AirBnB in San Francisco: Opportunities & Possibilities

GROUP 6

11/24/2019

BAX 401 Introduction to Business Analytics

Zunfeng Huang
Mitesh Ranmal Jain
John Elmer Loretizo
Mayank Mani
Shalini Mishra
Prakriti Rastogi

Table Of Contents

Executive Summary

Airbnb is one of the classic examples of product disruption. Founded in 2008, it became the first online platform that allows hosts to put up their apartments and homes on the site to be rented by guests. Today, with more than 2 million listings in over 191 countries, Airbnb's biggest challenge has been to help the hosts increase the number of bookings that will, in turn, help the company thrive in the long run. In this report, we focused on two key items that most consumers consider when booking a listing – price, and ratings, the factors that affect each one separately and jointly. To this end, we built two separate models for price and ratings respectively and found that prices are usually affected by the amenities and neighborhood while ratings are primarily affected by host responsiveness. Upon further analysis, it was discovered that being a super host drives out higher ratings for users that allows them to leverage and charge higher nightly rates. With that in mind, we have come up with the following recommendations:

- *Property-centered improvements.* Factors affecting price are centered on property attributes. Thus, it is recommended that the company educate hosts on property acquisition, improvement as well as maintenance.
- *Host-User relationship improvements*. Key factors impacting ratings are centered on how established communication is between a host and a user. Thus, knowledge about best practices in the hospitality industry may be imparted to hosts.
- *Super host business model.* Super host is a factor that's overlapping for both price and ratings. To this end, recommendations above helps hosts improve as super hosts.

Further analysis revealed that conversion of hosts to super hosts provide a monthly increase in revenue of 2%. Thus it may be favorable to highlight this incentive to the hosts.

Airbnb: Disruption in the Hospitality Industry

The hospitality industry is one of the most lucrative and consistently growing sectors over the last 10 years. With a gross booking of $185 billion in 2019[1] and a CAGR of 6%, hotels are the outstanding segment in the industry. This growth has attracted disruptions and new players alike. Concepts like serviced apartments, hostels, home-owned properties, and time-sharing have sprung up and re-designed how people evaluate their options when planning a vacation or travel. Airbnb has emerged as one of the most favored avenues for travelers. Operating as a marketplace platform and without directly providing a product to consumers, Airbnb has tapped into revenues of lower-end hotel groups. Since its inception in 2008, its revenue has increased to $2.6 billion in 2018 and in the process, Airbnb has established itself as the largest peer-to-peer hospitality service.

Competition

Airbnb has few competitors following the same C2C model, with HomeAway and VRBO (merged as one) being one of the major competitors. In 2017, HomeAway had gross bookings amounting to $8.7 billion[2]. One thing that HomeAway[3] does differently from Airbnb is that it shows the hosts how much they can earn based on the number of bedrooms, bathrooms and location of the accommodation. Other notable competitors are Homestay.com, Wimdu, Couchsurfing, Roomorama, etc.

---

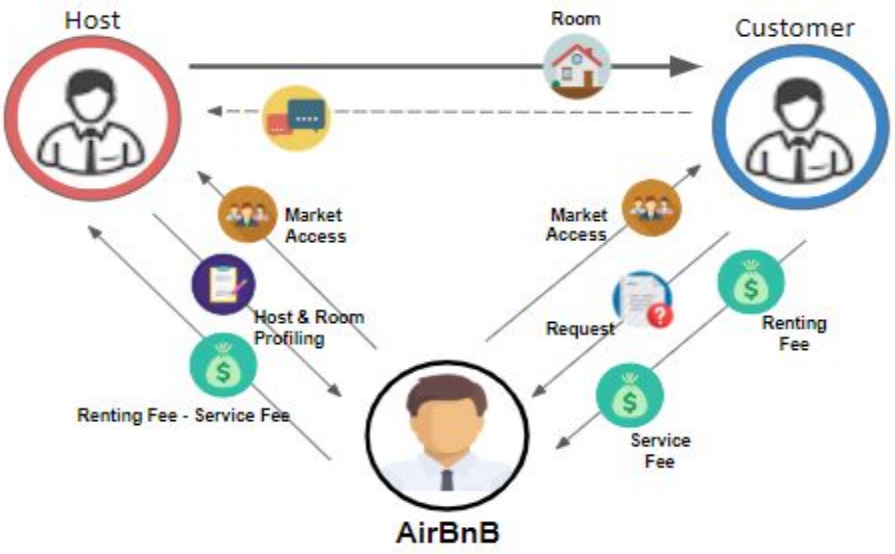
[1] https://www2.deloitte.com/content/dam/Deloitte/us/Documents/consumer-business/us-consumer-2019-us-travel-and-hospitality-outlook.pdf

[2] https://www.forbes.com/sites/briansolis/2018/09/12/homeaway-delivers-exceptional-customer-experiences-by-focusing-on-what-not-to-do/#236f97424e4c

[3] Check Appendix

Products & Services

Airbnb follows a Customer to Customer Business(C2C) model. This type of business model facilitates business between two sets of customers. In Airbnb's case, it's the hosts and the guests. Airbnb acts as an online marketplace where these customers can find each other otherwise difficult to locate



There are three main products and services provided by Airbnb

i. **Experiences** - Airbnb listings are unique in a way that they are provide novel activities and designs hosted by locals. Hosts offer their guests local knowledge through guided tours, meetups, special events etc. Hosts offer their guests special knowledge, unique skills, and inside access to local places and communities that otherwise guests creating lasting connections and treasured memories.

ii. **Stays** - Airbnb has listings around the world. Many hosts provide their properties with unique customizable experiences under affordable prices

iii. **Adventure-** Multi-day trips led by local experts - activities, meals and stays included

Existing Channels

Airbnb lets the hosts decide the price for their listings. They suggest hosts research for comparable listings in your city or neighborhood to get an idea of market prices.

For additional fees, hosts have the flexibility to add it in their total price or in their price settings.

Along with this Airbnb also has a feature called "Smart Pricing", which automatically increases or decreases the prices set by the hosts, when there is an increase or decrease in demand[4].

There is no fee for signup and posting a listing. Service fees for hosts, generally 3%, are charged once the host receives a reservation.

Airbnb has community standards on safety, security, and reliability, and hospitality which helps hosts earn great guest reviews. Also, options for guidance and tips to increase demand are provided.

Industry Practices

Traditionally, hotels priced their rooms on the basis of location, seasonality, demand, amenities and more importantly, star ratings. 5 star rated hotels have a more streamlined process for pricing their services. They rely on not just their own historical data, but also compare with other hotels

[4] https://www.airbnb.com/help/article/1168/how-do-i-turn-smart-pricing-on-or-off

with similar characteristics. However, in practice, most hotels go with competitive pricing, sometimes negotiating with customers, if required. The basic characteristics hotels consider for creating rate fences are:

Physical characteristics like Room-type, view, amenities etc. ,controlled availability, customer type,transaction type, service type. Most of the aforementioned "differentiators" rarely apply to Airbnb's business model. Other peer-to-peer companies that have built pricing tools are eBay and Uber. But eBay simply suggests a price based on what other similar products have sold for and lets the seller decide their own price. Uber, on the other hand, uses very sophisticated techniques for calculation of price but does not involve either of the end users in the decision process. Even though Airbnb has a unique business model, it still has to let demand drive prices in order to maximize revenue.

Data Description

We have data for ~8.1K Airbnb listings in San Francisco which are active since 2015. Each row corresponds to a listing in SF identified by a unique listing ID.

Each row contains listing information on property type, occupancy, availability, listing price, neighborhood, additional fee etc.We have no customer data

We also have data for the host of the listing, columns such as host ID(their unique ID), number of listings under them, response rate, etc. Summarized review and ratings data are also available for these

Nearly 20% of the listings had incomplete data for ratings and reviews, we have removed those listings from our analysis dataset.

Exploratory Data Analysis

We removed outliers from our data which were indicative of wrong data entry. For e.g. we had 6 records of bed type- couch with a much higher listing price than real beds.

We found out that 99% of the accommodations had "real" beds. Most of the Airbnbs were in the Mission neighbourhood, while Airbnbs in Presidio Heights had the highest median prices($275) which was $60 more than the next best - Marina with a median price of $211.5. Russian Hill, Pacific Heights and Financial Districts round of the top 5 neighbourhoods based on median price. We also found that properties with higher price points usually received better ratings.

Superhost Vs Host: We found that a superhosts get 10 times more reviews than a normal host and they are booked more often too. Surprisingly, their listings have a median price less than normal host by ~3$. Given the minimum difference, there is a huge traction in terms of guests' responses.[5]

Problem Formulation

Airbnb caters to two types of customers - guests and hosts. For guests, the ease of booking and how fairly the listings are priced plays a significant role in driving their loyalty with Airbnb. Similarly, how often a host listing is booked and rental fee affects the association of a host with Airbnb. Hence, we realized listing price and ratings are features that affect both hosts and guests in one way or another. In order to stay on top, Airbnb has to frame strategies around these to stay ahead in their game compared to its competitors.

[5] Charts added in Appendix

In our EDA exercise, we observed that the guests were willing to pay more for places with more reviews. Even having one review as compared to none made a huge difference in setting the perception about a listing.

We also found that listings on the higher end of the price spectrum tend to have higher ratings on average. Hence, in order to better understand the effect of price and ratings on value addition for Airbnb, we took the following steps:

1. Built a linear regression model to understand which factors significantly affect pricing.
2. Built a random forest classifier model to discern those aspects of a listing that affect the ratings received on the portal.

In an ideal setting, price and ratings would be independent of each other. However, in real life, when a guest is providing reviews, they would look at it from the perspective of whether the price was justified or not. Even while booking their stay, guests check the reviews and judge whether the price would be worth the amenities.

Model Results

Model 1:

To understand the relationship between the listing price and various listing variables, we ran a linear regression model with price as a dependent variable.

Given the categorical nature of most of our variables, we dummy encoded i.e. created n-1 columns for each. While running preliminary checks before regression, we noticed that our data is highly skewed. After log transformation, the required normality condition was passed.

We found the top 20 variables based on their p-value and we interpreted the beta values to understand their effects on price.

Model 2:

We couldn't use a simple linear regression model as our ratings data was extremely left-skewed. After a failed normality test in the preliminary analysis, we decided to use decision tree algorithms to understand major variables affecting ratings. We used the random forest classifier model to control overfitting.

On comparing the set of top 10 important features between the two models, variables like 'host_is_superhost' variable, a categorical variable mentioning whether a host is super host or not (Y/N), host_listings_count(total listings assigned under a host), cleaning fee, etc turned out to be common. Keeping in view the business model of Airbnb and our problem statement, we decided to focus on the 'host_is_superhost' variable to design our solutions.

Superhost badge is awarded to hosts who meet certain criteria: 80% 5-star reviews, no-host initiated cancellation for at least a year, at least 10 guests and a response rate of at least 90%.

Recommendation:

With the realization that certain factors have a direct effect on the revenue and rating of each listing, we analyze avenues to increase the price and rating for the properties and what measures can be taken to improve them:

i. To increase prices:

- Reduce the minimum listing nights. The model showed that properties that allow guests to stay for shorter durations can command higher prices than those which do not.

Customers treat such properties as being more flexible and suited to short trips. Airbnb can encourage the hosts to shorten the minimum number of nights allowed for their listings.

- Improve Amenities. Amenities like wi-fi, ac and tv, security system, linen, child-friendly, basic cooking equipment, and washer/dryer are valued by customers and have a positive impact on the price of the property. The key takeaway is that hosts can update such utilities and start charging higher prices.
- Focus on hotspots. Geography plays an important role in determining the price of a listing. If a host has multiple properties and plans to invest in new ones, it is recommended that they invest in hot spot areas such as Downtown, San Francisco or the Financial District.

To this point, one can see that attributes related to the property are the same factors that influence the listing price. Thus, centering on programs that educate hosts right from acquisition down to the maintenance of properties can help hosts boost prices.

ii. To improve ratings:

- Responsiveness: Hosts who respond frequently to guests receive better ratings than non-responsive hosts. Airbnb can encourage the practice of replying to the guests and educate the hosts with the appropriate responses for customer queries. The response rate can be monitored under the “Standards” submenu in the “Stats” section of the profile. Auto-suggested responses like Gmail can come in handy and improve the response-frequency as well as the ratings of the hosts.

- Number of listings: We found that hosts who have multiple properties tend to have higher ratings than single-listing hosts. The reason for this may be attributed to the experience of hosts with more listings against those who have. It provides for a greater avenue for hosts to learn and adapt to consumer preferences that can be applied to all their listings.

In the same manner that prices are centered on property improvement, one can see that the host-consumer relationship is a critical aspect in improving a listing's ratings. Thus, training on the best practices in hospitality can an advantage that Airbnb can impart to its hosts – a practice that hotels have been doing from the very beginning.

iii. Tiered Host Badge

There is a behavioral reason when it comes to guest's response on listings by super-host. Super host badge is a niche status among the hosts. For guests its an added assurance backed by Airbnb that the host can be trusted. Making it easier for hosts to become super hosts will reduce the importance of this status. Also, the stringent criteria makes it difficult for potential super hosts candidates to attain the status, reducing their chances of engagement with Airbnb.

Airbnb can instead add tiered badge system with incentives like added points for increased response rate , etc. There can be a bronze, platinum, gold badge for hosts thus facilitating easier movements for hosts who are relatively new and are loved by the guests. Also, at the same time the niche status is not dissolved by adding levels.

Conclusions

This report served as validation as to how explicit attributes present on a listing as well as for those related to hosts affect both prices and ratings. From the factors discovered as well as the recommendations given, we can see how improving listing prices are centered on the property itself. One can validly infer from this that any form of property improvement can help hosts leverage on increasing their prices. In the same manner, factors discovered and recommendations given for ratings are centered on host improvement. One of the factors revealed in this report is how becoming a super host affects both prices and ratings.

Thus any recommendation that strives to improve on both the price and ratings front would help in someone becoming a super host. Furthermore, the conversion of hosts to super hosts can boost monthly gross revenue by 2%[6] for the company. Should there be more sophisticated techniques that can be employed, it is very much welcomed to better enhance the results and output of the report. To end, this report can serve as a base guide for the company in enhancing programs for Airbnb hosts. It will serve as a manifestation of how the company is dedicated to helping regular household owners earn money on the platform – a core philosophy that will make Airbnb thrive for generations to come.

[6] More info on calculation in appendix

References

Dolan, R. J., & Simon, H. (1996). Power pricing. New York: Free Press

Kimes, S. E., & Wirtz, J. (2003). Has revenue management become acceptable? Findings from an international study on the perceived fairness of rate fences [Electronic version]. Retrieved from Cornell University, School of Hotel Administration site: http://scholarship.sha.cornell.edu/articles/946

Guo, X., Ling, L., Dong, Y. and Liang, L. (2019). COOPERATION CONTRACT IN TOURISM SUPPLY CHAINS: THE OPTIMAL PRICING STRATEGY OF HOTELS FOR COOPERATIVE THIRD PARTY STRATEGIC WEBSITES.

D. Hill, "How much is your spare room worth?," in IEEE Spectrum, vol. 52, no. 9, pp. 32-58, Sept. 2015. URL: shorturl.at/gELWX

Gutt, Dominik and Herrmann, Philipp, "Sharing Means Caring? Hosts' Price Reaction to Rating Visibility" (2015). ECIS 2015 Research-in-Progress Papers. Paper 54. ISBN 978-3-00-050284-2 Retrieved from: https://aisel.aisnet.org/ecis2015_rip/54

Neumann, Jürgen and Gutt, Dominik, (2017). "A HOMEOWNER'S GUIDE TO AIRBNB: THEORY AND EMPIRICAL EVIDENCE FOR OPTIMAL PRICING CONDITIONAL ON ONLINE RATINGS". Retrieved from: https://aisel.aisnet.org/ecis2017_rp/65

Appendix

Linear Regression Model

(Intercept)	4.16E+00	1.92E-01	21.664	< 2e-16	***
neighbourhood_cleansedCastro/Upper Market	4.28E-01	4.38E-02	9.767	< 2e-16	***
neighbourhood_cleansedHaight Ashbury	3.81E-01	4.48E-02	8.49	< 2e-16	***
neighbourhood_cleansedMarina	5.52E-01	5.31E-02	10.405	< 2e-16	***
neighbourhood_cleansedMission	3.80E-01	4.09E-02	9.28	< 2e-16	***
neighbourhood_cleansedNob Hill	4.30E-01	4.96E-02	8.673	< 2e-16	***
neighbourhood_cleansedNoe Valley	4.15E-01	4.48E-02	9.271	< 2e-16	***
neighbourhood_cleansedNorth Beach	5.34E-01	5.97E-02	8.95	< 2e-16	***
neighbourhood_cleansedPacific Heights	6.52E-01	5.51E-02	11.835	< 2e-16	***
neighbourhood_cleansedPotrero Hill	4.19E-01	4.84E-02	8.653	< 2e-16	***
neighbourhood_cleansedRussian Hill	5.16E-01	5.54E-02	9.298	< 2e-16	***
neighbourhood_cleansedSouth of Market	5.27E-01	4.52E-02	11.638	< 2e-16	***
neighbourhood_cleansedWestern Addition	4.49E-01	4.26E-02	10.545	< 2e-16	***
room_typeHotel room	-5.46E-01	5.49E-02	-9.946	< 2e-16	***

room_typePrivate room	-3.87E-01	1.69E-02	-22.919	< 2e-16	***
room_typeShared room	-1.11E+00	5.10E-02	-21.745	< 2e-16	***
accommodates	9.77E-02	6.60E-03	14.807	< 2e-16	***
bedrooms	1.91E-01	1.17E-02	16.311	< 2e-16	***
minimum_minimum_nights	-8.15E-03	4.84E-04	-16.848	< 2e-16	***
with_air_conditioning	2.38E-01	2.22E-02	10.693	< 2e-16	***

availability_30	5.02E-03	6.64E-04	7.557	4.92E-14	***
neighbourhood_cleansedInner Richmond	3.41E-01	4.86E-02	7.014	2.65E-12	***
cleaning_fee	7.85E-04	1.13E-04	6.968	3.65E-12	***
neighbourhood_cleansedDowntown/Civic Center	3.05E-01	4.70E-02	6.486	9.72E-11	***
neighbourhood_cleansedTwin Peaks	4.40E-01	6.85E-02	6.417	1.52E-10	***
neighbourhood_cleansedInner Sunset	3.25E-01	5.23E-02	6.22	5.42E-10	***
extra_people	1.19E-03	1.94E-04	6.148	8.52E-10	***
neighbourhood_cleansedFinancial District	3.99E-01	7.07E-02	5.644	1.76E-08	***
neighbourhood_cleansedBernal Heights	2.44E-01	4.33E-02	5.634	1.87E-08	***

with_wifi	3.27E-01	6.38E-02	5.128	3.05E-07	***
neighbourhood_cleansedChinatown	3.58E-01	7.01E-02	5.103	3.48E-07	***
availability_365	-2.66E-04	5.38E-05	-4.934	8.35E-07	***
neighbourhood_cleansedGlen Park	3.26E-01	6.91E-02	4.722	2.41E-06	***
beds	-4.63E-02	9.89E-03	-4.687	2.85E-06	***
calculated_host_listings_count	-7.73E-04	1.71E-04	-4.535	5.91E-06	***
neighbourhood_cleansedPresidio Heights	5.24E-01	1.21E-01	4.339	1.46E-05	***
host_is_superhost	5.46E-02	1.28E-02	4.267	2.02E-05	***
neighbourhood_cleansedOuter Richmond	2.02E-01	5.10E-02	3.973	7.19E-05	***
property_typeResort	9.80E-01	2.54E-01	3.856	0.000117	***
minimum_maximum_nights	-7.32E-04	1.98E-04	-3.702	0.000217	***
guests_included	2.01E-02	5.54E-03	3.623	0.000294	***
bathrooms	-2.92E-02	8.14E-03	-3.582	0.000344	***
with_cable_tv	4.60E-02	1.30E-02	3.537	0.000409	***
maximum_nights	2.20E-04	6.33E-05	3.47	0.000525	***
property_typeHostel	-3.86E-	1.15E-	-3.3	0.0007	**

	01	01	66	69	*

Feature Importance Table for Random Forest Regression

colnames	importance	coef	p-value
calculated_host_listings_count	0.10932006	0.183	0
availability_365	0.07469601	-0.0043	0
host_is_superhost	0.0632261	3.3866	0
availability_90	0.04525052	-0.0122	0.31
cleaning_fee	0.04190786	0.0027	0.077
availability_60	0.0396693	0.0055	0.815
availability_30	0.03163466	-0.0143	0.555
maximum_nights	0.02459353	-5.692E-05	0.37
extra_people	0.02173554	0.0049	0.093
security_deposit	0.01980149	0.00009403	0.529
accommodates	0.01942523	0.3379	0.001
neighbourhood_cleansed_Downtown/Civic Center	0.01793499	0.464	0.318
beds	0.01717218	-0.342	0.022
calculated_host_listings_count_shared_rooms	0.0165772	-0.7136	0
bathrooms	0.01529125	-0.3542	0.003
calculated_host_listings_count_private_rooms	0.01494956	-0.2476	0
minimum_nights	0.01483812	0.00005685	0.37
calendar_updated_86 months ago	0.01482376	-69.2087	0
host_response_rate	0.0122713	0.0224	0.137
bedrooms	0.01214232	0.2049	0.243
property_type_Cottage	0.01186675	0.5568	0.778

calculated_host_listings_count_entire_homes	0.01135502	-0.1972	0
calendar_updated_2 weeks ago	0.01070712	4.6282	0

Total Listings: 8100

Assuming a 5% Conversion Rate: 227

Number of non Superhost: 4557

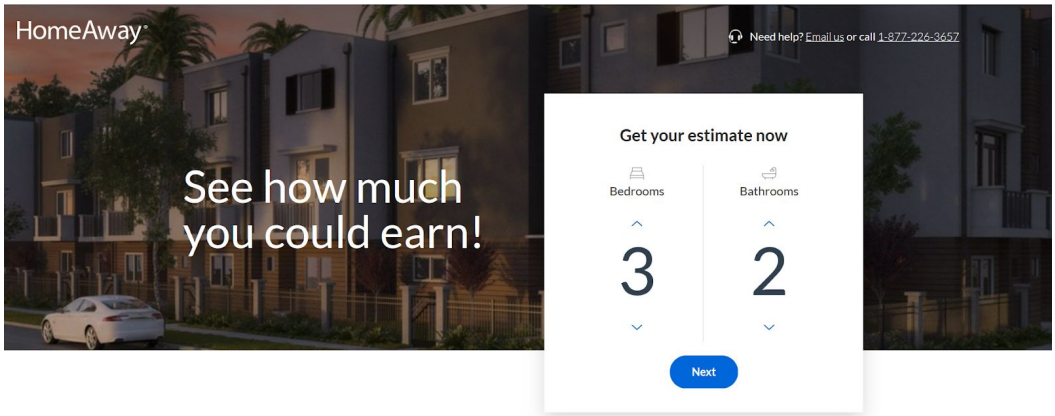
Superhost: 3546

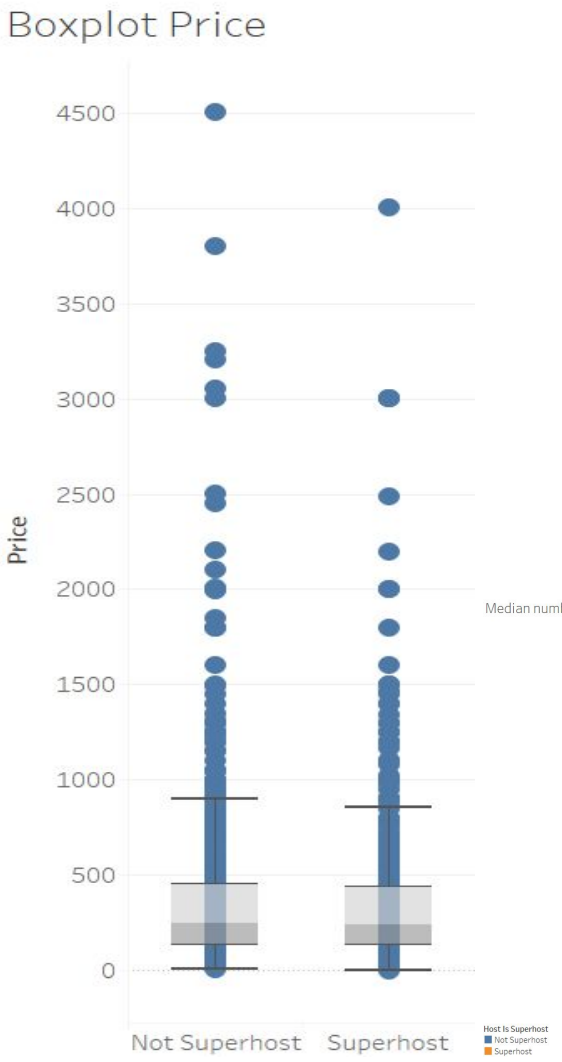
Host Type	Median Price	Number of Bookings	Projected Earnings	
Regular Host	153	3	459	
Super Host	150	30	4500	4041

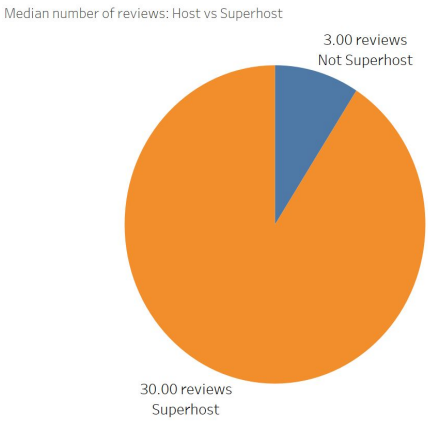
	Host Type	Company earnings	Grand Total
Status Quo	Regular Host	4555 * 153 * 2 = 1,394,442	3,522,042
	Super Host	3545 * 150 * 4 = 2,127,600	
After conversion	Regular Host	4328 * 153 * 2 = 1,324,368	3,587,568
	Super Host	3772 * 150 * 4 = 2,263,200	

We can see an estimated 2% increase in monthly revenue just by converting 5% regular hosts into superhosts.

3.


HomeAway
Need help? Email us or call 1-877-226-3657
See how much you could earn!
Get your estimate now
Bedrooms
Bathrooms
3
2
Next


Boxplot Price
4500
4000
3500
3000
2500
2000
1500
1000
500
0
Price
Not Superhost
Superhost
Host Is Superhost
Not Superhost
Superhost


Median number of reviews: Host vs Superhost
3.00 reviews
Not Superhost
30.00 reviews
Superhost

Different neighbourhoods command different prices: