**AIWR REPORT ASSIGNMENT 2**

Team members :

1. Sehag A - PES2UG20CS457
2. Setti Durga Poojitha - PES2UG20CS458
3. Shreyas Sai Raman - PES2UG20CS461
4. Varshini Jayasankar - PES2UG20CS567

**Introduction:**

Book recommendation System.

In today's digital age, with an abundance of books available in various genres and formats, it can be overwhelming for readers, especially students, to navigate through the vast sea of options and find books that align with their interests and preferences. This is where a book recommendation system comes in handy.

A book recommendation system is a powerful tool that leverages data analysis, machine learning, and artificial intelligence techniques to analyse and understand readers' preferences, behaviours, and reading patterns. It then provides personalised book recommendations that match their individual tastes, helping them discover new books they might not have come across otherwise.

**Problem Statement:**

We are building a model based collaborative filtering and a content based recommendation systems that have been trained on a huge dataset consisting of books and ratings. So the model will recommend books to the users based upon preferences. The recommendation also depends upon the type of the recommendation model used. We are also building a hybrid model which is a combination of model based and content based recommendation systems.

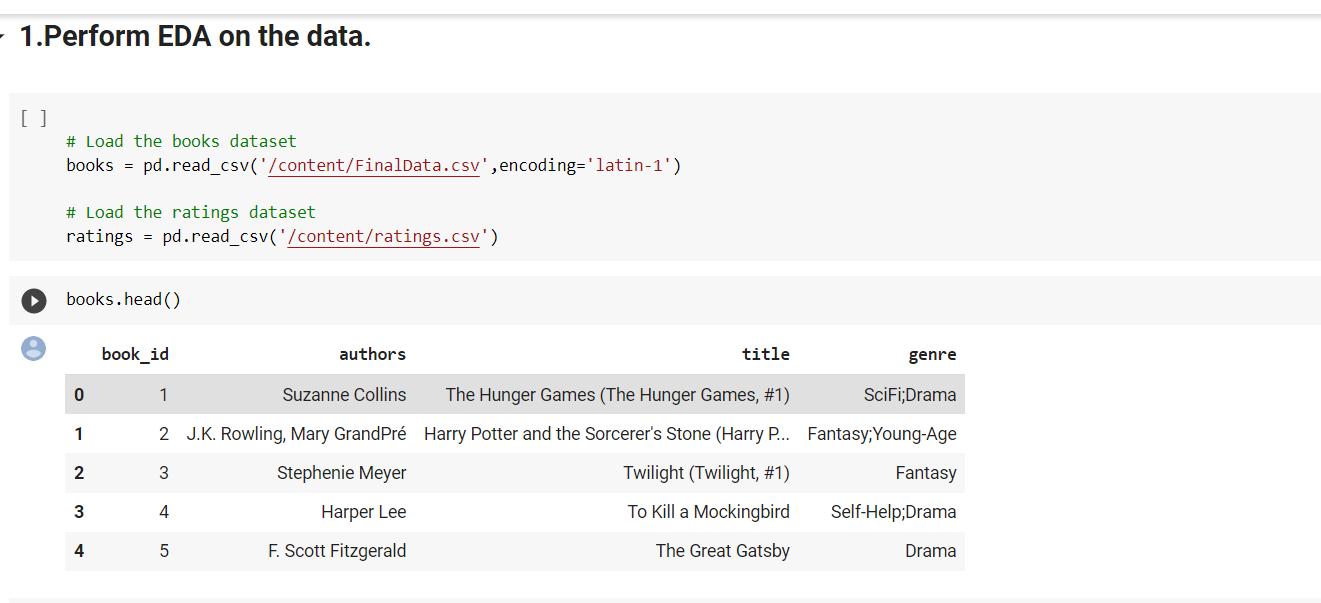
**Data Set Description:**

We have used a books dataset from kaggle which has around 7 csv files.Out of the 7 files,we have primarily used book.csv and ratings.csv files.The books.csv files consists of about 999 book titles along with their respective authors and many other columns such as book\_id,worktext\_review,work reviews etc.

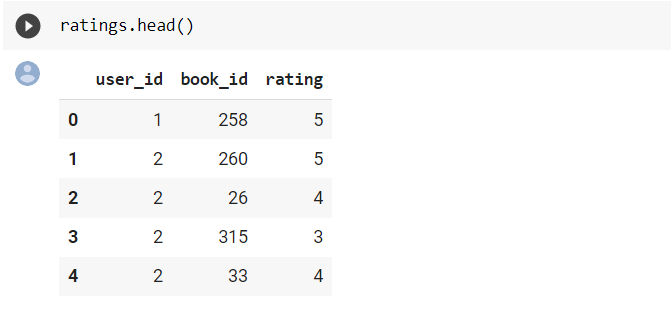
The ratings.csv file consists of ratings given by 56713 users for the books.

It consists of the attribute columns book\_id,user\_id and their corresponding ratings.

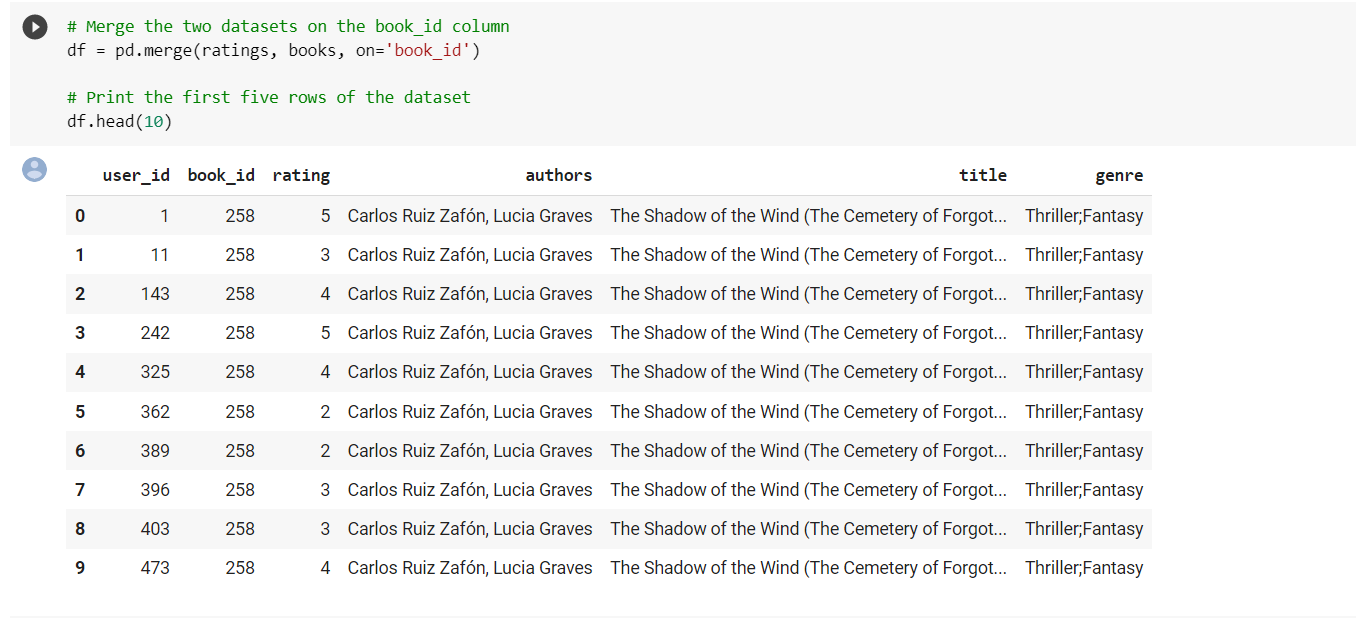
**1.EDA**

(a)Loading the dataset and viewing the first 5 rows of the books csv file after retaining only the necessary column attributes

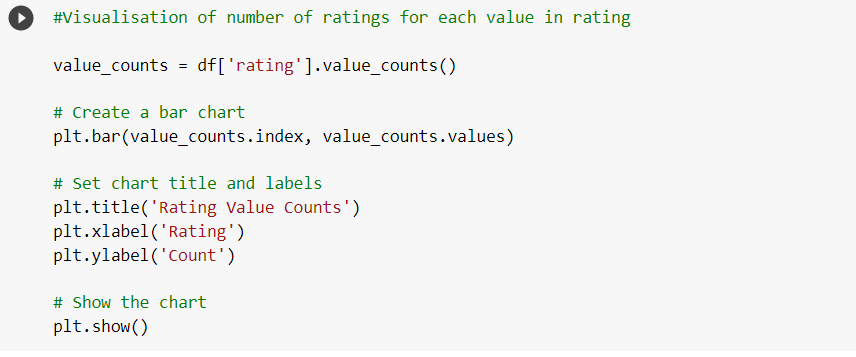
(b) Viewing the first 5 rows of ratings.csv file

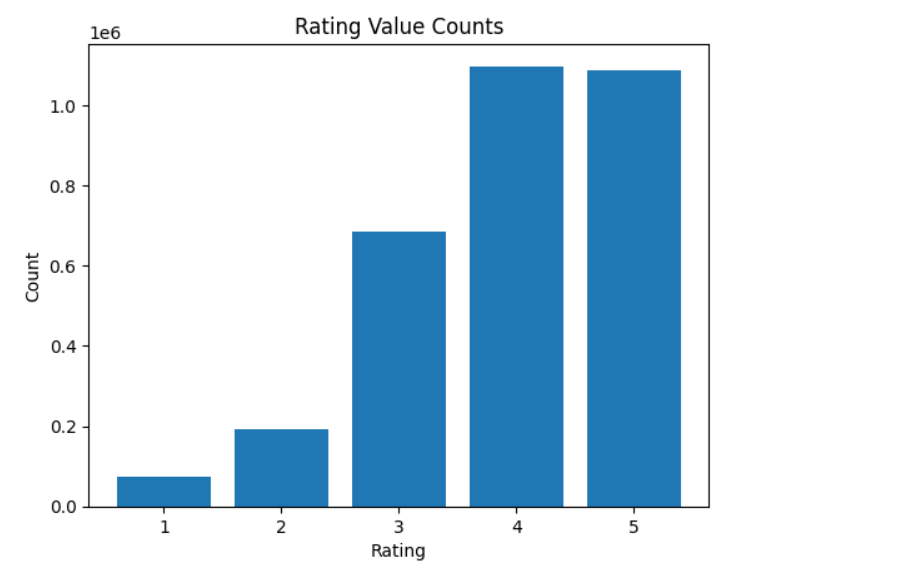


©.Merging the two datasets on the book id column and viewing the first few rows o the result



(d)Making bar chart to visualise the number of ratings for each value in string

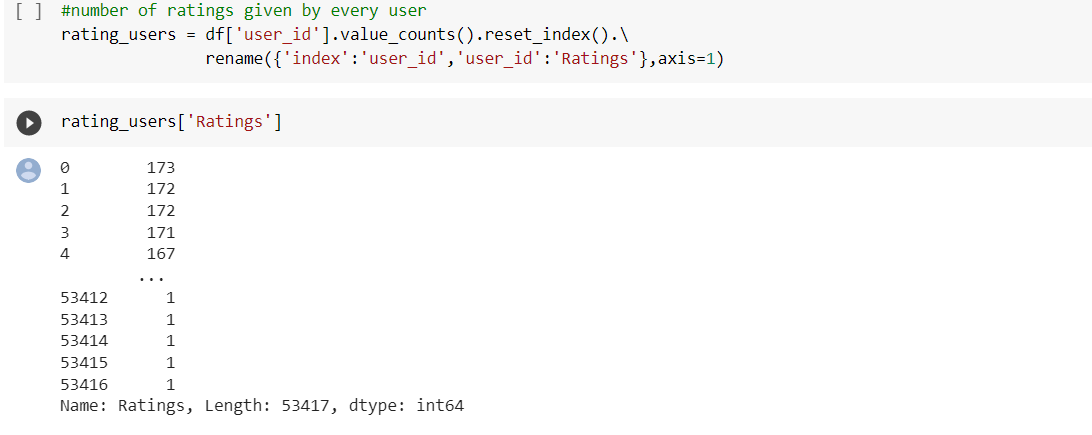




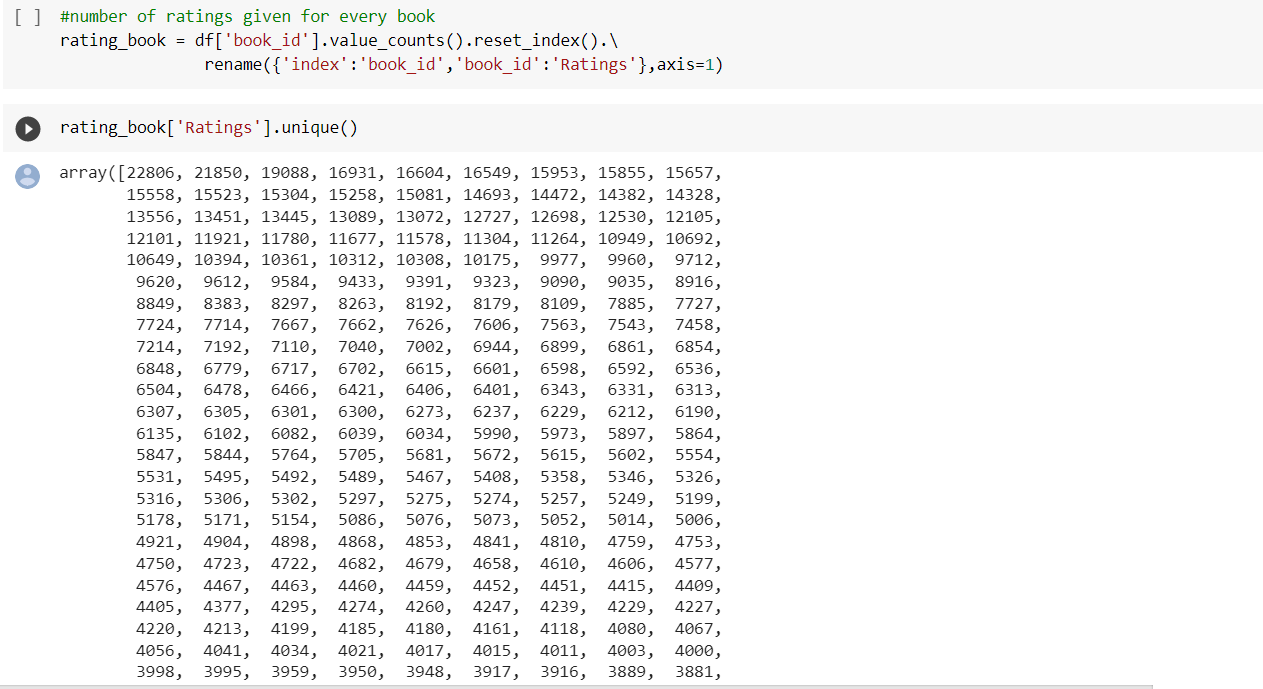
**2.DATA PREPROCESSING:**

We need to preprocess the data before we can use it to build our recommendation models.This involves removing any irrelevant columns, handling missing values, and converting categorical variables into numerical values.

(a)General preprocessing



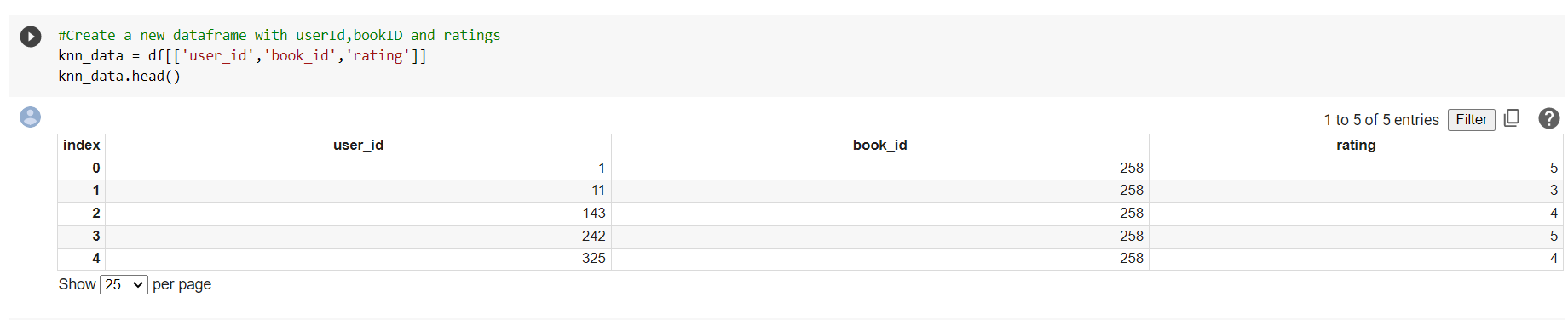
(b)Counting the number of ratings for each book



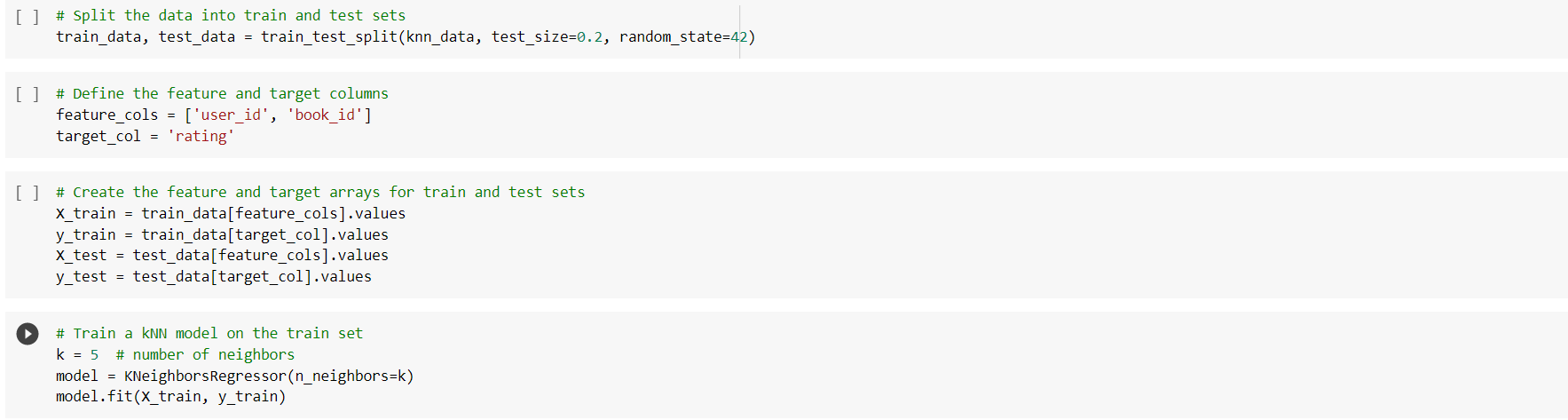
**RECOMMENDATION SYSTEMS**

**1.Model based collaborative filtering system(KNN)**

(a)Creating a new dataframe with userId,bookID and the ratings

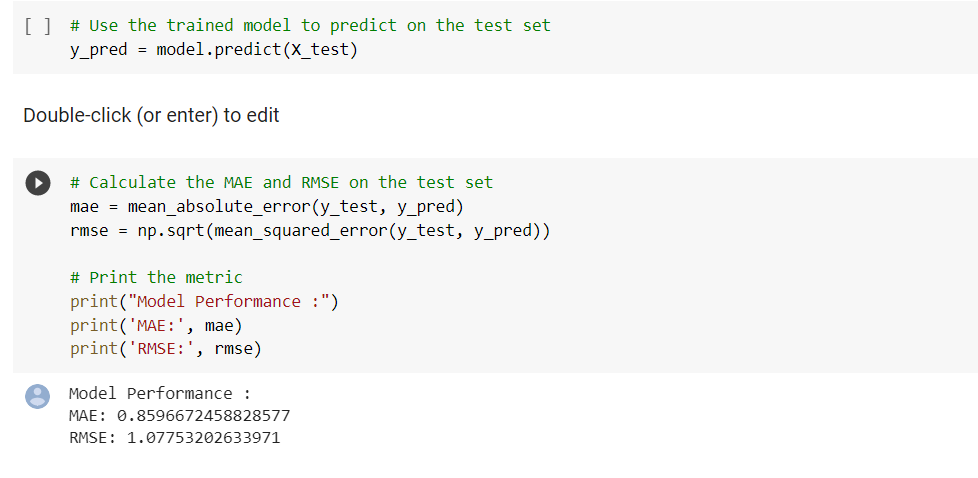


(b)Splitting the dataset into training and testing data and training the KNN model on the training dataset.Feature and Target arrays were also created for the train and test datasets.



(c) Using the trained KNN model to recommend on the test data set.

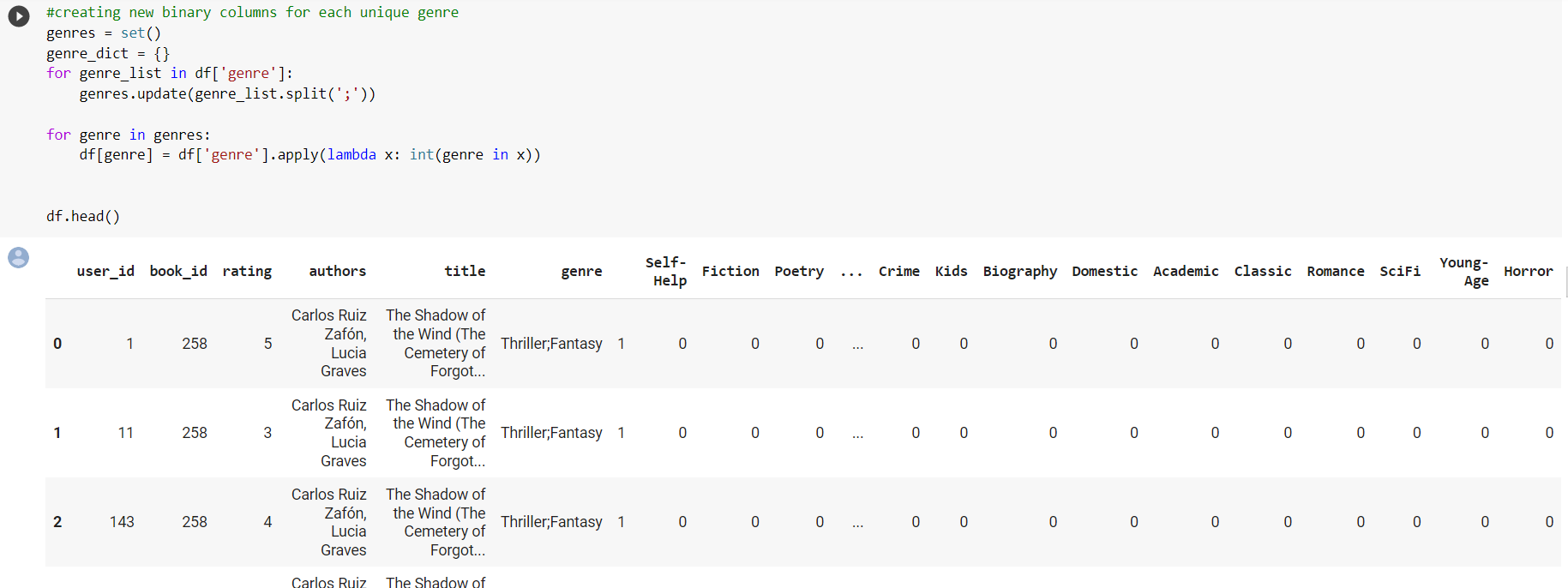
Calculating accuracy metrics such MAE and RMSE to test the accuracy of the model



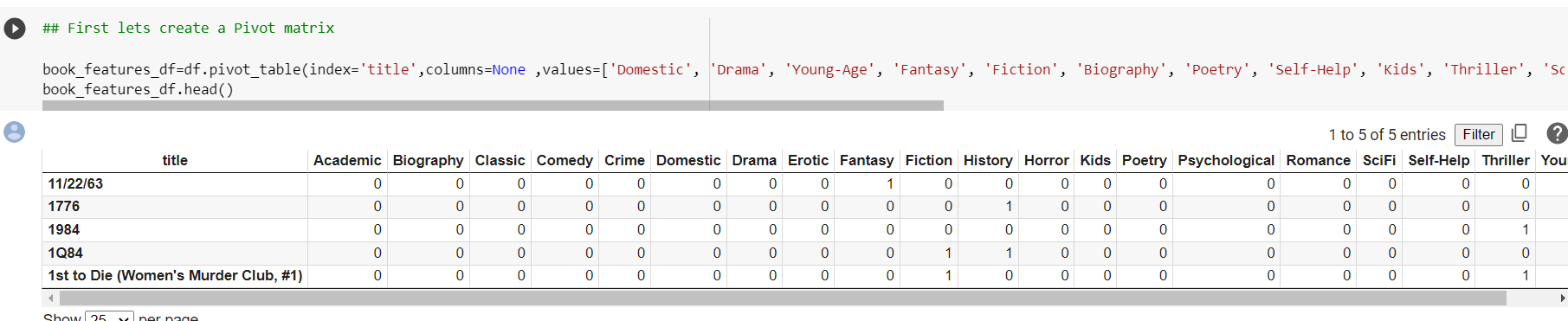
As you can see,the MAE value is 0.85966 and the RMSE value is 1.0775 which proves that the model performance is accurate.

**2.Content based recommendation system.**

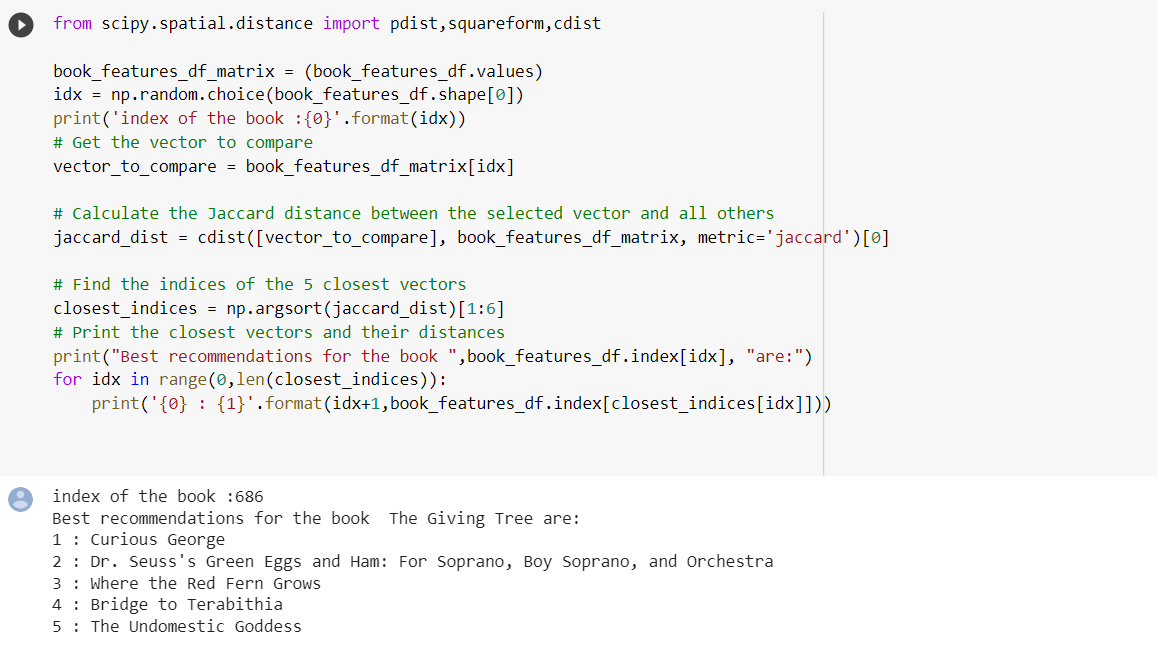
1. First we are creating a binary column for each of the unique genres(Fiction,poetry,crime,biography,etc)



(b)Creating a pivot matrix using the different genres.



(c)Building the content based recommendation system by calculating jaccard’s coefficient between the user’s preferences and the book’s features in the training data after converting them to vectors.After Jaccard’s coefficient is calculated,5 closest vectors which are books most similar to the user’s preferences are recommended.Their corresponding distances are also displayed.



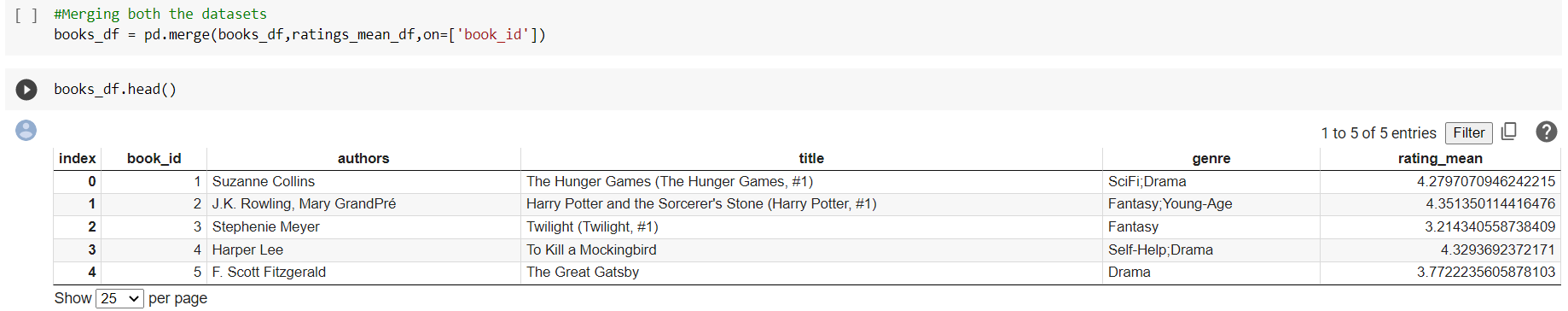
**3)Hybrid recommendation model-(Combination of Content based and model based collaborative filtering)**

(a)Loading the required csv files and reducing the number of records in the dataset for easier computation.

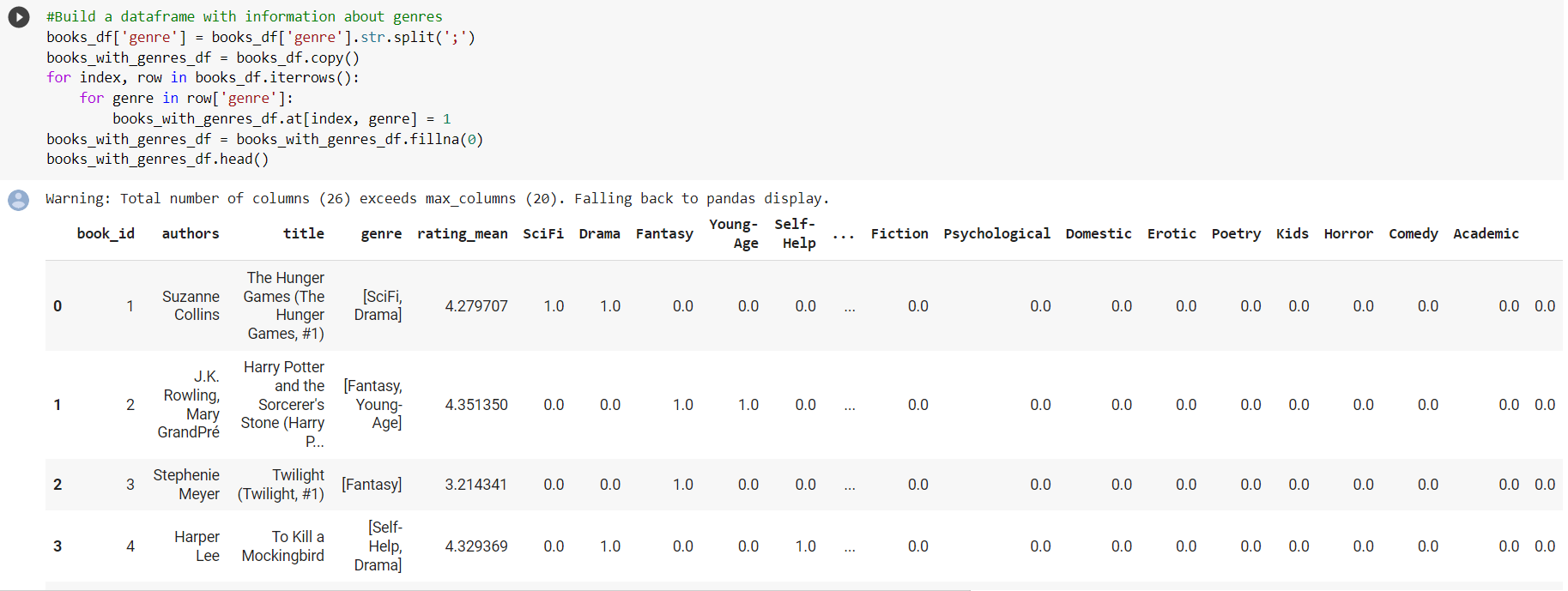
The average rating for each of the books is also calculated.



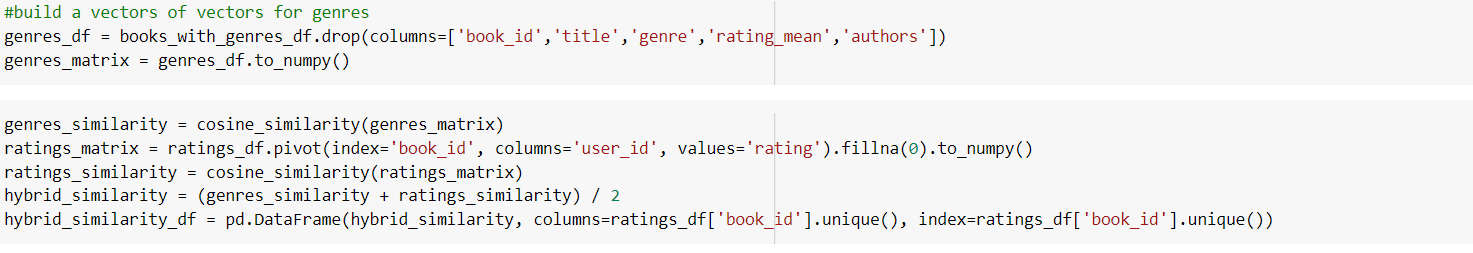
(b)Both the books.csv and ratings.csv files are merged on book\_id



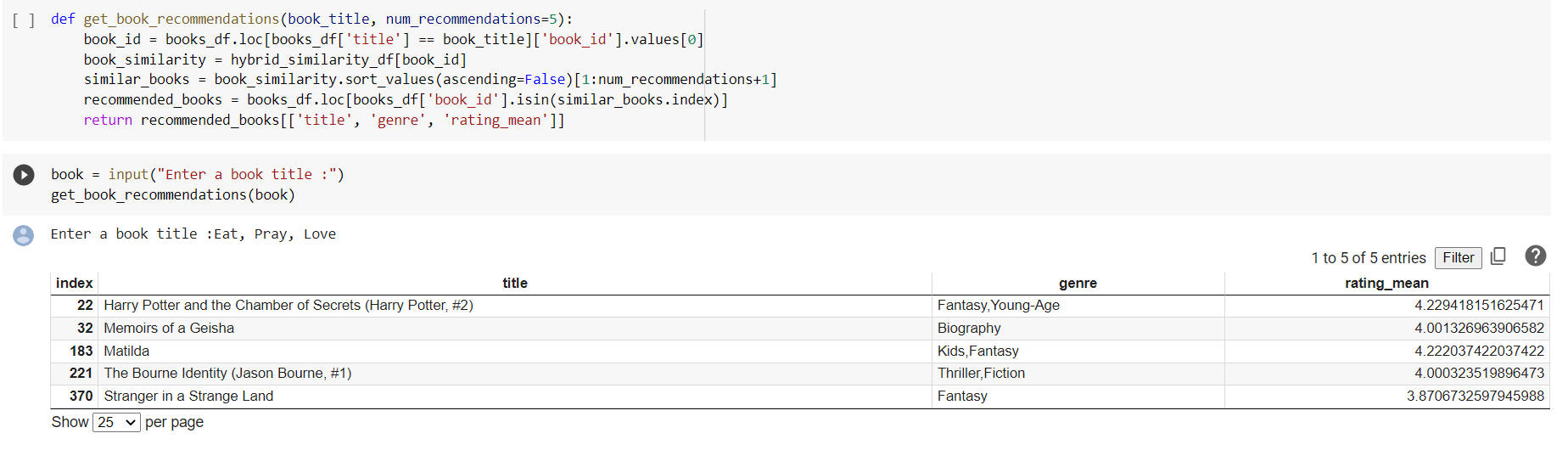
(c)Building a dataframe with information about the genres



(d)Converting the various genres into vectors and calculating the similarities between books

Using cosine similarity 

(e)Defining a method to make recommendations similar to the input book entered by the user.This is achieved after using cosine similarity to calculate the measure of similarities between the books



As you can see, after user gave the input :’Eat,pray,love’

The recommendation model recommends 5 books similar to the input book given by the user