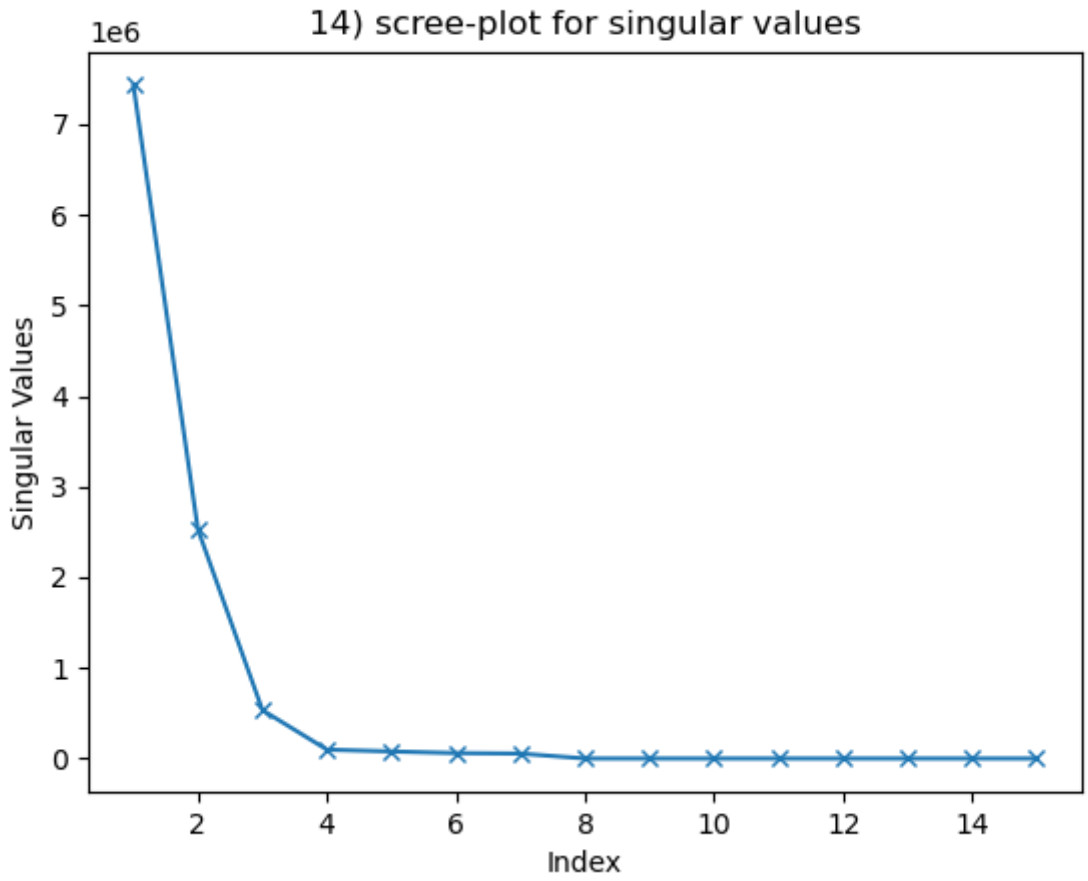
Answers to practical questions

Q13:

I found the following features to be categorical

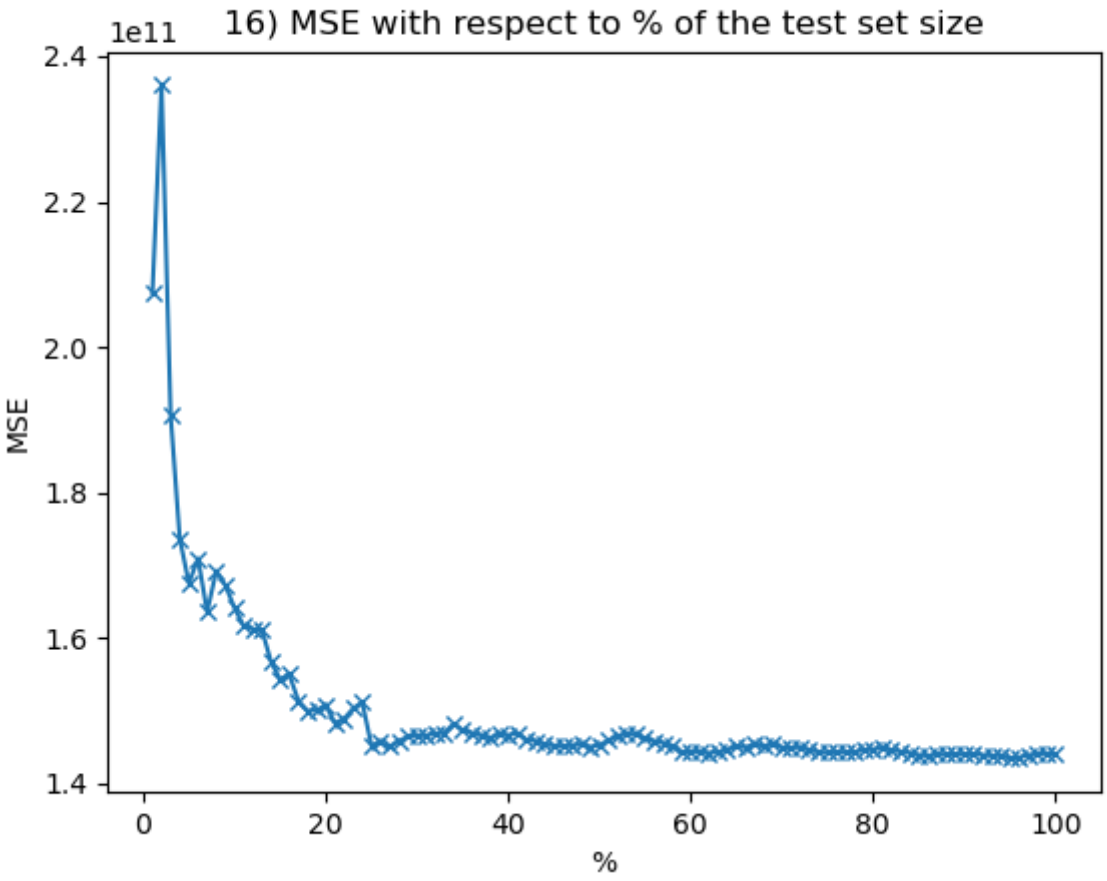
- **ID**: a unique number
- **Longitude/Latitude**: can be ordered but the order does not have meaning
- **Date**: since it's has meaning only over large period of time (years), and the only years available are 2014/2015, we can treat it as categorical data.

Q15:



The matrix is singular since it has values that are equal to 0
This means that the matrix is not invertible

Q16:

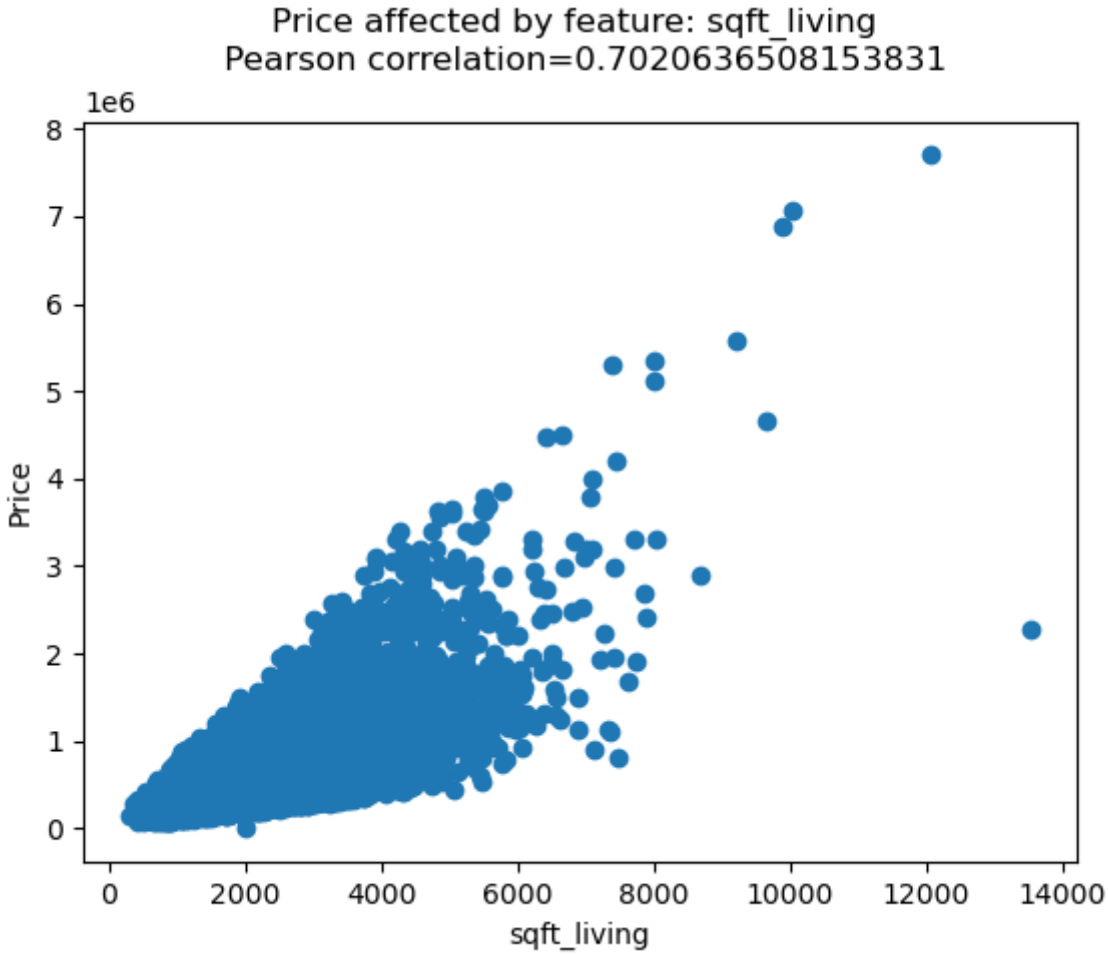


We can see that the MSE decreases with respect to decreasing the percentage of the test set size (increasing the train set size)

We can also see that increasing the percentage over ~25% does not change the does not change the MSE very much and it stays around a constant value.

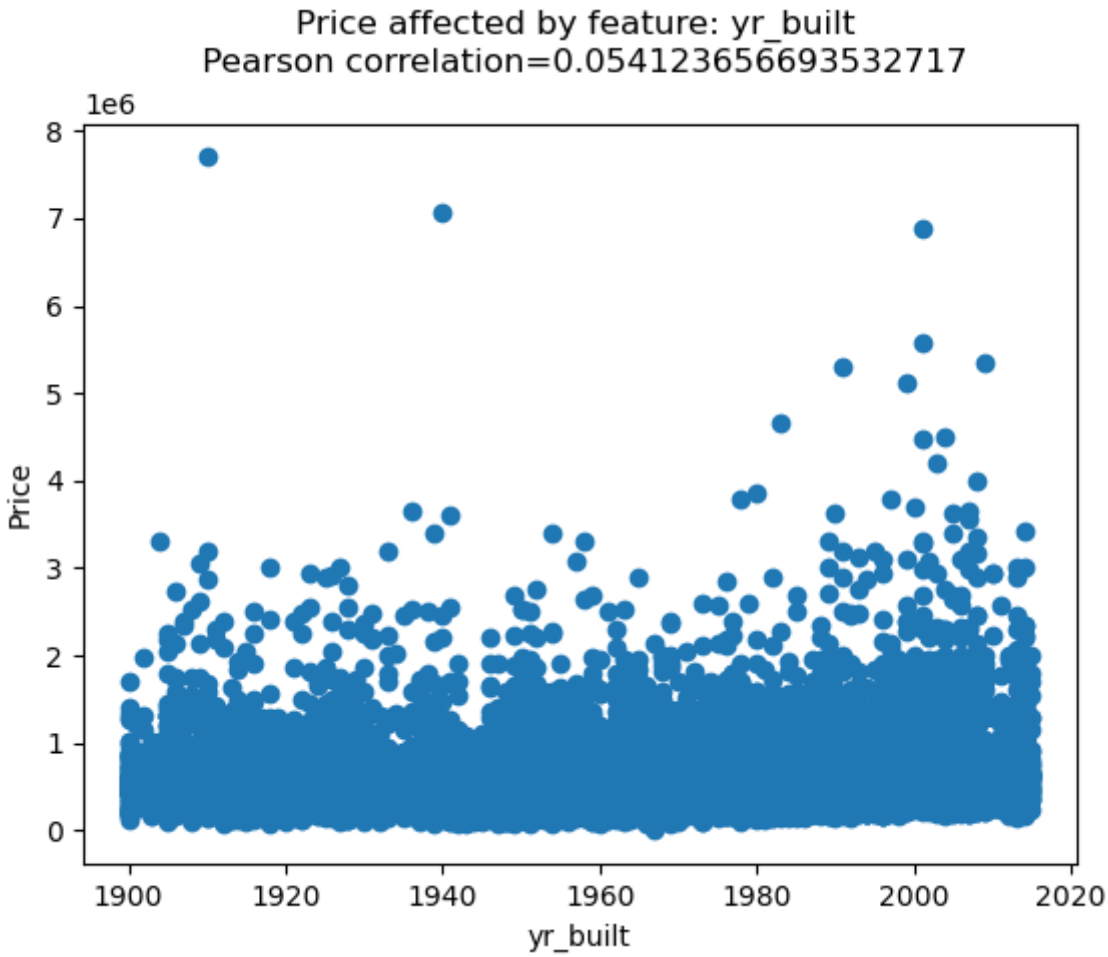
Q17:

Beneficial for the model:



We can see high correlation: the price increases when the as the size of the living room increases. Also the Pearson correlation is relatively high.
From this we can conclude that this type of data is beneficial for the model.

Not beneficial for the model:



We can see low correlation: the price does not have a particular trend as the size of the living room changes. Also the Pearson correlation is relatively low.
From this we can conclude that this type of data is not very beneficial for the model.