DSCI 551 – Spring 2022

Lab3, 20 points

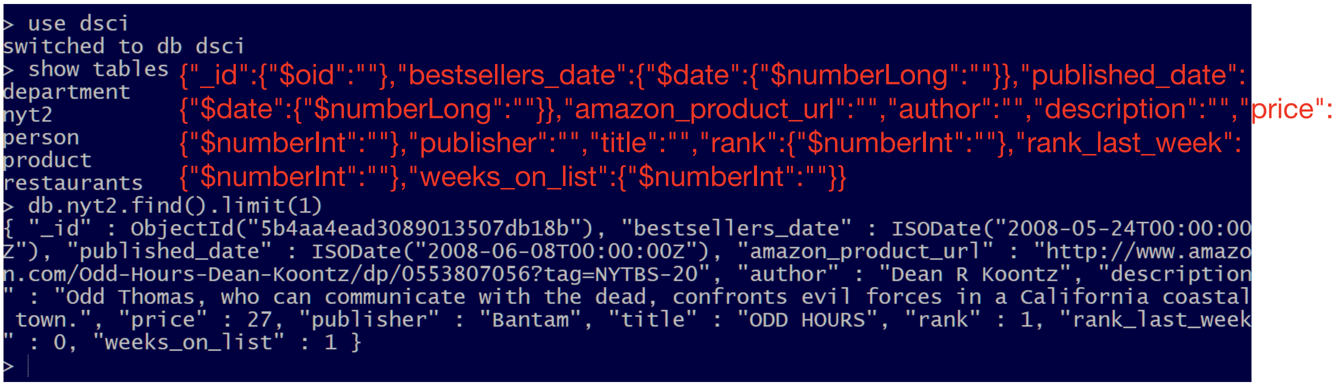
Note this lab is worth twice as much as lab1 and lab2.

Due: April, 8, Friday (end of day, 11:59pm)

Consider the New York best seller data set: nyt2.json. Import this data set into MongoDB using mongoimport as follows:

```
mongoimport --file nyt2.json --collection nyt2 --db dsci
```

Once imported, execute “mongo” and then “use dsci”, “show tables”, you should see the collection nyt2 listed.



```
> use dsci
switched to db dsci
> show tables
department
nyt2
person
product
restaurants
> db.nyt2.find().limit(1)
{ "_id" : ObjectId("5b4aa4ead3089013507db18b"), "bestsellers_date" : ISODate("2008-05-24T00:00:00Z"), "published_date" : ISODate("2008-06-08T00:00:00Z"), "amazon_product_url" : "http://www.amazon.com/Odd-Hours-Dean-Koontz/dp/0553807056?tag=NYTBS-20", "author" : "Dean R Koontz", "description" : "Odd Thomas, who can communicate with the dead, confronts evil forces in a California coastal town.", "price" : 27, "publisher" : "Bantam", "title" : "ODD HOURS", "rank" : 1, "rank_last_week" : 0, "weeks_on_list" : 1 }
>
```

```
{"_id":{"$oid":""},"bestsellers_date":{"$date":{"$numberLong":""}},"published_date":
{"$date":{"$numberLong":""}},"amazon_product_url":"","author":"","description":"","price":
{"$numberInt":""},"publisher":"","title":"","rank":{"$numberInt":""},"rank_last_week":
{"$numberInt":""},"weeks_on_list":{"$numberInt":""}}
```

Write a MongoDB script (using find, aggregate, update, etc.) for each of the following questions on the data set. Show the result of evaluating the script.

1. Find out how many books have “odd” in their titles (case insensitive) and a rank of at least 10. db.nyt2.count({"title": /odd/i, "rank": {$gte:10}}
2. Add a new attribute/field “read” for all books by ““John Grisham” and “Zadie Smith” and set their values to true. Show the response from MongoDB after executing your script.
3. Find out how many books which do not have the “read” field.
4. Find out the titles of books whose price is between 10 and 20 (inclusive). Output only the titles. Output the same title only once.
5. For each publisher, find out the maximum price of best sellers published by the publisher. Order the publishers by the descending order of the maximum price. Output the first 10 in the list

Submission: submit a word/pdf document listing the scripts and results.

2. db.nyt2.update({"author": {$in:["John ~", "Zadie ~"]}},
{$set: {"read": "true"}}, {multi: true})

3. db.nyt2.count() — db.nyt2.count({"read": {$exists: true}})
db.nyt2.count({"read": {$exists: false}})

4. db.nyt2.distinct("title", {"price": {$gte: 10, $lte: 20}})

5. db.nyt2.aggregate({$group: {"_id": {"$publisher"},

```
max_price={$max:"$price"}})
```