0.1 Question 2b

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

```
In [61]: sns.jointplot(
             x='TotalBathrooms',
             y='SalePrice',
             data=training_data,
             kind="kde",
             ratio=4,
             space=0,
             scatter_kws={
                 's': 3,
                 'alpha': 0.25
             }
         );
```

```
/srv/conda/envs/data100/lib/python3.7/site-packages/seaborn/distributions.py:437: UserWarning: The foll
  cset = contour_func(xx, yy, z, n_levels, **kwargs)
```
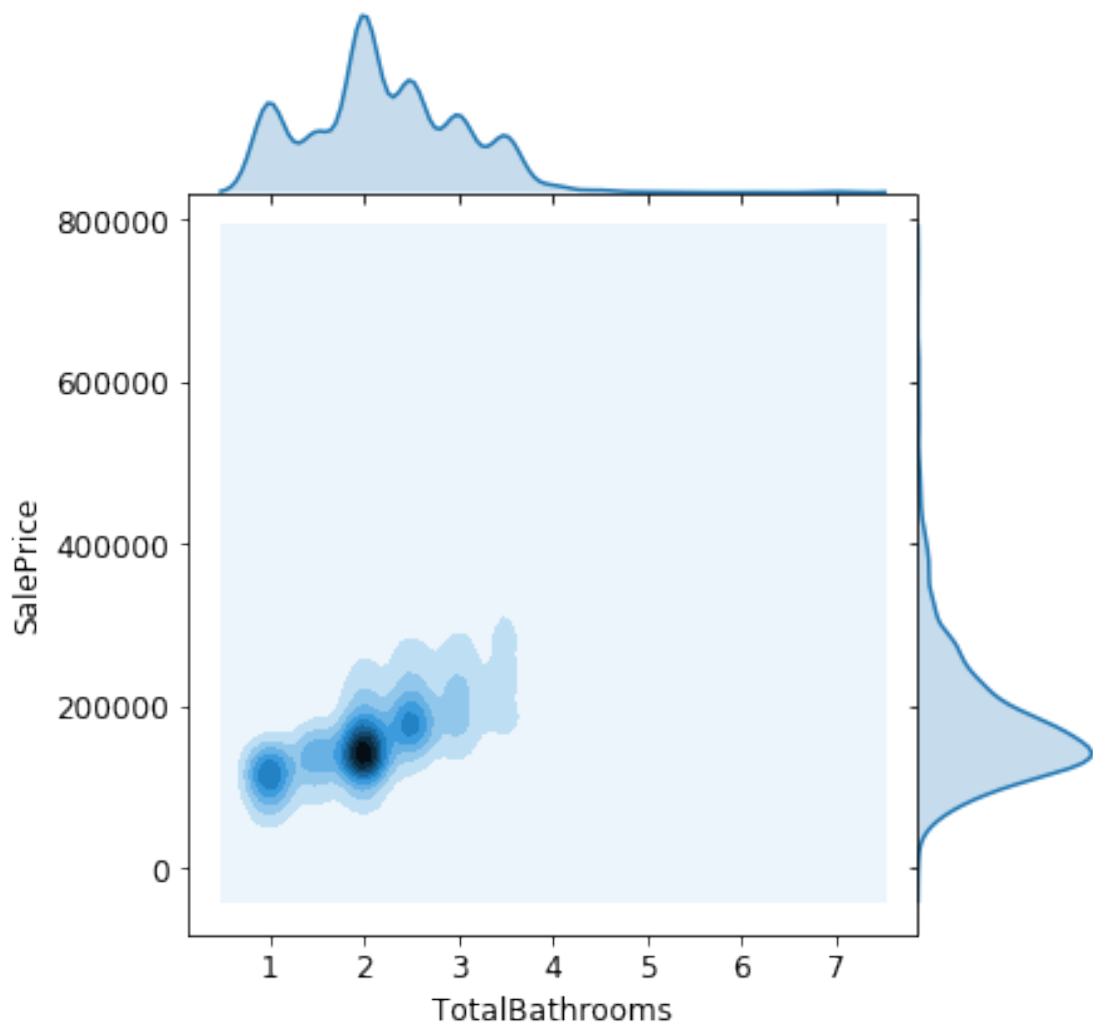
0.2 Question 5d

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.

Condition and neighborhood are features I would add to the model to improve accuracy and lower validation error. Both the condition and neighborhood that a home is located in play key roles in determining the selling price/value for a home in most cases, so including them in the model can potentially improve our accuracy.

0.3 Question 6a

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

Based on the plot above, there is no clear relationship between sale price and neighborhood. In some neighborhoods where there are high counts, such as NAmes, the range of housing prices varies widely. In addition, some neighborhoods do not have enough prices for us to determine a relationship.

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?

It is done intentionally. One of the six categories was Average, which is ommitable in this case because it is a redundant column. We can determine if a column was Average simply if the sum of the other columns is zero. Removing Average also allows us to have linearly independent columns.