**Installing MongoDB:** You will need to install a MongoDB server in order to complete this part of the homework. If you are using the recommended Ubuntu 20.04 VM, you can use the following command to install it:

* sudo apt-get install mongodb
* sudo systemctl enable mongodb

If you are on another operating system, there is documentation for installing a MySQL server here:

* <https://docs.mongodb.com/guides/server/install/>

**Installing the PyMongo connector:** If you are on the Ubuntu 20.04 VM, you can install the Python connector for MongoDB using pip. If you do not have pip installed already, you can install it using:

* sudo apt-get install python3-pip

Then, you can install the PyMongo connector using:

* pip3 install pymongo

**Importing Dataset:** Download this .json file containing books data. If you are on Linux, you can use the wget command to download it via the command line.

* wget <https://raw.githubusercontent.com/ozlerhakan/mongodb-json-files/master/datasets/books.json>

Then, import the .json file into your MongoDB database. This command worked for me:

* mongoimport books.json

The entirety of this assignment should be done in a single Jupyter notebook. I’m not sure if it’s possible to use Colab for this assignment. **Use the MongoDB Python connector** to write the following queries and show their output. **You must print your query results using the pprint.pprint() function as shown in the Lec11 – PyMongo notebook.**

To turn in the assignment, upload your Jupyter notebook to Blackboard. **Your notebook must have the following naming format:**

1. Print just the titles of the first 5 books in sorted order by title. Use sort() and limit(). (10 pts)

2. Print the title and author list of the first 5 books in sorted order by title that have exactly two authors. Use $size on the authors field. (10 pts)

3. Print just the title of each book for which "W. Frank Ableson" is one of the authors. Use $all. (10 pts)

4. Print just the titles in sorted order of every book with the word Web in the title. Use $regex. (10 pts)  
  
  
5. Print the number of distinct authors. Use distinct() and len(). (10 pts)