

## Report

### My dataset:

The dataset that I chose is called `goodreads_books.json`. It is a dataset on books and their Goodreads reviews. I chose this dataset because I like books, and my final project is related to books, therefore, I thought it might be interesting to start looking at datasets that may come in handy for the final. This dataset has quite a few fields including title, description, the number of text reviews, the number of ratings, the average rating, the language code, the number of pages, etc. It is a pretty thorough dataset with a lot of information, however, a couple of issues that I came across were: first, everything is in strings, including the number of pages, the number of reviews, ... which means that for my queries, I needed to consider everything as strings, and `$lt` or `$gt` seemed to work most of the time, it doesn't seem to have always worked... Another thing I did not like was how they dealt with author. The dataset has an author field which then has an array, but inside this array, there is an `"author_id"` which is a number, and a `"role"` which is most of the time empty, but it does not have the name of the author there. It seems like if it is anywhere at all, it is in the `"popular_shelves"` field which has an array of with elements like genre (sci-fi, romance, ...), `"to-read"`, and sometimes authors like `"king-stephen"` which is not great if I want to send back the author of the book. The `"popular_shelves"` is also quite messy without recurring order from book to book so you never know what place in the array you want to look at.

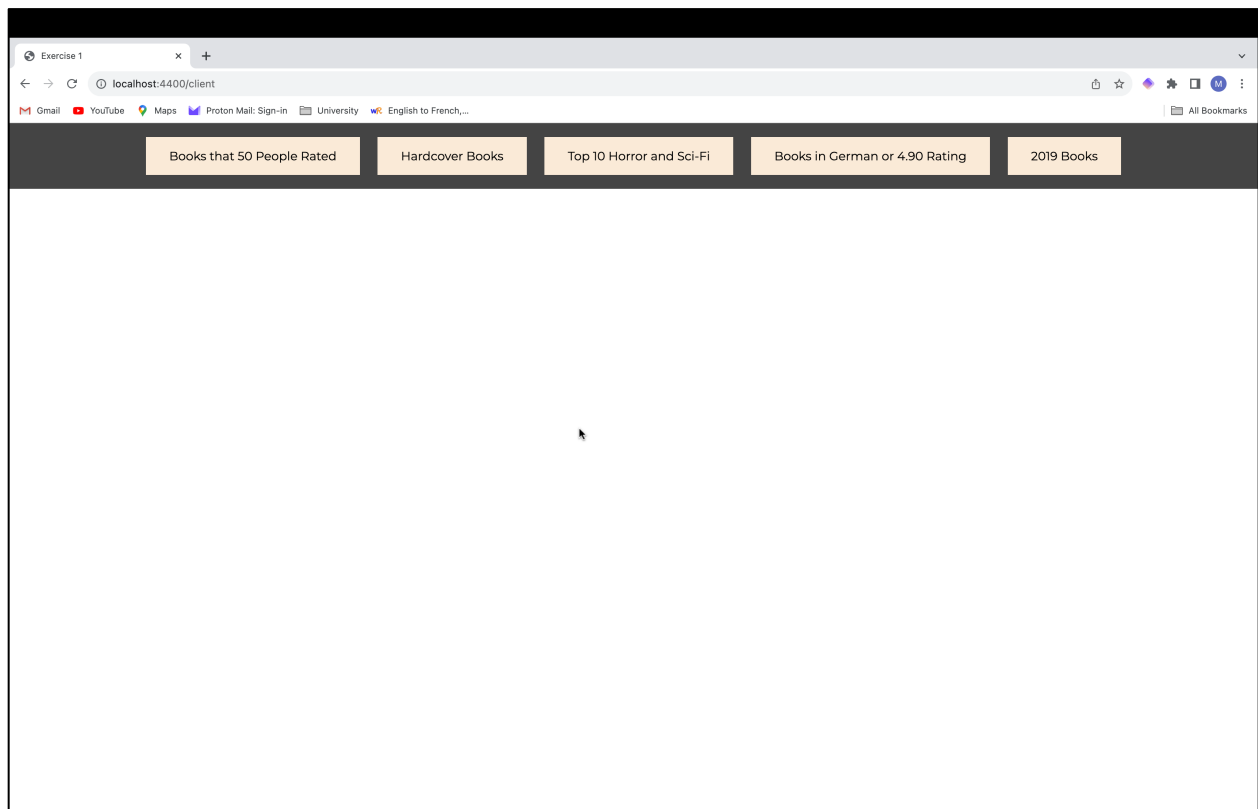
### Queries:

1. My first query looks at all the books that have 50 ratings and less than 3.20 for their average rating. It sends back the books' titles, descriptions, number of ratings (`"ratings_count"`), and average ratings (`"average_rating"`). It also sorts the results based on their `average_rating`, starting with the highest ones all the way to the lowest. This query is where I tried to use `$lt` on strings. It seems to have worked alright despite not really working when I tried to use `$gt`: which is quite confusing. I also learned that the descriptions are quite long.
2. For my second query, I looked at books that were not eBooks and that had a hardcover which are found in the `"is_ebook"` and `"format"` fields in the dataset. For this one, I didn't send back too much, simply the format (`"Hardcover"`) and the title of the book and then, I sorted the books by their titles in alphabetical order. This query and the first one were quite useful to really understand how the `$and` works. I didn't realise that the `$and` is basically an array and we check if every element of this array is applied to each object. I also hadn't realised there was a `"format"` field in the dataset before doing this query.
3. For this third query, that is when I tried to look at the genre inside the `"popular_shelves"` field. I look at 10 books that are both sci-fi and horror and I display their average rating and their title. I also sort them by their ratings. Given that the `"popular_shelves"` has an array inside, I wasn't sure how to look at the `"name"` inside each element of the array. I first thought that maybe I needed a for loop to go through all the elements, but I tried what I have in my code, and it worked. That was quite interesting and useful to learn. I

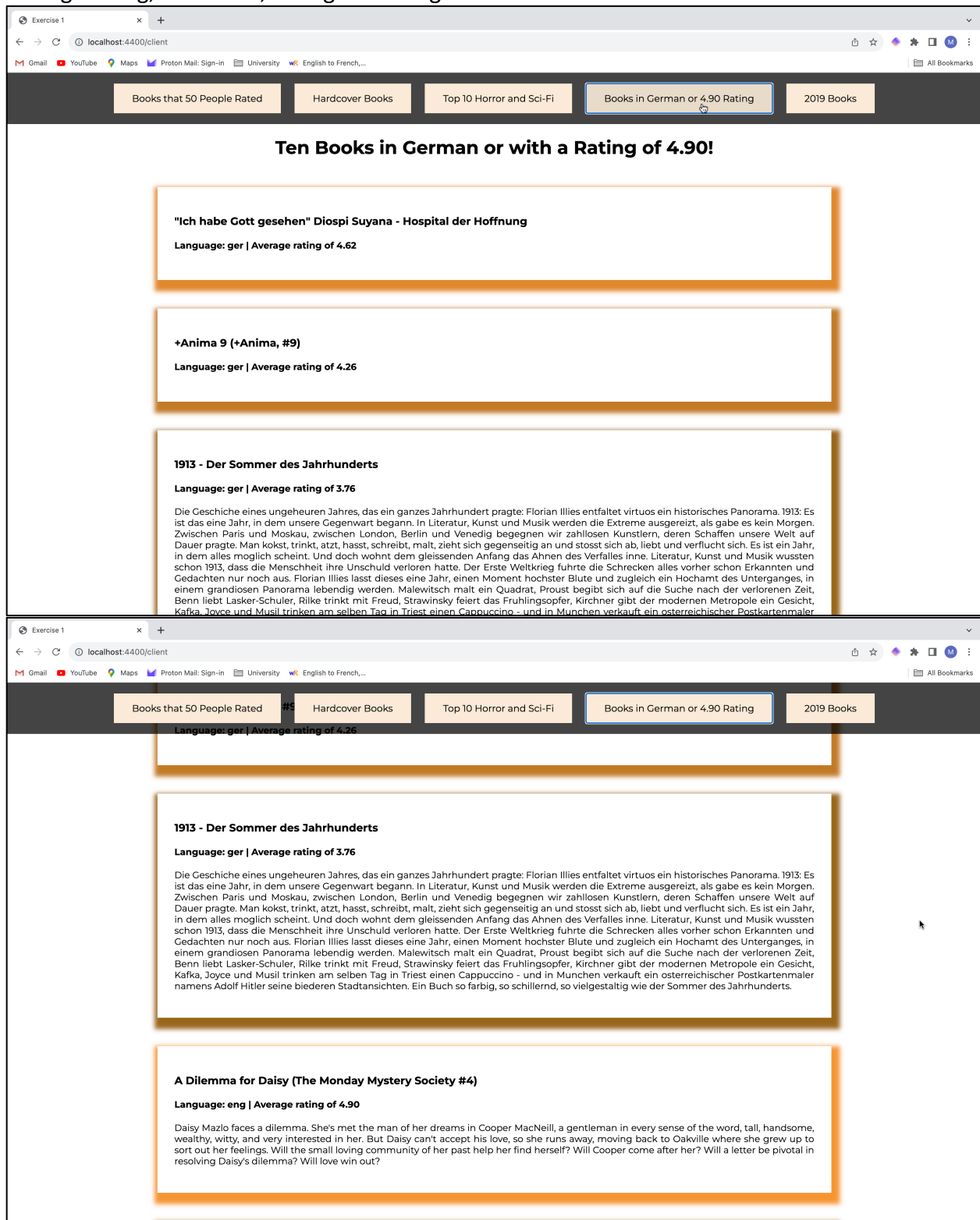
also decided to look at limit to only display 10 results which can come in pretty handy to find the top 10 horror, sci-fi books.

4. For the fourth query, I look at books that are either in German, or that have an average rating of 4.90. I display the title, average rating, language code and description. After making the query, what I used to sort was tricky. I wanted to have at least a couple of German books and one or two 4.90 books to be sure that it works, but if I sort the results by average ratings, I only get the 4.90 results, so I tried to sort by title and I do have at least one 4.90 and a whole bunch of books in German. What I learned from this query is that sometimes we have to tweak a couple things in our queries to get results that more or less align with what we want to see. I used the limit as well to not have too many results.
5. In my fifth query, I look at the publication date of the books and I sort them by alphabetical order. This query was a little annoying to do because, like every other element from my database, the publication dates are strings, which means that when I tried to use \$gt: the query returned me books that were published way before the date I was setting. The original goal was to make a query that would return books that were published in the past 5 years for example. Since it wasn't working the way I wanted I modified it to give me back 5 books from 2017.

When the user arrives on the page, they can click on different buttons with different search criteria.



For the fourth button, that corresponds to the fourth query where the books are arranged in alphabetical order, and the shadow around each book description is based on the books average rating, the better, the lighter orange:



For the third button, ten books are selected if they are horror and sci-fi. They are then arranged based on their average rating with their place on the right.

The screenshot shows a web application interface with a navigation bar and a list of books. The navigation bar has five buttons: "Books that 50 People Rated", "Hardcover Books", "Top 10 Horror and Sci-Fi" (which is highlighted with a blue border), "Books in German or 4.90 Rating", and "2019 Books". Below the navigation bar, the title "Top 10 Horror and Sci-Fi Books!" is displayed. The list of books is as follows:

Rank	Book Title	Average Rating
1	Back from the Dead: The True Sequel to Frankenstein	4.70
2	The Essential Supernatural [Revised and Updated Edition]: On the Road with Sam and Dean Winchester	4.69
3	Tokyo Ghoul: re, Vol. 5	4.59
4	Gemina (The Illuminae Files, #2)	
5		
6		
7	Lazarus Code	4.54
8	Fall of Man (Lycan Fallout #2)	4.52
9	Tokyo Ghoul, vol. 12	4.51
10	Maelstrom (Whyborne & Griffin, #7)	4.50

For the first button, the books are chosen if 50 people rated them. They are then arranged by rating which also defines the colour of the div. Higher rating is green, lower turns red.

The screenshot shows a web application interface with a navigation bar at the top containing five buttons: "Books that 50 People Rated", "Hardcover Books", "Top 10 Horror and Sci-Fi", "Books in German or 4.90 Rating", and "2019 Books". The "Books that 50 People Rated" button is selected and highlighted with a blue border.

The main content area displays a list of books under the heading "Books that 50 People Rated!". The books are sorted by rating, with higher ratings in green and lower ratings in red.

- The Secret of the Scarlet Stone (Gabby Girls Adventure, #1)**  
50 people rated this book | Average rating of 3.49  
The Adventure Begins...My name is Gabrielle Martin. And before I discovered that the scarlet pendant I received on my thirteenth birthday was not just any ordinary pendant but a crazy weird talisman and lifeline to my supernatural powers; the only major thing that I cared about was starting my freshman year at Vineswell Academy. But that was way before finding out that everything that I had known about my life was a lie...essentially smoke-and-mirrors designed to hide the truth about my powerful supernatural legacy.
- Zuma Exposed**  
50 people rated this book | Average rating of 3.46  
This is the book President Jacob Zuma does not want you to read. From Shaik to 'The Spear', award-winning investigative journalist Adrian Basson reveals the truth behind Jacob Zuma's presidency of the ANC and South Africa. From one bad decision to another, this explosive, roller-coaster account traces the unravelling of a likeable but deeply flawed leader who came to power as victim, not visionary. Basson forensically unpacks the charges against Zuma and reveals a president whose first priority is to serve and protect his own, rather than the 50 million people he was elected to lead. To be published on the eve of the ANC elective conference in Mangaung, this is essential reading for any South African who cares about the country they live in.
- The Blackpool Highflyer (Jim Stringer, #2)**  
50 people rated this book | Average rating of 3.44  
'Genuinely gripping ... A brilliant evocation of Edwardian working-class life - the sort of thing DH Lawrence might have written had he been less verbose or been blessed with a sense of humour.' Peter Parker, Evening Standard The second Jim Stringer adventure, The Blackpool Highflyer is a superbly atmospheric thriller of sabotage, suspicion and steam. 'Unique and important ... There is no one else who is writing like Andrew Martin
- appianate. C'è Claire, con i suoi uomini sempre sbagliati, e la certezza che la vita non le darà mai l'unica cosa che ha sempre voluto. Nora, la casalinga perfetta, che ha tenuto nascosto un segreto alle amiche per più di venticinque anni. Maggie, che è sposata infelicitemente con Ray da più tempo di quanto voglia ricordare. E poi Georgie, altezzosa e supponente, che ha fatto a modo suo più di quanto le sarebbe convenuto. Ma stasera la complessa trama di mariti, amanti e segreti che le ha tenute insieme sta per dipanarsi. E una delle quattro donne non ha intenzione di esserci. A questo punto, le cose non potranno più essere le stesse...**
- Jenna & Jonah's Fauxmance**  
50 people rated this book | Average rating of 3.15  
Fans of romance don't need to look any further than the fauxmance brewing between teen idols Charlie Tracker and Fielding Withers-known on their hit TV show as Jenna and Jonah, next-door neighbors flush with the excitement of first love. But it's their off-screen relationship that has helped cement their fame, as passionate fans follow their every PDA. They grace the covers of magazines week after week. Their fan club has chapters all over the country. The only problem is their off-screen romance is one big publicity stunt, and Charlie and Fielding can't stand to be in the same room. Still, it's a great gig, so even when the cameras stop rolling, the show must go on, and on, and on, and on... Until the pesky paparazzi blow their cover, and Charlie and Fielding must disappear to weather the media storm. It's not until they're far off the grid of the Hollywood circuit that they realize that there's more to each of them than shiny hair and a winning smile.
- Too Short A Blessing**  
50 people rated this book | Average rating of 3.14  
"You can't cling to the past forever, Sara!" Why wouldn't Jonas Chesney just leave her alone with her memories? Why was this arrogant new neighbor so determined to intrude on Sara's emotions - and arouse her passion? After her fiancé's sudden death more than a year before, Sara had resolved never to fall in love again. But she'd reckoned without the strong-willed Jonas Chesney... To her own surprise, Sara found herself responding to his kisses - responding with a fervor so intense it overshadowed all her memories.
- The New Mom's Survival Guide: How to Reclaim Your Body, Your Health, Your Sanity, and Your Sex Life After Having a Baby**  
50 people rated this book | Average rating of 2.76

My data visualisation is simple and based on colours as I was working with books and text. The way I used the data to influence the visualisation was mainly with the ratings that were mapped to colours in order to modify some aspects of the divs containing the textual data. I also used the number of results given back to me with the query for the second example I show above to do a countdown of the top ten books. This dataset was interesting to look at, but I found it difficult to be creative with how to visualise that data because of the few numbers that could be used as values to draw elements. Another thing was that not every object of the dataset had all the elements I was looking at. Some don't have a language code, description, rating, ... which means that some objects would show up in my results not because they respect my query per se, but because they do not have information where I am looking and are still returned.