CS 327E Project 5, due Thursday, 03/25.

This project makes use of the Open Food Facts dataset. Before beginning the assignment, read the data dictionary for this dataset so that you know what the various fields mean.

The goal of this assignment is to gain some practice writing and executing various CRUD operations against a MongoDB database. As you work through the questions, it is a good idea to consult the official documentation when you have questions about which operators or methods to use when formulating your answers.

Open a terminal window in JupyterLab and download the dataset from Google Cloud Storage. Run the following commands to download and extract the dataset:

```
gsutil cp gs://cs327e-open-access/open_foods.zip .
unzip open_foods.zip
```

The extracted data contains the mongodb dump file `products.bson` along with a metadata file `products.metadata.json`.

In the same terminal, restore the dump file by running the command:

```
mongorestore -d open_food -c products dump/open_food/products.bson
```

The restore command may take up to 5 minutes to complete. It creates a database `open_food` with a `products` collection and populates the collection with 309370 documents from the Open Food Facts dataset. There are several indexes on this collection, which take some time to create.

Create a new Python Jupyter notebook and name it `project5.ipynb`.

Translate the following SQL queries to Mongo's query language. Place each query into its own notebook cell and run each cell individually.

Q1.
```
select count(*)
from products;
```

Q2.
```
select _id, product_name
from products
where categories = 'Snacks, Sweet snacks, Confectioneries,
                   Candies, Chews';
```

Q3.
```
select _id, code, product_name
from products
```

```
    where last_modified_t >= 1601856000;

Q4. select count(*)
    from products
    where packaging = 'Plastic';

Q5. select _id, code, creator, product_name, brands
    from products
    where manufacturing_places = 'Austin, TX'
    and stores = 'Whole Foods';

Q6. select _id, creator, product_name, brands
    from products
    where brands = 'Trader Joes' and product_name is not null
    order by product_name;

Q7. select _id, product_name, brands
    from products
    where brands in ("m&m's", "mars", "Mars", "oreo", "starburst")
    order by product_name
    limit 5;
```

- Insert a new document into the `products` collection. The document must have a minimum of 5 attributes with non-NULL values. Read back the document you just created.

- Update the document you created in the previous step and then read it back.

- Delete the document you updated in the previous step and then query the collection to verify that it’s been deleted.

CS 327E Project 5 Rubric

Due Date: 03/25/21

Download and extract the open food facts dataset to your jupyter notebook instance. **-3** no dataset or incorrect dataset found in Jupyter instance	3
Create a new Python Jupyter notebook named `project5.ipynb`. **-3** incorrect file name	3
Implement queries Q1 - Q7. **-7** for each missing, incomplete or incorrect query **-3** for each missing or incorrect output	70
Run an `insert` followed by a `find` to read back the newly inserted document. **-3** missing, incomplete or incorrect insert **-3** missing, incomplete or incorrect find **-2** for missing or incorrect output	8
Run an `update` followed by a `find` to read back the newly updated document. **-3** missing, incomplete or incorrect update **-3** missing, incomplete or incorrect find **-2** for missing or incorrect output	8
Run a `remove` to delete the document you inserted in the previous step. **-3** missing, incomplete or incorrect remove **-3** missing, incomplete or incorrect find **-2** for missing or incorrect output	8
`project5.ipynb` pushed to your group's private repo on GitHub. Your project **will not** be graded without this submission.	**Required**

`submission.json` submitted into Canvas. Your project **will not** be graded without this submission. The file should have the following schema:
`{`
`    "commit-id": "your most recent commit ID from GitHub",`
`    "project-id": "your project ID from GCP"`
`}`
Example:
`{`
`    "commit-id": "dab96492ac7d906368ac9c7a17cb0dbd670923d9",`
`    "project-id": "some-project-id"`
`}`
	Required

Total Credit:	**100**