CMSC 462 – Introduction to Data Science
Assignment 5

Total Points – 25   **Due:** **December 8, 2024**

1. You will use MongoDB for this Assignment. First, download MongoDB in your computer. Then do the following.
2. For this assignment, it is best that you use Python.
3. 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).
4. Next, write a program to add five movies that you have watched this year, or you would like to watch to the collection "movies". Make sure that you assign **unique movie IDs** and specify the genres (genres need not be completely accurate).
5. Corresponding to the movies that you added, write a program to add some suitable ratings to the collection "ratings" and some suitable "tags" to the collection "tags". Make sure that you use a unique userid for yourself.
6. For the following items, **you must use <u>Aggregation Pipeline</u>. 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.
    e. Develop code to find the most popular tag.
7. What to submit?
    a. Jupyter Notebook file that contains all the above code.
    b. Summarize all the data you added.
    c. A document that summarizes what you learnt while doing the Assignment.

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/