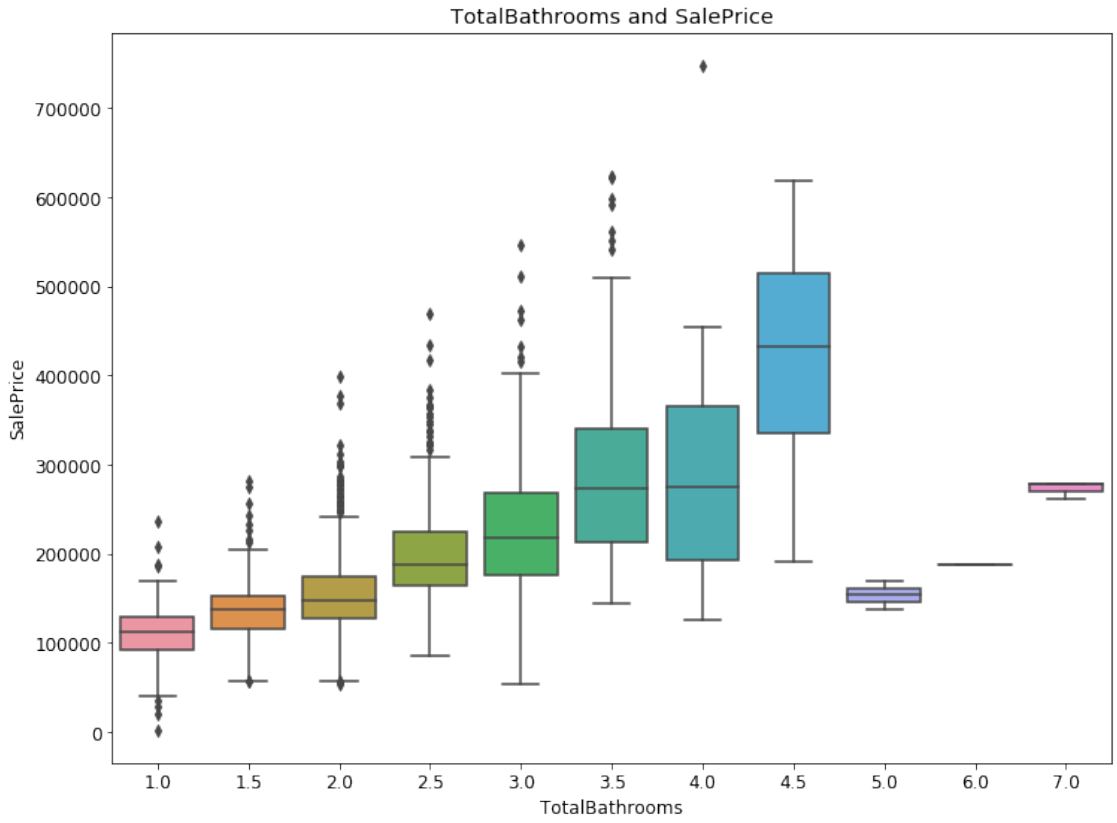
0.1 Question 2b

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

```
In [15]: sns.boxplot(x = "TotalBathrooms", y = "SalePrice", data = training_data_with_bathrooms)
         plt.title("TotalBathrooms and SalePrice")
```

```
Out[15]: Text(0.5, 1.0, 'TotalBathrooms and SalePrice')
```



0.2 Question 5d

What changes could you make to your linear model to improve its accuracy and lower the validation 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.

In order to improve linear model accuracy and lower validation error, we would have to consider changing the complexity of the model.

1. Cross Validation We can take a look at running k -folds cross validation on our training data which includes splitting the data into K different partitions and using K-1 partitions to train the model while using the last to come up with a vaidation error. We repeat this for every partition and use the average partition error to evuluate different models against this one.
2. Chaning the complexity of models through Regularizations We can change the weighting of each feature by using Lasso or Ridge techniques to limit the complexity of the model and this can reduce the variance of the model.

Finally another approach is to intuitively add another data point that may be useful in predicting house prices. Adding and effective data point can reduce the model bias and increase model variance. In combination with cross validation we can make sure we are not compromising test error for training error.

0.3 Question 6a

Based on the plot above, what can be said about the relationship between the houses' sale prices and their neighborhoods?

There is no relationshup between the count of houses in a neighborhood and the price of houses in a neighborhood, however there seems to be different mean house price in different neighborhoods. For example the average house sale price in OldTown is below 175,000 while the average house sale price in StoneBR is around 300,000. There is no distinct relationship but there is a correlation between a different neighborhood and house price in that neighborhood,and there is a lot of overlap between neighborhoods but there is some differences between some neighborhoods.

0.4 Question 8a

Although the fireplace quality variable that we explored in Question 2 has six categories, only five of these categories' indicator variables are included in our model. Is this a mistake, or is it done intentionally? Why?

One of the fireplace variables take the place of the dummy variable that stands as the standard. The Theta_5 to Theta_9 measures the differnce between the dummy variable (or the standard) and the change in condition. If we include all the variables, the all add to 1, making it an issue of invertibility for the matrix.