Exercise 1

Jackson Dunnigan
Cart 451

Part A:

For my MongoDB dataset queries, I chose the "Amazon Top 50 Bestselling Books 2009–2019" dataset from Kaggle. This dataset provides a comprehensive record of the bestselling books from 2009 to 2019, including details such as the title, author, genre, user ratings, price, and year of publication. Analyzing this data using MongoDB allowed me to delve into various aspects, including the distribution of genres, patterns in user ratings, and the relationship between authors and the pricing of their books. I initially was planning on using the "NASA - Nearest Earth Objects" dataset, but ran into many hurdles, and incomplete information making it very hard to query.

https://www.kaggle.com/datasets/sootersaalu/amazon-top-50-bestselling-books-2009-2019/data

Part B:

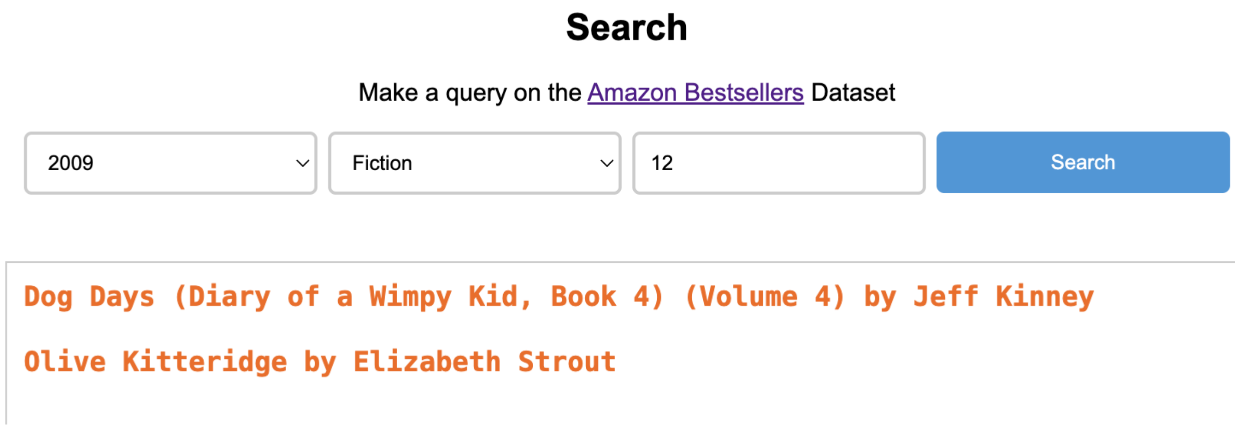
1) For my initial Mongo DB query, my objective was to retrieve information about books released in a particular year using the find() function. In this case, I specifically chose 2010, I was surprised with how many books I didn't recognize.
2) The goal of my second query was to separate books by genre. I wanted it to be in more specific genres such as: fantasy, science fiction, romance etc. but the database's "Genre" tag was somewhat underwhelming only having the options, "Fiction" or "Nonfiction." Due to this I decided to just separate the two using the find() function once again.
3) My third query was an attempt to average out data from all the entries using the aggregate function. I did this by grouping entries with the $group and $avg tags respectfully. It was quite frustrating to get it working but once I did, the rest of this

project became easier because I could now specifically average out or sort any set of data. I specifically chose to average out user ratings, finding it interesting that on average fiction books had better reviews with an average of 4.64 opposed to nonfiction's 4.59.

4) My fourth query sorted books by which one has the most reviews. This was done using the aggregate function as well as the $sort tag. I was trying to implement the $project tag to simplify the data into just author, book name, and review amount. Doing this query I discovered that the most reviewed book was "Where the Crawdads Sing" with 87841 reviews, followed by "The Girl on the Train" with 79446 reviews.

5) My fifth and final query aggregated the average price of a book per author using the aggregate function as well as $avg and $group tags similar to my third query. After this, I implemented the $sort tag to sort the prices from highest to lowest. For the highest average price per author, I got the "American Psychiatric Association," (which isn't technically an author) but has an average of $107. For the lowest average price I got a 10 way tie (I won't list for ease of reading) with $13.

Part C:



Search

Make a query on the Amazon Bestsellers Dataset

2009 | Fiction | 12

Dog Days (Diary of a Wimpy Kid, Book 4) (Volume 4) by Jeff Kinney

Olive Kitteridge by Elizabeth Strout

When creating my server, I made a query that takes 3 parameters: Year, Genre, and Price, using the find() function to retrieve the proper books. The visualization displays the author and book title, using the year and book price as start and end points for a css animation that warps the text's color.