Statistical Data Analysis

df_train[['SalePrice', 'BedroomAbvGr']].describe()

	SalePrice	BedroomAbvGr
count	1457.000000	1457.000000
mean	180942.138641	2.866163
std	79521.569966	0.816595
min	34900.000000	0.000000
25%	129900.000000	2.000000
50%	163000.000000	3.000000
75%	214000.000000	3.000000
max	755000.000000	8.000000

Split BedroomAbvGr into two different groups. The mean number of bedrooms is 2.87 so it is best to group homes that have 2 bedrooms and 3 bedrooms above grade.

two_bedroom = df_train.SalePrice.loc[df_train.BedroomAbvGr == 2]

three_bedroom = df_train.SalePrice.loc[df_train.BedroomAbvGr == 3]

Calculate T-Test for the two independent variables

import scipy

scipy.stats.ttest_ind(two_bedroom, three_bedroom)

```
Ttest_indResult(statistic=-5.247681020072384, pvalue=1.8309779204141817e-07)
```

The t-test observes homes with 2 bedrooms and 3 bedrooms above grade. Since the p-value is less than 0.05, we reject the null hypothesis. The average sales price for a 2 bedroom houses is different from that of a 3 bedroom houses.