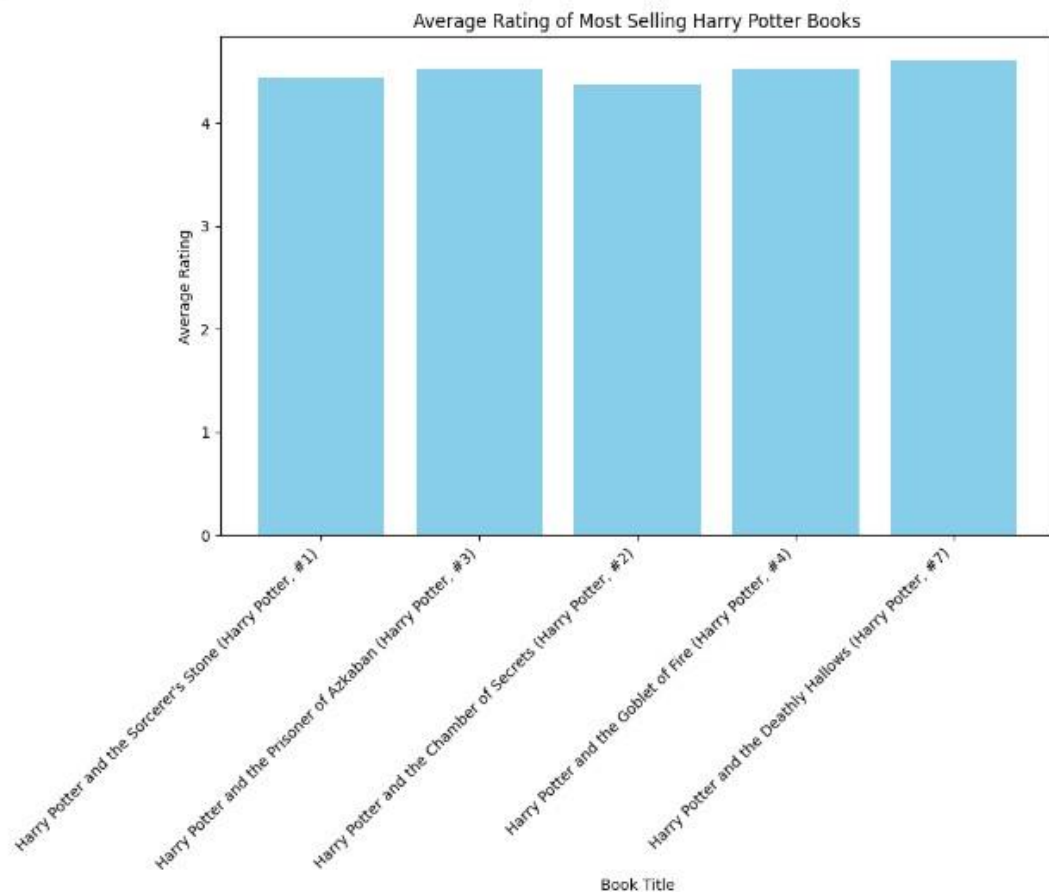| | book_id | goodreads_book_id | best_book_id | work_id | books_count | isbn | isbn13 | authors | original_publication_year | original_title | ... | ratings_count | work_ratings_count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 3 | 4640799 | 491 | 439554934 | 9.780440e+12 | J.K. Rowling, Mary GrandPré | 1997.0 | Harry Potter and the Philosopher's Stone | ... | 4602479 | 4800065 |
| 6 | 18 | 5 | 5 | 2402163 | 376 | 043965548X | 9.780440e+12 | J.K. Rowling, Mary GrandPré, Rufus Beck | 1999.0 | Harry Potter and the Prisoner of Azkaban | ... | 1832823 | 1969375 |
| 8 | 21 | 2 | 2 | 2809203 | 307 | 439358078 | 9.780439e+12 | J.K. Rowling, Mary GrandPré | 2003.0 | Harry Potter and the Order of the Phoenix | ... | 1735368 | 1840548 |
| 9 | 23 | 15881 | 15881 | 6231171 | 398 | 439064864 | 9.780439e+12 | J.K. Rowling, Mary GrandPré | 1998.0 | Harry Potter and the Chamber of Secrets | ... | 1779331 | 1906199 |
| 10 | 24 | 6 | 6 | 3046572 | 332 | 439139600 | 9.780439e+12 | J.K. Rowling, Mary GrandPré | 2000.0 | Harry Potter and the Goblet of Fire | ... | 1753043 | 1868642 |

5 rows × 23 columns

```python
most_selling_books
```

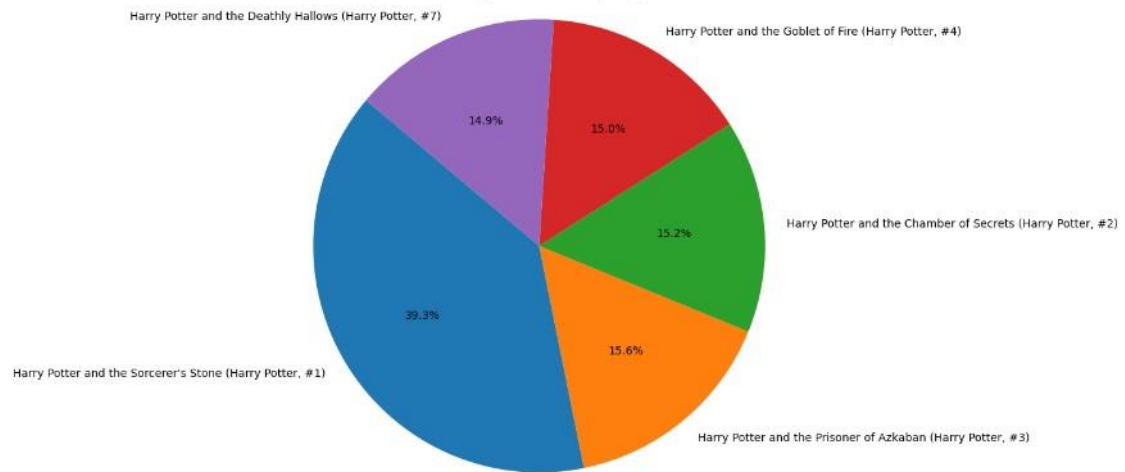| | book_id | goodreads_book_id | best_book_id | work_id | books_count | isbn | isbn13 | authors | original_publication_year | original_title | ... | ratings_count | work_ratings_count |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 3 | 4640799 | 491 | 439554934 | 9.780440e+12 | J.K. Rowling, Mary GrandPré | 1997.0 | Harry Potter and the Philosopher's Stone | ... | 4602479 | 4800065 |
| 6 | 18 | 5 | 5 | 2402163 | 376 | 043965548X | 9.780440e+12 | J.K. Rowling, Mary GrandPré, Rufus Beck | 1999.0 | Harry Potter and the Prisoner of Azkaban | ... | 1832823 | 1969375 |
| 9 | 23 | 15881 | 15881 | 6231171 | 398 | 439064864 | 9.780439e+12 | J.K. Rowling, Mary GrandPré | 1998.0 | Harry Potter and the Chamber of Secrets | ... | 1779331 | 1906199 |
| 10 | 24 | 6 | 6 | 3046572 | 332 | 439139600 | 9.780439e+12 | J.K. Rowling, Mary GrandPré | 2000.0 | Harry Potter and the Goblet of Fire | ... | 1753043 | 1868642 |
| 11 | 25 | 136251 | 136251 | 2963218 | 263 | 545010225 | 9.780545e+12 | J.K. Rowling, Mary GrandPré | 2007.0 | Harry Potter and the Deathly Hallows | ... | 1746574 | 1847395 |

5 rows × 23 columns

```
plt.figure(figsize=(10, 6))
plt.bar(most_selling_books['title'], most_selling_books['average_rating'], color='skyblue')
plt.title('Average Rating of Most Selling Harry Potter Books')
plt.xlabel('Book Title')
plt.ylabel('Average Rating')
plt.xticks(rotation=45, ha='right')
plt.show()
```

```
plt.figure(figsize=(8, 8))
plt.pie(most_selling_books['ratings_count'], labels=most_selling_books['title'], autopct='%1.1f%%', startangle=140)
plt.title('Distribution of Ratings for Most Selling Harry Potter Books')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
```

Distribution of Ratings for Most Selling Harry Potter Books



```
average_rating
```

4.482727272727273