# HOUSING INTERNSHIP PROJECT

houses are one of the necessary need of each and every person around the globe and therefore housing and real estate market is one of the markets which is one of the major contributors in the world's economy. It is a very large market and there are various companies working in the domain. Data science comes as a very important tool to solve problems in the domain to help the companies increase their overall revenue, profits, improving their marketing strategies and focusing on changing trends in house sales and purchases. Predictive modelling, Market mix modelling, recommendation systems are some of the machine learning techniques used for achieving the business goals for housing companies. Our problem is related to one such housing company. A US-based housing company named Surprise Housing has decided to enter the Australian market. The company uses data analytics to purchase houses at a price below their actual values and flip them at a higher price. For the same purpose, the company has collected a data set from the sale of houses in Australia. The data is provided in the CSV file below. The company is looking at prospective properties to buy houses to enter the market. We are required to build a model using Machine Learning in order to predict the dataset

technical Requirements: • Data contains 1460 entries each having 81 variables.

Data set contains both training data set and testing dataset. We need to do EDA for both the data sets.

Checking the null values for training dataset.there are some of the variables such as Alley,LotFrontage,MasVnrType…etc

We use the mean imputation method to fill the null values.

Test data doesnot contains the null values.

##Data visualization

We use scatter technique to plot the visualization between quality and Saleprice

We use scatter technique to plot the visualization between year and Saleprice

We use scatter technique to plot the visualization between yearRemodAdd and Saleprice

We use scatter technique to plot the visualization between TotalBsmtSF and Saleprice

We use scatter technique to plot the visualization between  1stFlrSFand Saleprice

Etc..

#Modelling

We use the train test split for modelling and training

We first use the Linear Regression for training the dataset and predicting the testing data

Next we use the random forest regressor for training the dataset and predicting the data on the model

Next we use the DecisionTree regressor for training the dataset and predicting the data on the model.

Next we use the kneighbors regressor for training the dataset and predicting the data on the model

Of all these models,RandomForest Regressor has the better r2_score ..

So we use the hyper parameter technique to improve our model.

We use the Randomized Search CV method to hypertune the model.

This improved our accuracy..

At last we save the best model.