

## Continuous Internal Assessment 1

### CS 2009 - Exploratory Data Analysis

#### 1. Use the movies dataset and answer the questions.

- a. What is the average imdb score for movies?
- b. How many null values are there in the 'Age\_certification' column.
- c. How many numeric valued columns are present?
- d. Write the code to display the movies released after the year 2000. How many movies were released after the year 2000?
- e. After dropping all the null values in the dataset, how many rows remain?

#### Use the data after dropping the null values to answer the below questions

- f. Out of these rows, how many movies released after 2020?
- g. What is the average tmdb\_popularity for movies containing 'Comedy' as a genre? –
- h. In how many movie descriptions, the word 'novel' appears?
- i. How many unique production countries are present in the data?
- j. Average IMDB score for action movie – 7.22

#### 2. Use the student\_1 and student\_2 excel files to answer the below question:

Write the code for the following:

1. Get columns Registration number and Program in student\_1 which should come from student\_2
2. Save this as a separate dataframe.
3. Create a column called attendance. The column should contain the number of hours each student is present.