Amazon Database

**Dataset:**

Amazon.csv and Amazon.xlsx

The same dataset was given as .csv file and .xlsx. Since I used MySQL workbench, I imported Amazon.csv into the workbench as amazon table. The table has the following fields:

Name - string

Author - string

User Rating - float

Reviews - int

Price - int

Year - int

Genre – string

The table has 550 records and no missing values.

**Methods:**

Before importing the amazon.csv into MySQL workbench as amazon table, I created a database called amazon database. I created sql scripts to perform the queries.

-- 1. Create a database amazon

CREATE DATABASE amazon\_database;

USE amazon\_database;

-- 2. Categorize the 550 books to fiction and non-fiction

SELECT Genre, COUNT(Genre) AS Total

FROM amazon

GROUP BY Genre;

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 1 Categorization by genre

The above picture shows that there are 310 Non-Fiction and 240 Fiction.

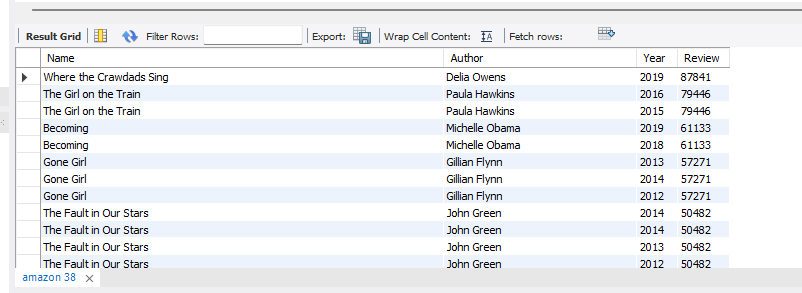
-- 3. Top 50 best sellers

SELECT Name, Author, Year, Reviews AS Review

FROM amazon

ORDER BY Review DESC

LIMIT 50;



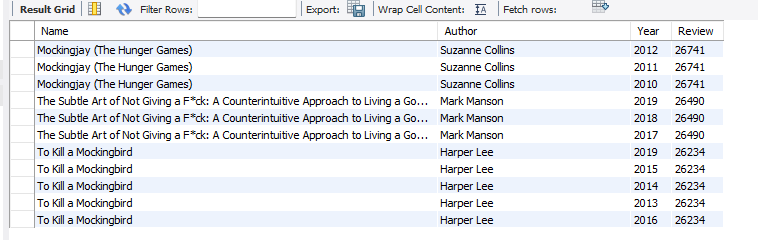


Fig. 2 Top 50 best sellers. Top is the beginning and bottom is the end.

The sample above shows that the best seller is Where the Crawdads Sing with 87841 reviews, followed by The Girl on the Train published in 2016 and 2015 with a total review of 79446. The last five is To Kill a Mockingbird (published 2019, 2015, 2014, 2013, 2016) with 26234 reviews.

-- 4. Books with rating greater than 4 release in 2019

SELECT \*

FROM amazon

WHERE CAST(REPLACE(`User Rating`, ',', '.') AS DECIMAL(3,1)) > 4.0 AND Year = 2019;

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 3 Books with rating greater than 4 and released in 2019

There are 40 books that meet the two conditions, rating is greater than 4 and year 2019. The result shows the records that the rating is more than 4.0 and year is 2019.

-- 5. Books which have greater than 10k reviews last 2018

SELECT \*

FROM amazon

WHERE Reviews > 10000 AND Year = 2018;

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 4 Books with over 10k reviews and released in 2018

There are 41 of such books that have over 10k reviews and released in 2018.

-- 6a. Most expensive books

SELECT \*

FROM amazon

WHERE Price = (

SELECT MAX(Price) FROM amazon

);

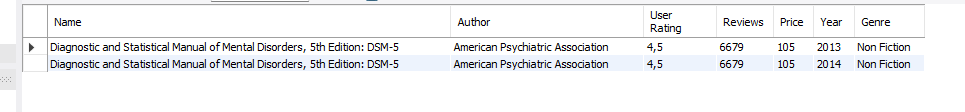


Fig. 5 Most expensive books

The most expensive books are shown above. Price is 105 but published 2013 and 2014 respectively.

-- 6b. The cheapest books

SELECT \*

FROM amazon

WHERE Price = (

SELECT MIN(Price) FROM amazon

);

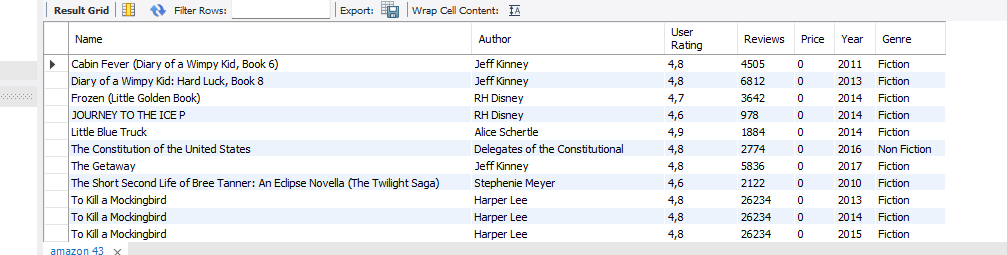


Fig. 6 The cheapest books

There are 43 books as shown above. Their price is 0.

-- 6c. Find the authors with the highest total number of books

SELECT Author, COUNT(\*) AS `Total Number of Books`

FROM amazon

GROUP BY Author

ORDER BY `Total Number of Books` DESC;

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 7 Authors with the highest books

There are 44 of such records, and it shows that Jeff Kinney is at the top with 12, followed by Gary Chapman with 11.

-- 6d. Which year has the highest number of books

SELECT Year, COUNT(\*) AS `Total Number of Books`

FROM amazon

GROUP BY Year

ORDER BY `Total Number of Books` DESC;

A screenshot of a computer

AI-generated content may be incorrect.

Fig. 8 Years and total books published.

The table above shows that there are 50 books each year.