

### **Visualize Your Data to Make Sure You Know What They Are**

Before you start any kind of analytical or statistical modeling, it's important to look at your data to make sure you understand what they represent, to look for outliers, and to examine whether they contain any obvious effects or relationships you should take into account when designing your models. If your data are already in table form (as they will be after you have extracted them from the MySQL database), the fastest way to do this is by using Tableau.

First, make sure you have the following fields of information in your database query output:

Watershed property ID  
(optional: location ID)  
city  
state  
zipcode  
property\_type: apartment or house  
property\_type: number of bedrooms (1 or 2)  
property\_type: kitchen  
property\_type: shared  
Watershed property current monthly rent  
comparable\* short-term rental 10th percentile rent  
comparable\* short-term rental 90th percentile rent  
comparable\* short-term rental example rent  
comparable\* short-term rental example occupancy rate (for 2015)

\*comparable means the short-term rental that has the same location and property type as the Watershed property

Next, now that you know what data you have available, use that information to help you prioritize which arms of your SPAP you should focus on in your data visualization and subsequent analysis steps. Since this project has a short timeline there will not be time to collect more data, but if the project had a longer timeline and a deeper scope, this would be a good point to talk to the rest of your team about any missing data you need collected to achieve your objectives.

Then, simply connect Tableau to the excel worksheet you pasted your database results into, and complete the following exercises:

- 1) Do any of your rent measures have extreme outliers? (Hint: remember to un-aggregate your data for this step.)

**An important note: for the purposes of this Capstone project, please do not exclude any data from your analysis because you think it might be an outlier. Although some data values might pop out to your eye, we do not have enough data replicates to be sure that any data point is truly invalid. Thus, best practice in this situation would be to consider using a median rather than an average to**

**group data together, and to include all the data until more information can be collected.**

- 2) Do average and median short-term rental prices tend to vary in different locations, or according to different property types? If so, it is likely that you will have to take this into account in your models in some way.
- 3) Are there geographic locations or property types whose short-term rental prices cover a particularly wide range (in other words, the difference between their 10<sup>th</sup> and 90<sup>th</sup> percentile rental rates are particularly large)?
- 4) If you haven't determined this already through examining your data within the database, how many different geographic locations are the Watershed properties located in, and how many different property types are there?

The more you understand the details of your data, the better prepared you will be to create accurate forecasts of what your data would likely be in other situations or at other times, so it's good practice to run these types of quick analyses at the beginning of any business analytics project. Especially when used in combination with a SPAP, they will help you efficiently narrow in on what features of your data are most important, and will ensure that you don't miss any important trends or relationships. We do not want you to spend a lot of time on these exercises, but we recommend that you look at your data in enough different ways that you have a sense for what values your data span, and so that you know what factors are likely to influence rental prices.

The short quiz at the end of this week will make sure you have identified the most important trends in your data that will be useful for designing your short-term rental price optimization strategy next week.

Happy Tableauing!