Housing Report

The housing data provided is for US-based housing company named 'Surprise Housing' who 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 dataset contains 1460 rows and 81 columns with null values present.

Steps taken:
1. We check the data inside the dataset to understand the type of data.
2. We understand the data to treat it for the null values that are present.
3. After the treatment for the null values, we encode the data to produce columns all into numerical type.
4. We see few features that are either unique to each column such as 'Id' or features that are same for all the rows of the data that add no value to the learning.
5. Such data is removed so as to reduce the features for the dataset.
6. Now we do an extensive research using a correlation matrix to grade each column to the weightage to the target column.
7. We remove the features that has a correlation value of less than 0.1.
8. This process reduces our data set to 55 columns.
9. Now we have reduced our dataset to the features that produces a weightage to the target column and hence ready to train the model.
10. On the basis, we better understand what value each housing features produces in the price predictions.

Conclusion:
From the data we can see that various qualities of the house such as the Overall quality, garage and living area are of the most importance for the people in terms of the price of a house. On the other hand, features such as the basement quality, bath quality, Alley produces a reverse correlation of near to 0 correlation and hence, least importance.

This concludes that house with great living area, garage and overall quality should be given more importance while procuring whereas the other factors that has low importance can be used to reduce the price of the house while buying.