**Final Project : TEAM G**

# Overview Of the Project :

Movie Suggestion based on watched history.

# Data used :

Data files from module 8

1. Moviemetadata.csv
2. Ratingssmall.csv
3. Watched\_list.csv

# Database:

# Tables Used :

## Kaggle\_metadata:

|  |  |
| --- | --- |
| id | Movie id |
| genres | Genres of the movie |
| imdb\_id | IMDB id of the movie |
| Title | Movie Title |

## ratings:

|  |  |
| --- | --- |
| movieId, | Movie id |
| rating (Average) | Movie Rating |

## watched\_list:

|  |  |
| --- | --- |
| imdb\_id | IMDB id of the movie |
| Genres | Genres of the movie |
| title | Movie Title |
| rating | Movie Rating |

kaggle\_metadata.id joins one to many with ratings.movieId originally

and by using average rating we make it one to one relationship within ratings and the metadata.

We will not host common database anywhere.

Will be using the csv files as main source of our data.

ERD diagram - g\_mvs\_ERD.png

SQL files - g-mvs-CreateTables.sql

List down the columns which are going to be used in the machine learning.

# Data preparation:

Will read data from csv file.

Clean data using Python techniques and have data ready in data frame.

# Machine Learning :

We will be using K- Means clustering on the training and testing dataset.

Training dataset will be watched\_list.csv

# Tableau Dashboard:

Need to do brainstorming

Tableau Dashboard story for movie suggestions

Movie Suggestion based on watched History

3 inputs and get 3 Suggestion on movie output.

## Inputs on UI

1. Title
2. Star ratings
3. Each title has genre defined in metadata so we can show the same.

## Output on UI

Three movie suggestions