Final Project on Data Analysis Using Python

**Overview**

This assignment is focused on implementing all the learning that has taken place throughout the course. The design of a final project and the implementation as well as testing a Python program from scratch is a necessary skill in order to move forward in becoming proficient in the use of Python. This project uses skills learned throughout the course in order to solve a problem.

**Problem**

If you’re an avid reader like me, you constantly struggle to find a new book to read. The difficulty comes in finding something that suits your tastes and will be a good book to read. There are just so many books out there; it can be overwhelming to think about finding that one book that will satisfy the urge to curl up with a page-turner. The difficulty lies in the sheer number of books as well as the desire to know if the book was judged well by fellow bookworms. That’s where databases come in handy.

**Solution**

This library database is a collection of 10,000 books that have been reviewed 6 million times as well as tagged by numerous users. It details the book names, their authors, all the ratings, tags given to the books by the users to describe them more accurately, and metadata such as publication year. The program is designed to query the library to help users search for whichever books they would like to read next. The program was built using [this](https://github.com/zygmuntz/goodbooks-10k) previously compiled dataset. The *ratings* csv file contains the data for each rating given to the books. The csv file called *to\_read* provides the suggestions from previous readers on whether the books are worth reading or not. *Books.csv* is the compilation of all the books in the database as well as their metadata, including year published, ID, ISBN, author, etc.). Another file, *tags.csv*, contains all the tags given to the books by others, which corresponds to the books by using tag\_id and book\_id.

The program has four functions. When started, a menu will display the options:

1. **Search for all the books in the library that were published between a range of years.** The user will be asked to input a beginning year of the range and an end year. Then the program will search for all the books in that range and display the titles as well as the average rating for each book.
2. **Search for an author’s name.** When the user inputs an author’s name, the program will search through all the authors and return a list of books written by authors with similar names.
3. **Search for a book’s title.** The user can search for a book title using a partial or full title. The program will return a list of all books that match the user input.
4. **Search for all books within a genre.** The user will beprompted to input a genre to search for. Once one has been decided on, the program will search for all tags that match the genre and list all the books with those tags, as well as the average ratings of each book.
5. **Quit program.** The program will quit and close the terminal.