Files:

Dataset: 'train.csv' contains the price of houses and various other features of houses along with it. A 'Data Dictionary.xlsx' is provided along with to help understand what those variables are.

Analysis: It is carried out in 'Advanced Regression – House Price Prediction.ipynb'.

Problem Statement:

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 value and flip them at a higher price. For the same purpose, the company has collected a data set from house sales in Australia. The data is provided in the csv file below.

The company is looking at prospective properties to buy to enter the market.

You are required to build a regression model using regularization, so as to predict the actual value of the prospective properties and decide whether to invest in them or not.

The company wants to know:

- Which variables are significant in predicting the price of a house
- How well those variables describe the price of a house

Also, determine the optimal value of lambda for ridge and lasso regression.

Business Goal:

You are required to model the price of houses with the available independent variables. It will then be used by the management to understand how exactly the prices vary with the variables. They can accordingly manipulate the strategy of the firm and concentrate on areas that will yield high rewards. Further, the model will be a good way for management to understand the pricing dynamics of a new market.