

CMSC 462 – Introduction to Data Science

Assignment 5

Total Points – 25

Due: **May 10, 2023**

1. You will use MongoDB for this Assignment. First, download MongoDB in your computer. Then do the following.
2. Write code to create a collection 'myMovies', and add 5 movies to the database that you last watched, having property: name, genre, rating.
3. Write code to return
 - a. all movies in the database,
 - b. find one movie by name
 - c. find top 3 high rated movies
4. Write code to add review to 2 of the movies as 'review' property, and sets / changes rating attributes of one of the movies from other 3.
5. Please download movies, tags and ratings files. Write a program to read the given 3 different csv files (movies, ratings, tags), and insert all the records into 3 different collections (movies, ratings, tags).
6. For the following questions, **you must use Aggregation Pipeline. If you use any other method no credit will be given.**
 - a. Develop code to find number of movies released per year.
 - b. Develop code to find number of movies per genre.
 - c. Develop code to find number of movies per rating.
 - d. Develop code to find number of movies tagged.

For doing this assignment, it may be easier to setup a virtual environment - (<https://pypi.org/project/virtualenv/>)

Use PyMongo - <https://pypi.org/project/pymongo/>

Links:

- <https://docs.mongodb.com/manual/administration/install-community/>
- <https://docs.mongodb.com/manual/installation/>
- <https://docs.mongodb.com/drivers/pymongo/>
- <https://www.mongodb.com/developer/quickstart/python-quickstart-aggregation/>
- <https://www.analyticsvidhya.com/blog/2020/08/how-to-create-aggregation-pipelines-in-a-mongodb-database-using-pymongo/>
- <https://www.mongodb.com/docs/manual/core/aggregation-pipeline/>
- <https://www.mongodb.com/basics/aggregation-pipeline>
- <https://www.mongodb.com/docs/v6.0/core/aggregation-pipeline/>
- <https://www.mongodb.com/docs/manual/reference/operator/aggregation/count/>