# **Assignment Name - Analytics basics**

**Problem Statement -**

**Answer the following questions to the best of your knowledge including the concepts taught to you in the level. You can find the IMDB\_data in the folder containing this file.**

1. Write R code using data “IMDB\_data” to
   1. Load CSV in R by skipping second row.
   2. Extract the unique genres and its count and store in data frame with index key.
   3. Convert the required data types
   4. Sort the genre by its name
   5. Create new variable whose values should be square of difference between imdbrating and imdbvotes.
2. Write Python code using data “IMDB\_data” to
   1. Load CSV in R by skipping second row.
   2. Extract the unique genres and its count and store in data frame with index key.
   3. Convert the required data types
   4. Sort the genre by its name
   5. Create new variable whose values should be square of difference between imdbrating and imdbvotes.
3. Define problem category for below problem statement  
    “A chemist wants to find some interesting patterns in which patients are behaving upon administering the drug

ANS:

Problem statement: No Response variable

In this situation we have to find a pattern how patients are behaving upon giving drugs. But there is no response variable (Response variable: No known output). We have to find a structure from internal category of data. Taking consideration of different types of behavior we have to create groups which is unsupervised learning. By using unsupervised learning we have to find inherent structure and prepare group of clusters. Each cluster will hold people of same behavior or same characteristics. So we can encounter in which pattern how patients are behaving.

1. How will you select suitable machine learning algorithm for a problem statement
2. Define one problem statement for Education industry?

# **Evaluation Basis**

This project will be evaluated on following basis -

1. **Comprehensibility of the answer :** You have to answer the given question with full explanation, background information and examples(if required).
2. **Quality of code** : Write the best possible implementation from your side ensuring the coding practices and conventions taught to you during training.

## **Deliverables from Candidate**

1. A Doc or Pdf file containing your answers to theory questions
2. Code files - R and Python separate.
3. Notes to explain your code logic (if required)

**Warning** - Do not submit incomplete or wrong assignments. They will result in negative skill score. Also, you will not a get a second attempt at the assignment! So whatever score you get will be the final score