# **MA207 - Introduction to Statistics:Lab #7: Hypothesis Tests and Confidence Intervals on Large Data SetsDay 1 - AirBnB Data**

In this multi-day lab we will work with several large data sets. This lab will span two class days. Complete this part of the lab prior to our next class period. For each day you will receive a large data set and will be asked to answer several specific questions related to the data set. You will submit your completed lab after you have combined the work for both days. For each problem you will need to be able to:

* Filter and sort the data (pivot tables are helpful),
* Find appropriate descriptive statistics for parts of the data set,
* Determine the appropriate statistical tool for the question being asked and complete all of the calculations necessary to answer the questions, and
* Interpret and explain the meaning of your computations in context.

## The Seattle AirBnB Data:

Today’s data set comes from the company [airbnb.com](https://www.airbnb.com/). This company allows homeowners to rent out all or part of their home to travelers that are visiting their city. AirBnB has become a very popular alternative to renting hotel rooms or hostels. Furthermore, it tends to allow travelers more privacy, more flexibility, and possibly more amenities than traditional lodging. As with many businesses of this type, its rise in popularity hasn’t been without it’s [issues](https://www.entrepreneur.com/article/284610).

Download the AirBnB data set from the Moodle page (you may safely press “Ignore Links” if Excel gives you a warning when you open it). This data set contains a sample of 3,318 AirBnB rentals during 2015 in the greater Seattle area. Make yourself familiar with all of the headings in the data set. We are using the 2015 data as a sample for all Air-BnB data in recent years. This data was gathered from the site: [insideairbnb.com](http://insideairbnb.com/) on Nov. 1, 2016. One way to explore the data visually is to use insideairbnb.com’s visual inspection tool found [HERE](http://insideairbnb.com/seattle/) (you may or may not find this useful in the following analysis, but it is certainly interesting and fun to play with).

**Your Tasks:**

Use appropriate statistical tools to answer the following questions. Provide sufficient detail in your write-up so that anyone with a similar statistics background will understand what you have done. You do not need to turn in your Excel document.

1. Give a range estimate for the average daily price of a private room at an AirBnB in the Queen Anne neighborhood of Seattle. Hint: Use a pivot table to sort by host\_neighborhood.
	1. Which statistical tool are you using? (ex: paired means confidence interval, two proportion hypothesis test, etc.) Why are you using this tool?
	2. Clearly report your numerical results of the statistical test.
	3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.
2. Find the two neighborhoods that have the most AirBnB rentals. Is there a statistical difference between the average daily price for AirBnB rentals in those two neighborhoods?
	1. Which statistical tool are you using? Why are you using this tool?
	2. Clearly report your numerical results of the statistical test.
	3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.
3. Is there a statistical difference in daily price between the rentals that have a flexible vs. moderate cancellation policy?
	1. Which statistical tool are you using? Why are you using this tool?
	2. Clearly report your numerical results of the statistical test.
	3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.
4. Is there a statistical difference between the proportion of rentals in the Capitol Hill neighborhood that have a cleaning fee as compared to the rentals in the University District that have a cleaning fee?
	1. Which statistical tool are you using? Why are you using this tool?
	2. Clearly report your numerical results of the statistical test.
	3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.
5. Write a question of your own based on this data. Clearly state your question and then answer it with a hypothesis test or confidence interval.
	1. Which statistical tool are you using? Why are you using this tool?
	2. Clearly report your numerical results of the statistical test.
	3. Clearly write a conclusion from your statistical test that ties your numerical results back to the context of the problem.