“Machine Learning Model Predicts HDB Flat Resale Price to Within $39,179 of the Actual Price!”

- The State Times

Problem Statement

I am looking to buy a HDB resale flat. Given a list of flat attributes such as street name, type of HDB flat, what is the resale price?

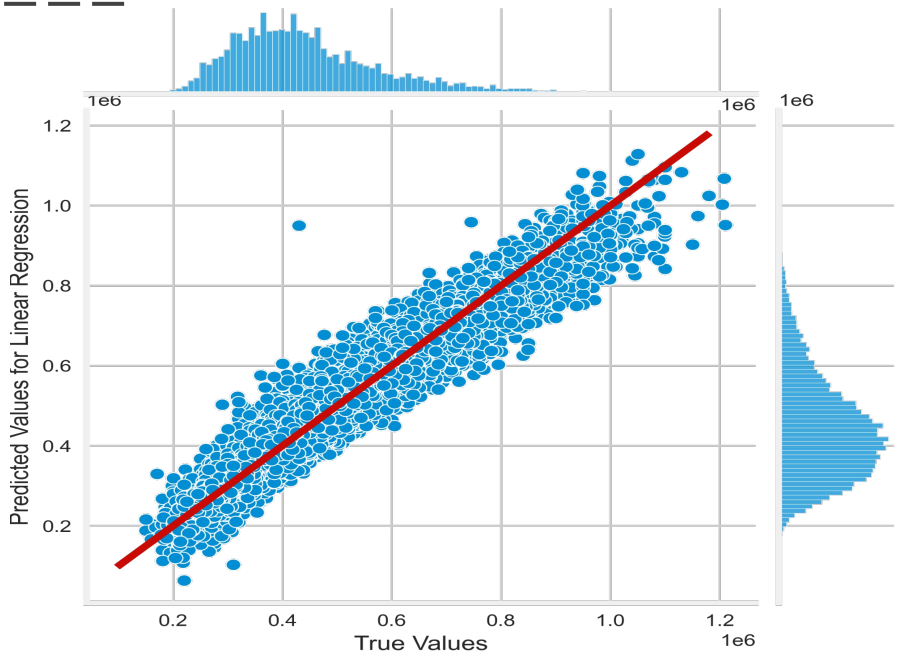
The Journey to find the Best Model

- **Establish a baseline for comparison**
  - Dummy regressor generates predictions based on the mean
- **Automatic Feature Selection**
  - Best features are those maximize information gain
  - Minimizes total entropy of the model
- **Percentile of Features Selected**
  - 99%, 90%, 75%, 50%

Models, Models and more Models

- **The following models have been fitted 4 times each on the 50%, 75th%, 90% and 99% of best features**
  - Linear Regression
  - L1-regularized Linear Regression
  - L2-regularized Linear Regression
- **The best model is the non-regularized linear regression using 50% of the best features**

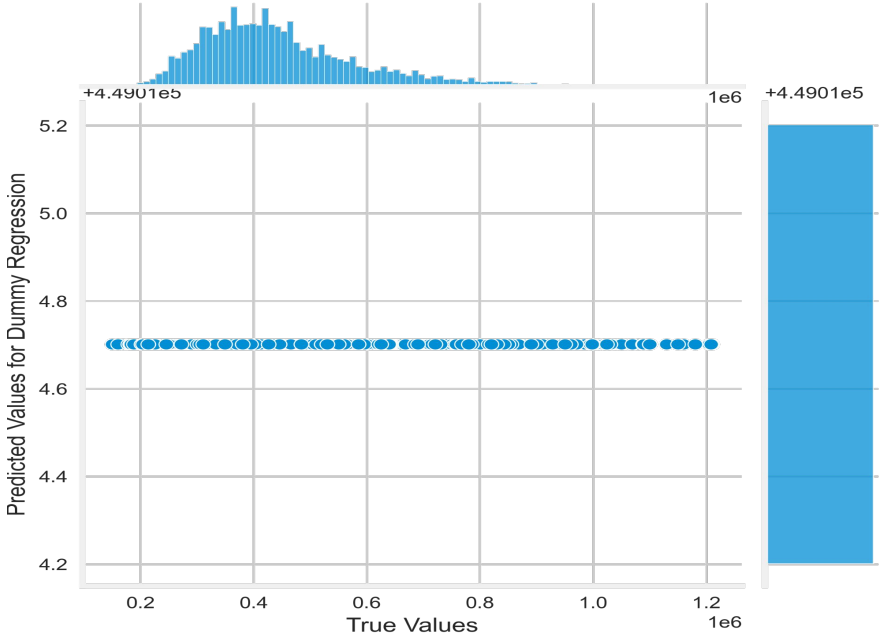
Visualisation of Best Model Performance



Best Model	
R-Squared	0.9261
RMSE	39,179

Fig 1. A prediction error plot of predicted values of resale price against the true values of resale price

Visualisation of Dummy Model Performance



Dummy Model	
RMSE	144,095

Fig 2. A prediction error plot of predicted values of resale price against the true values of resale price

Conclusion

Given a list of flat attributes such as street name, type of HDB flat, the best model is able to predict the resale price to within an error of $39,179

Recommendations