Polynomial Regression for House Price Prediction

Methodology:

- **Data Cleaning:** Checking for null values and based on their number either dropping them or replacing with mean, median, mode based on the type and description of data.
- **Data Visualization:** This step helps understand the understand the data in a visually. We can understand normality of the data as well. This helps us to decide whether to normalize the data.
- **Feature Selection:** Based on the Pearson correlation between the labeled column and rest of the features. In general, a very great correlation should have an absolute value greater than 0.75. When the labeled column is depended on multiple columns, the correlation with one column may be less. But combined features may have higher effect.
- **Model Selection:** Based on the data visualization and data correlation, we need to select a model that would best suit. In Polynomial Regression Degree selection is very important, as giving higher value leads to overfit and less to underfit. So, by trying different value helps to find the optimum value.
- **Evalution:** In this case we are using RMSE to determine the accuracy of the predicting model.

Code:

https://colab.research.google.com/drive/1bPkJOW_PHMhHaoNwuaseDUgH_ZfRPq6Y?usp=sharing