Statistics: Airbnb, San Francisco listings

**I. Background**

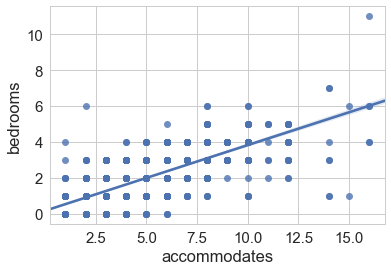
When you want to go on a vacation or you have to move to a different city, you need to plan many things. One of the most important things you will need is a place to stay. There are houses with unused rooms and other unoccupied housings that people ignore. Airbnb is allows people to lease or short term rent such lodgings.

**II. Problem and Customer**

There are listings in airbnb, especially in a big city such as San Fransisco. But how will we know which listing to choose? To figure this out, my project will focus on which neighborhoods and listings are the most popular, and what factors are associated with them. I looked at five different questions and analyzed them using statistical techniques.

**III. Questions**

1. Is there a linear relationship between the number of accommodates and the number of bedrooms?



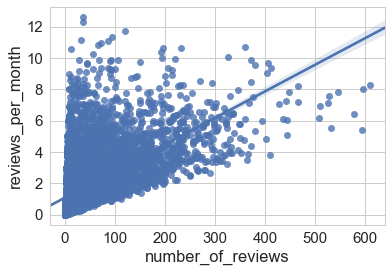
**Null Hypothesis**: There is not a significant linear relationship between the number of accommodates and the number of bedrooms.

**Alternate Hypothesis**: There is a significant linear relationship between the number of accommodates and the number of bedrooms.

**Results**: coef = 0.7595, pval = 0

I used a Correlation Test so that I can learn if there is a linear relationship between the number of accommodates and the number of bedrooms. The results show that the Pearson correlation coefficient in the correlation test is .7595, which is closer to 1 and the p value is less than .05 so we reject the null hypothesis. This means that there is a linear relationship between accommodates and bedrooms. Knowing this shows that people tend to not care about the number of bedrooms. If there are a lot of people, they tend to get a house with more bedrooms rather than try to save money by getting the same about of bedrooms for different number of accommodates. This means that people in San Fransisco care about the overall experience rather than just using the house as a place to sleep.

1. Is there a linear relationship between the number of reviews per month and the number of reviews?



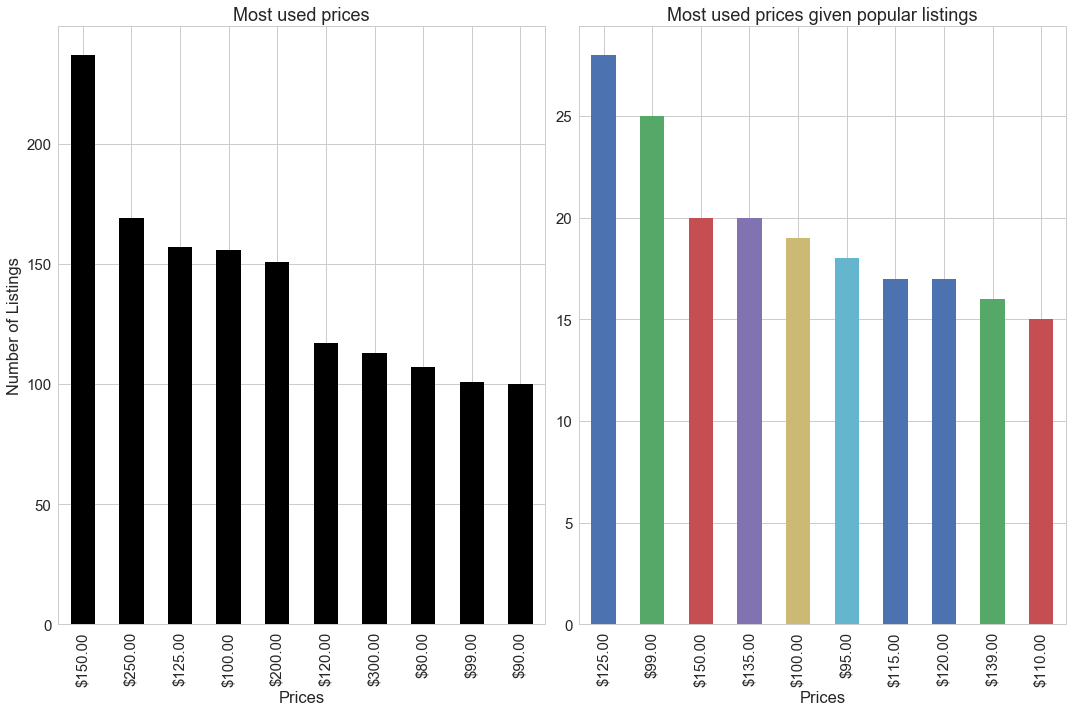
**Null Hypothesis**: There is not a significant linear relationship between the number of reviews per month and the number of reviews.

**Alternate Hypothesis**: There is a significant linear relationship between the number of reviews per month and the number of reviews.

**Results**: coef = 0.6388, pval = 0

I once again used a Correlation Test so I can learn if there is a significant linear relationship between the number of reviews per month and the number of reviews. The results show that the Pearson correlation coefficient in the correlation test is .6388, which is closer to 1 and the p value is less than .05 so we reject the null hypothesis. This means that there is a linear relationship between the number of reviews per month and the number of reviews. I tested this relationship because it is an indicator of popularity. Popular listings should have a larger number of reviews and a larger reviews per month whereas unpopular listings will have a lower number of both.

1. Does having lower prices make a listing more popular?

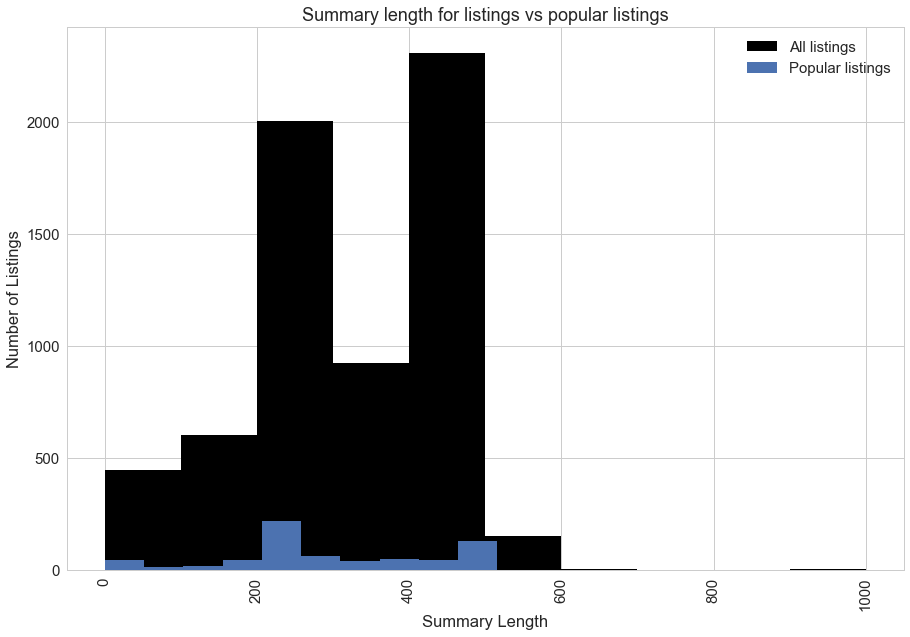


**Null hypothesis**: Popular listings have the same prices as non-popular listings  
**Alternate hypothesis**: Popular listings don't have the same prices as non-popular listing

**Results**: Ttest\_indResult(statistic=-4.2071077218452446, pvalue=2.9719290963239736e-05)

I used a 2 Sample T-test so I can learn if popular listings have the same prices as non popular listings. The p value is less than .05 so we can reject null hypothesis. This means that popular listings don’t have the same prices as non-popular listings. I then found that the average price of popular listings is $189 while the price of non popular listings is $496. This shows that cheaper listings are usually more popular.

1. Does having lower a larger summary length make a listing more popular?

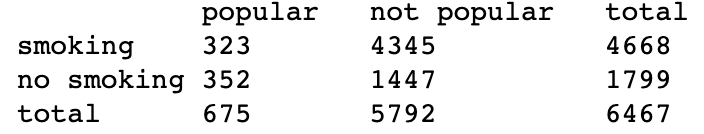


**Null hypothesis**: Popular listings have the same summary lengths as non-popular listings  
**Alternate hypothesis**: Popular listings don't have the same summary lengths than non-popular listings

**Results**: Ttest\_indResult(statistic=-0.88781476250156488, pvalue=0.37492464455180641)

I used a 2 Sample T-test so I can learn if popular listings have the same summary length as non-popular listings. The p value is greater than .05 so the null hypothesis can’t be rejected. This means that popular listings have the same summary lengths as non-popular listings. The results show that summary length is not a factor in determining popularity of listings.

1. Does having lower a larger summary length make a listing more popular?



**Null hypothesis**: A listing not allowing smoking and being popular are independent  
**Alternate hypothesis**: A listing not allowing smoking and being popular are not independent

**Results**: Power\_divergenceResult(statistic=222.1804342714089, pvalue=6.7821335266394548e-48)

I used a Chi Square test so I can learn if a listing not allowing smoking and being popular are independent. The p value is less than .05 so we can reject the null hypothesis. This means that a listing not allowing smoking and being popular are not independent. We also learn that the smoking policy is related to a listing’s popularity. The above table shows that a majority of smoking listings are not popular.

**IV. Conclusion**

I learned many details regarding the five main statistical questions asked. Firstly, people who use Airbnb for San Fransisco listings care about the overall experience rather than just using the house as a place to sleep. Secondly, popular listings have a larger number of reviews and a larger reviews per month. Thirdly, cheaper listings are more popular than expensive ones. Fourthly, the summary length is not an indicator of popularity, so a bigger or smaller summary does not indicate popularity. Lastly, I learned that smoking is a reasonable indicator for popularity.