30401 Mathematics and Statistics (Module II) Group 7 Project

ABSTRACT

This paper investigates the causes of price variations between various accommodations under the AirBnB management, trying to individuate different confounders. The data provided some general trends and a city-by-city prospectus. We found that costs depended on some variables: understanding this heterogeneity is important to improve pricing, advertising, guidance for landlords and customers.

INTRODUCTION

Starting from the Dataset, we identified 5 clusters selecting variables with logical relationships between them: “placement”, “host”, “feedback”, “location”, and “services”. After having verified the effective validity of the clusters, we carried out a first more in-depth general study. We proceeded as follows:

1. General comparison between the variables of the individual clusters;
2. General comparison between the variables of the different clusters;
3. General comparison between the different variables and the price;

After that, we repeated the same operations city by city to obtain more specific conclusions.

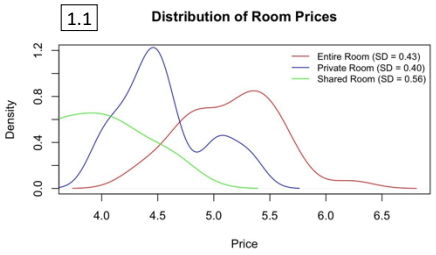
GENERAL TRENDS

How the different clusters affect price

An initial analysis of the clusters has revealed some general trends. The type of room is closely linked to the services offered: entire home/apartment and private rooms are characterized by a stricter cancellation policy, the absence of the possibility of instant booking and the presence of a cleaning fee. Vice versa, large-scale accommodations based on shared rooms have a more flexible cancellation policy, a greater possibility of instant booking and a less present cleaning fee, probably to try to fill all the seats. This trend also appears with respect to the number of beds, bathrooms, bedrooms and accommodates: the higher the number, the stricter is the cancellation policy, the possibility of instant booking is less frequently present, and the cleaning fee is much more frequent. However, this only held true for properties with no more than 5 reviews, as those, again, with more reviews likely fall under the category of dormitories or large-scale accommodations.

Linking together our previous findings, we proceeded to verify the impact of the variables on price. Entire houses have a higher price than private rooms, which have a higher price than shared rooms. Upon analyzing the standard deviation of this variable, we found that the standard deviation of shared rooms is higher than that of private rooms, indicating that the prices of shared rooms vary more widely than those of private rooms or entire home therefore more concentrated (1.1). On the other hand, private rooms have a lower standard deviation and a higher mean. This observation reflects the difference in the prices. The insights gained from this analysis could be useful for individuals looking to enter the platform by managing an apartment with several rooms or renting out the entire apartment. The choice must be made obviously considering the other cofounders found inside this research. Some property types, timeshare and vacation home, for every city have generally the highest price, while dorm and hostel the lowest. The cancellation policy has a strong impact on log price: houses with super strict cancellation policy have a higher log price than strict, which is usually higher than moderate and flexible, this trend, however, differs across cities.

1.1



Upon analyzing the variable 'accommodates' and creating subsets for each neighborhood, we discovered that those neighborhoods with an average log price of 4.9 or higher typically have an average number of accommodates of at least 3. This finding suggests, as we confirmed graphically, that the most expensive neighborhoods are also the ones with the highest number of accommodates. This phenomenon may be indicative of higher costs for hosts to maintain their facilities, as well as a higher demand for these locations among users. Furthermore, the central positioning of these neighborhoods within the city further supports this observation. To conclude, we found some cofounders with no impact on the price: the host response rate, the time an accommodation has been available and the date of the first review. Meanwhile, the effect of instant booking will be explored city by city, since it is ambiguous.

Description and amenities are important

The following features justified higher prices for both accommodations, rarity, and cost-related services. Below we report the most significant observations about them.

Based on the data provided, it’s evident that TV is the most prominent amenity with over 70% of accommodations offering it, namely Boston (72.4%) Chicago (79.27%), DC (79.57%), LA (75.63%), NYC (64.08%), and SF (72.01%). The amenity has a positive impact of as high as 0.6 when combined with the service Cable.tv, making a remarkably positive impact on the overall value of the property.

Another crucial amenity is air conditioning, with over 89.78% of accommodations offering it in Chicago, 83.96% in NYC, and 78.55% in Boston. It has a positive impact from 0.38 to 0.50 in Boston. Contrary to what we might expect, air conditioning appears to be an amenity that increases the price of properties only in northern cities, perhaps because considered as a base service in California.

The data, then, suggests that hosts consider families as an important demographic, with over 52% of accommodations in Boston, 61% in Chicago, 61.37% in DC, 56.71% in LA, 41.99% in NYC, and 48.07% in SF catering to this demographic adding value between 0.37 and 0.46.

Furthermore, a positive social-related amenity is the suitability for events, with 7.42% of accommodations in LA offering it. The amenity had a positive impact of 0.44 therefore indicating a shift in price when it’s destinated to events.

Continuing, analyzing the not-so-common impacting features that underlined a premium accommodation we find gyms, with a percentage ranging from 7.28% in NYC to 14.52% in DC; doormen with only 8.16% of accommodations in NYC and an impact of 0.48; pools, present in 7.23% of accommodations in Chicago with a high impact of 0.46; and kitchen since it is a costly and paid service and brings value with a positive impact of 0.36 and 0.24 for LA and SF, respectively. There are, then, multiple amenities that affected the price of only some cities in particular and cannot be grouped for their impact: this was the case, for example, in DC where, probably due to the business and institutional character of the place, the dryers had a greater impact on the value of the property while self-check-in with lockbox was an indicator of a lower price, perhaps indicator of less hospitality and lower costs for the host.

Another example is indoor fireplace, a peculiar feature of the accommodations in SF and LA which, although not considered as a business or premium feature, added uniqueness to the offer with a positive impact of 0.33. Furthermore, in LA it emerged that amenity smoking allowed reduced the value of the accommodation, demonstrating once again how focusing on a single narrow slice of the market did not bring added value to the property and resulted in lower prices.

Looking at negative impacting amenities, however, we find primarily “pets living on the property” that despite being a rare feature, with a percentage ranging from 8.65% in Boston to 17.25% in LA, appears to have a negative impact on the overall value of the property up to 0.37 in DC. The same goes for the amenity “free parking on premises” which, even though it is most of the time present in less premium accommodations located far from the city center, was a not-so-common feature with a percentage ranging from 9.69% in NYC to 22.26% in Boston and a negative impact of -0.17. Concluding, while locks on bedroom doors, most present in private rooms and perceived as a necessity in dangerous zones, is a common characteristic with a percentage raging from 22.13% in SF to 25.16% in NYC and a negative impact of -0.26 on the price, translation missing is another popular feature with a percentage ranging from 29.96% in Boston to 40.52% in Chicago and having a negative impact of -0.27 on the value of the property, showing how communication is fundamental to maximize the potential value of the property.

We proceeded with the analysis of description, a fundamental field used by users to benchmark and understand in few lines the pros and cons of the accommodation. Our study indicated how different factors of the descriptions has a different impact on the log price: listings including descriptions of “bedrooms” (+0.32) "views", “modern”,

Boston Amenity	Percentage	Impact
TV*	72.4%	0,59
Air.conditioning	78.55%	0,50
Family	52.51%	0,46
Gym	11.79%	0,42
Doorman	8.42%	0,39
Pets.live.on.this.property	8.65%	- 0,12
Free.parking.on.premises	22.26%	- 0,17
translation.missing	29.96%	- 0,38
Lock.on.bedroom.door	23.79%	- 0,38
*Cable.TV	57.63%	0,074

Chicago Amenity	Percentage	Impact
TV*	79.27%	0,46
Pool	7.23%	0,46
Family	61.98%	0,43
Air.conditioning	89.78%	0,37
Gym	13.42%	0,36
Dog	9.36%	- 0,10
Lock.on.bedroom.door	26.22%	- 0,21
Pets.live.on.this.property	17.16%	- 0,30
Cat	6.16%	- 0,40
translation.missing	40.52%	- 0,42
*Cable.TV	45.25%	0,13

DC Amenity	Percentage	Impact
TV*	79.57%	0,44
Family	61.37%	0,38
Kitchen	92.74%	0,34
Suitable.for.events	7.0%	0,29
Dryer	80.75%	0,25
Gym	14.52%	0,25
Self.Check	20.04%	- 0,12
Lockbox	10.48%	- 0,15
translation.missing	35.39%	- 0,27
Dog	6.36%	- 0,30
Pets.live.on.this.property	11.43%	- 0,37
*Cable.tv not relevant (0.017)		

LA Amenity	Percentage	Impact
Family	56.71%	0,46
TV*	75.63%	0,46
Suitable.for.events	7.42%	0,44
Indoor.fireplace	22.68%	0,41
Kitchen	88.71%	0,36
Dog	10.86%	- 0,14
Smoking.allowed	7.66%	- 0,17
Cat	5.42%	- 0,19
Lock.on.bedroom.door	22.67%	- 0,21
Pets.live.on.this.property	17.25%	- 0,21
translation.missing	37.86%	- 0,39
*Cable.TV	44.8%	0,14

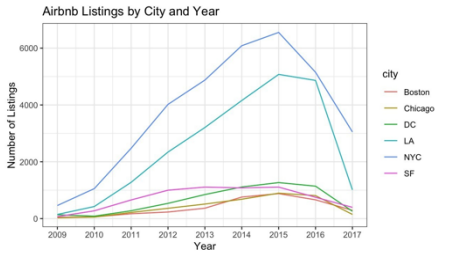
NYC Amenity	Percentage	Impact
Doorman	8.16%	0,48
Family	41.99%	0,41
Air.conditioning	83.96%	0,38
TV*	64.08%	0,38
Gym	7.28%	0,35
Free.parking.on.premises	9.69%	- 0,15
Pets.live.on.this.property	10.09%	- 0,21
translation.missing	37.4%	- 0,26
Lock.on.bedroom.door	25.16%	- 0,26
*Cable.TV	39.28%	0,12

SF Amenity	Percentage	Impact
Family	48.07%	0,37
TV*	72.01%	0,34
Indoor.fireplace	17.72%	0,33
Bathtub	7.48%	0,25
Kitchen	86.87%	0,24
Dog	9.36%	- 0,03
Pets.live.on.this.property	15.6%	- 0,11
Cat	5.49%	- 0,15
Lock.on.bedroom.door	22.13%	- 0,27
translation.missing	33.34%	- 0,29
*Cable.TV	47.2%	0,11

"open", "outdoor" spaces, and proximity to the "heart" of the city or a "village" are, in fact, usually associated with higher prices. Additionally, we found that descriptions characterized by certain words such as "steps", "beds", "entire", and "re", are related to higher prices of accommodation.
On the other hand, the presence of certain words such as "shared" (-0.29), "train" (-0.14), or "friendly" (-0.12) are linked to a more negative impact on the prices. Upon further examination of the cluster "location", we directed our attention towards the mode of amenities in each individual neighborhood. Our findings reveal that 78% of the amenities that stood out were related to "Wireless Internet". Upon delving deeper into the remaining amenities, we discovered a significant concentration of "Heating" in Central-East America, specifically in cities such as NYC, DC, Chicago, and Boston.
In summary, some unique amenities, while not excessively increasing the price of the accommodation, positively differentiated it from the others. Furthermore, it was recognized how hosts with greater attention to convenience, security, and clear communication were able to increase the value of their offer by creating a more desirable and lucrative listing on Airbnb. Finally, these results suggest that hosts should be aware and careful of their accommodation descriptions, avoiding emphasizing features that could lower the perceived value of their offer and highlighting those specific amenities that could lead to higher prices.

The link between the number of listings and the cities' policies

As it is shown by the graph, there was a boom in the service and demand followed by a rise in the offer between 2011 and 2015 with NYC being the leader of the trend. The trend appeared to be influenced by the policies implemented by the different local governments. In 2015, in fact, for example, San Francisco implemented new regulations requiring hosts to register their housing to the city and obtain a permit, followed by Chicago in 2016, and Washington DC in 2018.



Similarly, in 2018, Los Angeles and Boston passed new ordinances forcing hosts to obtain licenses and requiring them certain safety standards.
Again, NYC has had strict regulations against Airbnb rentals for years, making it illegal for most hosts to rent out their properties for less than 30 days: these regulations may have had a positive impact on the trend of listing as their percent change after the first boom dictated by the spreading of the platform remained relatively stable over the years, with a 7.69% increase in 2015 and a 21.41% decrease in 2016.
It's important to note that other factors, such as demand from travelers and housing stock availability, may also have contributed to shifts in Airbnb listings. Additionally, regulations and policies may change over time, leading to fluctuations in the trend.

A Simpson's paradox

In our analysis we also focused on the relationship between the property type and the room type with respect to premium houses, houses with a log price higher than the third quantile. The result resembled the Simpson's paradox. In fact, in all the cities in the most frequent type of property type, rooms were almost only entire rooms, while private and shared ones disappear completely. When we focused on this relationship city by city this trend disappeared. Note: doing this operation, we looked at the third quantile of each city, and thus in each analysis the parameter changed. We are aware of that, but we still found it relevant.

City	Property type	Entire home/apt	Private room	Shared room
All cities	Apartment	0,93	0,07	-
	Condominium	0,91	0,09	-
	House	0,93	0,07	-
Boston	Apartment	0,70	0,28	0,02
	Condominium	0,67	0,32	0,01
	House	0,33	0,67	-
Chicago	Apartment	0,67	0,32	0,01
	Condominium	0,49	0,49	0,02
	House	0,39	0,58	0,03
DC	Apartment	0,83	0,14	0,03
	Condominium	0,72	0,26	0,02
	House	0,45	0,53	0,02
LA	Apartment	0,66	0,30	0,04
	Condominium	0,52	0,46	0,02
	House	0,49	0,47	0,04
NYC	Apartment	0,51	0,46	0,03
	Condominium	0,65	0,35	-
	House	0,32	0,65	0,03
SF	Apartment	0,60	0,39	0,01
	Condominium	0,64	0,36	-
	House	0,52	0,47	0,01

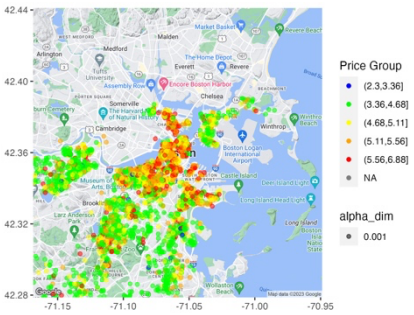
CITY BY CITY ANALYSIS

General approach

Wanting to group the data according to the geographical position, we realized that the variable 'neighborhoods', for reasons related to their size and boundaries, was the best indicator of location. Once sorted by mean price, the neighborhoods were further analyzed. For each other variable to be investigated, taking into consideration the mean price, the three most and least expensive neighborhoods were identified and compared based on them: if some particular cost trends, linked to the examined cofounders, had arisen, the analysis would have proceeded by enlarging the sample of neighborhoods to 10 to verify whether the trends still occurred. If so, therefore, it could be deduced that these cost differences were attributable to the analyzed variables and their influence on the price. Below we report the most interesting observations related to the neighborhoods.

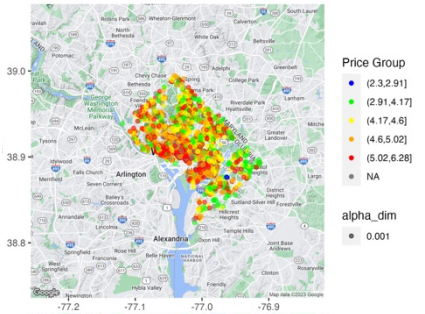
Boston

It was evident that for the best 10 and worst 10 neighborhoods, the mean price for accommodations with a moderate cancellation policy was higher than the ones with other cancellation policy types. Moreover, for the best 10 neighbourhoods, accommodations for which instant bookable was not available had a higher mean price than the ones that offered it.



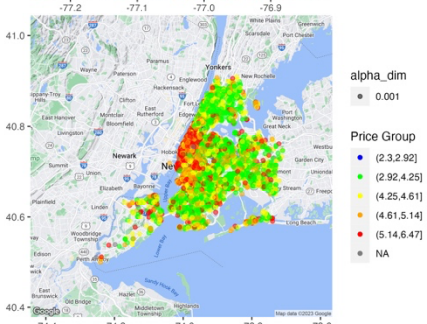
Washington DC

Differently from Boston, for the best 10 neighbourhoods, instantly bookable accommodations had a higher mean price than the ones for which it was not possible. Furthermore, accommodations of the same group, where the cleaning fee was not present, were characterized by higher average prices than the ones where it was present.



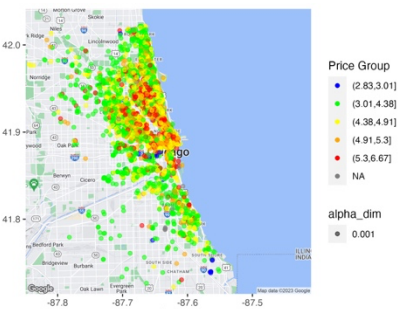
New York

In NYC, the number of accommodations offered by verified hosts was higher than those offered by hosts without verification (60% vs. 40%). Additionally, for the top 10 neighborhoods, the price of both types of accommodations fluctuated. However, the price of accommodations of verified hosts tended to cluster around the median price for each neighbourhood more often.



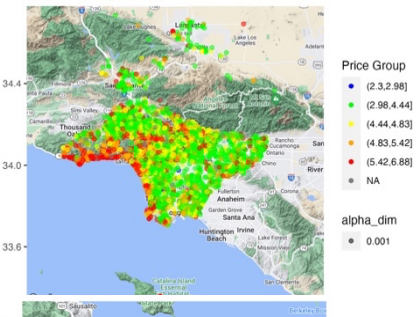
Chicago

Neighborhoods with average log price higher than 5 all have an average review score higher than the ones with a price lower than 5. Quality pays off! The higher the number of services offered, the more the score of the reviews rises as well as the price of the accommodations. This last trend also occurs in Boston and DC.



Los Angeles

For the best 10 neighborhoods in LA, the mean price for accommodations with a strict cancellation policy was higher than the one of housing without it.



San Francisco

If we divide the total number of reviews of each neighborhood for the number of accommodations of each neighborhood, we can see that the higher is the ratio, the lower the mean price of the neighborhood is. It could be hypothesized that the more reviews an accommodation has, the more likely there are negative reviews, thus forcing the host to lower the price of the accommodation.