Cloud Computing

Assignment 2

Name: Samaa Mohamed Shawky

DEP.: Cyber security

ID: 2206200

Analysis code:

```
import pandas as pd

# Load the dataset
df = pd.read_csv("books.csv")
# Perform data cleaning and preprocessing
# Handle missing values
df.dropna(inplace=True)

# Remove unnecessary columns
df.drop(['book_id', 'goodreads_book_id', 'best_book_id', 'work_id', 'isbn', 'isbn13', 'image_url', 'small_image_url'], axis=1, inplace=True)

# Filter dataset for Harry Potter series
harry_potter_df = df[df['authors'].str.contains('J.K. Rowling') & df['title'].str.contains('Harry Potter')]

# Check the filtered dataset0
harry_potter_df.head()

# Find the most selling books within the Harry Potter series
most_selling_books = harry_potter_df.sort_values(by='ratings_count', ascending=False).head()

# Display the most selling books
most_selling_books[['title', 'authors', 'ratings_count', 'average_rating']]

# Calculate the average rating of the Harry Potter books
average_rating_hp = harry_potter_df['average_rating'].mean()

# Display the average rating
print(f"Average rating of Harry Potter books: {average_rating_hp}")
print(most_selling_books[['title','ratings_count','books_count']])
```

Analysis Result:

```
Average rating of Harry Potter books: 4.550000000000001
                                                 title  ratings_count  \
1    Harry Potter and the Sorcerer's Stone (Harry P...        4602479
6    Harry Potter and the Prisoner of Azkaban (Harr...        1832823
9    Harry Potter and the Chamber of Secrets (Harry...        1779331
10   Harry Potter and the Goblet of Fire (Harry Pot...        1753043
11   Harry Potter and the Deathly Hallows (Harry Po...        1746574

    books_count
1           491
6           376
9           398
10          332
11          263
```