Talha Hussain

Data Science

***Summary for All Added Data in Homework Assignment 5 MongoDB using Movies Dataset and Ratings Dataset and Tags Dataset and What I Learned While Doing This Assignment:***

**Summary:**

In terms of summarizing the data that I added, I took information from the Internet Movie Database IMDb Website <https://www.imdb.com/?ref_=nv_home> for five different movies I have watched that are not included in the database. First, “Yugioh! 3D Bonds Beyond Time” is the title of the first movie that I added that came out in the year 2010. In addition, this movie that I inserted falls into the genres “Animation” and “Fantasy” with an Internet Movie Database IMDb Rating Score of 6.3 which I converted to the 1-5 Rating Scale that is the rating scale used in the Commas-Separated Value CSV File “ratings.csv” and got an Internet Movie Database IMDb Rating Score of 3.2. Second, “Pokémon Detective Pikachu” is the title of the second movie that I added that was released in the year 2019. Moreover, this movie that I added falls into the genres “Action”, “Adventure”, “Comedy”, “Family”, “Mystery”, and “Science Fiction” with an Internet Movie Database IMDb Rating Score of 6.5 which I converted to the 1-5 Rating Scale that is the rating scale used in the Commas-Separated Value CSV File “ratings.csv” and got an Internet Movie Database IMDb Rating Score of 3.3. Third, the title of the third movie that I added is “Sonic the Hedgehog” that came out in the year 2020. In addition, this movie that I inserted falls into the genres “Adventure”, “Comedy”, “Fantasy”, and “Science Fiction” with an Internet Movie Database IMDb score of 6.5 which I converted to the 1-5 Rating Scale that is the rating scale used in the Commas-Separated Value CSV File “ratings.csv” and got an Internet Movie Database IMDb Rating Score of 6.5 which I converted to the 1-5 Rating Scale that is the rating scale used in the Commas-Separated Value CSV File “ratings.csv” and got an Internet Movie Database IMDb Rating Score of 3.3. Then, the title of the fourth movie that I added is “Pokémon the Movie: Secrets of the Jungle” that was released in the year 2020. Moreover, this movie that I added falls into the genres: “Adventure”, “Animation”, “Anime”, and “Children” with an Internet Movie Database IMDb Rating Score of 6.3 which I converted to the 1-5 Rating Scale that is the rating scale used in the Commas-Separated Value CSV File “ratings.csv” and got an Internet Movie Database IMDb Rating Score of 3.2. Finally, the title of the fifth and last movie that I added is “Tom & Jerry” that came out in the year 2021 and this movie that I inserted falls into genres “Adventure”, “Animation”, and “Comedy” with an Internet Movie Database IMDb Rating Score of 5.2 which I converted to the 1-5 Rating Scale that is the rating scale used in the Commas-Separated Value CSV File “ratings.csv” and got an Internet Movie Database IMDb Rating Score of 2.6. Furthermore, last but not least, I then ensured that for all the five movies that I added, each movie entry has four tags that are different, unique, and appropriate.

**What I Learned While Doing This Assignment:**

While working on “Assignment 5” that involved using MongoDB, I learned a lot about how a document database worked for querying and processing data in order to retrieve particular documents from a collection as a result of using a query filter. By applying query filters to the “movies\_collection”, “ratings\_collection”, and the “tags\_collection”, I was able to quickly find the documents that contains the criteria that I am searching for in the movies dataset, the ratings dataset, and the tags dataset. In addition, while doing “Assignment 5” that required me to use MongoDB, I learned how using an Aggregation Pipeline in my Python Jupyter Notebook Code efficiently helps me save time in terms of using aggregates such as the sort aggregate to decrease the number of documents that I needed to process which resulted in my program to run faster and therefore made me realize that performance wise, an Aggregation Pipeline is the most beneficial method when doing aggregation in MongoDB. Furthermore, I discovered that the Aggregation Pipeline system in MongoDB is a powerful tool when working with data to produce results that are valuable because using an Aggregation Pipeline made it uncomplicated to group the data, sort the data, do calculations on the data, and examine the data in detail.