数据clean后的新处理摘要：

1. 删除了State 和 Country，都是US MA没意思
2. 将所有Factor选项里的blank换成了Text的“Not Sure”这样可以将NA也当成一个factor变量了
3. 对于数字变量的空白基本上用了mean代替了NA，但是删除了所有空白的y变量的一整行（大概十几个）
4. 有一些时间变量的空白比如：start as host，这个直接用了Most Frequent Year代替了NA，理解为大概率上大部分房主会撑到这个时候，其实最合理的是根据这个多factor的Normal distribution用 Random函数生成，不想搞了most frequent也行吧。
5. 想和大家探讨一下用Mean代替NA是不是欠妥，尤其是Square Feet这样的明显影响价格的变量，统一用一个Mean会使得整个coefficient降低。