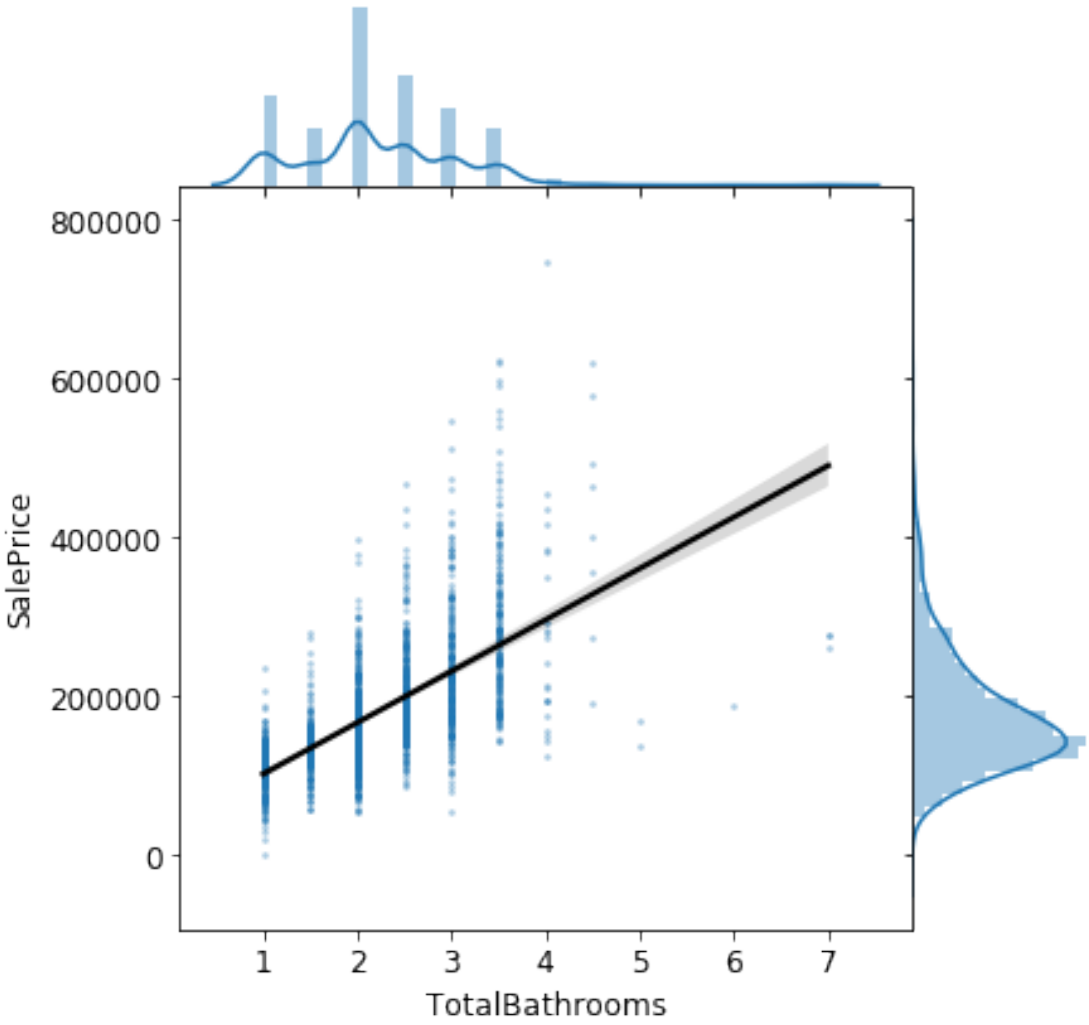
Notebook

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 [209]: sns.jointplot(
              x='TotalBathrooms',
              y='SalePrice',
              data=training_data,
              stat_func=None,
              kind="reg",
              ratio=4,
              space=0,
              scatter_kws={
                  's': 3,
                  'alpha': 0.25
              },
              line_kws={
                  'color': 'black'
              }
          );
```



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.

One way to improve its accuracy and lower the test error is to add much more data. Variables such as geographical locations or quality of neighborhoods can help improve the accuracy of our prediction. Another way is to choose variables that play a bigger role in determining what we are trying to find, such as using the location of the property to predict sale prices or how many floors and total square feet of the property rather than using number of bathrooms.