

### Assignment Questions:

1. Define problem category

Problem Statement	Problem Category
Which Campaign Themes are Likely to Generate the Most Enthusiastic Leads?	
A banker wants to find customers who are most likely to buy the new loan product he is launching	
Client want to build parking lot at multiple places with specific budget and reduce risk	
A chemist wants to find some interesting patterns in which patients are behaving upon administering the drug	

2. How will you select suitable machine learning algorithm for a problem statement?
3. How will you derive meaningful features from the data?
4. Write R code using data "IMDB\_data"
  - a. Load CSV in R by skipping second row.
  - b. Extract the unique genres and its count and store in data frame with index key.
  - c. Convert the required data types
  - d. Sort the genre by its name
  - e. Create new variable whose values should be square of difference between imdbrating and imdbvotes.
5. Explain secondary data?
6. Write R code to impute missing values using mean?
7. Why do we go for data scaling?
8. How do you deal with outliers?
9. Write R code to delete observations with missing values?
10. Explain difference between inner join and left outer join?