```python
#%%
import re
import pickle
import operator
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
from collections import Counter
from scipy.sparse import csr_matrix
from pandas.api.types import is_numeric_dtype
from sklearn.neighbors import NearestNeighbors
from sklearn.feature_extraction import DictVectorizer
from sklearn.metrics.pairwise import cosine_similarity
from sklearn.feature_extraction.text import TfidfVectorizer

import warnings
warnings.filterwarnings('ignore')
#%%
books = pd.read_csv(r'Books.csv', delimiter=';', encoding='ISO-8859-1', on_bad_lines='skip')
users = pd.read_csv(r'Users.csv', delimiter=';', encoding='ISO-8859-1', on_bad_lines='skip')
ratings = pd.read_csv(r'Book-Ratings.csv', delimiter=';', encoding='ISO-8859-1', on_bad_lines='skip')


print('Books: ', books.shape)
print('Users: ', books.shape)
print('Book Rating: ', ratings.shape)

#%%
books.head()
#%%
# Pre-Processing

print("Coluns: ", books.columns)
```

```
#%%
books.drop(['Image-URL-L', 'Image-URL-M', 'Image-URL-S'], axis=1, inplace=True)
books.head()
#%%
books.isnull().sum()
#%%
books.loc[books['Book-Author'].isnull()]
#%%
books.loc[books['Publisher'].isnull()]
#%%
books.at[187689, 'Book-Author'] = 'Other'
books.at[118033, 'Book-Author'] = 'Other'
books.at[128890, 'Publisher'] = 'Other'
books.at[129037, 'Publisher'] = 'Other'
#%%
books.loc[books['Book-Author'].isnull()] #Null authors are removed
#%%
books['Year-Of-Publication'].unique()
#%%
pd.set_option('display.max_colwidth', -1)
#%%
books.loc[books['Year-Of-Publication'] == 'DK Publishing Inc']
#%%
books.loc[books['Year-Of-Publication'] == 'Gallimard']
#%%
# Change Year of Publication to actual Year and not Publisher Name
books.at[209538, 'Publisher'] = 'DK Publishing Inc'
books.at[209538, 'Year-Of-Publication'] = 2000
books.at[209538, 'Book-Title'] = 'DK Readers: Creating the X-Men, How It All Began (Level 4: Proficient Readers)'
books.at[209538, 'Book-Author'] = 'Michael Teitelbaum'


books.at[221678, 'Publisher'] = 'DK Publishing Inc'
books.at[221678, 'Year-Of-Publication'] = 2000
books.at[221678, 'Book-Title'] = 'DK Readers: Creating the X-Men, How Comic Books Come to Life (Level 4: Proficient Readers)'
books.at[221678, 'Book-Author'] = 'James Buckley'


books.at[220731, 'Publisher'] = 'Gallimard'
books.at[220731, 'Year-Of-Publication'] = 2003
books.at[220731, 'Book-Title'] = 'Peuple du ciel, suivi de les bergers'
books.at[220731, 'Book-Author'] = 'Jean-Marie, Gustave Le ClÃ?Â©zio'


#%%
#Converting year of publication in Numbers
books['Year-Of-Publication']=books['Year-Of-Publication'].astype(float)

print(sorted(list(books['Year-Of-Publication'].unique())))
#%%
#Replacing invalid years with max year
count = Counter(books['Year-Of-Publication'])
[k for k, v in count.items() if v ==max(count.values())]
#%%

```