AirBnB Boston and the "Sharing" Economy

MIDS--INFO--W18

April 27,2017

Robert Deng, Blake Williamson and Collin Reinking

Context

The rise of "sharing economy" business models like Uber, GrubHub and AirBnB have confronted citizens and lawmakers with new questions. Our project seeks to investigate the space that AirBnB has created, it's relation to the sharing economy and to better characterize it.

Focusing Questions

1. Is AirBnB really an example of the "sharing economy"?
2. How do the the listings that "sharers" post compare to those that "businesses" post?
3. What features of a listing create value on AirBnB?

Source Data

Our data was scraped off the airBnB website by a private citizen who sought to facilitate the investigation of how airbnb is affecting neighborhoods and the hotel industry, it can be found at www.insideairbnb.com.

- http://data.insideairbnb.com/united-states/ma/boston/2016-09-07/visualisations/listings.csv
    - 95 columns x 3,858 rows
    - Contains the data that is in a JSON inside of the html for each listing on airbnb.
    - The JSON contains the data used to generate individual listings.
    - The 95 columns are the different attributes of the listing including information about the host, listing location, amenities listed by the host, and reviews.

- http://data.insideairbnb.com/united-states/ma/boston/2016-09-07/data/calendar.csv.gz
    - 4 columns x 1.3 million rows
    - Scraped on 9/6/16, with a year looking forward to 9/6/17.
    - One row for each day for each listings (365 rows per listing).
    - Each row represents whether a listing is available to rent on a specific date.
    - If the listing is available, the price is given given for that date.

Initial Exploration and Data Preparation

Given the magnitude of data fields available we encountered some initial challenges with understanding our data and dealing with pre-conceived assumptions. Additionally, we discovered ~21% of listings were missing data for various fields. Subsequent calculations in PANDAS ignores rows with null values hence these listings were not factored into our analysis.

After a painstaking process of evaluating the 95 variables in the listings data, we generated a list of 29 variables (see Data_Read_Clean_Augment_Write notebook) that we wanted to use to investigate a new set of focused questions, many of them arising out of surprising findings from the listings.csv data. We were astonished to a considerable number of hosts with more than 10 listings and one host with more than 100.

Our data set provided some real challenges for cleaning which can be explored fully in our Data_Read_Clean_Augment_Write notebook, but one highlight was was the challenge of filling in missing zip code data. In the end we were able to apply a lambda function that called a function that we we wrote that made use of the geopy package in order to use longitude and latitude date to look up zip codes.

One of the major challenges of our calendar data is that we only had pricing data for days when listings were “available” (and not reserved for rental). In order to approximate the revenue generated by listings occupied for certain days we filled in the null values in the price column with a mean transformation of the price column of the data, grouped by the listing_id and the day of the week. We then created a new column that represented the daily revenue by containing 0.0 if the listing was “available” and the value from the price column if it was not occupied.

Because our data was forward looking we recognized that only dates in the near future had a reasonable chance of being reserved. To limit our analysis to that near future time we created, from the calendar data, quarterly revenue and occupancy columns that summed the values for each listing in 90-day groups. Unless otherwise stated, we will use quarter one data for future projections of revenue.

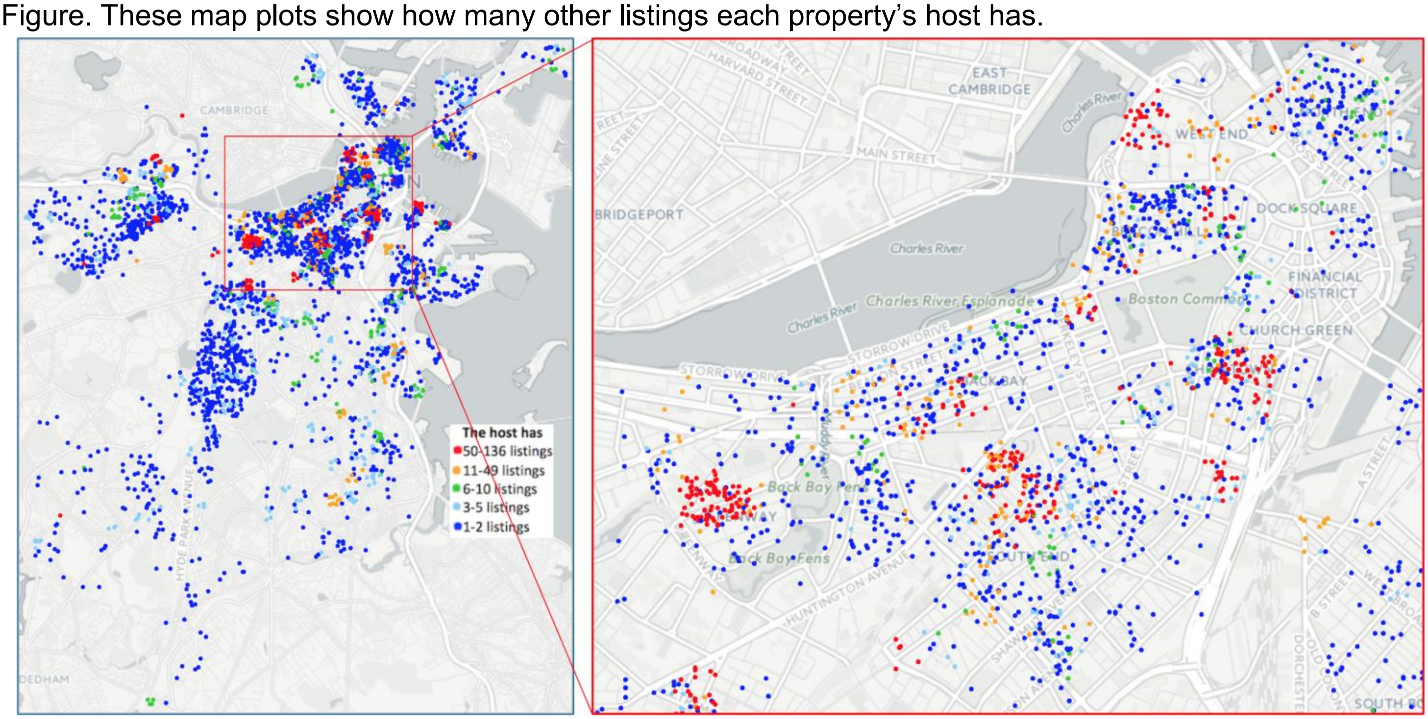
Finally we created dummy variables from our data regarding what amenities were included in each listing. The amenities “lists” were originally strings that had to be cleaned using a regular expression and then split. We then created a list from the set of all amenities in all listings. For each amenity in that list we created a dummy variable that would be “True” for every listing that contained that amenity and we eliminated one variable from each set of dummy variables since n-1 variables are included in regression analysis. That data datable was then split by row with each row finally occupying a single column of the listings data frame.

Exploratory Questions

Is AirBnB an example of the “Sharing Economy”?

If AirBnB is an example of people “sharing” their extra space for a small fee then we would expect to see that most “hosts” (people create listings on AirBnB) would only have one or two listings. Perhaps some hosts end up with a second property through marriage, inheritance or work arrangements.

Figure. These map plots show how many other listings each property’s host has.



An interactive version of this map can be found in our project folder: Listings_Colored_by_Host_Listing_Count.html

In our data set, only 62% of the 3,585 listings are listed by hosts with 2 or fewer listings. Thus 38% are what we might consider outside of the traditional “sharing” economy. Interestingly enough, 11% of all listings belong to hosts with 50 or more properties in Boston. These are no mom and pop operations!

So who are these high volume hosts? Meet “Kara,” she is our highest lister with 136 properties. After reaching out to AirBnB to comment on this we simply heard back that the appropriate term is that Kara may be a ”serviced apartments provider”. Next in line meet sounder.com, a company whose model is to provide “hotel quality service.” Other high listing volume hosts include property management companies and self described “real estate investors.

What distinguishes “businesses” from “sharers” on AirBnB?

Within the AirBnB ecosystem, we opted to segment our users for further analysis:

- Business Audience: 3+ listings owned by one host ID (38% of all listings)
- Sharer Audience: 2 or less listings owned by one host ID (62% of all listings)

The following analysis is a deeper dive into these audiences and their preferences. Businesses are looking to accommodate volume and the masses- accommodating on average 3.3 people per listing (14% more compared to casuals). The trend continues with ancillary offerings as well as businesses offer 15% more bathrooms, 6% bedrooms and 8% more beds respectively.

In return, prices are also parallelly adjusted, with businesses charging $186 on average per nightly listing, 10% more than casuals at $166. Professionals also have a higher minimum nights required at 3.8 vs. 2.5 nights on average, potentially helping boost their occupancy rates. Cleaning fees are no joke either with professionals charging *twice* what casuals charge - $67 vs $35. Later analysis suggests that this is a shrewd move.

Despite the lower average number of reviews per listing for casuals, surprisingly, the overall review score rating for casuals outperformed professionals (93 v 90). Casuals consistently receive fewer, but better reviews than professionals. This further sheds light on the contrasts between the sharing economy and professional experiences as casual experiences may be received better, or perhaps higher standards exist for businesses.

Average review score per audience

user_type	Businesses	Sharers
number_of_reviews	21.13	17.78
review_scores_rating	89.67	93.25
review_scores_accuracy	9.19	9.57
review_scores_cleanliness	9.17	9.31
review_scores_checkin	9.44	9.77
review_scores_communication	9.43	9.78
review_scores_location	9.27	9.5
review_scores_value	8.9	9.33

In terms of how a user books a listing, sharers appear to be very cautious and do not offer as much instant bookability as professionals do (84% of sharers do not offer instant booking). For good reason, sharers likely want to vet the traveling user if they’re staying in their personal home. Professionals are slightly more inclined to simplify the booking process as they value customer acquisition and higher occupancy rates.

Cancellation policies follow similar strand of thought - once professionals have you booked, they want to keep you as nearly 7 in 10 businesses are strict / super-strict. A cancelled listing would mean potentially lost revenue and lost time to advertise the listing for their *primary source of income*. In contrast, 67% of sharers are either flexible or moderate on cancellations, given this isn’t their primary source of income.

Attributes, % of Audience

user_type	Businesses	Sharers
user_count	1351	2234
instant_bookable_f	78%	86%
instant_bookable_t	22%	14%
cancel_pol_flexible	11%	38%
cancel_pol_moderate	20%	29%
cancel_pol_strict	63%	33%
cancel_pol_super_strict_30	6%	0%

However, there is a tradeoff between the two experiences in room type. Most casual sharers prefer to rent out specific private rooms, emphasizing the shared portion in the "shared economy space." 41% of these casual listers rent out private rooms, compared to 33% for the professional crowd. 66% of the professional listers will however give you the entire home or apartment.

Room Type, % of Audience

user_type	Businesses	Sharers	Total Listers
user_count	1351	2234	3585
Room_type_Entire_home/apt	66%	56%	2127
Room_type_Private_room	33%	41%	1378
Room_type_Shared_room	1%	3%	80

In terms of amenities, both audiences appear to provide the comfortable living essentials: internet, heat, kitchen, washer/dryer, and TV.

Amenities by Audience, % of Audience > 70%

user_type	Businesses	Sharers
user_count	1351	2234
Internet	99%	95%
Wireless_Internet	97%	94%
Heating	96%	94%
Kitchen	92%	91%
Dryer	86%	82%
Air_Conditioning	85%	75%
Essentials	82%	85%
TV	81%	73%
Washer	70%	68%

To identify specific audience attributes, Professionals appear to offer the luxury building experience, with more people providing cable TV, gyms, family-friendliness, elevators, 24-HR check-in services, irons, doormen, and pools.

user_type	Businesses	Sharers	Total Listers	% Difference (Businesses)
user_count	1351	2234	3585	-
Cable_TV	64%	41%	1770	23%
Gym	23%	7%	479	16%
Family/Kid_Friendly	61%	47%	1888	14%
Elevator_in_Building	32%	19%	868	13%
24-Hour_Check-in	42%	31%	1247	11%
Iron	56%	48%	1828	8%
Doorman	12%	5%	268	7%
Pool	9%	2%	160	7%

In contrast, sharers tend to provide a slightly more safer environment with items such as: smoke detectors, first aid kits, buzzers, fire extinguishers, and carbon monoxide detectors. Parking on premise skews 8% higher for sharers and breakfast, 4%; all which contribute to the sharing economy sentiment of “mi casa, su casa.”

Regarding pets, our original hypothesis was that sharers are more inclined to allow pets. 15% of these sharers tend to have pets living in their homes yet only a subsequent 14% listings actually allow visitors to bring their pets. A similar amount of businesses allow guest pets on the premise (13%), with only 7% of businesses with pets on the property. Ultimately, traveling with pets and finding places that permit them is still a pain-point for both audiences.

user_type	Businesses	Sharers	Total Listers	% Difference
Smoke_Detector	74%	86%	2911	12%
First_Aid_Kit	23%	34%	1063	11%
Buzzer/Wireless_Intercom	18%	27%	839	9%
Free_Parking_on_Premises	18%	26%	831	8%
Pets_live_on_this_property	7%	15%	429	8%
Fire_Extinguisher	39%	47%	1582	8%
Dog(s)	3%	8%	229	5%
Carbon_Monoxide_Detector	65%	70%	2442	5%
Breakfast	6%	10%	316	4%
Pets_Allowed	13%	14%	486	1%

In effort to understand the difference between the audiences’ review score ratings (sharers 93 vs businesses 90), the sharer experience may be closer to the actual “bed and breakfast” experience where users can interact with the locals and engage with local culture. Sharers are not only “sharing” parts of their homes, but potentially an experience with guests. In contrast, business listings excel in providing luxury amenities, but may not have the local interaction. Professional listings may be solely evaluated on the living amenities as opposed to sharers, who can add a personal touch in *addition* to their amenities for higher ratings.

From the audience analysis, we learn professional listing performance is objectively behind sharers by 3 points. An effort could be made to educate the businesses on what the sharers are doing well on, yet they offer such different experiences like apples and oranges and each have their own trajectories. Each audience should seek to preserve their offering and product improvements can be made to help them reach their specific goals.

Further analysis could be done within each audience to identify product growth opportunities to cater to their unique experiences. Sample recommendations could be:

- A professional account option to offer options in luxury amenities for incremental revenue could be implemented in the roadmap to cater to a sliding scale pricing.
- Sharer accounts could leverage ways to offer personalized host experiences to allow visitors to better visit local culture.

What Generates Value?

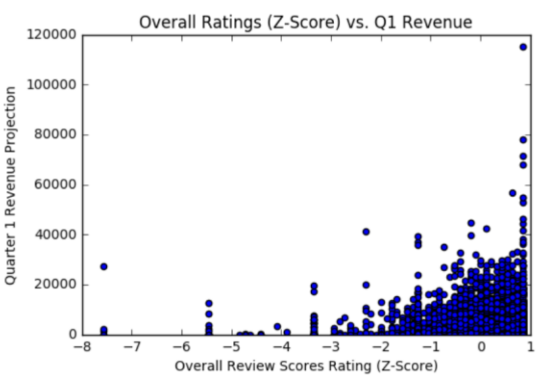
On a more macro scale, we want to understand which listing features create value through regression analysis. The question of value truly resides in the eye of the beholder. Specifically, the two “value” metrics we based our analysis on was the Overall Ratings metric and a projection of listing revenues for the first quarter of our data.

Why these two metrics? To a certain extent, these metrics represent a fundamental tension between supply and demand. On the demand side, with reviews, users are communicating their relative satisfaction with given listings and stays at those listings in order to collectively guide future travelers (e.g., wisdom of the crowds). Airbnb hosts would be wise to understand the perceptions of their potential clients as expressed by the types of features that satisfy or dissatisfy travelers, the locations and other general attributes. On the supply side, hosts want to know what types of features are most highly correlated and associated with higher profitability. However, given the data limitations at this time we were only able to explore the metric of the association of features to projected revenues thereby limiting our analysis to business listings.   Ultimately, the success or failure of Airbnb's business model, and, in fact, the sharing economy writ large, depends on the intersection of users' willingness to pay for features that hosts' are willing to provide. Alas, we hearken back to the "invisible hand".

The below exploration posits that the underlying market conditions of Airbnb in Boston can be observed through the intersections and/or divergences in features that are correlated with higher/lower ratings vs. revenues. Accordingly, the central assumption explored is the idea that if Airbnb were a perfectly balanced market, one would expect a very strong and close alignment between these two metrics. *The analysis begins with an explanation of the method employed to ask the question of what generates value, briefly highlights some of the key features associated with value, succinctly provides a model for potentially evaluating and beginning to explore the results, and closes with a discussion of the limitations of this analysis and potential next steps for discovery.*

Approach

We began our analysis by understanding the relationship between revenue and normalized rating scores. Our analysis confirmed that there is a statistically significant relationship, however, it is not very explanatory (only 1.6% of projected quarter 1 revenue can be explained by the normalized scores).




```
                            OLS Regression Results
==============================================================================
Dep. Variable:             q1_revenue   R-squared:                       0.017
Model:                            OLS   Adj. R-squared:                  0.016
Method:                 Least Squares   F-statistic:                     18.01
Date:                Tue, 25 Apr 2017   Prob (F-statistic):           2.40e-05
Time:                        16:00:04   Log-Likelihood:                -10359.
No. Observations:                1016   AIC:                         2.072e+04
Df Residuals:                    1014   BIC:                         2.073e+04
Df Model:                           1
Covariance Type:            nonrobust
===============================================================================================
                               coef    std err          t      P>|t|      [95.0% Conf. Int.]
-----------------------------------------------------------------------------------------------
Intercept                 6428.6870    203.620     31.572      0.000      6029.121  6828.253
review_scores_rating_zscore  864.0907  203.620      4.244      0.000       464.525  1263.656
==============================================================================
Omnibus:                      398.754   Durbin-Watson:                   1.593
Prob(Omnibus):                  0.000   Jarque-Bera (JB):             1648.644
Skew:                           1.844   Prob(JB):                         0.00
Kurtosis:                       8.034   Cond. No.                         1.00
==============================================================================
```

It is also interesting to note that there is a non-normal distribution of ratings- meaning that there is an inflationary impact at work. This was somewhat consistent with our expectations that these two metrics would be related. After all, when travelers make a decision to book a property one would think the relative rating of a listing would shape the future decisions of travelers.

Next, we began the analysis to understand the impact of features to revenue and ratings. Given the combinatorial challenges associated with finding the most relevant predictors with 99 variables explored in this section, we opted for the LASSO (Least Absolute Shrinkage and Selection Operator). By constraining the absolute size of the coefficients we are able to arrive at some of the most significant predictors of ratings as well as revenue projections. This LASSO regression was applied independently to both questions.

NOTE: the utilization of dummy variables requires n-1 variables and a base case. Our base is an (property type) 'Apartment' with a (amenity) 'Carbon Monoxide Detector' and an (room type) 'Entire home/apt'.

LASSO: Factors Impacting Perceptions in Ratings

Overview

Our first regression model aims to understand which listing factors and amenities correlate the most with ratings satisfaction . After running eight-fold cross-validated LASSO models to describe normalized ratings scores, we arrived at a model that describes 82% of variability in the test data. Now, this number is quite high due to the presence of contributory ratings within the data set. We opted to include these because their relative weight and importance lends insight into the perceptions of value and importance placed on these by reviewers.

Results

The following coefficient table ranks the most important factors based on estimated value. Across the entirety of listings, the value and cleanliness ratings respectively rate as the first and second most important predictors of value (roughly 30% of the magnitude of coefficients are contained in these two variables). Interestingly, listers in the sharing economy slightly outperform businesses in cleanliness (9.31 v 9.17) and value (9.33 v 8.9), despite the fact that businesses charge a cleaning fee nearly 2X as high as sharers. Furthermore, businesses tend to offer more premium amenities i.e. Cable TV, gyms, doorman, elevators yet the value ratings are lower by a larger margin. Travelers find more value from staying with the sharers (multiple subjective interpretations), and sharers can clean better than businesses with lower cleaning fees, potentially out of personalized sense of ownership.

Unsurprisingly, location is an important factor towards people’s overall satisfaction as roughly 25% of the total predictors in this model relate to that variable (either directly through ratings or through neighborhoods). (See below for top 5 and bottom 5 predictors.)

Type of Variable	coefficient_name	estimated_value
Rating	review_scores_value	0.364668
Rating	review_scores_cleanliness	0.240849
Neighborhood	West End	0.172262
Amenity	Other pets	0.150278
Amenity	Dryer	0.143614
...	...	...
Amenity	Laptop Friendly Workspace	0.004803
Amenity	Smoke Detector	0.004662
Neighborhood	Jamaica Plain	-0.017787
Amenity	Kitchen	-0.066066
Neighborhood	Mission Hill	-0.083571

In terms of amenities, what is quite interesting is to note that pet friendliness (as identified by the three categories of ‘Other Pets’, ‘Pets Allowed’ and ‘Pets Live on this Property’) would amount to the third strongest coefficient in identifying overall relative ratings. From the audience analysis, we learned that only ~14% of all listings offer pet allowance. While this could be a recommended feature to be add on, many liabilities and legal compliance needs to be considered.

Lasso: Factors Impacting Revenue

This model is remarkably worse (comparatively) at explaining the revenue, as it captures 32% of the variability in the test data. Nevertheless, it is quite interesting that this much variability can indeed be explained by mostly categorical variables independent of advanced sentiment or imagery analytics. Also, it is noteworthy to mention, that unlike the earlier data set, this portion of analysis is specifically targeting business revenue projections.

Category	coefficient_name	estimated_value
Neighborhood	South End	3247.992058
Neighborhood	Beacon Hill	3101.895147
Neighborhood	Downtown	2689.539947
Neighborhood	Condominium	2355.920559
Neighborhood	North End	2337.882215
...	...	...8
Neighborhood	South Boston Waterfront	-1929.708139
Amenity	Free Parking on Premises	-1933.584277
Room Type	Private room	-1986.674635
Neighborhood	Roslindale	-2117.963458
Room Type	Shared room	-2203.705522

Overall there are an incredible amount of variables that are flagged as predictors to generate this model (48 to be exact). This data suggests that there are some very desirable locations for travelers to Boston and that those are the key features that users are focused on. In fact, the six strongest predictors of revenue are purely geographic. At the bottom end of the spectrum it appears that type of accommodation and neighborhoods explain lots of the variability associated with lower projected revenues. Most of the neighborhoods ranked at the top are closer to downtown and further away from the suburbs, which is a natural bias in the model.

Notably, consistent with our previous analysis, cleanliness rates highly. Interestingly enough, the perception of cleanliness is correlated with higher revenues. What is somewhat fascinating, and counter to our expectations, is that the presence of a cleaning fee appears to generate $204.00 of additional revenue (NOTE: When calculating our revenue projections we did not include the cleaning fee into that analysis).

From audience comparisons, cleaning fees are naturally higher for business listings (2X) because of longer minimum nights of stay (4 vs 2.7). However, the review scores for cleanliness are not as high as sharers. Even though you pay more for cleanliness, it’s not being reflected in the reviews. It’s time for businesses to clean up their acts!

Intersection and Divergence of Results

Businesses exist to deliver value to customers who are willing to pay. Therefore when exploring these results we formulated the following model to approach the features analysis.

		Impact to Revenue	
		+	-
Impact to Reviews	**+**	**Sweet Spot**	**Travelers Crushing It**
	-	**Hosts Crushing It**	**Less Please**

Sweet Spot

The first category is comprised of the items that customers rate highly and also exhibit a willingness to pay. Items in this category are things that hosts ought to consider focusing their efforts on as travelers appear willing to pay for these and also are willing to provide higher ratings for these types of features, thereby creating a virtuous cycle of higher ratings and higher revenues.

The biggest feature we would like to highlight is the fact that **cleanliness** consistently rates among traveler's most desired features and, also, they appear willing to pay for it. Surprisingly in this data, it suggests that the presence of a cleaning fee actually has a positive impact on revenue (NOTE: cleaning fees were not used in calculating the revenues). Hosts pay attention!

Category	Predictor	Effect
It's the extras that count	Breakfast	$512
	Bed & Breakfast	0.01187
Security	Doorman	$10.6366
	Doorman	0.07367
It's the extras that count	Dryer	$853.199
	Dryer	0.14361
Connected	Wireless Internet	$1753.83
	Laptop Friendly Workspace	0.0048
Clean is King	cleaning_fee (presence of)	$204.854
	review_scores_cleanliness	$165.267
	review_scores_cleanliness	0.24085

Less Please

Interestingly enough, this analysis did not uncover any strong predictors of items that people have expressed a desire for less of and are not willing to pay for. This intuitively jives as one would expect these offerings to be filtered out, or never consumed in the first place..

Traveler's Crushing It

Category	Predictor	Effect
Location, location, location	Brighton	-442.198838
	Brighton	0.063317
Must love dogs	Dogs	-1323.67834
	Other pets	-910.560084
	Other pets	0.150278
	Pets Allowed	0.016994
	Pets live on this property	-335.340299

	Pets live on this property	**0.044513**
Overall good value?	**review_scores_value**	**-125.643053**
	review_scores_value	**0.364668**

Three interesting divergences arrive in the data: two localized and one perhaps with broader implications. First, what's going on with Brighton, MA? Travelers who stay in this locale tend to have more favorable ratings, however, based on the projected revenues of our commercial listers, staying in Brighton appears to be correlated with lower revenues. There are a number of potential explanations beyond the limits of this analysis. Could Brighton be a pretty well received location and the hosts of Airbnb haven't caught onto their ability to effectively price this neighborhood? Could the bias in our segmentation approach negatively impact the revenue projections for hosts in this area? Perhaps Brighton is filled with positive and optimistic college students who just don't have the means to pay? Or even, is it possible that it is the next hipster paradise just waiting to be discovered by business?

The second divergence and perhaps the most surprising, in terms of magnitude, has to do with the impact of pets on the revenues. In terms of reviews, the presence and association with pets appears to have an overwhelmingly positive correlation with relative review ratings. However, when it comes to the revenue projections the story could not be quite the opposite. In fact, it appears that pets have roughly a -$1,245 combined effect on a property's projected revenues. Is this because hosts just aren't aware of how much people value their pets? Are pets a polarizing issue when it comes to making an Airbnb stay decision? Once again, does the cut in our data negatively impact our projections of expected revenue?

In both of these situations it is quite intriguing to consider the possibility that travelers are free-riding on hosts' listings. In both situations, the data might possibly suggest that hosts have an opportunity to charge more than what they are currently charging. Or, it may also be possible that an overage of supply exists driving down the total revenues per listing. Either way, it appears that travelers are definitely getting a good deal with respect to these two features.

The macro-issue perhaps uncovered in the data is that the AirBnB marketplace in Boston is favorable to travelers somewhat at the expense of hosts. This suggestion stems from the divergence between perceived value ratings and willingness to pay for that perceived value. Theoretically, one might expect value ratings and willingness to pay for that associated value to have zero impact on revenue and reviews. Admittedly part of the challenge is that the overall review ratings are partially calculated using these values. Nevertheless, the magnitude of their delta may suggest the degree to which such imbalance exists. If this divergence points to underlying market conditions, it could perhaps inform investment strategies of businesses or perhaps be of interest to AirBnB as they seek to maximize the growth of their network, necessitating a long term win for both sides.

Hosts Crushing It

Further perhaps bolstering our curiosity is the fact that there are no examples in our data set that suggest that hosts are able to charge a premium for features that travelers don't seem to favor. This could possibly further advance the claim that traveler (buyer) power is greater than hosts (supplier) power in Boston.

Limitations

Overall this analysis was intended to be a starting point on an extremely robust and interesting data set. There are many limitations in this analysis, but nevertheless, it is a fascinating starting point.. Foremost is the opacity in the calendar data which was discussed at length earlier. This invariably impacted the outcomes of our analysis in ways that were beyond our reach. Another key limitation is the strong underlying correlations in our data. Within our analysis we treated every amenity as an independent variable, but there are likely clusters of offerings that form within the data. Meaning, there are commercial properties in downtown Boston that have doormen, gyms, pools, and other amenities that are treated as independent observations that likely skew the outcomes of the feature selection. Future analysis may consider a form of cluster analysis to build distinct groupings to understand the ideal “type” of property as opposed to raw features.

Another limitation is the fact that Boston is somewhat treated as a monolithic entity. As the data indicates, revenues and ratings necessarily vary quite dramatically between different neighborhoods, different urban/suburban contexts, different types of properties, etc. We attempted to normalize our data by neighborhoods, however, the data set was not large enough to effectively conduct that analysis. We did not have exact addresses either, which hindered our ability to effectively combine square footage and other data to come up with efficiency metrics which would be very meaningful to scale the data for future hosts.

Also, perhaps the most key limitation in this analysis is that it does not deal with the imagery nor the unstructured textual data associated with the listing and its reviews. These items are central to the decision making process of future travelers (customers) and would necessarily be a part of future analysis in trying to better characterize the market and the impacts of different features on revenues and reviews.

Conclusion

We first sought to reveal some underlying and fundamental observations from our data that shed light on the central question of whether or not AirBnB was an example of a “sharing” economy. Through the course of that analysis we proposed that there are two key segments of hosts present in the data- Businesses and Sharers. By simply segmenting the groups based on the total number of listings by the host, we uncovered clear statistical differences between the two groups. By comparing the two segments side-by-side, we revealed clear, statistically significant differences between the two groups of hosts- those who are sharing their space for a reasonable fee and those who are using AirBnB has a booking agent for their full time business.
We then transitioned into some advanced analysis to better understand perceptions of value from both business hosts (projected revenue) and travelers (overall ratings) and to better understand the areas of intersection and divergence. We proposed a model for beginning to evaluate the relative tensions that exist in the underlying data and proposed that it is possible the data may suggest a macro-fundamental trend that the AirBnB market is perhaps favorable to travelers at the expense of business hosts. Finally, we discussed some of the key limitations from this analysis.