Best Books Ever Dataset

source: https://zenodo.org/records/4265096
addicional: https://github.com/scostap/goodreads_bbe_dataset

The dataset has been collected in the frame of the Prac1 of the subject Tipology and Data Life Cycle of the Master's Degree in Data Science of the Universitat Oberta de Catalunya (UOC).

The dataset contains 25 variables and 52478 records corresponding to books on the GoodReads Best Books Ever list (the larges list on the site).

Original code used to retrieve the dataset can be found on github repository:
github.com/scostap/goodreads_bbe_dataset

The data was retrieved in two sets, the first 30000 books and then the remainig 22478. Dates were not parsed and reformated on the second chunk so publishDate and firstPublishDate are representet in a mm/dd/yyyy format for the first 30000 records and Month Day Year for the rest.

Book cover images can be optionally downloaded from the url in the 'coverImg' field. Python code for doing so and an example can be found on the github repo.

The 25 fields of the dataset are:

Attributes	Definition	Completeness
bookId	Book Identifier as in goodreads.com	100
title	Book title	100
series	Series Name	45
author	Book's Author	100
rating	Global goodreads rating	100
description	Book's description	97
language	Book's language	93
isbn	Book's ISBN	92
genres	Book's genres	91
characters	Main characters	26
bookFormat	Type of binding	97
edition	Type of edition (ex. Anniversary Edition)	9
pages	Number of pages	96
publisher	Editorial	93
publishDate	publication date	98
firstPublishDate	Publication date of first edition	59
awards	List of awards	20
numRatings	Number of total ratings	100
ratingsByStars	Number of ratings by stars	97
likedPercent	Derived field, percent of ratings over 2 starts (as in GoodReads)	99
setting	Story setting	22
coverImg	URL to cover image	99

```
| bbeScore | Score in Best Books Ever list | 100 |
| bbeVotes | Number of votes in Best Books Ever list | 100 |
| price | Book's price (extracted from Iberlibro) | 73 |
```


—

Attributes	Definition	Completeness
bookId	Book Identifier as in goodreads.com	100
title	Book title	100
series	Series Name	45
author	Book's Author	100
rating	Global goodreads rating	100
description	Book's description	97
language	Book's language	93
isbn	Book's ISBN	92
genres	Book's genres	91
characters	Main characters	26
bookFormat	Type of binding	97
edition	Type of edition (ex. Anniversary Edition)	9
pages	Number of pages	96

publisher	Editorial	93
publishDate	publication date	98
firstPublishDate	Publication date of first edition	59
awards	List of awards	20
numRatings	Number of total ratings	100
ratingsByStars	Number of ratings by stars	97
likedPercent	Derived field, percent of ratings over 2 starts (as in GoodReads)	99
setting	Story setting	22
coverImg	URL to cover image	99
bbeScore	Score in Best Books Ever list	100
bbeVotes	Number of votes in Best Books Ever list	100
price	Book's price (extracted from Iberlibro)	73