1. Import the packages
2. Import the file
3. We concatenated the train and test files – skip this test if we don’t have sep train and test files
4. Feature Engineering:
   1. Identify the useless columns (which have 0 std dev or which have all names) – delete them
   2. Identify the character columns – check if it is having any useful information and you create a categorical variable out of it (like in titanic dataset, in the name col, we could extract the title; in the cabin column (which was a character column) we could extract the deck level)
   3. To create new column
5. Impute the missing values
   1. Delete the rows having missing values
   2. Generally continuous data has to be imputed by mean / median, categorical data has to be imputed by most frequent value.
   3. Fillna fn
   4. Imputer class
   5. Group by factor column and then transform
6. Handling outliers (if there are only 2-3 outliers, then may be we can delete those rows)
7. If there are multiple outliers – then may be the distribution is skewed, we have to handle it:
   1. Either take log to convert it into normal
   2. Discretize the distribution
8. Convert the categorical columns into numerical
   1. Ordered : 0,1,2,3
      1. Cat.codes
      2. Map
      3. Replace
   2. Nominal - Create dummies