**BIG DATA ANALYTICS**

**LAB 6**

**AIM:** Connecting to NoSQL database/s and querying to provide analysis using API like aggregation, etc. To be able to successfully import/export from/to CSV.

**1.** **Run Jupyter Notebook and using MongoDB kernel, do the same exercise shown in the screenshot by mongo client below.**

**Solution: -**

I have done this exercise in mongodb shell since I am not able to run imongo kernel in jupyter notebook.



**2. Write the insert method to store the following document in MongoDB.**

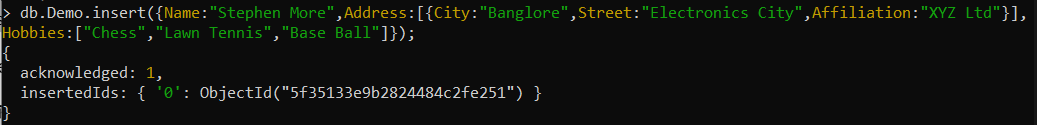
Name: "Stephen More"

Address: { "City" : "Banglore", "Street" : "Electronics City", "Affiliation" : "XYZ Ltd" }

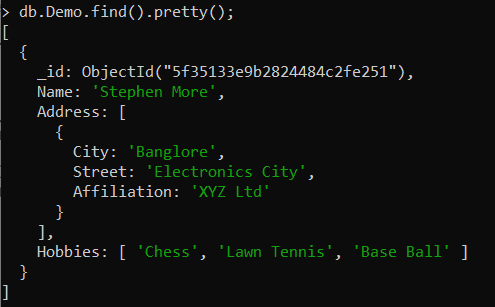
Hobbies: Chess, Lawn Tennis, Base Ball

**Solution: -**

1. Write the following script in mongo shell.

****

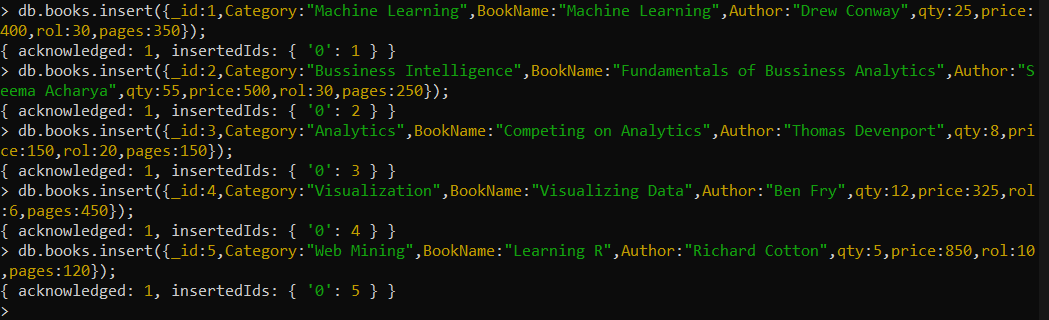
1. Check the Collection if the data is inserted or not.



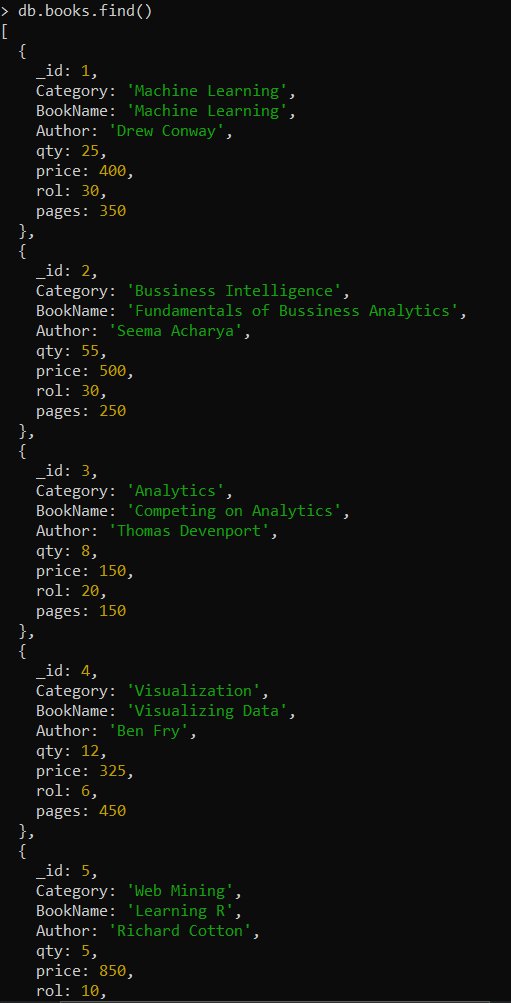
1. **To practice Map Reduce programming in MongoDB.**

**Solution:**

Step 1: Insert 5 documents as shown below in a collection named 'books'.



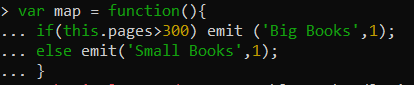
Step 2: Confirm the presence of above documents in the “books” collection run the following command in mongo shell.



Step 3: Write map and reduce functions to split the books into the following two categories:

1. Big Books (b) Small Books

Books which have more than 300 pages should be in the big book category. Books which have less than 300 pages should be in the small book category.

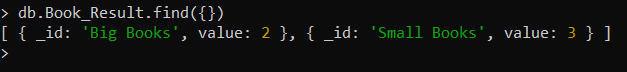


Step 4: Count the number of books in each category.



Step 5: Store the output as follows as documents in a new collection, called “Book\_Result”. Book Category Count of the books.



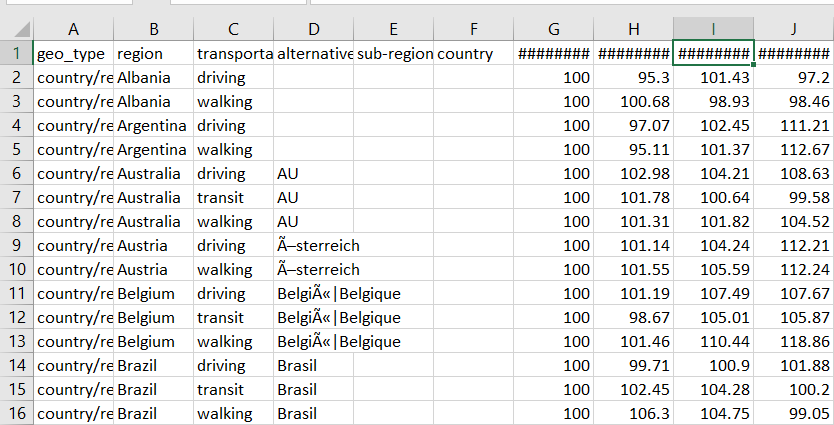


1. **To practice import, export and aggregation in MongoDB.**

**Solution:**

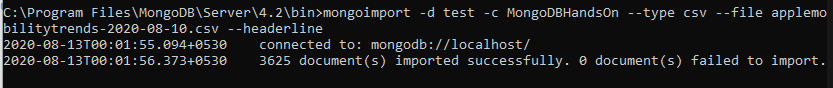
Step 1: Pick and public dataset from the site www.kdnuggets.com Convert it into CSV format. Make sure that you have at least two numeric columns.

* Downloaded and used ‘applemobilitytrends-2020-08-10’ dataset.



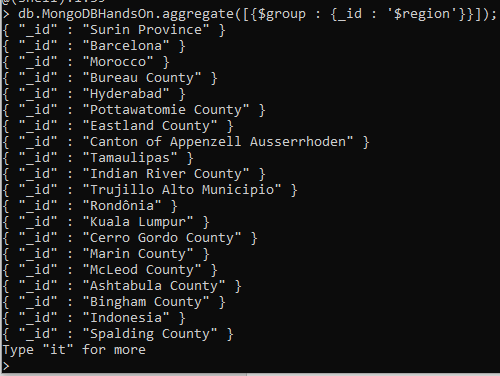
Step 2: Use MongoImport to import data from the CSV format file into the MongoDB collection named “**MongoDBHandsOn**” in the test database. Use the following command.

Mongoimport -d test -c MongoDBHandsOn --type csv --file applemobilitytrend-2020-08-10.csv –headerline



Step 3: Identifying a grouping column.

* We can group that column which has some fixed values. Like the ‘region’ column.



Step 4: Compute the sum of the values in the first numeric column.



Step 5: Compute the average of the values in the second numeric column.



1. **Exercise Python to MongoDB Connectivity using Jupyter Notebook.**

**Solution:**

