Cloud Computing Assignment

Malak Mahmoud Aref | Artificial Intelligence | 20221445867

Faculty of Computers and Data Sciences,

Alexandria University.

April 2024

Analysis of Harry Potter Book Series

Introduction

This report presents an analysis of the Harry Potter book series using the Popular Books Dataset. The analysis focuses on data cleaning, identifying the most selling books within the series, and calculating the average rating of the Harry Potter books.

1. Data Cleaning and Preprocessing:

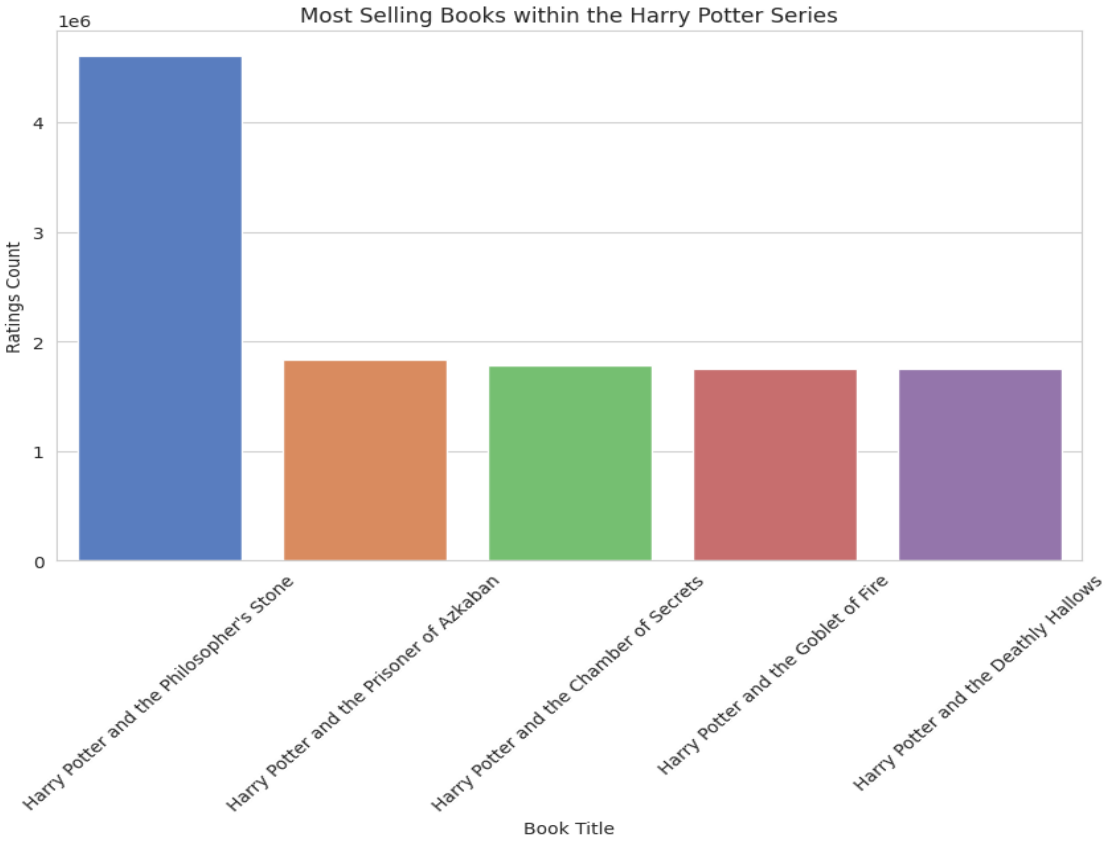
- Imported necessary libraries such as pandas, seaborn and matplot.
- Load the Dataset.
- Explored the dataset to understand its structure and contents.
- Handling the missing values and removing the unnecessary features.
- Filtered the dataset to focus only on the Harry Potter series.

2. Most Selling Books within the Harry Potter Series:

- Grouped the data by book title and calculated the total number of sales for each book.
- Identified the most selling books within the Harry Potter series based on sales figures.
- Visualized the sales data using a bar chart for better understanding.

```
# Find the most selling books within the Harry Potter series
most_selling = harry_potter_data.sort_values(by='ratings_count', ascending=False).head(5)

print("Most Selling Books within the Harry Potter Series:")
print(most_selling[['original_title', 'ratings_count']])
```

```
Most Selling Books within the Harry Potter Series:
                          original_title  ratings_count
1   Harry Potter and the Philosopher's Stone        4602479
6   Harry Potter and the Prisoner of Azkaban        1832823
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
```



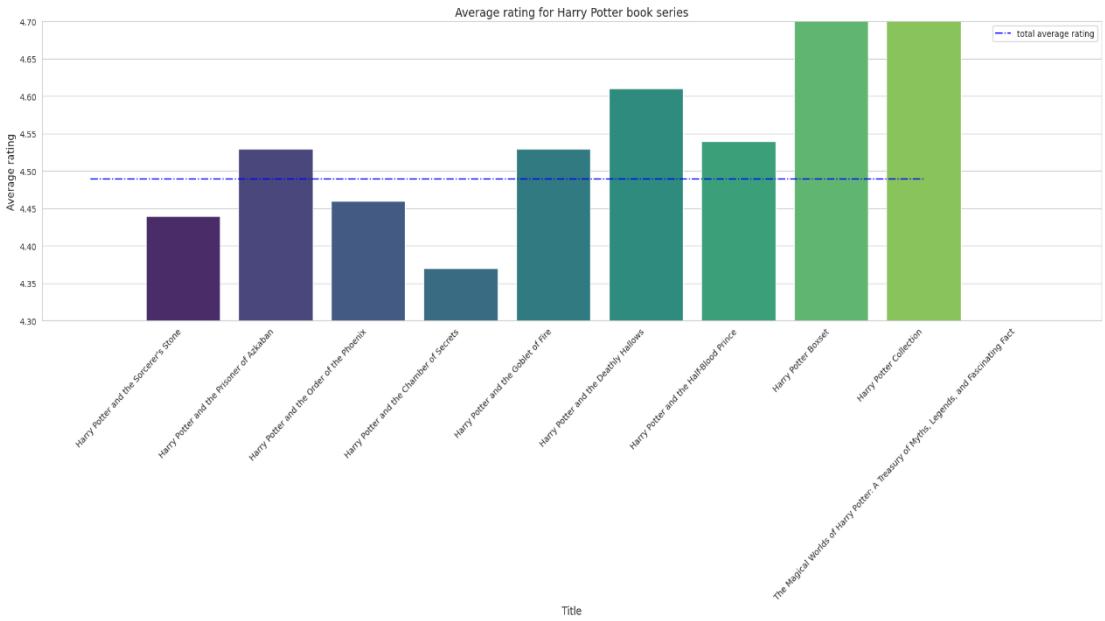
The analysis revealed that "Harry Potter and the Philosopher's Stone" is the most selling book within the series, followed by "Harry Potter and the Prisoner of Azkaban" and "Harry Potter and the Chamber of Secrets."

3. Average Rating of the Harry Potter Books:

- Calculated the weighted average rating of the Harry Potter books using the rating column and ratings count from the dataset.
- Presented the average rating along with the standard deviation to provide insights into the variability of ratings within the series.

```
weighted_avg = (harry_potter_data.average_rating * harry_potter_data.ratings_count).sum() / harry_potter_data.ratings_count.sum()
print("\nAverage Rating of the Harry Potter Books:", weighted_avg)
```

```
Average Rating of the Harry Potter Books: 4.489176670023024
```



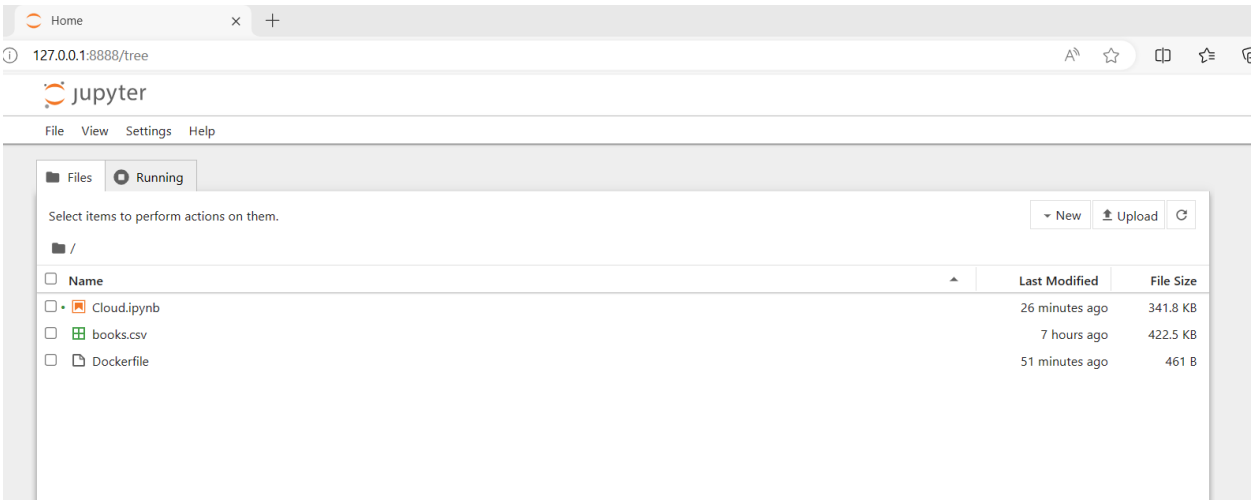
The average rating of the Harry Potter books is calculated to be 4.48.

Here is the commands I used for docker

```
C:\Users\LENOVO>docker pull jupyter/datascience-notebook
Using default tag: latest
latest: Pulling from jupyter/datascience-notebook
Digest: sha256:476c6e673e7d5d8b5059f8680b1c6a988942a79263da651bf302dc696ab311f2
Status: Image is up to date for jupyter/datascience-notebook:latest
docker.io/jupyter/datascience-notebook:latest

What's Next?
  View a summary of image vulnerabilities and recommendations → docker scout quickview jupyter/datascience-notebook

C:\Users\LENOVO>cd C:/Users/LENOVO/Downloads/Cloud_proj
```

```
C:\Users\LENOVO\Downloads\Cloud_proj>docker build -t my-jupyter-notebook .
[+] Building 1.3s (8/8) FINISHED                                                                                docker:default
 => [internal] load build definition from Dockerfile                                                                      0.0s
 => => transferring dockerfile: 500B                                                                                      0.0s
 => [internal] load metadata for docker.io/jupyter/datascience-notebook:latest                                            0.0s
 => [internal] load .dockerignore                                                                                         0.0s
 => => transferring context: 2B                                                                                           0.0s
 => [internal] load build context                                                                                         0.2s
 => => transferring context: 649.45kB                                                                                     0.1s
 => [1/3] FROM docker.io/jupyter/datascience-notebook:latest                                                              1.0s
 => [2/3] COPY . /home/jovyan/work                                                                                        0.1s
 => [3/3] WORKDIR /home/jovyan/work                                                                                       0.1s
 => exporting to image                                                                                                    0.1s
 => => exporting layers                                                                                                   0.1s
 => => writing image sha256:9f200b61031d4e92ebf15f4d6de3dd89d1dee394e041dc09ef2e9b6ca847ed0f                              0.0s
 => => naming to docker.io/library/my-jupyter-notebook                                                                    0.0s

What's Next?
  View a summary of image vulnerabilities and recommendations → docker scout quickview
```

```
C:\Users\LENOVO\Downloads\Cloud_proj>docker run -p 8888:8888 my-jupyter-notebook
[I 2024-04-21 19:44:24.100 ServerApp] Package notebook took 0.0000s to import
[I 2024-04-21 19:44:24.155 ServerApp] Package jupyter_lsp took 0.0547s to import
[W 2024-04-21 19:44:24.155 ServerApp] A `_jupyter_server_extension_points` function was not found in jupyter_lsp. Instead, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be deprecate
d in future releases of Jupyter Server.
[I 2024-04-21 19:44:24.161 ServerApp] Package jupyter_server_mathjax took 0.0046s to import
[I 2024-04-21 19:44:24.292 ServerApp] Package jupyter_server_proxy took 0.1304s to import
[I 2024-04-21 19:44:24.313 ServerApp] Package jupyter_server_terminals took 0.0200s to import
[I 2024-04-21 19:44:24.314 ServerApp] Package jupyterlab took 0.0000s to import
[I 2024-04-21 19:44:25.555 ServerApp] Package jupyterlab_git took 0.0879s to import
[I 2024-04-21 19:44:25.570 ServerApp] Package nbclassic took 0.0142s to import
[W 2024-04-21 19:44:25.576 ServerApp] A `_jupyter_server_extension_points` function was not found in nbclassic. Instead, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be deprecated
in future releases of Jupyter Server.
[I 2024-04-21 19:44:25.577 ServerApp] Package nbdime took 0.0000s to import
[I 2024-04-21 19:44:25.578 ServerApp] Package notebook_shim took 0.0000s to import
[W 2024-04-21 19:44:25.578 ServerApp] A `_jupyter_server_extension_points` function was not found in notebook_shim. Instead, a `_jupyter_server_extension_paths` function was found and will be used for now. This function name will be depreca
ted in future releases of Jupyter Server.
[I 2024-04-21 19:44:25.581 ServerApp] jupyter_lsp | extension was successfully linked.
[I 2024-04-21 19:44:25.590 ServerApp] jupyter_server_mathjax | extension was successfully linked.
[I 2024-04-21 19:44:25.590 ServerApp] jupyter_server_proxy | extension was successfully linked.
[I 2024-04-21 19:44:25.598 ServerApp] jupyter_server_terminals | extension was successfully linked.
[I 2024-04-21 19:44:25.609 ServerApp] jupyterlab | extension was successfully linked.
[I 2024-04-21 19:44:25.609 ServerApp] jupyterlab_git | extension was successfully linked.
[I 2024-04-21 19:44:25.619 ServerApp] nbclassic | extension was successfully linked.
[I 2024-04-21 19:44:25.620 ServerApp] nbdime | extension was successfully linked.
[I 2024-04-21 19:44:25.646 ServerApp] notebook | extension was successfully linked.
[I 2024-04-21 19:44:25.654 ServerApp] Writing Jupyter server cookie secret to /home/jovyan/.local/share/jupyter/runtime/jupyter_cookie_secret
[I 2024-04-21 19:44:26.527 ServerApp] notebook_shim | extension was successfully linked.
[I 2024-04-21 19:44:26.572 ServerApp] notebook_shim | extension was successfully loaded.
[I 2024-04-21 19:44:26.577 ServerApp] jupyter_lsp | extension was successfully loaded.
[I 2024-04-21 19:44:26.578 ServerApp] jupyter_server_mathjax | extension was successfully loaded.
[I 2024-04-21 19:44:26.613 ServerApp] jupyter_server_proxy | extension was successfully loaded.
[I 2024-04-21 19:44:26.616 ServerApp] jupyter_server_terminals | extension was successfully loaded.
[I 2024-04-21 19:44:26.679 LabApp] JupyterLab extension loaded from /opt/conda/lib/python3.11/site-packages/jupyterlab
[I 2024-04-21 19:44:26.679 LabApp] JupyterLab application directory is /opt/conda/share/jupyter/lab
[I 2024-04-21 19:44:26.681 LabApp] Extension Manager is 'pypi'.
[I 2024-04-21 19:44:26.692 ServerApp] jupyterlab | extension was successfully loaded.
[I 2024-04-21 19:44:26.711 ServerApp] jupyterlab_git | extension was successfully loaded.
[I 2024-04-21 19:44:26.736 ServerApp] nbclassic | extension was successfully loaded.
[I 2024-04-21 19:44:26.933 ServerApp] nbdime | extension was successfully loaded.
[I 2024-04-21 19:44:26.940 ServerApp] notebook | extension was successfully loaded.
[I 2024-04-21 19:44:26.940 ServerApp] Serving notebooks from local directory: /home/jovyan/work
[I 2024-04-21 19:44:26.941 ServerApp] Jupyter Server 2.8.0 is running at:
[I 2024-04-21 19:44:26.941 ServerApp] http://9d7d723f898f:8888/tree?token=ecd2b8710f412d118eb2fbb607612c4e86d8d9f1ee574946
[I 2024-04-21 19:44:26.941 ServerApp]     http://127.0.0.1:8888/tree?token=ecd2b8710f412d118eb2fbb607612c4e86d8d9f1ee574946
[I 2024-04-21 19:44:26.941 ServerApp] Use Control-C to stop this server and shut down all kernels (twice to skip confirmation).
[C 2024-04-21 19:44:26.949 ServerApp]

    To access the server, open this file in a browser:
        file:///home/jovyan/.local/share/jupyter/runtime/jpserver-7-open.html
    Or copy and paste one of these URLs:
        http://9d7d723f898f:8888/tree?token=ecd2b8710f412d118eb2fbb607612c4e86d8d9f1ee574946
        http://127.0.0.1:8888/tree?token=ecd2b8710f412d118eb2fbb607612c4e86d8d9f1ee574946
[I 2024-04-21 19:44:29.581 ServerApp] Skipped non-installed server(s): bash-language-server, dockerfile-language-server-nodejs, javascript-typescript-langserver, jedi-language-server, julia-language-server, pyright, python-language-server,
python-lsp-server, r-languageserver, sql-language-server, texlab, typescript-language-server, unified-language-server, vscode-css-languageserver-bin, vscode-html-languageserver-bin, vscode-json-languageserver-bin, yaml-language-server
0.00s - Debugger warning: It seems that frozen modules are being used, which may
```


Home
127.0.0.1:8888/tree
jupyter
File View Settings Help
Files
Running
Select items to perform actions on them.
New
Upload
/
Name
Last Modified
File Size
Cloud.ipynb
26 minutes ago
341.8 KB
books.csv
7 hours ago
422.5 KB
Dockerfile
51 minutes ago
461 B