

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;
```

Result Grid			Filter Rows:
	Genre	Total	
▶	Non Fiction	310	
	Fiction	240	

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;
```

Result Grid					Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
	Name	Author	Year	Review				
▶	Where the Crawdads Sing	Delia Owens	2019	87841				
	The Girl on the Train	Paula Hawkins	2016	79446				
	The Girl on the Train	Paula Hawkins	2015	79446				
	Becoming	Michelle Obama	2019	61133				
	Becoming	Michelle Obama	2018	61133				
	Gone Girl	Gillian Flynn	2013	57271				
	Gone Girl	Gillian Flynn	2014	57271				
	Gone Girl	Gillian Flynn	2012	57271				
	The Fault in Our Stars	John Green	2014	50482				
	The Fault in Our Stars	John Green	2014	50482				
	The Fault in Our Stars	John Green	2013	50482				
	The Fault in Our Stars	John Green	2012	50482				

Result Grid					Filter Rows:	Export:	Wrap Cell Content:	Fetch rows:
Name	Author	Year	Review					
Mockingjay (The Hunger Games)	Suzanne Collins	2012	26741					
Mockingjay (The Hunger Games)	Suzanne Collins	2011	26741					
Mockingjay (The Hunger Games)	Suzanne Collins	2010	26741					
The Subtle Art of Not Giving a F*ck: A Counterintuitive Approach to Living a Go...	Mark Manson	2019	26490					
The Subtle Art of Not Giving a F*ck: A Counterintuitive Approach to Living a Go...	Mark Manson	2018	26490					
The Subtle Art of Not Giving a F*ck: A Counterintuitive Approach to Living a Go...	Mark Manson	2017	26490					
To Kill a Mockingbird	Harper Lee	2019	26234					
To Kill a Mockingbird	Harper Lee	2015	26234					
To Kill a Mockingbird	Harper Lee	2014	26234					
To Kill a Mockingbird	Harper Lee	2013	26234					
To Kill a Mockingbird	Harper Lee	2016	26234					

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;

Name	Author	User Rating	Reviews	Price	Year	Genre
5,000 Awesome Facts (About Everything!) (National Geographic Kids)	National Geographic Kids	4,8	7665	12	2019	Non Fiction
Becoming	Michelle Obama	4,8	61133	11	2019	Non Fiction
Brown Bear, Brown Bear, What Do You See?	Bill Martin Jr.	4,9	14344	5	2019	Fiction
Can't Hurt Me: Master Your Mind and Defy the Odds	David Goggins	4,8	16244	18	2019	Non Fiction
Difficult Riddles For Smart Kids: 300 Difficult Riddles And Brain Teasers Families ...	M Prefontaine	4,6	7955	5	2019	Non Fiction
Dog Man: Brawl of the Wild: From the Creator of Captain Underpants (Dog Man...	Dav Pilkey	4,9	7235	4	2019	Fiction
Dog Man: Fetch-22: From the Creator of Captain Underpants (Dog Man #8)	Dav Pilkey	4,9	12619	8	2019	Fiction
Dog Man: For Whom the Ball Rolls: From the Creator of Captain Underpants (Do...	Dav Pilkey	4,9	9089	8	2019	Fiction
Educated: A Memoir	Tara Westover	4,7	28729	15	2019	Non Fiction
► Giraffes Can't Dance	Giles Andreae	4,8	14038	4	2019	Fiction
Girl, Stop Apologizing: A Shame-Free Plan for Embracing and Achieving Your Goals	Rachel Hollis	4,6	7660	12	2019	Non Fiction

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;
```

	Name	Author	USDA Rating	Reviews	Price	Year	Genre
▶	12 Rules for Life: An Antidote to Chaos	Jordan B. Peterson	4,7	18979	15	2018	Non Fiction
	Becoming	Michelle Obama	4,8	61133	11	2018	Non Fiction
⋮	Dear Zoo: A Lift-the-Flap Book	Rod Campbell	4,8	10922	5	2018	Fiction
	Educated: A Memoir	Tara Westover	4,7	28729	15	2018	Non Fiction
	Fahrenheit 451	Ray Bradbury	4,6	10721	8	2018	Fiction
	Fire and Fury: Inside the Trump White House	Michael Wolff	4,2	13677	6	2018	Non Fiction
	First 100 Words	Roger Priddy	4,7	17323	4	2018	Non Fiction
	Giraffes Can't Dance	Giles Andreae	4,8	14038	4	2018	Fiction
	Girl, Wash Your Face: Stop Believing the Lies About Who You Are So You Can B...	Rachel Hollis	4,6	22288	12	2018	Non Fiction
	How to Win Friends & Influence People	Dale Carnegie	4,7	25001	11	2018	Non Fiction
	Last Week Tonight with John Oliver Presents A Day in the Life of Marlon Bundo ...	Jill Twiss	4,9	11881	13	2018	Fiction
amazon 41 x							

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
);
```

	Name	Author	User Rating	Reviews	Price	Year	Genre
▶	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5	American Psychiatric Association	4,5	6679	105	2013	Non Fiction
	Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5	American Psychiatric Association	4,5	6679	105	2014	Non Fiction

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
);
```

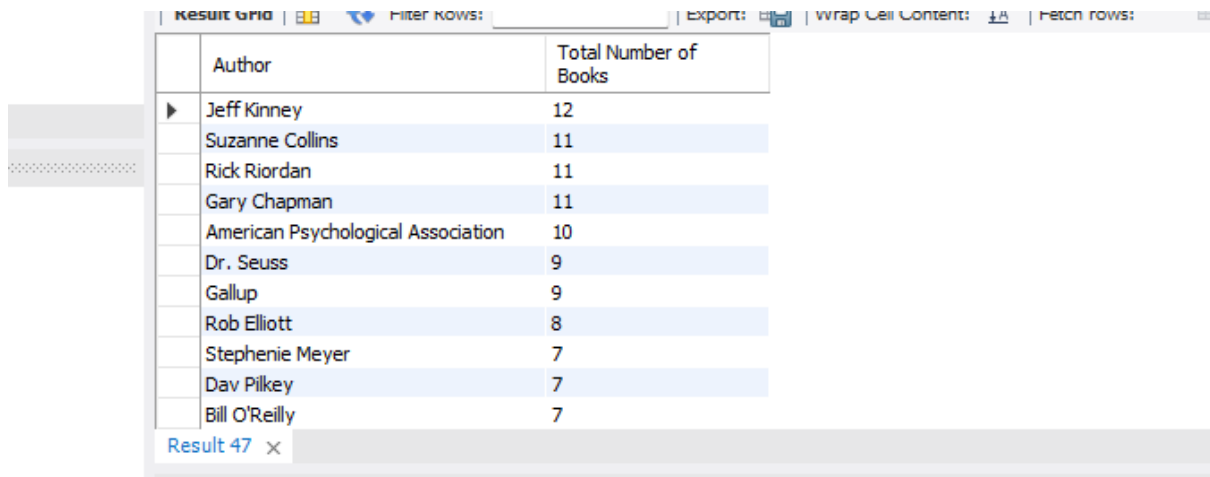
Result Grid							
Filter Rows:							
Export:							
Wrap Cell Content:							
	Name	Author	User Rating	Reviews	Price	Year	Genre
▶	Cabin Fever (Diary of a Wimpy Kid, Book 6)	Jeff Kinney	4,8	4505	0	2011	Fiction
	Diary of a Wimpy Kid: Hard Luck, Book 8	Jeff Kinney	4,8	6812	0	2013	Fiction
	Frozen (Little Golden Book)	RH Disney	4,7	3642	0	2014	Fiction
	JOURNEY TO THE ICE P	RH Disney	4,6	978	0	2014	Fiction
	Little Blue Truck	Alice Schertle	4,9	1884	0	2014	Fiction
	The Constitution of the United States	Delegates of the Constitutional	4,8	2774	0	2016	Non Fiction
	The Getaway	Jeff Kinney	4,8	5836	0	2017	Fiction
	The Short Second Life of Bree Tanner: An Eclipse Novella (The Twilight Saga)	Stephenie Meyer	4,6	2122	0	2010	Fiction
	To Kill a Mockingbird	Harper Lee	4,8	26234	0	2013	Fiction
	To Kill a Mockingbird	Harper Lee	4,8	26234	0	2014	Fiction
	To Kill a Mockingbird	Harper Lee	4,8	26234	0	2015	Fiction

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;
```



The screenshot shows a database query result grid with the following data:

Author	Total Number of Books
Jeff Kinney	12
Suzanne Collins	11
Rick Riordan	11
Gary Chapman	11
American Psychological Association	10
Dr. Seuss	9
Gallup	9
Rob Elliott	8
Stephenie Meyer	7
Dav Pilkey	7
Bill O'Reilly	7

The interface includes a toolbar with options like 'Filter Rows', 'Export', 'Wrap Cell Content', and 'Fetch Rows'. The result is labeled 'Result 47'.

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;
```

Result Grid			Filter Rows:	Export:	Wrap Cell Cont
	Year	Total Number of Books			
▶	2016	50			
	2011	50			
	2018	50			
	2017	50			
	2019	50			
	2014	50			
	2010	50			
	2009	50			
	2015	50			
	2013	50			
	2012	50			

Result 48 x

Fig. 8 Years and total books published.

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