JOB-A-THON - September 2021 Organized by Analytics Vidhya

1. **Approach:** Through EDA of data set variation of target variable was understood with predictor.
   Then multiple models (Linear regression, Decision Tree, Random Forest, Gradient Boosting Regressor, Decision Tree with SVD transformation of data) were built and best model was selected.
   Model selection was done using cross validation error. Performance metric used was mean square log error.

2. **Data Pre-processing / feature engineering:** Extracting weekday from date. Dummification of categorical data. First modelling was done without date. Then weekday data was extracted from date and again modelling was done. On comparing the two models later one found to perform better.

3. **Final Model:** *"Gradient Boosting Regressor".* Multiple decision tree models were built by varying the max depth. *Gradient Boosting Regressor* was found to perform best on cross validation score. Later score on test data was also compared for different decision tree models and other models mentioned above. At the end *"Gradient Boosting Regressor"* was found to perform best among all.