MongoDB Query Questions & Answers

 # Homework Assignment 03
DS 4300 - Spring 2025

- **EC Due Date: Feb 16, 2025 @ 11:59pm**
- **Regular Due Date: Feb 18, 2025 @ 11:59pm**
- Upload to GradeScope (no question/solutions to Match)

```
# Set up your connection to Mongo DB here.
import pymongo
from bson.json_util import dumps


uri = "mongodb://michael:Michael0323@localhost:27017"
client = pymongo.MongoClient(uri)
mflixdb = client.mflix
```

Question 1:

Give the street, city, and zipcode of all theaters in Massachusetts.

```
data = mflixdb.theaters.find({"location.address.state": {"$in": ["MA"]}},
                             {"_id":0, "location.address.street1": 1, "location.address.city": 1,
"location.address.zipcode":1})
print(dumps(data, indent=2))
```

Question 2:

How many theaters are there in each state? Order the output in alphabetical order by 2-character state code.

```
data = mflixdb.theaters.aggregate([{"$group": {"_id": {"State": "$location.address.state"},
"Count": {"$sum": 1}}},
                           {"$sort" : {"_id": 1}}])

print(dumps(data, indent=2))
```

Question 3:

How many movies are in the Comedy genre?

```
data = mflixdb.movies.count_documents({"genres": "Comedy"})

print(dumps(data, indent=2))
```

Question 4:

What movie has the longest run time? Give the movie's title and genre(s).

```
data = mflixdb.movies.find_one({}, {"title": 1, "genres": 1, "runtime": 1, "_id": 0},
sort=[("runtime", -1)])
```

MongoDB Query Questions & Answers

```
print(dumps(data, indent=2))
```

Question 5:

Which movies released after 2010 have a Rotten Tomatoes viewer rating of 3 or higher? Give the title of the movies along with their Rotten Tomatoes viewer rating score. The viewer rating score should become a top-level attribute of the returned documents. Return the matching movies in descending order by viewer rating.

```
data = mflixdb.movies.find({"year": {"$gt": 2010}, "tomatoes.viewer.rating": {"$gte": 3}},
    {"title": 1, "tomatoes.viewer.rating": 1,"_id": 0}).sort("tomatoes.viewer.rating",-1)

print(dumps(data, indent=1))
```

Question 6:

How many movies released each year have a plot that contains some type of police activity (i.e., plot contains the word "police")? The returned data should be in ascending order by year.

```
match = {"$match": {"plot": {"$regex": "police", "$options": "i"}}}
group = {"$group": {"_id": "$year", "movie_count": {"$sum":1}}}
project = {"$project": {"year": "$_id", "movie_count": 1, "_id": 0}}
sort = {"$sort": {"year": 1}}

data = mflixdb.movies.aggregate([match, group, project, sort])

print(dumps(data, indent=1))
```

Question 7:

What is the average number of imdb votes per year for movies released between 1970 and 2000 (inclusive)? Make sure the results are order by year.

```
match = {"$match": {"year": {"$gte": 1970, "$lte":2000}}}
avg = {"$group": {"_id": {"release year": "$year"}, "Avg Votes": {"$avg": "$imdb.votes"}}}
sort = {"$sort" : {"_id": 1}}

data = mflixdb.movies.aggregate([match, avg, sort])

print(dumps(data, indent=1))
```

Question 8:

What distinct movie languages are represented in the database? You only need to provide the list of languages.

```
data = mflixdb.movies.distinct("languages")

print(data)
```