**1. Introduction**

Nowadays, with all the advantages that technology offers us, many businesses are virtualized like for example the books that before we used to read them physically but now, we can even read them online. Considering the previous factors, in this report I will analyse a book dataset which has different details of users, books, rating, among others.

Recommendation systems are used and applied in different businesses to attract the attention of the customers or users and offer them items that could be of their interest. Also, Market Basket Analysis is developed to offer them deals or understand the customer behaviour; that is why in the first part of this report, I will be developing the answers of Machine learning implementing different techniques and explaining how they work and why were they applied. Besides, Data Visualization techniques will be applied crating an interactive dashboard for seniors (+65 years old), in which with simple visualizations, I will summarize the important characteristics.

**2. Data**

The data was taken from Kaggle in the following link: <https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset/data> (Kaggle, 2023)

**2.1 Characterization of the dataset**

The dataset was compound of 3 excel files and I decided to work with 3 data frames with different dimensions since in every Recommended system and visualizations I selected the necessary features for their analysis and not losing the information they have inside.

* df\_book has 271360 rows and 8 columns.
* df\_rating has 11499780 rows and 8 columns.
* df\_users has 278858 rows and 3 columns.

**2.2 Data Dictionary**

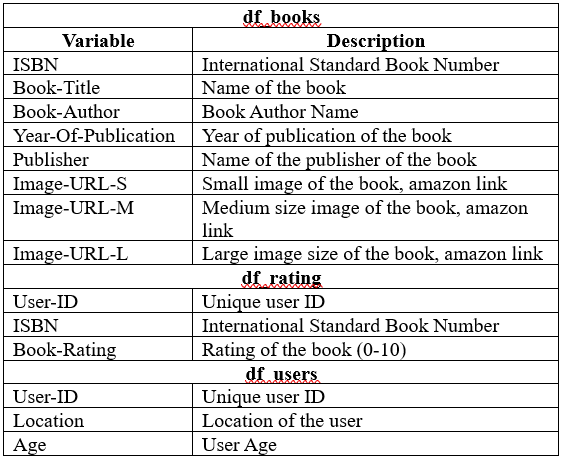


Figure 1: Data Dictionary of the dataset

As we can see in Figure 1, the datasets share common variables to adjust the necessary information for the different questions.

**2.3 Data Preparation**

After applying Data Preparation and Data Cleaning techniques, I got the next decisions and information:

* Drop 3 columns in df\_book where they had no information about year of publication.
* The 3 data frames don’t have duplicates
* As we can see in Figure 2, Null values present in “Book-Author” and “Publisher” were replaced by “Unknown” since they represented less than 1 percent in df\_books and “Age” column was dropped since it had around 39% of missing values which can cause variance to the results.

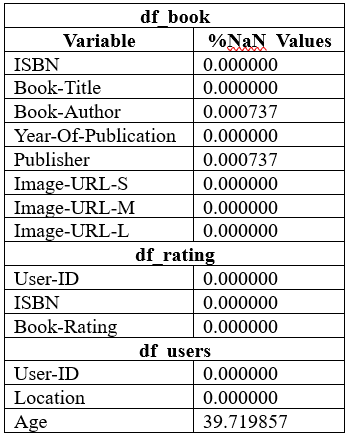


Figure 2: Percentage of Null values in the data frames

**3. Machine Learning**

**3.1 Question 1**

**3.1.1 Discuss and explain the purpose of a recommendation system for online retail business in machine learning.**

A recommendation system in online retail has the purpose to suggest customers or users services or products that they would be interested to get, buy or read according to previous data like previous sells, streaming services, demographics among others characteristics analysed in which using Machine Learning Recommendation Systems we can predicts what are the interests of the customers and give us the option to offer it to them, the final purpose is to give more value to the company making the user be more interested on us and for the user the benefit is the time since we are offering products or services he/she is interested.

**3.1.2 Briefly compare Content and Collaborative filtering using any dataset of your choice (Datasets used in the class tutorials or exercises are not allowed to use in this CA2).**

**Content Based Recommended System**

According to Kulkarni, A.B. (Kulkarni et al., 2022) Content-based filtering is used in recommending products or items very similar to those being clicked or liked. User recommendations are based on the description of an item and a profile of the user’s interest. Content-based recommender systems are widely used in e-commerce platforms. It is one of the basic algorithms in the recommendation engine. Content-based filtering can be triggered for any event; for example, on click, on purchase, or add to cart.

To start working with this system I merged df\_book with df\_rating creating columns of Average\_Rating and rating count, but as we can see in Figure 3, many people read books and don’t rate them showing as zero. So, I didn’t consider the ratings as zero for this analysis.

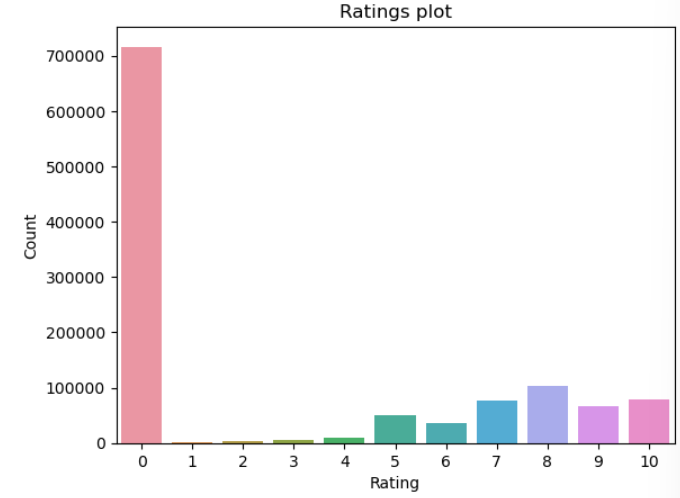


Figure 3: Ratings plot according to the number of users

To start I calculated the weighted rating with the next formula in Figure 4:

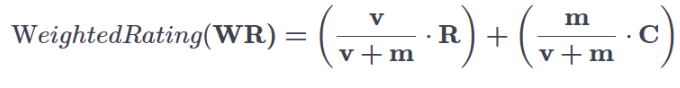


Figure 4: Weighted Rating Formula

Where:

v: number of people that rated the books (Rating\_Count)

m: minimum of rated books required to be listed in the chart

R: Average Rating of the books

C: mean rated count of the books across the dataframe

In my results I got that:

* C = 7.53 on a scale

**3.1.3 Train and test machine learning models for the user-user or item-item collaborative filtering. Justify your recommendations for the considered scenario by providing a conceptual insight.**

**2) Perform Market Basket Analysis on the chosen dataset by using Apriori and FP growth algorithms. Can you express major divergence between these models? Compare and contrast the machine learning results obtained based on both algorithms. (50, 50 = 100 marks)**

**Data Visualization**

**3) Create an interactive Dashboard aimed at older adults (65+) with specific features to summarise the most important aspects of the data and identify through your visualisation why this dataset is suitable for Machine Learning models in an online retail business. Explain how your dashboard is designed with this demographic in mind.**

www.kaggle.com. (2023). Book Recommendation Dataset. [online] Available at: <https://www.kaggle.com/datasets/arashnic/book-recommendation-dataset/data>

Kulkarni, A.B., Adarsha Shivananda, Kulkarni, A. and V Adithya Krishnan (2022). Content-Based Recommender Systems. Apress eBooks, pp.63–87. doi:https://doi.org/10.1007/978-1-4842-8954-9\_3.