Improving Service to the Emerging Needs of Airbnb Customers

Heather Swanke

Lewis University

**Introduction**

Since 2008, Airbnb has been breaking the status quo in the hotel industry. They've popularized a new model of person to person hoteling. The market they serve has grown drastically in the past 10 years and they’ve evolved from a California-based start up to a multinational company that operates in 191 countries and more than 81,000 cities. The US still remains as Airbnb’s largest market with 660,000 listings as of last year (Hartmans, 2017). However, as they continue to grow, it's important to keep in mind how they're serving their customers and how they can improve. To provide a comprehensive perspective on US listings, I will analyze Airbnb listing data for 5 major US cities: Austin, Chicago, Los Angeles, New York, and San Diego. In this paper, I will utilize customer review data to examine listings where Airbnb is failing to meet customer expectations and understand trends within and across cities. I will also examine data based on listing availability to determine the key attributes for in demand listings in each city. I propose that in order to better serve customers in the Airbnb US market, this will require a twofold approach of improving reviews and promoting in demand listing features to maximize availability.

Airbnb provides services and accommodations to customers through individuals who register themselves as a host and post their properties online. Airbnb does not charge hosts a listing fee to post their property. Instead, the company generates revenue off of a flat rate fee of 3% on reservation transactions according to the “Earn money as an Airbnb host” webpage on Airbnb. The site dictates that hosts and customers are expected to adhere to general community standards to ensure safety, security, and fairness for everyone in the Airbnb community. Beyond these key factors, the unique experiences offered by individual hosts and properties varies widely. Airbnb celebrates the uniqueness that they provide. However, with this variability, it’s still important to ensure that customers are able to find what they’re looking for in each city. While Airbnb is already doing very well in terms of their profitability and viability as a company, I will examine how they can improve on their successes.

**Data Description**

Inside Airbnb (not directly affiliated with Airbnb) has compiled data from public information posted on Airbnb.com regarding listings and reviews, and the data compilation can be found on Inside Airbnb’s “Get the data” page. Listing information is available for international cities as well as dozens of smaller cities and towns in the US. I chose to focus on US locations so that the information would be more directly comparable. I also chose to focus specifically on large cities which demonstrate a more competitive and dynamic market for Airbnb rather small cities which have limited demand for tourist and business travelers and individually only represent a small portion of Airbnb’s bottom line. Austin, Chicago, Los Angeles, New York, and San Diego all rank within the top 12 cities for residential population according to 2017 population estimates calculated by the United States Census Bureau. On InsideAirbnb.com, each city has its own individual tables available with listing details as well as review details. The Airbnb data for all of this is quite current and was scraped from the website and compiled between October 2018 and November 2018 (specific compilation dates vary by city). The data set also includes data points spanning back to 2008. The table dedicated to review details is organized by property and the corresponding reviews of individuals who stayed there, however, this only includes qualitative text content and not individual numeric ratings chosen by the reviewers. The reasoning as to why rating numbers chosen by individual reviewers are not published will be addressed in more detail in the legal and ethical issues section of the paper. In the listing details table, information is organized by individual listing ID numbers with one row per listing. The overall number of properties available varies significantly by city:

|  |  |
| --- | --- |
| City | Number of Listings |
| Austin | 12037 |
| Chicago | 7666 |
| Los Angeles | 43763 |
| New York City | 50041 |
| San Diego | 11768 |

In the columns of each listing details table, there are nearly 100 attributes which provide a variety of specifics on each listing including numerical data, categorical data, location data (with longitude & latitude coordinates), and textual data. The column headings are consistent across data tables for all cities. It is also important to note that the listings included in each data set are exclusively within city limits and do not include outlying suburban areas. You would not find information about a Romeoville property listing within the Chicago data set.

When submitting a listing review on Airbnb, the customer is only allowed to submit a review if they have in fact stayed at that listing. Reviewers are able to submit a review score for 6 key categories of the listing: accuracy, cleanliness, check-in, communication, location, and value. They evaluate each category on a scale from 1-5 stars (with half star options available). Based on the categorical scores, an overall review score is calculated. Online, the page for each listing displays the average scores for each category as well as the average overall rating. In the data set at hand, this translates to 7 columns of averaged review details for each individual property listing. The review star system has been converted to a scale of 2-10 for each of the 6 categories. The overall review score is on a scale from 20-100. In the average reviews for each listing, values are rounded to the nearest whole number. So, for the category reviews, this means that there’s 9 distinct options, whereas the overall reviews have a more precise numerical scale.

**Methodology**

I will analyze the Airbnb data using pivot tables and bar charts to get an idea of how percentages and volume of a specific measure may change within certain categories of data in each market. This will be a useful lens to understand not only details of the individual city, but also complete side-by-side analysis of how the cities compare. I will also utilize Tableau box plots, scatterplots, and line graphs in Tableau to look for correlations between different measures. Due to the geographic nature of the data, I will also analyze attributes using symbol maps in Tableau. While these attribute trends might also be discovered by looking at numbers aggregated by neighborhood, I feel that a symbol map would be an appropriate and efficient way to understand how certain data points may change across the city or concentrate in a central downtown area.

Additionally, I will analyze the data with machine learning methodology in Weka. There are multiple availability measures present in the dataset (availability per 30 days, 60 days, 90 days and 365 days). I chose to use availability per 365 days as a key output variable when creating a J48 decision tree model. I felt that this measure is the best overall indicator of availability as it would include a balance of seasonal highs and lows. Availability for a shorter time period would have more volatile variability, and it would be unclear if the timeframe measured was just the seasonal high or seasonal low. On the initial trial of developing a tree model, availability per 90 days appeared at the top of the tree as a key determinant predicting availability per 365 days. I believe this mostly just demonstrates a degree of multicollinearity amongst the availability measures rather than providing new insight on which listing attributes have an impact on availability. Subsequently, I removed the additional availability measures from the data file in Weka. This left me with over 40 variables to examine. The full list of variables included in my Weka analysis is listed on the appendix.

The numerical data type for many attributes as well as the large number of total attributes included were the primary reasons why I chose to complete a J48 decision tree as opposed to association rule mining and clustering methodologies in Weka. Many of the variables are numeric, and with the exception of converting the output variable, I felt it would be more accurate to analyze the data in its original format rather than convert everything to nominal groups for association rule mining. Furthermore, clustering methodologies provide percentage values for every attribute included in the cluster analysis. I felt this type of model would include a lot of noisy data results to sift through with some variables more relevant than others for each cluster. Since it does not include an output variable, a clustering model also wouldn’t assist me in my goal of finding key determinants.

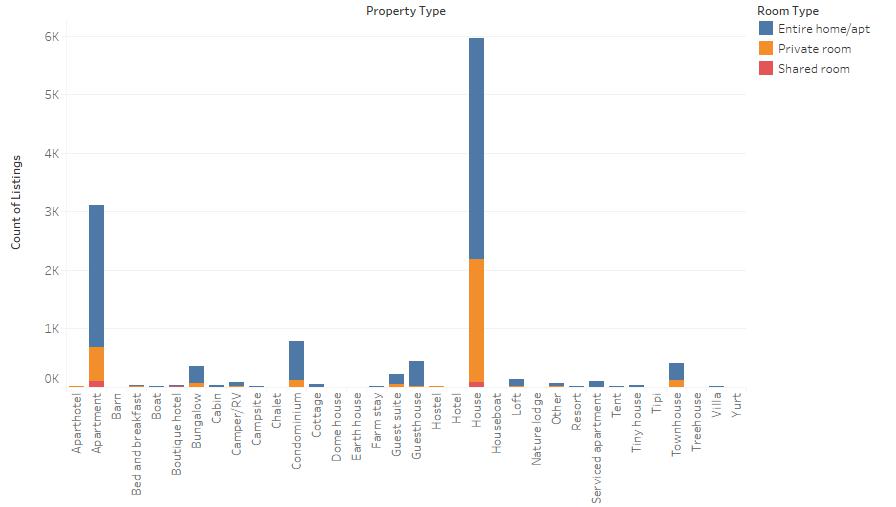
Finally, I will also query the listing data using SQL. I will utilize grouping and aggregate functions to better understand how specific attributes are connected.

**Visualizing Information and Data Mining for Business Decisions**

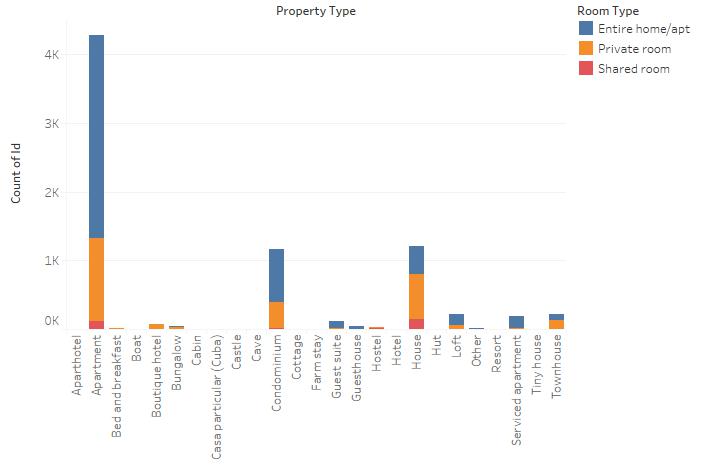
**Variety in Property Types**

Before digging into the results for review and availability data, I feel it would be worthwhile to provide context regarding the types of properties available within each urban market. For each individual city, I grouped the listings by property type to create a bar chart. Within each group, the stacked bars represent the three main room types: entire home/apt, private room, and shared room. In each chart, only the property types that have active listings in that particular city are listed along the x-axis. So, even property types that have seemingly invisible bars on the graph are present in the given data set, just in a very small proportion.

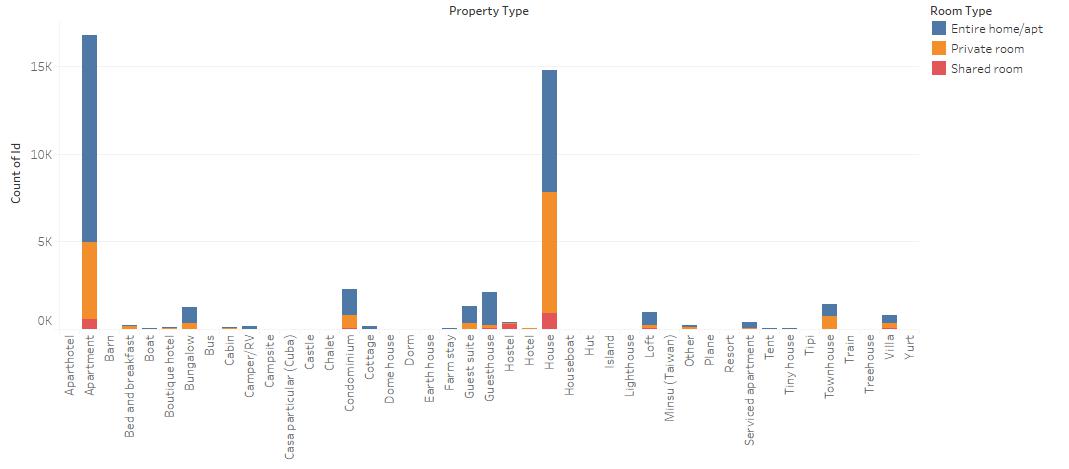
Austin:



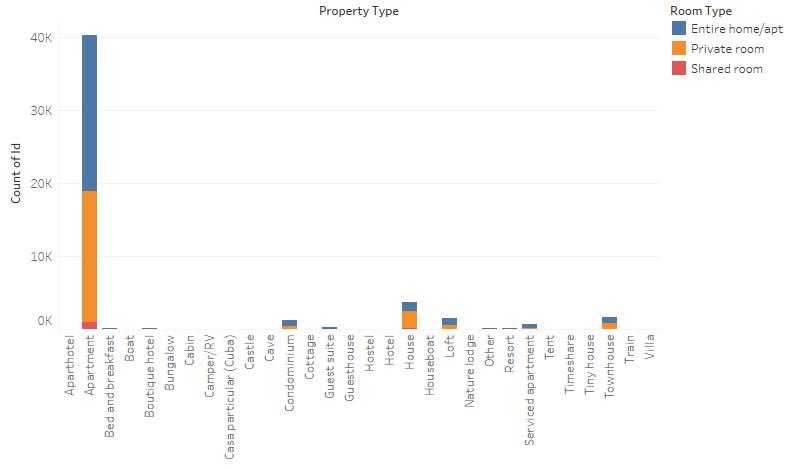
Chicago:



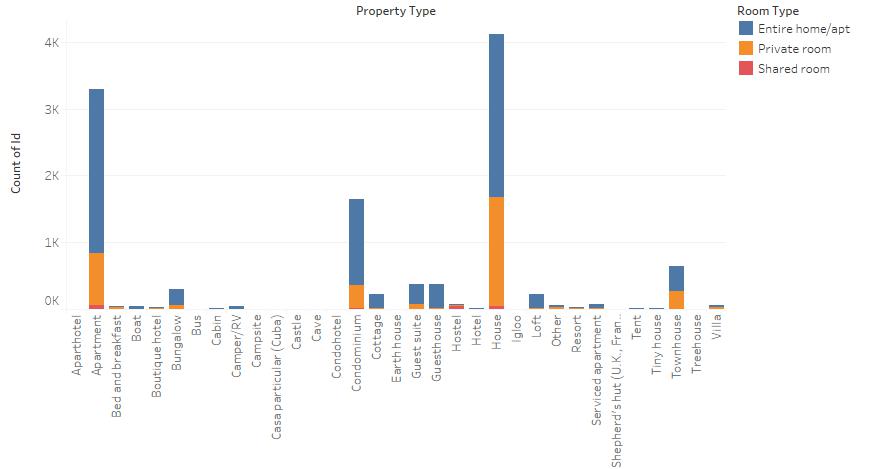
Los Angeles:



New York City:



San Diego:

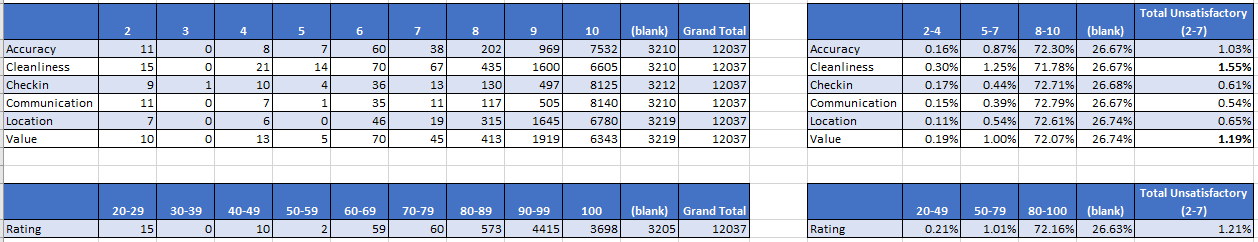


There are some important similarities and distinctions between the cities to explore. One similarity is that shared rooms demonstrate a very small market share of the listings. I was surprised to see that private rooms do make up a substantial portion of listings in each city. Apartments and houses consistently represented the top two most popular listing types, however, their proportions varied drastically by city. Austin was dominated by house listings. Chicago included primarily apartment listings as well as some house listings, and condominiums came at a close third in rank (1162 versus 1206 listings). Los Angeles and San Diego demonstrated similar proportions of apartments and houses, although Los Angeles slightly favored apartments and the market in San Diego slightly favored houses. New York City is unique in that the market is dominated almost exclusively by apartments. Of the 50,041 listings in the city, 40,369 of them (80.67%) are represented by apartments. I think the geographic nature of each city is a clear determining factor in what types of listings are available. New York is an extremely crowded city unlike Austin which has a lot more space. Additionally, the 5 cities also varied noticeably in the scope of properties that they offer. Los Angeles represented most variety with 43 different property types and Chicago was the least diverse with 25 property types. The unique “outdoor” property types that were even present in the cold Northeast climate of New York City (camper/RV, tent, tipi, yurt) were absent in Chicago.

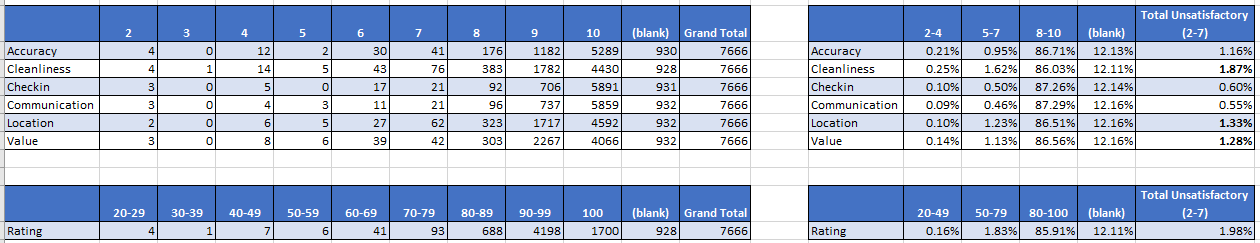
**Trends in Review Ratings by City**

To understand the details of where low reviews are taking place across categories and markets, I created Pivot Tables for each city looking at the review data broken down by categories and rating numbers with the total listing count as the value for each cell. Below this table, I developed a Pivot Table for the overall rating with the scores grouped in bins of 10 so that the data would be more visually comparable with the category data. To the right, I used the listing count in specific cells to calculate aggregate percentages of how certain scores were represented proportionately in the data. To provide perspective on the total percentage of scores with unsatisfactory ratings, I chose to include data from category rating scores 2-7 (scores 20-70 for the overall rating) which would be equivalent to 1-3.5 stars on the website. The percentages of unsatisfactory ratings represent the proportion of listings in each city which on average received scores in that range. The percent is not representative of the total percent of unsatisfied customers.

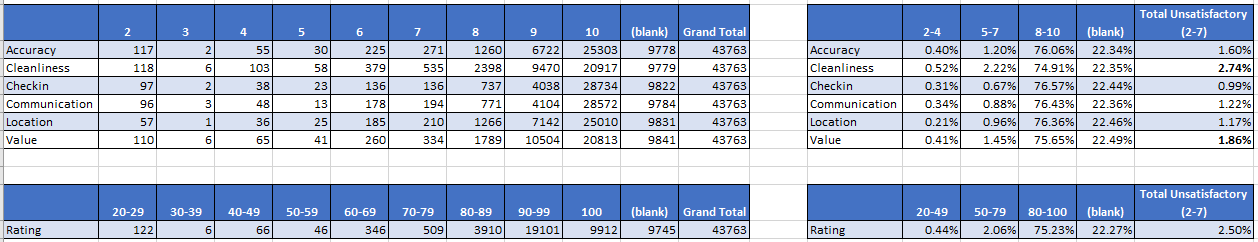
Austin:



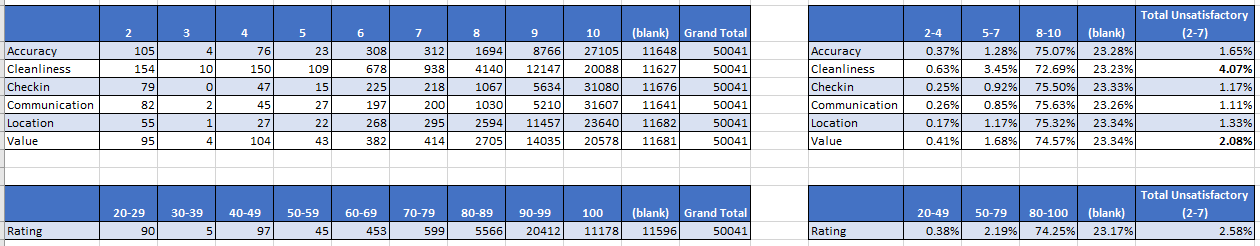
Chicago:



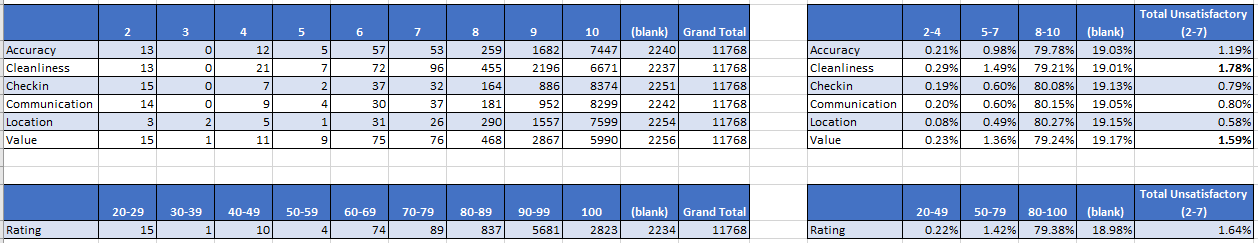
Los Angeles:



New York City:



San Diego:

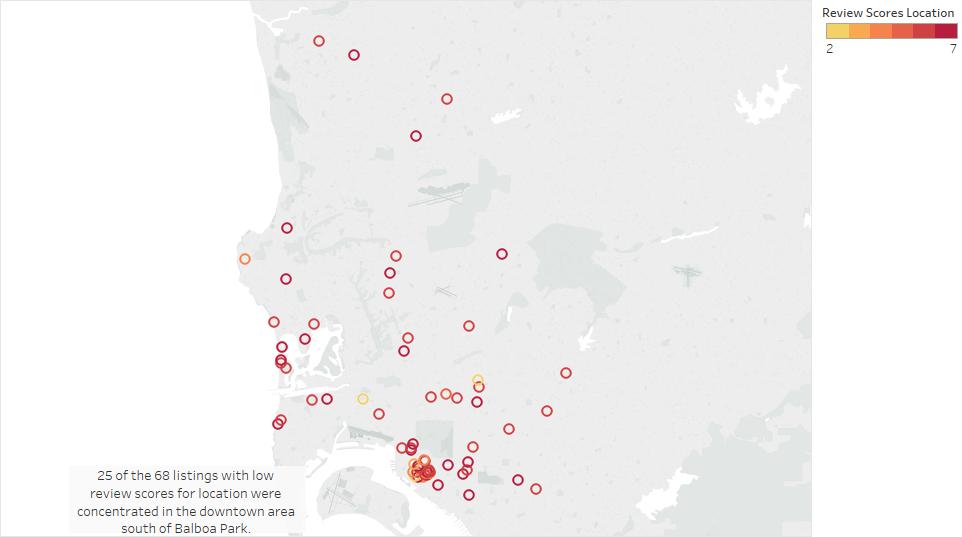


Since the cities being reviewed have a wide range in terms of the overall size of their Airbnb markets, using the raw numbers is not a viable association for comparing reviewers’ satisfaction. Instead, it is best to use the aggregated percentages on the right side of each spreadsheet when comparing one city to another. For the percentage of listings which received satisfactory ratings (average scores in the 8-10 for categorical ratings or scores 80-100 for overall rating), it is promising to see that this included 71.8%-87.3% of listings depending on the category and city. While there’s always room for improvement, this is a great baseline to start with. For the percentages of listings with unsatisfactory customer ratings, cleanliness and value consistently appeared with the highest percentages across cities. These are the two biggest areas of opportunity. Cleanliness was the biggest problem in New York, where 4.1% of listings received unsatisfactory ratings. The only location where another attribute outranked value was Chicago where location was the second highest value with unsatisfactory ratings. In contrast location represented the lowest unsatisfactory rating percentage for properties in San Diego. I was curious to see how the exact locations of these ratings might compare.

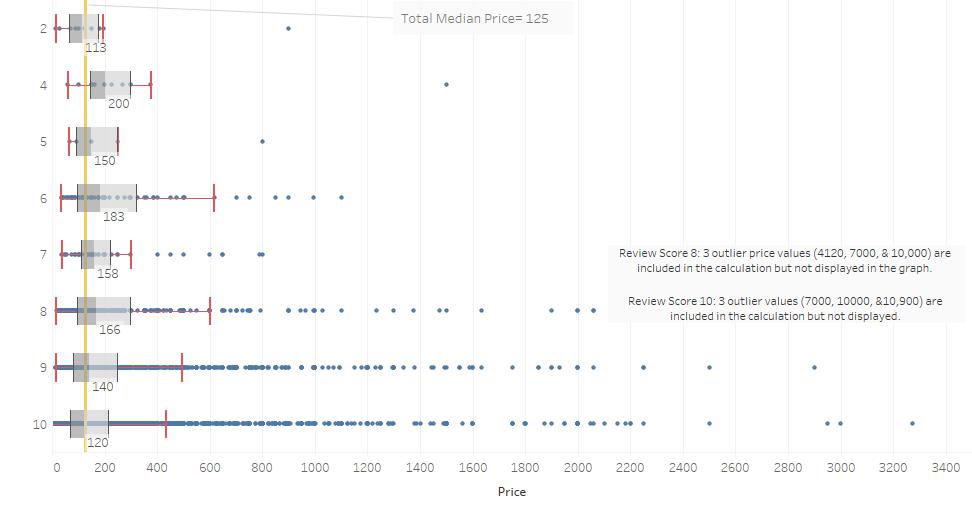
Chicago:



San Diego:

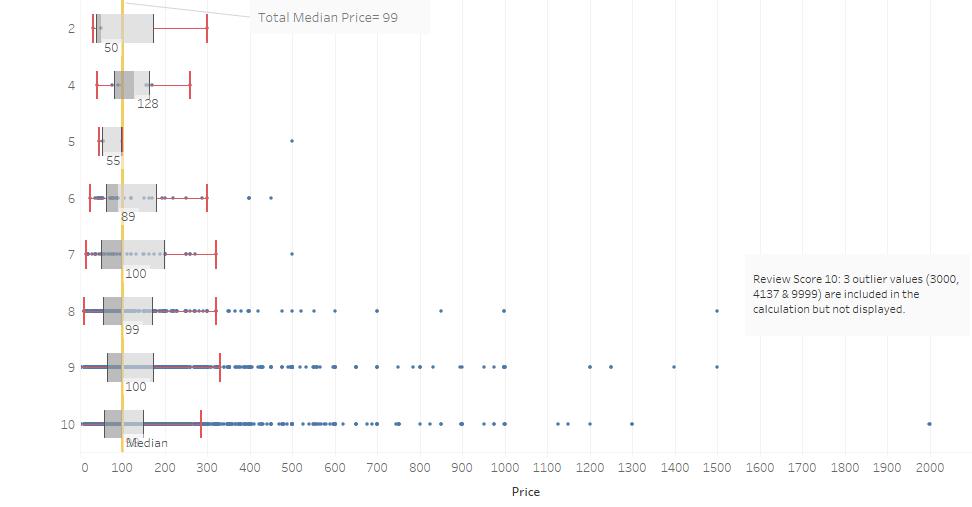


Of the 102 listings with poor location reviews in Chicago, they were primarily clustered on the south and west sides in neighborhoods that are less affluent. In contrast, of the 68 listings with poor reviews in San Diego, the largest cluster was 25 listings concentrated right in the downtown area. This seems to be the ideal location in San Diego, but there must have been something about this area that the selected reviewers did not enjoy. Unlike other categories, improving the location reviews is not something that can be achieved with the host making any physical changes to the property itself (with the exception of moveable properties like boats, RVs, or tents). The neighborhood along with the approximate location of Airbnb listings is available to prospective customers when they view a listing online. For security purposes, the exact street address is only shared with customers after they place a booking. With that in mind, I think the best solution to improving location reviews would be having hosts provide more detailed content about the location and neighborhood of their listings, especially in the Chicago market. This could also include the estimated travel time to reach key city landmarks. While the percentage of listings with unsatisfactory location reviews is already low, such detail would hopefully help customers make more informed decisions and select a listing with a location that meets their needs.

 Returning to another trend in the Pivot Table, improving value ratings is an area of opportunity across all cities. I suspected there may be a correlation between prices and value reviews and sought to explore whether certain properties are receiving low value ratings because their prices are too high. Are luxury listings still seen as a good value to customers? I utilized a Tableau box plot to take a closer look at prices based on value review ratings for value in Chicago and Austin. The Tableau box plot provides context on the aggregate quartile measures while also still providing visibility to the underlying data. In addition to calculating the box plot measures for each review value, I also included a reference line for the overall median of the dataset (marked in yellow). Due to the nature of the data points with a few extremely high outlier prices, I felt that calculating the median would be a better measure of central tendency for understanding this data compared to an average.

Austin:

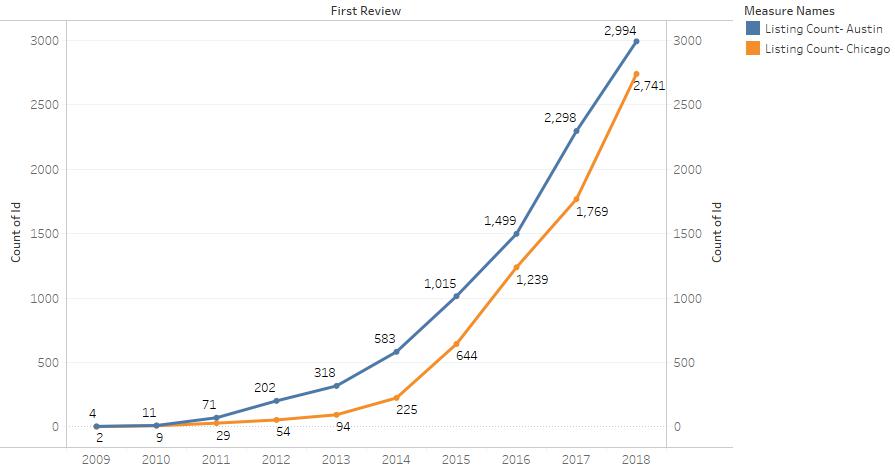
Chicago:



It is clear from the box plots that high prices were not correlated with low value reviews. In fact, the vast majority of the extremely high-priced luxury listings (including the outliers not displayed in the domain) received satisfactory value reviews. Many of the listings which received poor value reviews were on the low to mid-price range. Customers do find value in the higher priced listings as long as the amenities validate that price. This will be important for hosts to keep in mind as they are furnishing their listings and developing their online listing descriptions so that potential customers truly understand all that they have to offer for their price.

While improving location and value reviews may be more complex, improving the cleanliness of listings is straightforward: hosts should more thoroughly clean their property listings, most notably in New York City. To achieve this, hosts should plan ahead for requiring cleaning time this takes and if necessary, adjust check-in and check-out times as well as calendar availability dates to ensure that they have ample cleaning time between different bookings.

Additionally, I must also address the sparsity in the review data. While the proportion of blank reviews varies only by a fraction of a percentage amongst different categories within a city, it is still notable because it shows that some customers filled out partial reviews and provided feedback on only some categories. Depending on the city and category, review data was not available for 12.1%-26.7% of the listings available with Chicago at the low end and Austin at the high end. This made me curious as to whether Austin was a newer market for Airbnb with novel properties that had not been reviewed yet? The reverse was in fact true. The dataset does not directly measure the date that each individual listing became available and how long this property has remained unreviewed. However, looking at the attribute of first reviewed date provides some perspective on the timing of when Airbnb markets grew in each city.

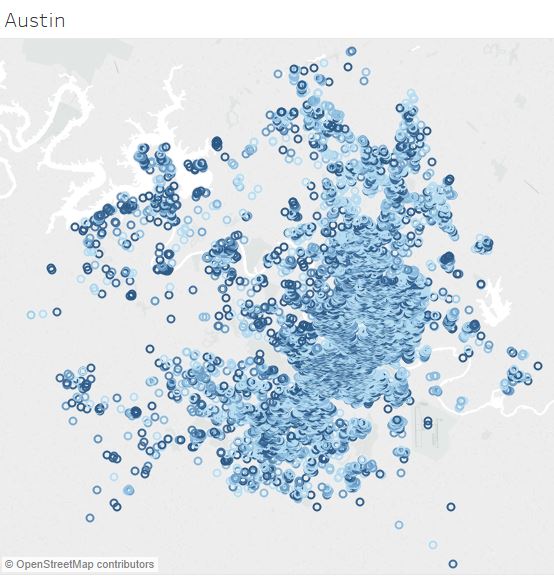
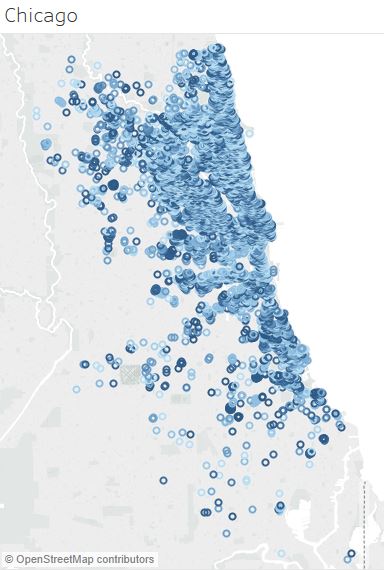


The count of how many listings were reviewed for the first time in each year demonstrates that the Airbnb market in Austin grew sooner and larger than the market in Chicago. A quick look at Airbnb’s history on the “Fast facts” webpage confirms that the annual SXSW conference in Austin (a large media, film, and music festival) was in fact where Airbnb officially launched as a company and made 2 bookings in 2008.

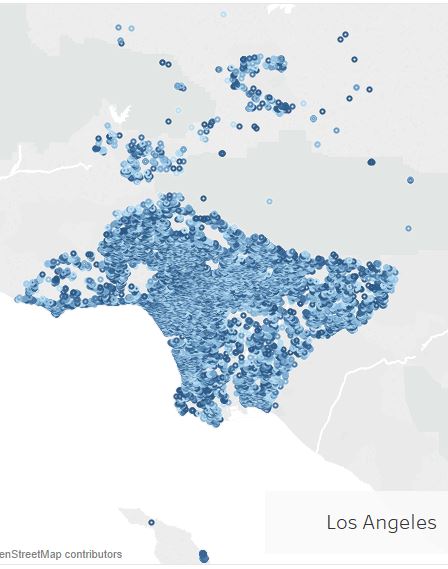
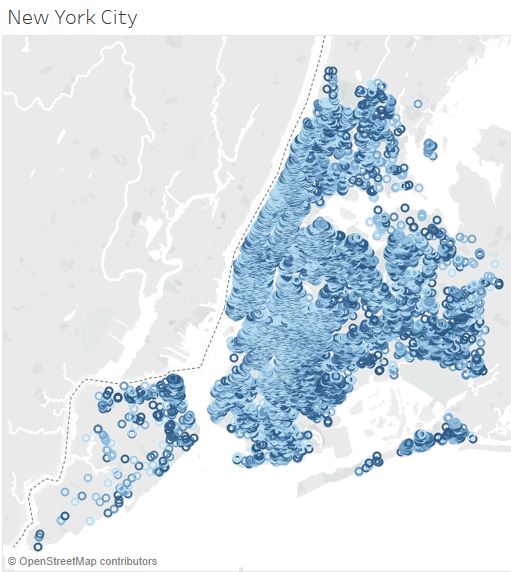
While Austin’s percentage of listings with unsatisfactory reviews was comparable to those in other cities, it’s interesting to see that Austin has the highest percentages of blank listing data as well as the lowest percentage of listings with satisfactory ratings (the opposite was true for Chicago). I would suspect that the non-reviewing customers in Austin could be primarily those that had a satisfactory experience but did not make it a priority to complete the review process. In a variety of business operations, it is often the case that customers with a negative experience are more likely to have something to say and want their voice and complaint to be heard. Satisfied customers that had a pleasant experience but weren’t wowed by a listing may be complacent about following through with a review. So, I would argue that Austin didn’t necessarily have fewer unsatisfactory listings, they more likely just had a higher amount of complacent satisfied customers.

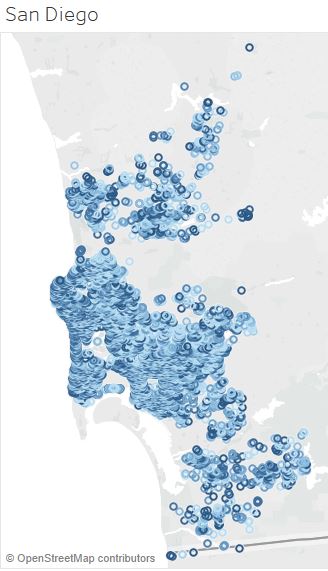
**Understanding Availability**

The availability of specific property listings is an indicator of how frequently it is generating revenue for both the host and Airbnb itself. It is a gauge of which listings are in high demand. Among the availability metrics, I’ve chosen to use availability per 365 days as the most comprehensive measure. To start, I examined listing availability based on location data to determine whether the high demand listings (marked in light blue) were clustered in certain neighborhoods or in the downtown area.

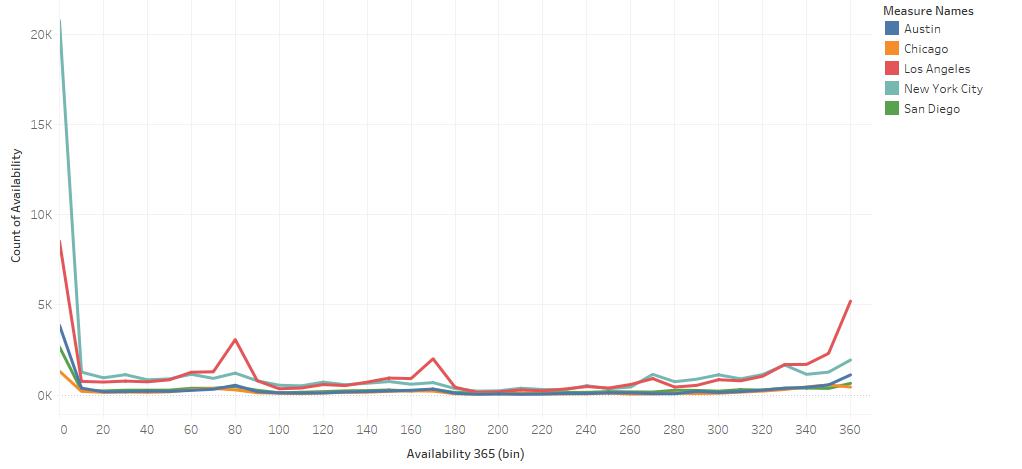
 





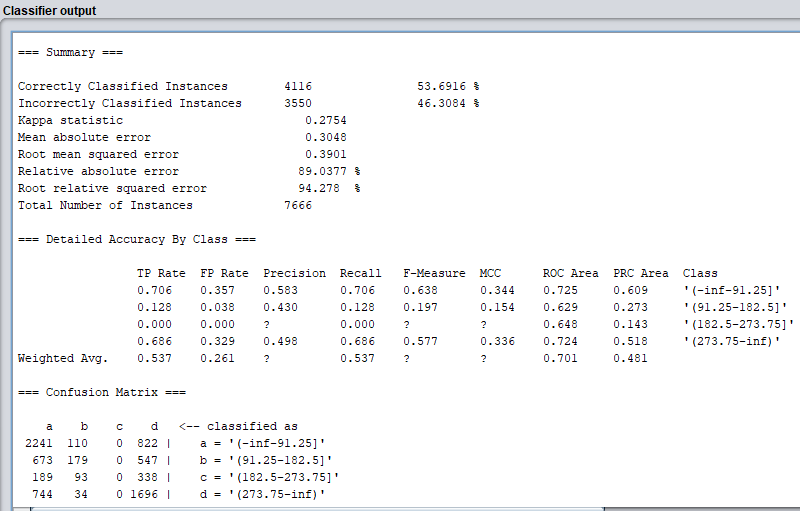
Comparing data across all 5 cities, it is clear that the in-demand listings with low availability are not concentrated in one specific area. There were hundreds of listings even in the outskirts of the city which had low availability. To better understand the proportion of Airbnb listings with low availability within each market, I also developed a histogram line chart with all of the cities to compare the volume of listings associated with each availability value. I used bin sizes of 10 to group the data.

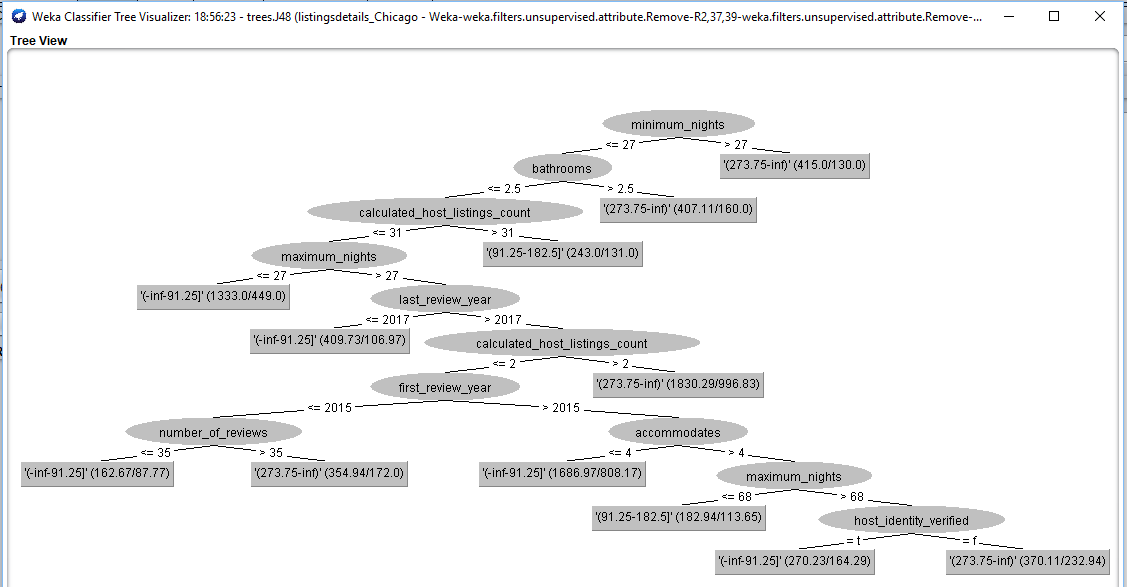


While New York City had the most drastic change from the first bin (availability of 0-9 days) to the second, all cities were similar in that they peaked at the first bin and then the listing count remained low in subsequent bins until a rise in the end with listings having more than 320 days of availability. When listings have low availability, it demonstrates that the resources here are being maximized. When other listings are sitting open for a few hundred days each year, those listings are not being used to their full potential.

**Business Intelligence and Data Analysis**

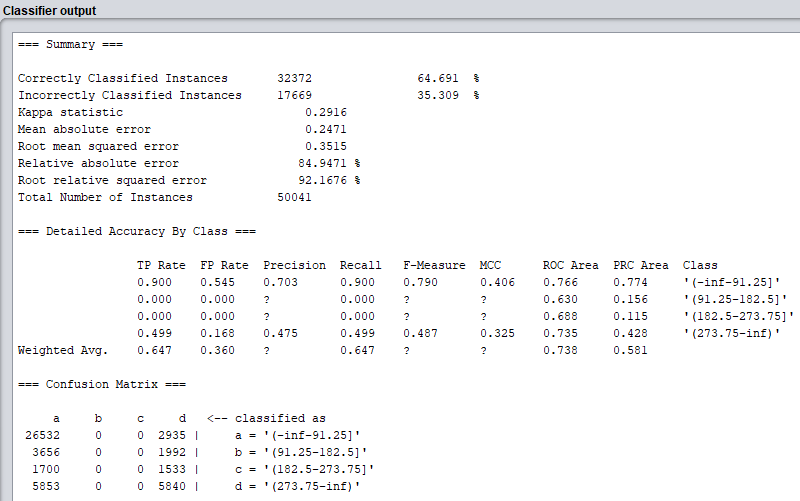
To better understand how specific attributes are associated with listing availability, I analyzed the listing details data in Weka using a J48 decision tree. Starting with Chicago, after making some adjustments to the model, I determined that grouping the output variable (availability per 365 days) into 4 nominal bins of equal width was the most successful model with a higher percentage of correctly classified instances. Applying the discretize filter divided the data into the following bins: 0-91.25, 91.25-182.5, 182.5-273.75, and 273.75-365 days. Looking at the Chicago market, there were 46 potential predictor variables for consideration (see appendix for full list). I updated the minimum number of objects parameter to 80. While models with a lower parameter cutoff had a higher percentage of correctly classified instances, these models included so many branches that the decision tree was not practical and barely even legible. After successfully running the updated model, I obtained the following summary results and tree visualization:

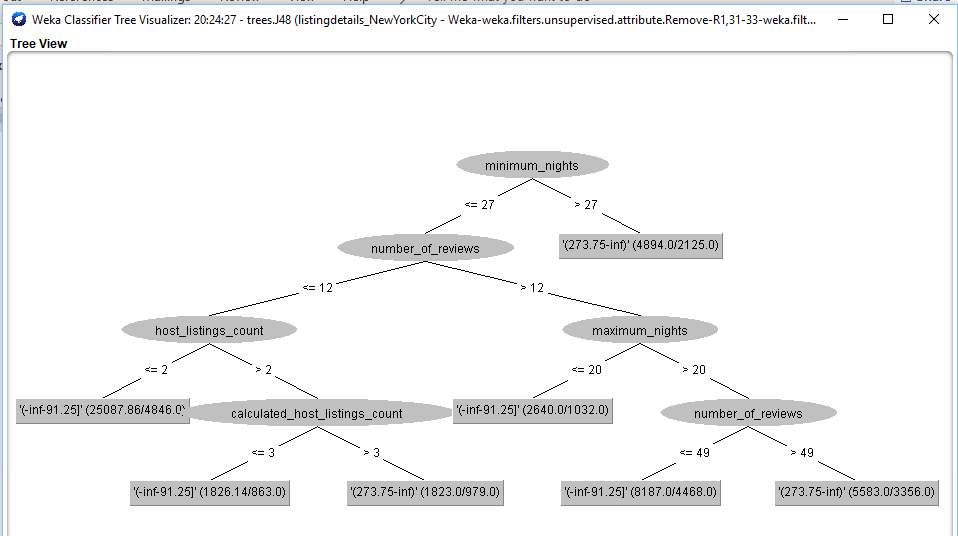




The model correctly classified 53.69% of the instances, so there is some limited applicability to the model, however I do think it would be worthwhile to examine the ranking of determinants and their output variable results. The first determinant in the tree is the minimum number of nights required to book the listing. Those requiring more than 27 nights correlated with the highest availability group. While securing a longer booking could potentially be more lucrative than a few very short bookings, it’s clear than those types of bookings are much harder to secure. Since Airbnb is designed for short term rentals for vacation and business travel, listings with booking requirements of more than 27 nights do not address Airbnb’s primary target audience. Conversely, properties with booking requirements of a maximum of 27 nights correlated with the lowest availability. Additionally, having 3 or more bathrooms also predicted the highest availability for listings (the data for this attribute included only whole and half numbers). Since this is an amenity that would typically only be available in a large home and not an apartment, I would conclude that Chicago Airbnb customers are not typically looking for listings of such a vast size and are instead targeting smaller properties. When the calculated host listing count was greater than 2 and less than 32, this also typically resulted in the highest availability group. Calculated host listing counts of 32 or great led to the second highest availability result. It may be the case that for Chicago listings where the hosts manage numerous other listings, these are not dedicating as much time to the booking of each individual listings.

Furthermore, I also took a look at the availability of New York City listings using a J48 decision tree. For this table of listings, I was unable to include the data for the 3 date-related attributes, so there were 43 potential predictor variables in the model. Once again, I found that grouping the output variable (availability per 365 days) into 4 nominal bins of equal width was the most successful model. For this data set, I updated the minimum number of objects parameter to 1700. This cutoff provided a good balance between the practically and accuracy of the model.





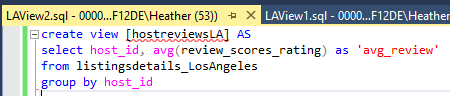
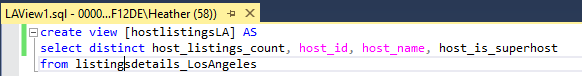
The model here correctly classified 64.69% of the instances, so it was more accurate compared to the predictive accuracy of Chicago’s model. Similar to the decision tree for Chicago, listings requiring bookings with minimum nights of more than 27 had the highest availability. When the calculated host listing count was 3 or less, this correlated with the lowest availability group. I found it surprising that listings with more than 49 reviews were associated with the highest availability group. While the cutoff off number was different in Chicago (greater than 35), both cities demonstrated similar conclusions. I had initially suspected that the listings with a more extensive review history would be more established and have lower availability. However, based on these results, I would presume having a high volume of reviews might increase the chance of having at least one very negative customer review. This could cause future customers to place a booking at that listing.

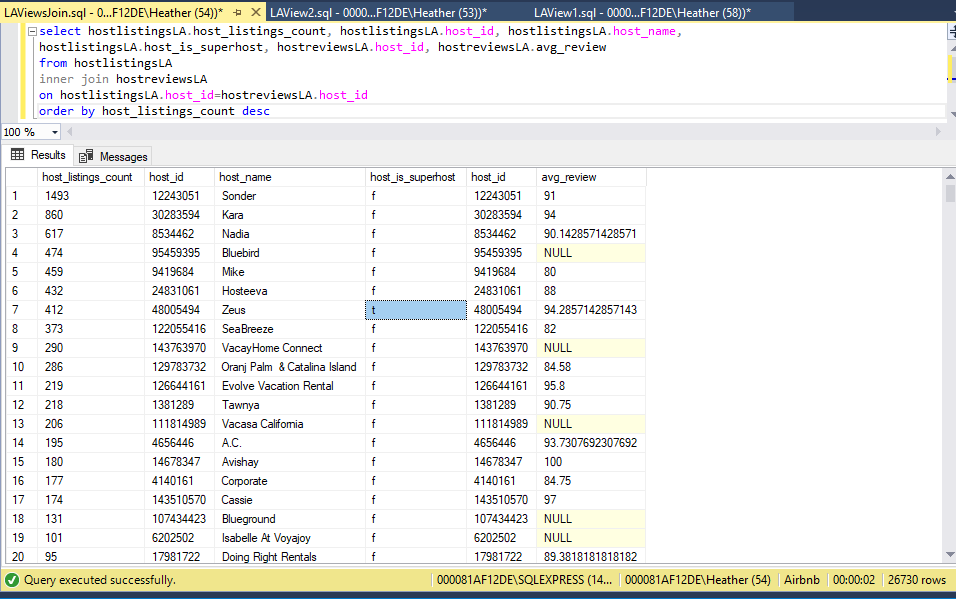
One of the key takeaways from this decision tree analysis is that Airbnb should encourage hosts to maintain a low minimum number of nights requirement for booking when possible. Keeping this minimum low would be beneficial to the success of the host and the company. Additionally, Airbnb should ensure that hosts who manage many properties are still dedicating attention to contact with potential customers at all of their individual bookings. They also need to be mindful of the customer service they provide to customers that actively stay at their listings, to hopefully limit the risk of negative listing reviews and its detriment to future potential customers.

**Database Management**

The analysis of attributes in Weka brought to my attention another area to explore: hosts that have a high listing count. What are the highest listing counts per host in each city? Are these hosts successful at operating an enterprise model? What do hosts with the highest number of listings in each city have in common? I utilized Microsoft SQL Server Management Studio 2017 to query the data for Los Angeles, Austin and San Diego. I wanted to consider the superhost status and average review rating scores for the top hosts in each city. It is important to note that if a host has a total listing count of 1,000 in a particular city, that does not necessarily mean that they operate 1,000 listings in that specific market. They may operate listings in a number of different markets. I used the overall review rating (ranging from 20-100) as a metric for this analysis. Regarding superhost status, this is a special designation for hosts who have met specific performance requirements during the most recent quarterly review period. According to “Superhost: Recognizing the best in hospitality” information available on Airbnb, superhosts are required to maintain at least a 4.8 overall rating out of 5 stars(equivalent to 96 on the 100 point scale used in the dataset), a minimum of 10 stays, 0 cancellations and a 90% response rate. While this status provides incentive perks to the hosts themselves, it is also very beneficial to consumers. Once the superhost badge is earned, it appears automatically on a host’s profile, so it is a clear distinction of exceptional experience and hospitality to potential customers.

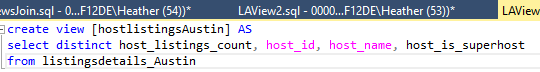
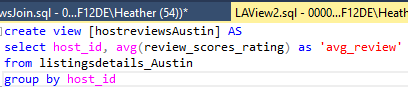
To start, I examined the Lost Angeles data set. I had two tasks to complete: filter the listing counts and superhost status for distinct hosts in the market and calculate average review scores grouped by individual hosts. To achieve both of these tasks in SQL, I had to create two separate views and join them together. The superhost attribute is a Boolean variable, and in the query results, I highlighted the t’s in blue for visual clarity. The top 20 hosts are displayed in the results field:

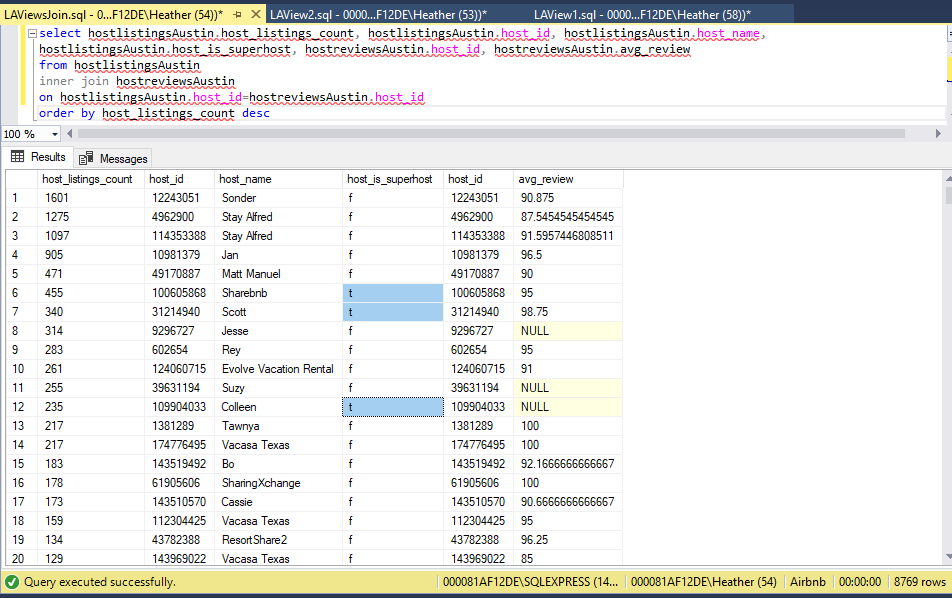




I was surprised to see that the top-ranking spot held by Sonder included almost 1,500 listings. This volume would require a large-scale operation to oversee. Of the top 20 hosts in Los Angeles with the highest listing counts, only 1 of them currently maintains superhost status. While achieving the minimum number of stays metric would be extremely easy for these hosts to accomplish, maintaining the rating, cancellation, and response requirements would be more difficult. While the review data here was somewhat sparse, for all of the hosts for which review data was available, their average review rating was in the satisfactory range of 80-100. In total (as noted by the number of rows in the query results), there are 26,730 distinct hosts represented in the Los Angeles market.

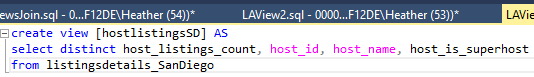
Moving on to Austin, I applied the same query structure of joining views for this data set.

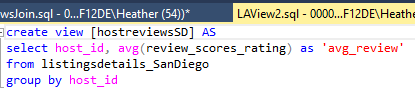


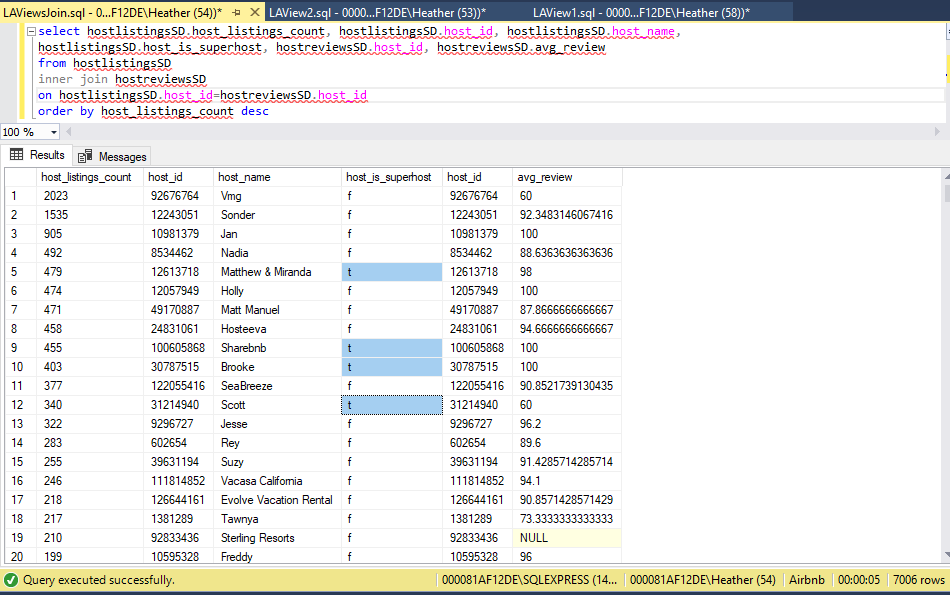


Although Austin is a much smaller overall market than Los Angeles, the highest host listing count here surpassed the top spot for Los Angeles. The top 20 group in Austin included 3 superhosts and a similar satisfactory range of scores. There is a total of 8,769 hosts operating listings in Austin.

When applying the query structure to San Diego, I achieved the following results:







While Austin and San Diego represent Airbnb markets of similar scale (12037 and 11768 total listings respectively), the top host listing count in San Diego was significantly higher. Looking at the progression of high listing counts for San Diego, it is interesting to see that 19 hosts in the San Diego market have a listing count of more than 200. There are a few superhosts in the San Diego group, but unlike Austin and Los Angeles, there are 3 hosts in this subset which received average overall ratings which were unsatisfactory (20-79). There are 7,006 individual hosts who operate listings in San Diego.

Comparing the data for all 3 cities, it is apparent that Austin and San Diego represent more concentrated markets for hosts with high listing counts. I would presume that in a much larger market like Los Angeles, there is also more competition. It would be more difficult for a host to achieve such a high volume of listings. Additionally, the biggest hosts are not necessarily the best. Only a few of them carry the superhost designation for superior hospitality. Where review data is available, most of the hosts in this subgroup do receive, on average, satisfactory review ratings, but there were 3 hosts in San Diego with average review scores which were unsatisfactory. I would recommend that Airbnb should take a closer look at these 3 “low-performing” hosts with on average unsatisfactory reviews. Based on their scale, they could potentially be accounting for a few hundred unsatisfied customers in San Diego.

**Business Data Warehousing**

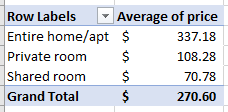
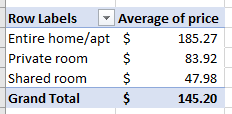
Considering a bigger picture than querying individual columns of data, it’s important to focus on how Airbnb should be structuring their data at the enterprise level. When data is simply just housed in a transactional (OLTP) system, this can support day-to-day operations. For example, this may be suitable if documenting check-ins at a host property. However, the listing details dataset from Inside Airbnb is more demonstrative of a data warehouse because it is more readily suited to provide strategic information. Available aggregates like the average reviews for each listing allows for searching for trends and clustering in the data. While the dataset utilized here is based on public online information, in the internal systems for Airbnb, it would also be important to have drill down capability on review data. While it would not be practical to insert all of the listing data updates and reviews in the business data warehouse in real time, the warehouse should allow for regular updates. Systematic updates are extremely important and already in use by Airbnb in tasks such as automated quarterly reviews of superhost standing.

From the lens of data warehousing, it is also worthwhile to address data quality. In this dataset, the information was compiled through text mining from a webpage. Specifically, with the New York City data, there were a few date attributes where the corresponding data did not fit that format. There were also some abnormal characters scattered throughout the table which had to be cleaned out for the data to be permissible for running in SQL. Internally at Airbnb, the data should feed from the user interface directly into an internal system rather than needing completing text mining of the external site. As noted by Ponniah (2010), two of the important ways in which improved data quality can benefit an organization include better customer services and reliable strategic decision making. In that way, improved data quality would benefit not only Airbnb themselves but also members of the Airbnb community.

**Operations**

Beyond how they operate their technical structures, Airbnb’s business operational strategies for competing in the marketplace also merit consideration. According to Heizer and Render (2014), there are three main ways that companies can achieve competitive advantage through operations: competing on differentiation, cost and response. Airbnb actually demonstrates all of these in at least one specific area of their business model. By providing unique property types such as yurts (a Central-Asian tent), castles, and tinyhouses, Airbnb is able to distinguish themselves from traditional hotel settings. On their site, they provide searchable filters for customers looking for specific property types as well as feature articles with detailed content on unique listings and allow customers to explore and discover new properties. Additionally, Airbnb is able to compete against traditional hotels on cost by distributing their property operating expenses to the hosts. In this way, Airbnb is able to avoid taking on the massive overhead expenses (amenities, staffing, etc.) that are incurred at a large hotel operation. Airbnb also appeals to budget-friendly travelers by offering varying degrees of privacy that come to customers at different price points.

Austin: Chicago:

As is evident for Austin and Chicago, there is strong clustering for the average pricing of each room type. Finally, Airbnb competes on response by tracking and publicizing their host response metrics of response time and response rate. Response time designated by 1 of 4 categories: a few days or more, within a day, within a few hours, and within an hour. By making these metrics public, Airbnb holds hosts accountable for providing timely customer service.

**Quantitative Methods for Business**

As listing prices determine revenue and directly impact Airbnb’s bottom line, it’s important understand how different property types and sizes may be susceptible to revenue fluctuations. To better understand these fluctuations, I took a deep dive look at 3 listings with the same property type (entire home/apt) and with similar nightly prices in the Chicago market. I wanted to analyze how properties of different accommodation sizes might compare in nightly total revenue as well as annual revenue. I selected three listings with a nightly price close to the average price in Chicago of $145.20. The discussion here will be limited to a revenue comparison as opposed to a profit comparison. In the scope of this dataset, there is not information available for the operating costs directly associated with maintaining an Airbnb listing (i.e. local licensing fees, labor hours for cleaning between guests) as well as the general costs of maintaining a property (i.e. utilities). For the nightly maximum revenue, I considered how much the listing would generate if a 1-night stay is booked for the maximum number of people that the property will accommodate. For this analysis to have practical application, it was necessary to select listings with minimum nights equal to 1. I selected the following records and attributes for analysis:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Id | Property Info | Accommodates | Nightly  Price | Cleaning Fee | Guests Included | Extra People | Availability 365 |
| 2238443 | Near North Side apartment | 4 | $150 | $80 | 2 | $75 | 15 |
| 3088192 | Logan Square apartment | 8 | $145 | $35 | 2 | $10 | 118 |
| 26669637 | West Town condo | 14 | $149 | $99 | 4 | $100 | 254 |

While many operations may have fixed and variable factors associated with the cost, these Airbnb listings have fixed and variable parts to the revenue. Based on this information, the revenue for a 1-night booking could be calculated as:

Nightly revenue = (nightly price) + (cleaning fee) + (extra people charge) x (# of extra people)

*fixed revenue variable revenue*

The fixed revenue will be generated based on any 1-night booking, but the variable revenue will depend on how many additional people are included in the booking beyond what is included in the standard nightly rate. The maximum nightly revenue for each listing would be as follows:

Near North Side apartment = $150 + $80 + $75(2) = $380

Logan Square apartment = $145 + $35 + $10(6) = $240

West Town condo = $149 +$99 +$100(10) = $1248

Depending on the rate charged for extra people, the listings which are able to host a larger number of people would have the biggest opportunity for nightly revenue. However, the annual revenue depends on not only the potential maximum nightly revenue but also the demand of the list (reverse of number of nights listing is available per year.

Near North Side apartment = $380 (365 – 15) = $133,000 annual potential

Logan Square apartment = $240 (365 – 118) = $59,280 annual potential

West Town condo = $1248 (365 – 254) = $**138,528** annual potential

While the large West Town condo had a much higher nightly revenue, due to limited demand for properties that accommodate so many people, the annual potential was actually rather similar to the small Near North Side apartment which can only accommodate a maximum of four guests. With this analysis, it is important to understand that while very large listings may initially seem more lucrative for revenue, demand must also be considered must also be considered to establish which properties would contribute most to the bottom line of Airbnb and its hosts.

**Legal and Ethical Issues**

The reviewing system on Airbnb is unique in that it allows both the host and the guest to review each other after the guest checks out. Under the original review system, reviews by either party would become public as they were submitted. As a result, the guest or host may be more inclined to give a positive or negative review in response to reading the other party’s review of them. In 2014, they launched a new system which Porges (2014) described as a double-blind submission. Regardless of who submits first, the host and the guest cannot read the other party’s review until they both have submitted. There is a 2-week window after checkout for review submission. After that, new reviews can no longer be submitted, and a pending review becomes public. Even after the reviews go live, hosts (as well as future potential customers) are only able to see the text commentary for the guest and not the individual star ratings that they provided. For hosts to remain in good standing, they need to maintain a certain average star rating. Not posting the individual star rating numbers limits the host from retaliating with negative commentary in response to poor rating numbers from an individual. This new system maintains privacy for star ratings provided by customers and promotes ethical in integrity of the reviews.

Additionally, I would be remiss if I did not address the legal pushback that Airbnb is facing from city legislation different parts of the United States. One example of a potential legal roadblock is the short-term rental ordinance that was approved Austin’s City Council. The ordinance was approved in 2016 and the rollout intends to phase out Type 2 short-term rentals by 2022. Widner (2018) explains these listings as “those in residentially zoned areas that are not operated by a homeowner who lives onsite.” A lawsuit was filed against this ordinance in 2016 which ultimately failed, but another lawsuit has been filed in an effort to overturn this local Austin law. In some other cities, apartment renters sometimes face legal consequences for being hosts and subleasing their apartment through Airbnb because it violates their lease agreements. However, as the Airbnb market in Austin consists primarily of homes, it’s a different scenario because of the hosts in questions are often homeowners which comes with rights for private property ownership. The basis for the second lawsuit is that the new regulation is an unconstitutional violation of homeowners’ private property rights. The legal litigation is ongoing, but if Austin City Council succeeds in implementing this regulation, it could have a major impact on the scale and profitability for Airbnb listings in Austin.

**Limitations**

While the pricing information included here is generally applicable for providing a scale of comparing properties within each city to one another, it is not indicative of the average price one would pay in a particular city on all 365 days of the year. Just like other hotels and airline companies, Airbnb property listings typically go up in price when the area is expecting an unusually high demand. For example, a quick average calculation of the Austin data set indicates that the average nightly price in the city is $270.60, but this number would be extensively higher for a night on New Year’s Eve or during the annual SXSW conference in March.

To address periods of high demand and help hosts maintain competitive pricing, Airbnb created a feature called smart pricing. Launched in November 2015, the smart pricing algorithm allows hosts to automatically adjust their prices based on seasonality, demand, and special local events. Hosts simply need to establish maximum and minimum price parameters and enable the smart pricing option. The Airbnb article “The Price is Right: How we used host feedback to build personalized pricing tools” describes how they developed this model to support a challenge that hosts were facing. Prior to this feature’s introduction, hosts would have to make manual individual adjustments if they wanted to raise or lower their prices on specific days. While smart pricing is certainly a beneficial feature, it reveals a limitation in the data set that these fluctuations are not addressed as it only indicates single values for nightly, weekly, and monthly prices rather than a range.

**Conclusions and Recommendations**

All in all, Airbnb has already demonstrated ways in which they’ve taken initiative to achieve success. To further improve review scores, Airbnb needs to ensure the categories of cleanliness and value are addressed. Hosts need to make sure that listings are clean and that even low-priced listings find a way to bring value to the customer. As context for property availability, it’s important to keep in mind that all measured markets have a spike with many properties only being available less than 10 days in a year. The attributes associated with high-demand low-availability listings include low minimum required nights for each booking and low to mid-sized host listing counts. Airbnb must be wary of hosts with a high volume of listings and ensure they’re still providing quality customer service even on a large scale, especially in San Diego. Especially in mid-size markets, they need to ensure that a monopoly is not being established. Additionally, hosts must also be mindful that any poor customer reviews could deter future customers. Finally, as hosts establish their pricing (manual parameters and smart pricing settings), it’s important that they understand how their choices could impact revenue.

**References**

Earn money as an Airbnb host. *Airbnb*. Retrieved from [https://www.airbnb.com/host/homes](about:blank)

Fast facts. *Airbnb*. Retrieved from [https://press.airbnb.com/en-uk/fast-facts/](about:blank)

Get the data. *Inside Airbnb*. Retrieved from [http://insideairbnb.com/get-the-data.html](about:blank)

Hartmans, A. (2017). Airbnb now has more listings worldwide than the top five hotel brands combined. *Business Insider*. Retrieved from [https://www.businessinsider.com/airbnb-total-worldwide-listings-2017-8](about:blank)

Heizer, J. & Render, B. (2014). *Operations Management: Sustainability and Supply Chain Management* (11th ed.). Pearson.

Ponniah, P. (2010). *Data Warehousing Fundamentals for IT Professionals* (2nd ed.). John Wiley & Sons Inc.

Porges, S. (2014). The Strange Game Theory Of Airbnb Reviews. *Forbes*. Retrieved from [https://www.forbes.com/sites/sethporges/2014/10/17/the-strange-game-theory-of-airbnb-reviews/#90187d536ef9](about:blank#90187d536ef9)

The Price is Right: How we used host feedback to build personalized pricing tools. *Airbnb*. Retrieved from [https://airbnb.design/smart-pricing-how-we-used-host-feedback-to-build-personalized-tools/](about:blank)

Superhost: Recognizing the best in hospitality. *Airbnb*. Retrieved from [https://www.airbnb.com/superhost](about:blank)

US Census Bureau (2018). Retrieved from [https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk](about:blank)

Widner, C. (2018). State challenges Austin’s short-term rental laws. *Curbed Austin*. Retrieved from [https://austin.curbed.com/2018/4/2/17189530/austin-short-term-rental-laws-suit](about:blank)

**Appendix**

Here are the attributes included when I analyzed the data in Weka:

1. host\_since\_year\*
2. host\_location
3. host\_response\_time
4. host\_response\_rate
5. host\_is\_superhost
6. host\_neighbourhood
7. host\_listings\_count
8. host\_has\_profile\_pic
9. host\_identity\_verified
10. neighbourhood\_cleansed
11. is\_location\_exact
12. property\_type
13. room\_type
14. accommodates
15. bathrooms
16. bedrooms
17. beds
18. bed\_type
19. square\_feet
20. price
21. weekly\_price
22. monthly\_price
23. security\_deposit
24. cleaning\_fee
25. guests\_included
26. extra\_people
27. minimum\_nights
28. maximum\_nights
29. calendar\_updated
30. availability\_365
31. number\_of\_reviews
32. first\_review\_year\*
33. last\_review\_year\*
34. review\_scores\_rating
35. review\_scores\_accuracy
36. review\_scores\_cleanliness
37. review\_scores\_checkin
38. review\_scores\_communication
39. review\_scores\_location
40. review\_scores\_value
41. requires\_license
42. instant\_bookable
43. cancellation\_policy
44. require\_guest\_profile\_picture
45. require\_guest\_phone\_verification
46. calculated\_host\_listings\_count
47. reviews\_per\_month

\* Attributes not included when analyzing New York City data set due to errors in these data columns.