

Housing project

Problem Statement:-

The problem statement of the project is that we have to find out exact value of houses in which the house is to be purchased so that the sale prices of houses can be ascertained and also understand the chances of prices in the market.

It's understanding:-

The project depicts that the houses play an important role in the people's life as their living. The houses are to be properly built and every necessary thing should be available. The market of real estate is dynamic, this can be due to the built quality and location where it is. Therefore, to understand the price range of a particular property according to its features and its location is very difficult. To ascertain the prices, according to its features and location, can be done by predicting it with the help of a data scientist who will analyze the data and ascertain the prices.

EDA steps and visualization, steps and assumption used to complete the project:-

EDA steps and assumptions used to complete the project:-

First the data is studied, analyzed, first five rows, last five rows, knowing about info of the data, its datatypes, columns of data, shape of data, null values present, etc. The data have null values present and the data is both categorical and object type. The null values present in the numerical data is imputed with the mean of their respective columns and the categorical data is encoded with label encoder. Now the correlation between the columns and with the target column is studied in which the column Utilities is showing no correlation with the target column so it is dropped out. The checking and removing of outliers present is done but after removing the outliers, the data is becoming biased as the data loss % is 62%, so the outliers present is not removed. The skewness present in the data is checked and removed too with the help of power transform. Now the dataset is ready for training and testing by the model before it do the scaling of the data by standard scaler.

The dataset which is used for training and testing by the model is first the data is studied, analyzed, first five rows, last five rows, knowing about info of the data, its datatypes, columns of data, shape of data, null values present, etc. The data have null values present and the data is both categorical and object type. The null values present in the numerical data is imputed with the mean of their respective columns and the categorical data is encoded with label encoder. In this data the PoolQC has no values in it therefore it is dropped out. Now the dataset is ready for training and testing by the model before it do the scaling of the data by standard scaler.

Visualization:-

In the dataset, I have used the heatmap for null values and for showing correlation between the columns.

For outliers, I have used the boxplot for checking it.

For skewness, I have used the histplot and kdeplot for checking it.

I have used the bargraph for the lasso and ridge model.

I have used the scatter plot to see the best fit line in the random forest regressor model graph as this model is the best predictive model for dataset.

Model dashboard and finalized model

- **Dashboard of model:-**

The model which are used to predict the house prices are:

Models used	accuracy score
Linear regression: state of 455	53% at random
Lasso:	35%
Ridge:	35%
Support vector regressor:	
Linear-	63%
Poly-	61%
Rbf-	61%
Random forest regressor:	83%

- **Finalized model:-**

The best model for predicting the house prices is the random forest regressor as the score of this model is 83% which is highest of all model described under dashboard of model.

Conclusion:-

After the deep study and analyzing, I come up with the conclusion that the data shared by the company for predicting the actual price of houses and also for the purpose of knowing the features which affect the price of houses is that:

The prices of houses that came to know after the prediction depicts that most of the house actual prices are less than the predicted prices means the company should invest on the purchase of houses in which actual prices are less than the predicted one's so that the company can make profits by selling them at higher prices.

The features that affect the house prices are the locations, built quality, how much nearest to the different essential services of the market such as schools, colleges, hospitals, food and cloth markets, the area of the property, the architecture of the house from outside and inside, etc. Somewhere these features or variables plays an important role in the predicting the price of houses.