Sheena Lee Villanueva

An Analysis on airbnb for investment opportunities

Introduction

Travel and tourism was the second most-popular choice for discretionary spending (Phocuswright). In 2019, travel expenditures totaled $1.1 trillion ($242.3B for lodging), around 80% of which were directly from spending on leisure travel by domestic and international travelers. (US Travel Association). When looking at countries that directly contributed the most to global GDP, the United States’ travel and tourism industry contributed the largest sum at $580.7 billion. Meanwhile, the administrative region of Macau generated the highest share of GDP through direct travel and tourism of any economy worldwide.

Most travelers stayed in hotels only several years ago. However, Airbnb changed that. From what started as an idea to rent out an air mattress to earn a bit of extra money to make rent, Airbnb has become a $31 billion company, with a global network of hosts earning $100 billion. Airbnb boasts of providing local, authentic travel experiences for their guests, having presence in over 100,000 cities around the world. They offer one-of-a-kind experiences and a place to stay for every trip - from treehouses for the weekend or an entire home for the family.

While travel has decreased significantly this year because of COVID-19, the industry is expected to bounce back after the pandemic. The lockdown, travel restrictions and sealed borders forced all sorts of travelers to cancel their trips; which is why there is an expectation that leisure travel will be back soon as people are desperate to set themselves free from quarantine. As such, there is a motivation to look into investing into the travel industry.

Research purpose

Knowing that growth is expected from the travel industry, and there are investment opportunities for lodging, is it a good idea to invest in an Airbnb hotel? If it is, what kind of investment opportunities are there and what would the target market be? This paper aims to look into a subset of historical Airbnb listings and bookings, and find out more about the distribution of revenue, reviews and popularity of Airbnb in Los Angeles, California.

Dataset

The data that will be used for this project was acquired through insideairbnb.com. This was sourced from publicly available information from the Airbnb site where some data have been analyzed, cleansed and aggregated where appropriate for public discussion. We will specifically look at the listings.csv file, where detailed listings data for Los Angeles scraped on January 2020 were obtained.

There are over 39,400 records and 106 column features, including but not limited to listing information, neighborhood/location information, property and room types, amenities, price, amenities, minimum and maximum stays and reviews.

For this project, to answer whether or not we should invest in an Airbnb hotel, we will look specifically at data revolving around revenue, reviews and popularity.

Literature Review

With Airbnb’s success, there are a lot of research done with its open data ranging from economic costs and benefits to motivation-based segmentation studies.

To explore the concept of disruptive innovation, Guttentag (2016) analyzed the motivations that attracted guests to Airbnb as well as its potential weaknesses. In his research, Guttentag (2016) measured performance expectations, experiential attributes, and price. Additionally, ratings along the same attributes were provided for hypothetical nearby budget hotels/motels, mid-range hotels, and upscale hotels. Airbnb predictably far outperformed all three hotel classes with regards to uniqueness and local authenticity. When comparing the cost of their most recent Airbnb stay with the three hotel categories, Airbnb also was considered much cheaper than even the budget hotels/motels. This finding supports the earlier suggestion that Airbnb is indeed generally a comparatively 214 inexpensive option compared to hotels, despite some recent analyses showing average Airbnb rates are higher than average hotel rates in some markets. Airbnb’s low cost and experiential benefits also very directly reflect some of the unique value proposition that Airbnb offers, which is integral to the notion of disruptive innovation.

Research done by Adamiak (2019) shows the size, structure, distribution, dynamics, and use of Airbnb accommodation in 167 countries. His study shows that active Airbnb listings are most numerous where the platform started its global expension: North America and Europe. The United States currently has 624 thousand active listings and accounts for 17.4% of the global offer. Adamiak (2019) also mentioned that almost one third of Airbnb listings is located in large cities; with these statistics in mind, this paper will be analyzing listings located in Los Angeles, California.

Preparing the Data

1. Data Cleaning

For data cleaning, duplicate and missing data were checked, as well as standardizing entries for string and binary columns. Columns that were not meaningful for this study such as columns for URLs, host’s personal information, specific location information, country code and name – as we know that the listings only include those that are in Los Angeles, dates, license information, and availability\_ data, were removed.

To make sure that all the records are unique, I checked for duplicates by listing id (Column A), as hosts can have multiple ids. Before handling null values, I changed entries such as “none” and “N/A” to NULL to represent missing data.

*Handling Missing Data*

For columns that have high percentage of missing data (90% null values), they were dropped and removed from the dataset. For those that have a small percentage of null values (less than 1%), I dropped the rows where these null values are as it will not affect the number of data that we have to use for analysis. Review scores are an important aspect for our research question, and they can vary greatly and it might be assuming too much if these were manipulated based on other fields. As such, although it is 20% of our data, I have decided to drop the rows without the columns starting with “review\_scores\_”. For the reviews\_per\_month, first\_review and last\_review, I have decided to drop these rows as it would not be used. As for the text columns such as neighborhood overview, notes, transit, access, interaction and house rules, I dropped these columns as there is no way to populate these. As neighborhood, neighborhood cleansed and neighborhood group cleansed describe the same information, I have decided to only keep neighborhood cleansed. After seeing if there were any instances of space that can be filled in description, the space column is also dropped. Finally, I checked if the hosts without response rate have other listings under them. However, since there is none, we will also drop the rows for the host\_response\_rate since we will need this during the predictive analysis.

For price and extra\_people, which are both monetary values, I removed the dollar sign and commas, and changed the datatypes from object to float. For host\_response\_rate, since it’s in X% form, I removed the percentage sign and divided it by 100, to keep it in its decimal form, and changed the datatype from object to float as well. As revenue is an important aspect in answering the research question, we have to make sure to remove columns where price==0. Additionally, for all columns that have string values, I made sure that all words are in lowercase.

1. Feature Engineering

Before I created any features for analysis, the following are the set of assumptions based on what was available or lacking from the dataset:

1. Each booking always has two guests unless the listing only accommodates one
2. Only half of the stays generate a review, so we assume that the listing has twice as much bookings as the number of reviews.
3. Because we don’t have the actual number of days stayed, we will take the minimum number of days allowed as the booking length.
4. The price indicated is price for a one night stay, and the additional price is added if a solo traveler books an accommodation that can accommodate more than 1

The following features are engineered from the assumptions stated above:

1. Bookings: the number of bookings based on 2x the number of reviews column
2. Estimated\_daily\_revenue (for two people):
   1. If guests\_included > 1, the price quoted includes two or more people, price = estimated\_daily\_revenue
   2. If guests\_included == 1 and accommodates == 1, price = estimated\_daily\_revenue
   3. If guests\_included == 1 and accommodates > 1, price + extra\_people = estimated\_daily\_revenue
3. Estimated\_revenue: estimated daily revenue multiplied by the minimum number of nights

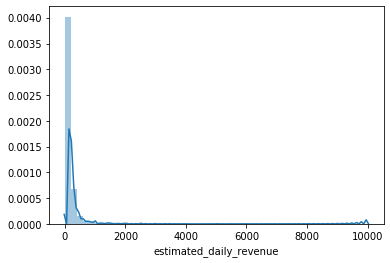
Exploratory Data Analysis

There are numerous ways in approaching the research question as there are several key factors that are considered in investment opportunities. In Guttentag’s paper described above, he mentioned that experiential attributes also play a part in travelers choosing Airbnb above hotels or relative motels. More than revenue, we can look at the reviews and

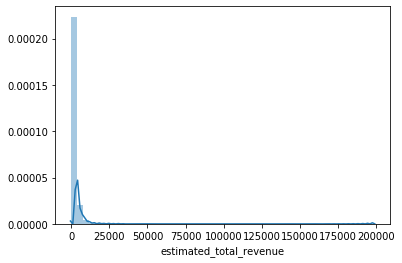
popularity of a listing, assuming that accommodations that have a great review or is more popular will attract more guests in the future. As such, the exploratory data analysis will be split into three sections: revenue, reviews and popularity.

1. Revenue

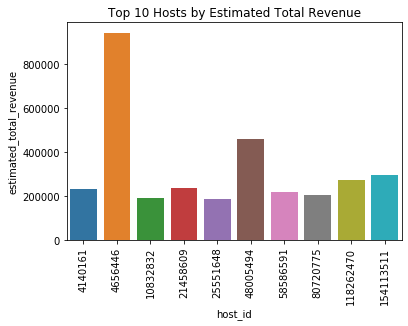
The easiest way to determine a good investment is through the revenue. We will look at the estimated\_revenue column that we made above to show the revenue per booking.



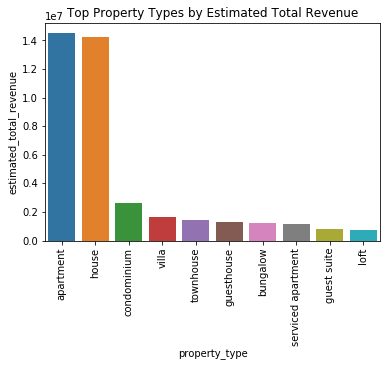
The figure above shows the distribution of estimated daily revenue. Although it is skewed to the right, the estimated daily revenue has a mean of $179/day and a 50th percentile of $115. The max as we can see goes up to $10,000/night, and this is an outlier that we may want to check further in our analysis.



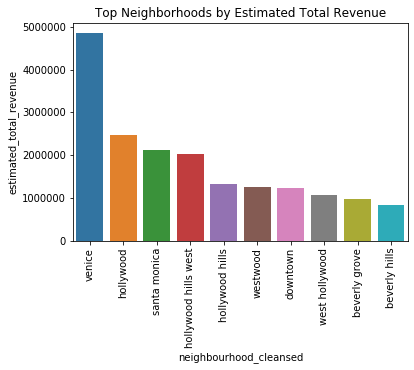
The estimated total revenue follows the same distribution with a mean of $1567 per stay.



The above figure shows us the top 10 hosts based on estimated total revenue. We can see that host 4656446’s total revenue is significantly greater than the other hosts.



The figure shows what type of property types have the most estimated total revenue. Since Airbnb started with renting out an airbed in the owners’ place, it makes sense that the majority of the listings that have bookings are for apartment and houses. It is also good to note what type of properties are majority in the city of Los Angeles.



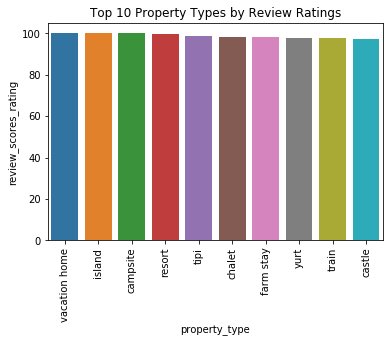
The figure above shows which neighborhood has the most estimated total revenue. We can see that the neighborhood in venice significantly generated more total revenue than other neighborhoods.

1. Reviews

Another aspect that we can look at are the review scores the listings get. Although revenue is a great indicator of a good investment, a listing can have a high daily price that could contribute to the total revenue. However, if the reviews are not good (i.e, not worth the price, bad value, etc), then the listing may not get bookings in the future or the price may have to go down. Similarly, for those with great reviews, the listings can actually increase the price.



As we can see from the figure above, a lot of the hosts have an average rating of 100%. As we deep dive later into specific top hosts, we can check if they also have high ratings.



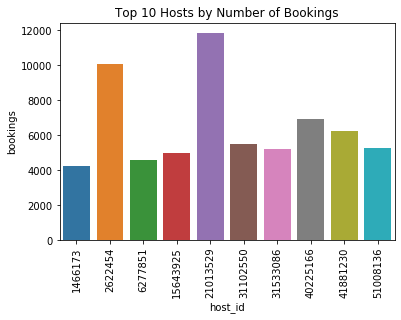
The above figure shows as that the reviews are only marginally better among the properties. It’s very interesting that a that the top property types are either very extravagant such as vacation homes, an island, resort, and castle or very “simple” such as campsite, tipi, farm stay, yurt and even train.



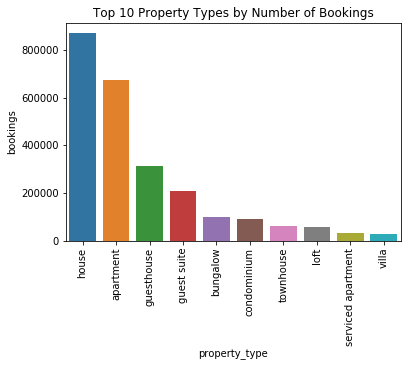
Looking at the ratings by neighborhood, it is interesting to see that the top 10 neighborhoods by review ratings are very different from neighborhoods that have the highest estimated total revenue.

1. Popularity

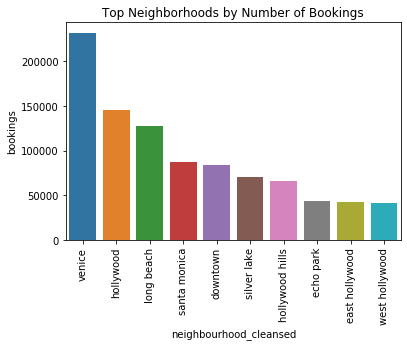
With popularity, we will look at the number of bookings per listing. It is important to know if there are enough demand for the listings that have top ratings and revenue.



The figure above shows the top 10 hosts by number of bookings, with host 21013529 and 2622454 having significantly greater number of bookings than any other host.



The figure above shows the top property types by number of bookings. It’s good to see that the same top 2 types are the same as the top 2 that we saw by estimated total revenue.



The figure above shows similar neighborhood results with those by estimated total revenue. We can see that there is a correlation between number of bookings and estimated total revenue.

*Top Host Analysis*

|  |  |  |  |
| --- | --- | --- | --- |
| Property Type | Estimated Total Revenue | Number of Bookings | Review Scores Rating |
| House | $899,206 | 2300 | 95.85 |
| Apartment | $20,708 | 220 | 96 |
| Condominium | $8,556 | 44 | 97 |
| Townhouse | $7,657 | 44 | 99 |
| Guest Suite | $5,145 | 28 | 99 |

The table above shows the analysis of Host 4656446 by property type. As we can see, the more number of bookings, the lower the review scores ratings are. However, for over 2000 bookings, a 95.85 rating is pretty good, especially with the revenue that Houses pull in.

|  |  |  |  |
| --- | --- | --- | --- |
| Neighborhood | Estimated Total Revenue | Number of Bookings | Review Scores Rating |
| Hollywood Hills West | $546,729 | 278 | 97.17 |
| Pacific Palisades | $91,623 | 120 | 98.5 |
| Westwood | $79,935 | 80 | 94 |
| Venice | $75,159 | 932 | 95.5 |
| Fairfax | $38,564 | 156 | 98.3 |
| Hollywood Hills | $35,883 | 154 | 98.71 |
| Bel-Air | $26,249 | 162 | 91.3 |
| Beverly Grove | $15,159 | 134 | 97 |
| Highland Park | $8,463 | 146 | 98 |
| Playa Vista | $7,781 | 46 | 100 |
| Malibu | $4,503 | 28 | 89 |
| Echo Park | $3,305 | 126 | 95.3 |
| West Los Angeles | $3,150 | 48 | 87 |
| Mount Washington | $3,131 | 30 | 97 |
| Manhattan Beach | $598 | 42 | 95 |
| Glendale | $426 | 38 | 98 |
| Long Beach | $384 | 56 | 96 |
| Mid-wilshire | $230 | 60 | 91 |

The table above shows the analysis of Host 4656446 by neighborhood. It’s very surprising to see that the number of bookings in Venice is more than 3 times that of Hollywood Hills West, but the estimated total revenue of the latter is more than 7 times of the former. Pacific Palisades might be a good neighborhood to look at as well, having a high 98.5 rating and over $90,000 estimated total revenue.

Conclusion and Further Studies

A lot of factors were taken into consideration to analyze whether it will be a good idea to invest in an Airbnb hotel in Los Angeles, California. As shown during the analysis, whether you are looking at revenue, ratings or popularity, there are property types and neighborhoods that stand out.

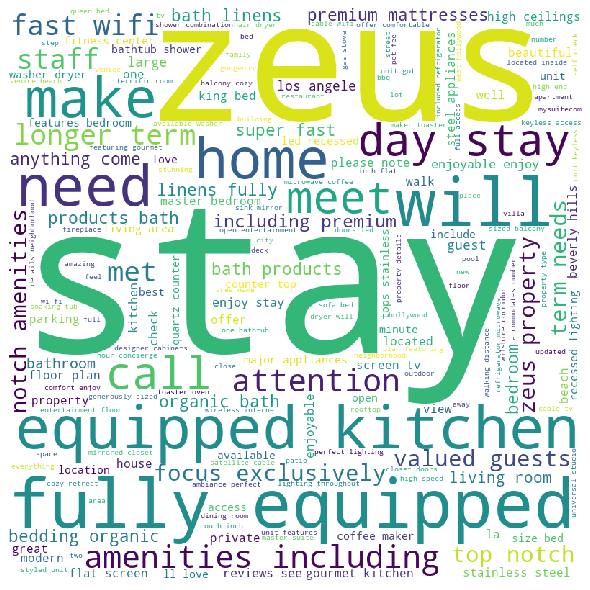
For property types, houses and apartments were most popular in terms of number of bookings and estimated total revenue. This information can be helpful if you plan on hosting, as most people stay or own houses or apartments, you can “invest” by opening up your house or apartment for short term stays. As for the neighborhoods, Venice and Hollywood are the areas with most bookings and estimated total revenue. When looking to invest in owning property to “rent” out, these areas can be taken into consideration.

It was interesting to see that the analysis of the properties and neighborhoods for review ratings were different. This might be because the less reviews a listing has, the higher the chances that the average will not be affected. As such, we can still take the recommendation by revenue and popularity, as more than 85% of our listings actually have a 90 or above score rating.

Finally, the top host analysis shows us interesting results when we zoomed in on the listings. Specifically for the neighborhood data, we saw that Hollywood Hills West, while having less bookings, have a significantly greater revenue intake.

For further studies, exploring time series data if available, and doing statistical analysis to predict demand of accommodation in the neighborhoods of Los Angeles will be beneficial for an investor as well. Additionally, if all active listings data is available, it will be interesting to see if it is possible to predict if a listing will be booked or not.

Appendix



The figure above shows the top words in the description for those who have the highest estimated total revenue.

References

Adamiak, C. (2019) Current state and development of Airbnb accommodation offer in 167 countries, Current Issues in Tourism, DOI: [10.1080/13683500.2019.1696758](https://doi.org/10.1080/13683500.2019.1696758)

Guttentag, D. (2016) Why tourists choose Airbnb: A motivation-based segmentation study underpinned by innovation concepts https://uwspace.uwaterloo.ca/bitstream/handle/10012/10684/Guttentag\_Daniel.pdf