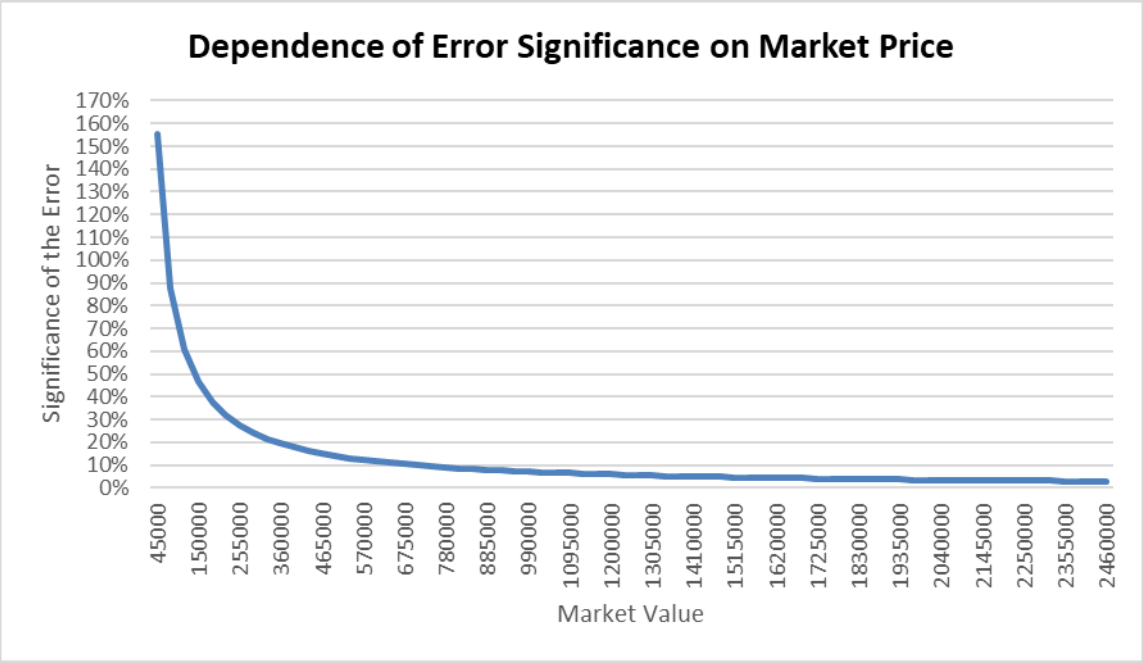
Question: Can we predict the market value of housing in 2013 based on data from 2011?

Answer: A regression model was built for the given question, which on average makes an error of about $70,000 in its predictions. It is acceptable to make predictions of the market value of more expensive housing (more than $1,000,000), but it performs poorly for more budget-friendly or average offerings.

Graphs:



The graph clearly shows that the error is extremely significant for budget and medium-range offers (10,000 - 456,000) and its significance greatly decreases with the increase in price.

Details:

A linear regression model was built to predict the market value prices of housing for the year 2013. This regression model was trained on 2011 data to predict the pre-known market values for 2013.

The R-square of the model is 0.59, which means it can explain ~60% of the variances. Adjusted R Square is almost identical to R Square, suggesting that the model does not have 'superfluous' variables.

After analyzing the t-statistics for each value, it can be concluded that each variable is statistically significant in the model.

Using the analysis of Holdout data (1000 random observations), which was excluded before creating the regression model, it was found that the Mean Absolute Difference between the actual and predicted was $70,000. This means that on average, the model makes a mistake of $70,000 in the market value of the housing.

Based on the data in the 'Data descriptors' sheet, it can be seen that the model did fairly well with the average, but it poorly identified the minimum and maximum values, which are quite far from the actual ones (significantly exceeding $70,000).

Also, the predicted values are much more evenly distributed