Project – Housing

- First, we import libraries for reading the dataset
- Now read dataset and assign it to a variable
- After reading the dataset, check for datatype, in which form data is present
- Some data are present in int form, and some in object form
- Let's check for null values
- Some nulls are present as shown in table
- But we check individually for all columns
- Now import encoder for encoding columns wherever required
- We will do encoding and filling null values together
- After encoding and filling null values, we will check for outliers
- Before it we plot distribution and boxplot
- Distribution plot for checking the trend in dataset
- Boxplot for taking a look on outliers
- Some data are left skewed and some are right skewed
- Some outliers are also shown
- Will treat outliers with the help of quantiles and inter quantile range
- Will remove all outliers wherever present
- Now plot heatmap for checking multicollinearity
- Two columns are showing some relation

- Let’s plot scatter plot for checking the relation
- Some relation is shown by scatter plot
- Delete one of column from them
- Now split dataset into two variables for further process
- First, we standardize data
- Then we will do train test split
- Import libraries for checking the error of model
- Because target variable having the continuous data, so we apply regression algorithm
- First model, Linear Regression
- Let’s apply regularization technique
- Same as linear regression, after regularization same result is coming out for all
- Second model, Decision Tree
- Third model, Random Forest
- Fourth model, SVM
- Fifth model, AdaBoost
- Apply hyperparameter tuning for AdaBoost
- Both are giving same result
- Let’s consider one of them and save it
- Save model, AdaBoost