Exploring the Role of Airbnb as a Rising Market Force in the Sharing Economy

Clara Chun

Jared Deely

Joseph Beltran

Judy Teng

Abstract

Sharing economies remain relatively new, and as such the role they are to play in their respective industry is uncertain. Here we ask whether the market catered to by Airbnb in relation to existing hotel business is competitive or complementary. Additionally, we examine the degree to which trends within the platform persist across location. Our study offers a cross-sectional comparison of Airbnb listings and the hotel industry in four major cities across the US, (Chicago, Los Angeles, New York, and San Francisco) as well as a deeper analysis of Airbnb pricing and accommodation patterns for each city. Ultimately, we find that Airbnb could continue to pose a threat to incumbent industry, by offering a variety of amenities at a relatively low price point, and that trends could be susceptible to variations in location depending on the nature of the local market, though further research is likely needed.

Introduction:

Peer-to-peer sharing economies continue to become increasingly relevant in today's market, as platforms such as Uber, Lyft, and in our case, Airbnb, pave the way for new approaches to interacting with existing industries. Unfortunately, the relative "newness" of these companies means that research on the subject has been left with little time to develop, leaving us to wonder about the roles of these rising corporations within existing markets. Our study seeks to compare existing trends between pricing in peer-to-peer and traditional markets (Airbnb and hotel) while also looking at trends between four major US cities to ask the question as to where Airbnb falls in relation to established industry and whether internal Airbnb trends are consistent across the locations of interest. We begin with a look at available research, moving to an explanation of methods, presentation of graphs, and finally a discussion of results and their implications.

Existing Knowledge (Briefly):

Literature suggesting the significance of these newcomers point toward their impact on existing offerings. According to one study, Airbnb's influence on Texas hotels takes the form of decreasing hotel revenue with increases in Airbnb supply that, while at first appear trivial, may prove significant when considering the construction costs of future hotel expansion[1]. Similarly, a study of New York lodgings concluded a loss of more than $450 million in revenue and $1 billion in developmental investments, supplementing the assertion of the Texas study that

[1] Zervas, Georgios and Proserpio, Davide and Byers, John, The Rise of the Sharing Economy: Estimating the Impact of Airbnb on the Hotel Industry (June 9, 2016). Boston U. School of Management Research Paper No. 2013-16

revenue losses, and the relative ease of expansion for Airbnb compared to that of the incumbent industry, could be hurting traditional markets [2].

However for many, there remains the question of whether these peer-to-peer platforms cater to a complementary, or competitive demographic. Tussyadiah outlines push and pull factors for consumer engagement with collaborative consumption (shared economies), with incentives such as sustainability, savings, and community engagement, and barriers like lack of trust or host reputation [3].

From the host perspective, Airbnb may prove an alluring prospect by allowing many to continue to live in areas where they fall below the median income, according to a study commissioned by the company and carried out by HR & A Advisors. This was the case for 60% of San Francisco owners renting out their homes on the site, with 41% of hosts either part-time employed or freelance [4].

Taking such information into consideration, we have set out to assess the role of Airbnb within the context of the urban temporary housing market. Mindful of the limitations of a cross-sectional study design, our exploratory works seeks to consider the extent to which pricing and consumption patterns persist across cities and then step back and examine plausibility of Airbnb as a market competitor to the existing market. We hope that our findings will help to guide further research in an attempt to address gaps in the literature.

Methods (Data Scraping):

---

[2] Dandapani, Vijay. *Airbnb and Impacts on the New York City Lodging Market and Economy.* New York City: Hotel Association of New York City, 2015. Print.

[3] Tussyadiah, Iis P. "An exploratory study on drivers and deterrents of collaborative consumption in travel." *Information and Communication Technologies in Tourism 2015*. Springer International Publishing, 2015. 817-830

[4] Geron, T. (2012). Airbnb had $56 million impact on San Francisco: Study. Forbes. http://www.forbes.com/sites/tomiogeron/2012/11/09/study-airbnb-had-56-million-impact-on- san-francisco/

Airbnb data was scraped from a combination of search results and listing specific pages on Airbnb's website. Prices and the identification numbers (ID's) assigned to each listing by the website were derived from the first 100 pages of search results for each of the four cities (San Francisco, Los Angeles, New York, and Chicago) in order to ensure a sufficiently large sample size. Then, using the ID's scraped in the first part, further information was gathered on the respective listings by sequentially visiting listings pages and scraping relevant data. From here, search rankings are presented on the basis of listing quality, trip experience, ease of booking, guest preferences (default to 1 and above), with 18 listings per page.

Airbnb's anti-scraping features lead to problems during the collection process. As for part 1, the listing prices proved easy enough to scrape, however during inspection, nodes containing some key pieces of information were found to be too deep, and were thus blocked from access. Fortunately, This information could be obtained by reading in and extracting key info from attribute values at levels high enough to be read into our script.

Data listed for the 100 search result URLs was limited to price and variables of no concern, requiring that we enter a subsequent scraping phase in which our script visited the listing specific page of each listing found among the first 100 pages of results to reveal more info. The high level of activity required during this portion of the script was enough to trigger anti-scraping measures and force the function to fail periodically. This prompted the introduction of random pauses, coded into the function to decrease the number of actions per minute (apm) and removed the need for the script to be run manually. Naturally this greatly increased the time taken for the script to run to completion and necessitated the use of a remote server to complete the scraping process.

Scraping Data (hotels)

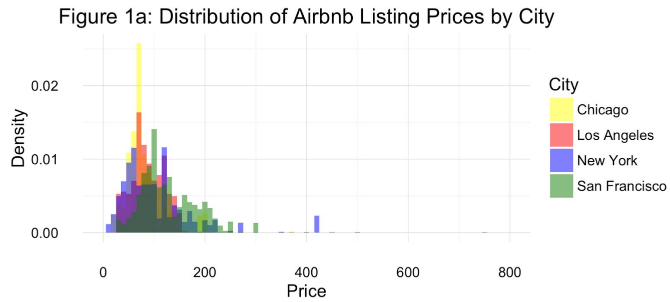
We wanted to see how Airbnb prices compared to hotel prices. Hence, we scraped data from Trivago which is a hotel information website where hotel prices from various booking websites are gathered and the best value for each hotel is displayed. We gathered the listings by searching the city name and scraped 25 observations of data per page. By using nodes that corresponded to the names, ratings, and prices, we were able to scrape a total of 100 observations for each city.

Cleaning Data

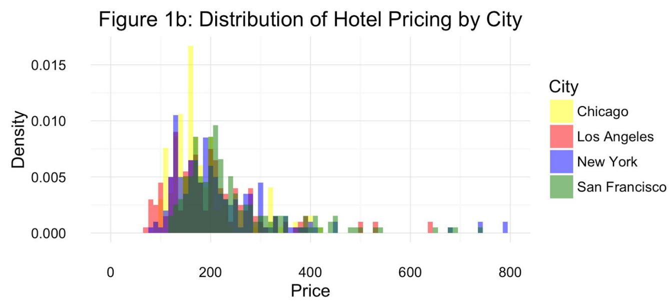
The goal of the cleaning process was to obtain two essential data frames that were foundational to our analysis. The first desired data frame consisted of information regarding hotels while the second included information on the Airbnb rental properties across the four cities. Let's begin with the first data frame. The scrapping process from trivago.com produced a set of two data frame for each city. To acquire the aggregate data frame with details on all four cities, we simply collapsed the eight data frames into one. The scrapping process for the Airbnb rental properties produced four xls files, SFdf, NYdf, LAdf, and Chicago.xls. SFdf consisted of 21 variables, whereas NYdf, LAdf, Chicago.xls contained 17, 15, 16 respectively. Since each data frame obviously contained a different amount of variables, we found the fifteen mutual variables and collapse the four data frames into one. To improve readability and simplify the analysis and graphing, we kept only the necessary numericals and replaced 'NA's' with either 0 or 'unspecified'. This final data frame is named 'updated_cleaned_four_cities.csv'.

Price Center and Spread for Airbnb and Hotel Industry:

Figure 1 and 2 compare pricing distributions for hotel and Airbnb listings. Prices are plotted horizontally while the vertical axis represents a density scale. It should also be noted that the window does not include listings above $800 for the sake of readability. The first pattern of note is that Airbnb listings are clustered about a much lower price point, $105.01, relative to an average hotel price of $215.29. Taking differences in centers into account, a relative standard deviation was calculated for all cities (0.7761683 and 0.6098457 for Airbnb and Hotels respectively), and for each city individually by reporting standard deviations as a proportion of the mean. These values are reported in the tables accompanying the figures. The relative standard deviations illustrate what may defy intuition upon mere inspection. While the distribution appears tighter for Airbnb listings, when we take into account the lower average price point, we find that for all cities, the relative deviation is actually higher for Airbnb. This observation persists at the city level in New York and certainly LA, however roles are reversed when examining Chicago and San Francisco. These patterns could have implications for the sorts of markets catered to by the services and perhaps challenges the idea of competing markets while simultaneously indicating variable trends by location; ideas explored further in our conclusion.



City	Relative SD (Airbnb)
Chicago	0.4735539
Los Angeles	1.1712121
New York	0.7234926
San Francisco	0.4258054

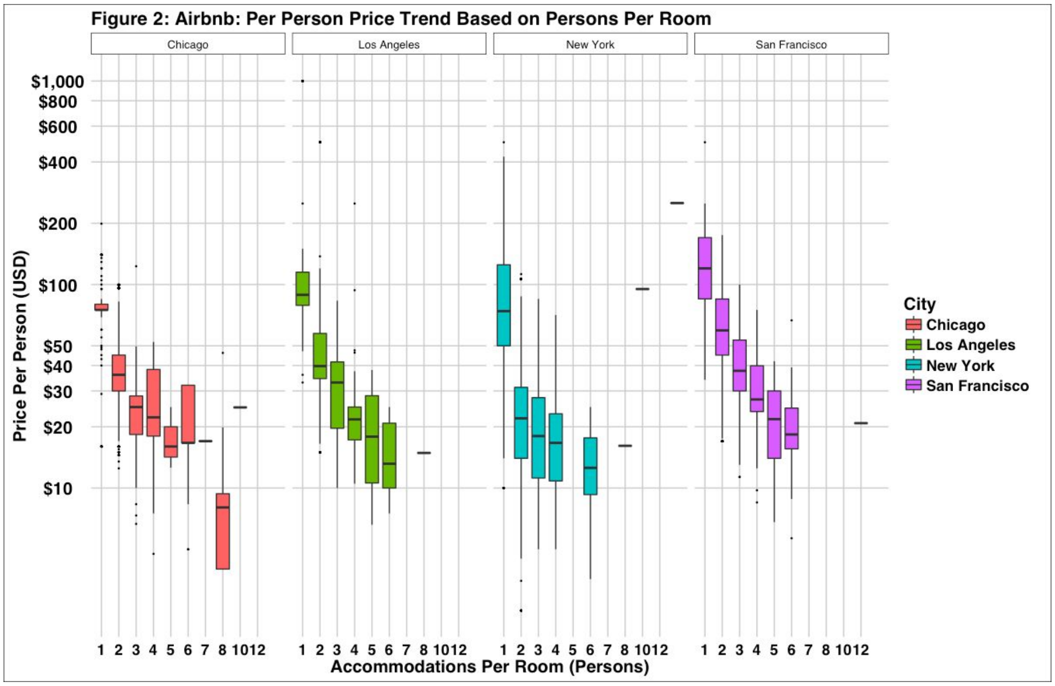


City	Relative SD (Hotel)
Chicago	0.5784049
Los Angeles	0.4704701
New York	0.5183986
San Francisco	0.7037612

Airbnb: Per Person Price Trend Based on Persons Per Room

From a consumer's point of view, an Airbnb rental is often considered as a more economical alternative to booking a hotel. For travelers on a budget, it seems obvious that choosing Airbnb offers a bigger bang for your buck. Our primary motivation for visualization two is to determine whether the previous statement is valid by detecting possible correlation between the rental price and the permitted accommodations per room.

In the 'updated_clean_four_cities' data frame, we have the variables price per night ('Price'), daily price per person ('Price.Per.Person'), and allowable persons per room ('Accommodations'). We wanted to know whether Airbnb is more economical on the individual level by assessing any correlation between price per person and allowable persons per room. We hypothesized that although the correlation between price per night and accommodation would be positive, the correlation between price per person and accommodations would be negative; in



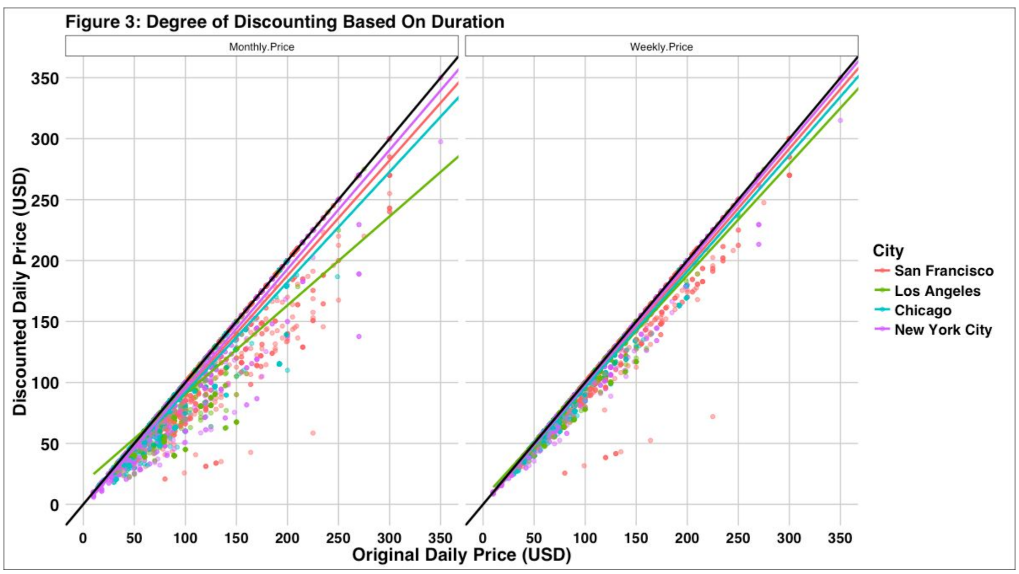
other words, the larger the room, the higher the price per night yet the lower the price per person.

Our hypothesis was somewhat correct. On graph one, 'Accommodations' was plotted against the variable, price per person. As per our hypothesis, all four cities experience a negative correlation. In consideration for travelers hoping to save, it seems renting in groups does save money since the price per person decrease as accommodations increase. Interestingly, the marginal saving does fluctuate as accommodations increase. Generally, marginal savings remain positive at a decreasing rate up to the sixth person; after that, marginal savings become negative.

Degree of Discounting Based on Duration

The practice of an extended stay is deeply rooted within the hotel industry. Differing from an overnight stay or a few days' stay, an extended stay hotel offers a discounted rate for patrons who stay longer than a certain period. Our main motivation for visualization 3 is to attempt to identify, if any, the existence and degree of such discounted rates amongst Airbnb rentals.

From the 'updated_cleaned_four_cities' data frame, we have the variables daily price ('Price'), weekly discount percentage ('Weekly.Discount(%)'), and monthly discount percentage ('Monthly.Discount(%)'). We first multiply (1- discount percentage) to the daily price rate to calculate the weekly and monthly discount rate. Next, we plot the daily rate against the discounted rate for each city and graph the least squares line which serves as an indicator to the degree of discounting. A line of slope one, which is provided as the black line on the graph, is synonymous to no discount rate while a slope less than one signifies a discount. The lesser the slope, the greater the discount. We then facet the plot according to duration to attain a side-by-side comparison of the degree in discounting dependent on length of stay. This way, we are able to see the difference in discounting rate across the four cities and across rental time.



We discover that, generally, the longer one stays at an Airbnb rental, the larger the discount rate. Of the four cities, Los Angeles offers the largest discount on both a weekly and monthly basis; New York City, on the other hand, offers the least discount in both durations. More specifically, the weekly discount rate for New York City is at almost zero percent, or no discount at all. We considered the use of a log scale on the y coordinate; however, this would render the black line useless. Thus, we must take the conclusions from the figure with a grain of salt due to the lower price ranges being slightly inflated.

Average Daily Price of Airbnb by Property Type

Due to the unique and diverse housing types that Airbnb offers, we thought it would be interesting to compare prices based on the types of housing among the four cities. Using the 'updated_cleaned_four_cities' data frame, we grouped the city name ('City') and property type ('Property_Type') and found the average of the daily price in a new column labeled 'dailyavg'. We plotted the different housing types on the x-axis and the average price on the y-axis. We only wanted to observe the housing type that all four cities had in common, so we filtered out any housing types that did not apply to all four cities.

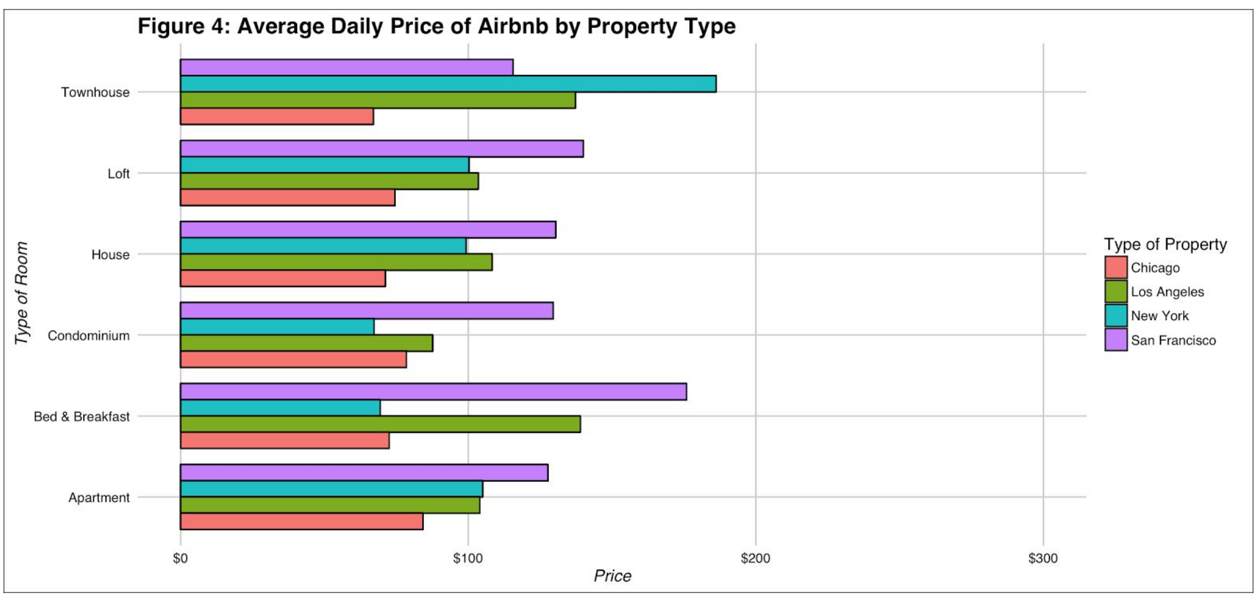


Figure 4 shows that New York has an unusually high average price for townhouses. From figure 1, it was observed that New York had a relatively high standard deviation. This high average price may be an explanation to New York's variability.

Density Plots of Check In / Check Out Times

Given our findings of Airbnb's unique variability in pricing and property types, we also wanted to explore to what extent check in and check out times might be reflective of the unique needs or lifestyles of the local environment. We were also interested in seeing whether or not the price range of a listing might act as a possible confound for this analysis, with more expensive listings perhaps being more stringent than their cheaper alternatives, regardless of location.

Figure 5a: Density Plot of Airbnb Check In Times

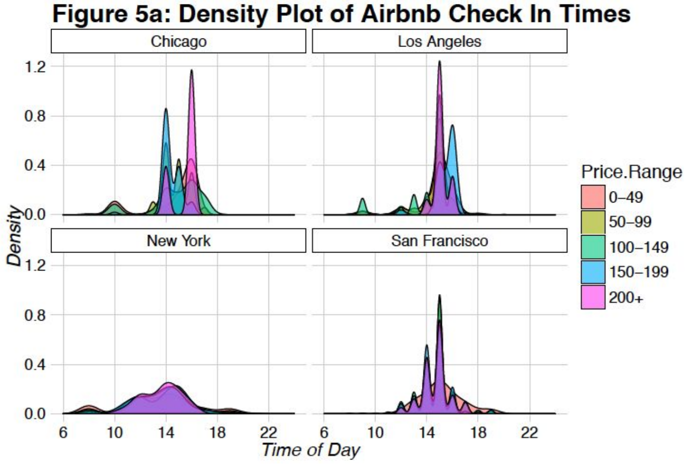
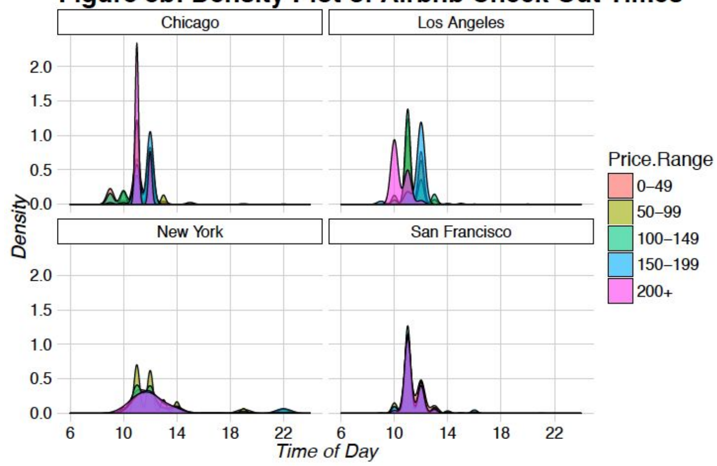


Figure 5b: Density Plot of Airbnb Check Out Times



Ultimately, we found similar trends for these times across different price ranges in terms of stringency, with slight variations for the most common time within a time range (for example, Chicago's 150-199 listings requiring you to check in earlier than the more expensive counterparts). However, it was interesting to also note things like Chicago's marked adherence to set check in and check out times in comparison to New York's greater flexibility and variety. Further work could be done to explore the possible factors contributing to these differences, whether it may be New York's reputation for being "The City That Never Sleeps," being more accommodating of more variable schedules; or the culture of a city that may be more attuned to a fixed routine or schedule. Either way, this could be an interesting direction with which to explore the ways peer to peer economies can be influenced by the communities they are based in. A column of `Price.Range` was mutated given Price values in our original cleaned data set, with replaced `Price.Range` values being mutated conditionally with if else statements specifying to which range a Price value should belong (to lie upon a categorical variable scale from which to fill results by).

Conclusions:

These findings ultimately suggest a rising organization that poses a potential competitive threat to established markets while simultaneously addressing the needs of new demographics dynamically based on the site of listing locations. The comparatively lower price of Airbnb listings could be perceived as the more economically viable option, however our analysis also suggests higher variability in price for the platform, although this is not necessarily the case for all cities individually. We however believe it's fair to attribute the higher cost of a hotel stay to the cost of staff wages that allow for greater onsite services that Airbnb listings cannot necessarily offer. On the other hand, the high variability of prices among listings could be due to a larger catalogue of accommodations. Hotels on the other hand have less freedom in the party sizes rooms can handle(as discovered through data cleaning), and offer similar services across the board, which could force competitive pricing and lead to increased clustering of cost to the consumer. This is potentially problematic when one also considers findings related to figure 2 in that as the size of the party accommodated increases, individual cost decreases. One could argue that while not directly competing with hotel business, by catering to a new market, Airbnb could prove an attractive option for visitors looking for a different sort of travel experience, who were previously beholden to a more homogenous market in terms of lodgings. While such analysis could again be supplemented by existing research on trends in educational and economic demographics, as well as what consumers find promising about sharing-economies [5] this may be a finding that warrants a second look in subsequent research endeavours.

In terms of city comparisons, findings do suggest some level of listing variability by location. LA poses the greatest savings opportunity for the consumer among the four cities studied, while New York offers barely any discount for prolonged stays. Similarly, the value placed on housing types is subject to change as one migrates from city to city as discussed in analysis of figure 4, and even factors such as check-in and check-out time stringency can depend on the local environment. The question of whether the consumer-host interaction is indeed mediated by local housing climate may warrant further study, perhaps even an analysis of the parity between Airbnb and permanent housing trends. These observations could lend credence to previous assertions that as a platform, Airbnb is very much integrated into local communities in that there appears to be a dynamic interplay between listings and the environment in which they are situated.

In the absence of longitudinal data, it is difficult to draw any definitive conclusions about Airbnb's influence on established markets or how between city patterns are projected to change. Observations made here could merely be snapshots of the rapidly developing infancy of a new market. However, we hope that our work can help to highlight where information is missing and provide a roadmap for a future, larger scale study.

[5] Tussyadiah, Iis P. "An exploratory study on drivers and deterrents of collaborative consumption in travel." *Information and Communication Technologies in Tourism 2015*. Springer International Publishing, 2015. 817-830