For this project, we have taken a dataset from Kaggle. This dataset is on Amazon's Top 50 bestselling books from 2009 to 2019. It keeps the record of 550 books in a .csv file.

Amazon's top 50 bestselling books

This is a Kaggle dataset in .csv format. It includes the information on name, author, user rating, reviews, price, year, and genre of 550 different books. So, data is arranged using the seven columns below.

Dataset entries

Name	Author	User Rating	Reviews	Price	Year	Genre
10-Day Green Smoothie Cleanse	JJ Smith	4.7	17350	8	2016	Non fiction
12 Rules for Life: An Antidote to Chaos	Jordan B. Peterson	4.7	18979	15	2018	Non fiction
1984 (Signet Classics)	George Orwell	4.7	21424	6	2017	Fiction
5,000 Awesome Facts (About Everything!) (National Geographic Kids)	National Geographic Kids	4.8	7665	12	2019	Non fiction

A Dance with Dragons (A Song of Ice and Fire)	George R. R. Martin	4.4	12643	11	2011	Fiction
...	...	...	...	...	...	...

The above table represents a book with various attributes detailing its characteristics and performance on Amazon. Let's discuss these columns as follows:

- `Name`: This column contains the title of the book
- `Author`: This column lists the author's name.
- `User Rating`: It shows the average Amazon user rating, which ranges from 3.3 to 4.9.
- `Reviews`: It indicates the number of reviews written by users on Amazon, with a minimum of 37 and a maximum of 87,800 reviews.
- `Price`: It provides the cost of the book, spanning from $0 to $105.
- `Year`: It specifies the year or years the book appeared on the bestseller list, covering the period from 2009 to 2019.
- `Genre`: Lastly, it classifies the book as either fiction or nonfiction.

Reading the Dataset

To begin working with the dataset, we need to read the data from a CSV file named data.csv. This file contains information about various books, structured in a tabular format. Each row represents a book and includes details such as the title, author,

user rating, number of reviews, price, publication year, and genre.

Define Book class

In this section, we will define a `Book` class that models the attributes of a book based on the dataset provided. The `Book` class will contain all the necessary details about each book, such as its title, author, user rating, number of reviews, price, publication year, and genre.

This class is designed to provide a structured way to manage and manipulate book data within our application.

Attributes:

    - `title`: The title of the book.
    - `author`: The author of the book.
    - `userRating`: The average user rating of the book.
    - `reviews`: The number of user reviews.
    - `price`: The price of the book.
    - `year`: The year the book appeared on the bestseller list.
    - `genre`: The genre of the book (either fiction or non-fiction).
- **Constructor**: Initializes a `Book` object with the provided values for each attribute.
- **Getters and setters**: These methods provide access to and modification of the book's attributes.

In the code above, we can have three java files used to read the dataset. Lets explore the objective of each file as follows:

- The `Book.java` file defines the `Book` class, This class represents a `Book` object with attributes for the title, author, user rating, reviews, price, year, and genre. It includes getters for each attribute and a printDetails method to print the details of the book in a formatted manner.
- The `DatasetReader.java` file is responsible for reading a CSV file and creating a list of `Book` objects. It handles the parsing of each line in the CSV, ensuring that each book has the required data fields, and skips malformed lines.
- The `driver.java` file contains the main method, which serves as the entry point of the program. It uses `DatasetReader` to read the dataset from the CSV file, and then iterates over the list of `Book` objects to print their details using the `printDetails` method of the `Book` class.

Tasks

1. Total number of books by an author
    - It takes the name of an author and dataset as input and returns the total number of books written by the author
2. All the authors in the dataset
    - Print name of all authors in the dataset
3. Names of all the books by an author
    - It takes the author as an input and returns all the books written by the author. Just for reference, Author is the second column, and Name (name of the book) is the first column in the dataset.
4. Classify with a user rating
    - It takes the rating as an input and returns all books with the user rating equal to rating.

5. Price of all the books by an author
    - It takes the name of the author as an input and returns the names and prices of all the books written by the author.