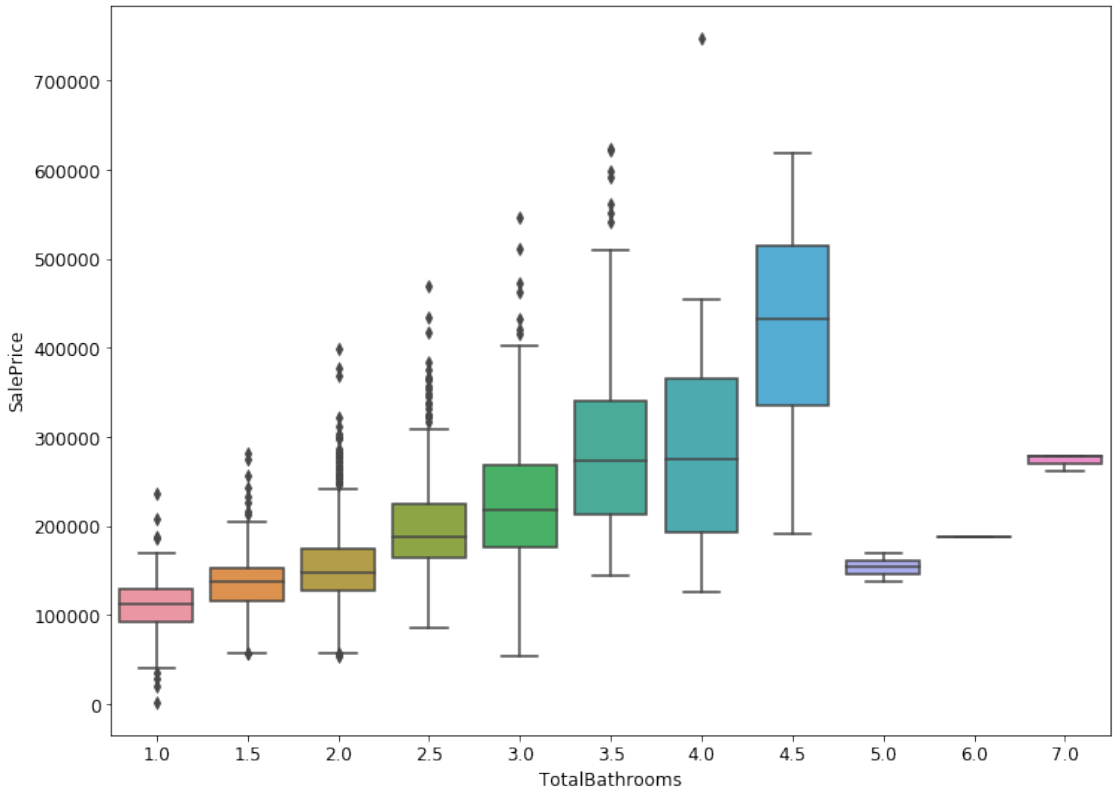
Notebook

March 30, 2020

0.1 Question 5

Create a visualization that clearly and succintly shows that `TotalBathrooms` is associated with `SalePrice`. Your visualization should avoid overplotting.

```
In [191]: sns.boxplot(
              x='TotalBathrooms',
              y='SalePrice',
              data=training_data
          );
```



Ideally, we would see a horizontal line of points at 0 (perfect prediction!). The next best thing would be a homogenous set of points centered at 0.

But alas, our simple model is probably too simple. The most expensive homes are systematically more expensive than our prediction.

0.2 Question 8d

What changes could you make to your linear model to improve its accuracy and lower the test error? Suggest at least two things you could try in the cell below, and carefully explain how each change could potentially improve your model's accuracy.

1. Using more parameters. Some of the parameters such as Lot_Area clearly have a positive influence on the price. Including these parameters will potentially improve the model's accuracy.

2. Including non-linear parameters. For example, from Question 5 we can see that the relationship between price and bathroom numbers is more than just linear. It would help if we include a parameter defined as 'square of bathroom numbers', potentially improve the model's accuracy.