1. Combine the training data and unseen data (where we have to predict the response variable) for cleaning purpose. We’ll create a new column in unseen data with the same name as the response variable in the training data.
2. Data Understanding and Wrangling

* Summary fns, structure of dataframe, check datat types
* Groupby categorical variables and check relationships

1. Feature Engineering
   1. Create useful features

* Create new features which might be useful for prediction of resonse variables.
* Extract important information from character columns and create new features
  1. Features Selection: Delete redundant / useless features
* Delete character columns
* Delete redundant columns
* Delete correlated columns
* Delete the columns which make no sense for prediction of response variable

1. Data Visualizations to visualize the relationship of each of the features with response variable
2. Data Cleaning and Preprocessing
   1. Missing Values Imputation
   2. Handling outliers
   3. Handling noisy data
3. Split the training data into training and unseen data back. We’ll use training data to train our model and deploy the model on the unseen data (to make the predictions)
4. Split the training data into train and test in the ratio 70%:30% or 80%:20%
5. Train the model using train data and test the model performance on the test data. Once we are satisfied with the model performance, we will finalize the model.
6. Deploy the model on unseen data, and make predictions