

# Assignment 2

Team Name on Kaggle

**trashBins**

Team Members

Ganesh Borle (201505587)

Sonam Gupta (201505506)

## **Approach :**

We can simply divide our approach in three sections.

1. Data Preprocessing
2. Model Usage
3. Final Results

### **1. Data Preprocessing**

After observing the data, one can find that it is the combination of numerical as well as categorical data. And also not every value is present for the given data frame / column. Hence we need to handle such cases.

For the categorical data, we will just replace it with the numerical data. For the missing data, we can use various techniques to fill in that data such as we can fill it with the mean of the data present in that column or median of the data, or simply can assign the value like unknown.

### **2. Model Usage**

We can try various model for the given problem, it would be simply choosing the one model over the other or using the combination of the two or more models to improve the accuracy. So, after trying out the different model, lasso and xgboost with some tuning gives the better accuracy. Hence, we have chosen those models.

### **3. Final Results**

As stated earlier, final results are based on the weighted average of the two models ( lasso and xboost ). The final accuracy we got is **0.11456**.