ANALYSIS OF DATA SETS FOR A BOOK STORE

Data set description:

The data set consists of the various parameters related to books being sold at a book store. Three tables have been provided. The tables provided are in the .csv format.

Download the data set for the problem statement via the below Dropbox link.

Link: https://www.dropbox.com/s/088v50nvtk39kg2/BX-CSV-Dump.zip?dl=0

The tables are as under:

BX-Users:

This file contains the list of the users, their age and where the books were collected. If that data is unavailable for any field then it is filled with NULL.

BX-Books:

It gives us the details about the book such as Book-Title, Book-Author, Year of Publication, Publisher, Image-URL and ISBN. Here ISBN will act as a unique code for a book. Invalid ISBNs have already been removed from the dataset. URLs linking to cover images are also given, appearing in three different flavors (`Image-URL-S`, `Image-URL-M`, `Image-URL-L`) i.e. Small, Medium and Large. These URLs point to the Amazon web site.

BX-Book-Ratings:

It contains the book rating information. Ratings are either explicitly expressed on a scale from 1-10 (higher values denoting higher appreciation) or implicitly expressed by 0.

Parameters

ISBN:

International Standard Book Number is a 13-Digit number assigned by standard book numbering agencies to control and facilitate activities within the publishing industry. ISBNs of the books sold at the store is mentioned in the table.

Book-Title:

The book titles of all the books have been mentioned in the table.

Book-Author:

The Author's name for each book sold is mentioned in the table.

Year of Publication:

The Year in which each book was published is mentioned in the table.

Publisher:

The publisher's name is mentioned for each book.

Problem Statement:

- Find out the frequency of books published each year. (Hint: Use Boooks.csv file for this)

- Find out in which year the maximum number of books were published.

- Find out how many book were published based on ranking in the year 2002. (Hint: Use Book.csv and Book-Ratings.csv)