

## Project Description:

In this particular project, we are using a dataset that contains information like, Address, Rooms, Type, Price, Seller etc and using that to predict the price of a given house.

However, before you go ahead and make a prediction, it is advised that you first pre-process the data, since it may contain some irregularities and noise. In addition, try various tricks and techniques in order to gain the best accuracy in your predictions.

## Column details(Meta description)

Suburb: Suburb

Address: Address

Rooms: Number of rooms

Price: Price in Australian dollars

### **Method:**

S - property sold;

SP - property sold prior;

PI - property passed in;

PN - sold prior not disclosed;

SN - sold not disclosed;

NB - no bid;

VB - vendor bid;

W - withdrawn prior to auction;

SA - sold after auction;

SS - sold after auction price not disclosed.

N/A - price or highest bid not available.

### **Type:**

br - bedroom(s);

h - house, cottage, villa, semi, terrace;

u - unit, duplex;

t - townhouse;

dev site - development site;

o res - other residential.

SellerG: Real Estate Agent

Date: Date sold

Distance: Distance from CBD in Kilometres

Regionname: General Region (West, North West, North, North east ...etc)

Propertycount: Number of properties that exist in the suburb.

Bedroom2 : Scraped # of Bedrooms (from different source)

Bathroom: Number of Bathrooms

Car: Number of cars spots

Landsize: Land Size in Metres

BuildingArea: Building Size in Metres

YearBuilt: Year the house was built

CouncilArea: Governing council for the area

Latitude: Self explanatory

Longitude: Self explanatory

## Part-1: data Exploration and Pre-processing

- 1) Load the given dataset
- 2) Print all the column names
- 3) Describe the data
- 4) Drop address, date, postcode, YearBuilt, latitude, longitude columns
- 5) Find the count of null value in each column
- 6) Fill the null value of property count, distance, Bedroom2, Bathroom, Car with 0
- 7) Fill Null value of land size and bidding area columns with Mean
- 8) Find the unique value in method column
- 9) Create a dummy data for categorical data

## Part-2: Working with Model

- 1) Create the target data and feature data where target data is price
- 2) Create a linear regression model for Target and feature data
- 3) Check if the model is overfitting or underfitting or it is accurate
- 4) If the model is overfitting then apply ridge and lasso regression algorithms
- 5) Extract slope and intercept value from the model
- 6) Display Mean Squared Error
- 7) Display Mean Absolute Error
- 8) Display Root mean Squared error
- 9) Display R2 score