

Analyzing Book- Crossing Data

Data set Description:

The Book-Crossing dataset consists of 3 tables.

BX-Users:

This file contains the list of the users, their age and where they are 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, 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.

Problem Statement:

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