

Airbnb “Superhost” Criteria: Does It Reflected Across All Listings?

*Analysis and evaluation of Airbnb’s criteria for Superhost badge,
focusing on hosts with multi-listings/properties in London*

By: Dinda Tisi Calista

Revisiting “Superhost” evaluation and star rating criteria in London area



- Superhost are the badge/status given by Airbnb to “experienced hosts” who satisfy several criteria^[1]
- However, the existing superhost criteria have limitation: for hosts with multi-listings, only one qualified property is needed. Even if their other properties got bad reviews and overall ratings, the host can still qualify for superhost status. To maintain customer’s trust, of course we need to avoid having this inconsistency.
- Therefore we are wondering if we should evaluate superhosts across all of their listings, or keep the existing criteria. The question we need to answer is: **Do hosts who have multi-listings got consistent review across all of their listings/properties?**
- *Please note that we are only interested in conducting this research in London area. This will be the scope of the analysis.*

[1]<https://www.airbnb.co.uk/superhost>

Datasets Introduction

- Data was collected on July 28, 2017, consisting of 64,155 listings in London. The CSV file was taken from <http://tomslee.net/airbnb-data-collection-get-the-data>
- The original dataset consists of 13 attributes. We will mainly use the attributes described below.

room_id	ID of the listing
host_id	ID of the host
room_type	“Entire home/apt”, “Private room”, or “Shared room”
price	The price (in \$US) for a night stay
reviews	The number of reviews that a listing has received.
overall_satisfaction	The average rating (out of five)

Cleaning & Extraction

The process was done using Python as Tableau can only handle limited rows. Steps taken:

- Remove records with `overall_satisfaction = 0`
- Filter out `room_id` with `< 7` reviews. This is important because superhost has minimum of 10 visits, and with average 70% of visit got reviews^[1], we can assume that a potential superhost has at least 7 reviews. Also, this help to minimize the possibility of having extreme rating value due to low number of review.

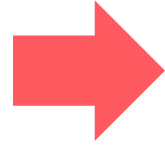
Dataset ready: 22,209 listings

Now the data is clean and only consist of rooms that has possibility to qualify for superhost evaluation.

[1] <http://tomslee.net/airbnb-data-collection-get-the-data>, <https://airhostsforum.com/t/how-many-of-your-guests-leave-reviews/2504/8>

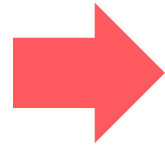
How Many Listings are Owned by Multi-Property Hosts and How Do They Perform?

- If there are only a few hosts with multi-listings, it is not necessary to continue this analysis.

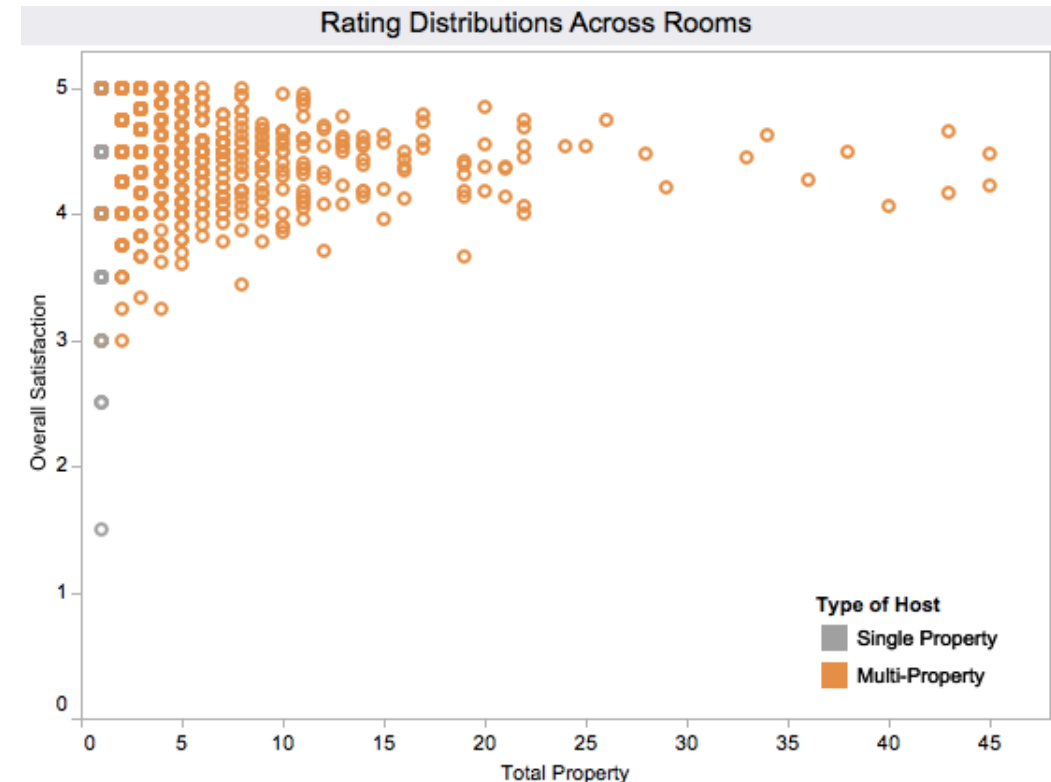


43.56% of all listings are owned by multi-property hosts

- We are also interested to find out about the distributions of overall satisfactions

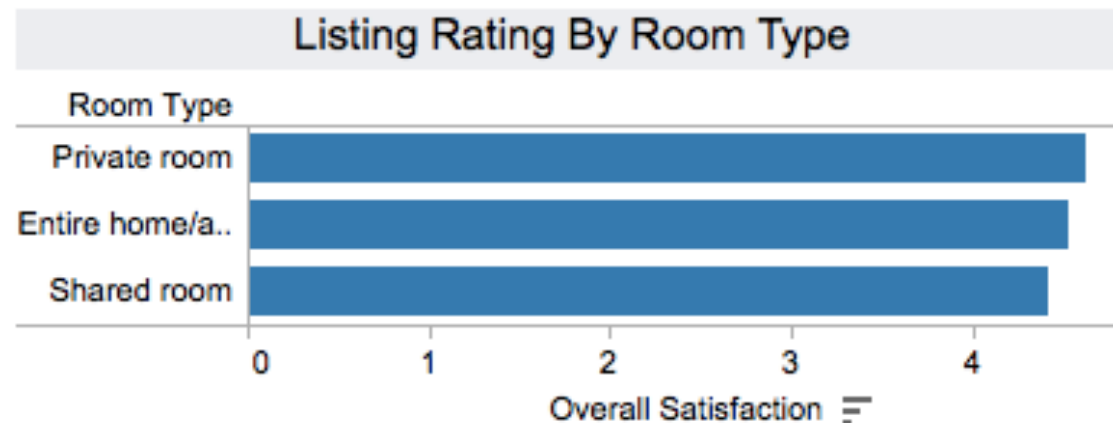


Insight: Overall satisfactions are more concentrated in listings owned by multi-property host, and not affected by the total number of properties owned.



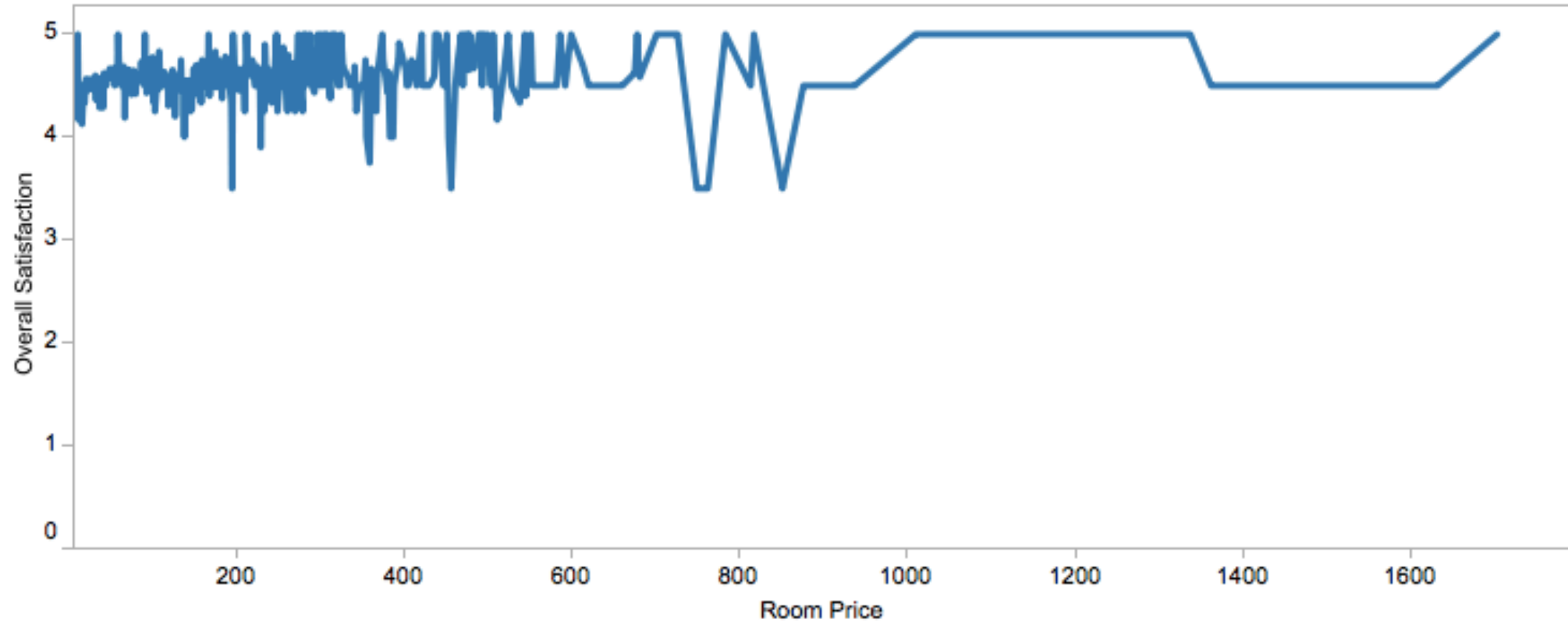
But How Does Overall Satisfaction Differ Across Room Types?

If room type affects the star ratings received, we should segment our analysis based on whether the host owns same or mixed room type.



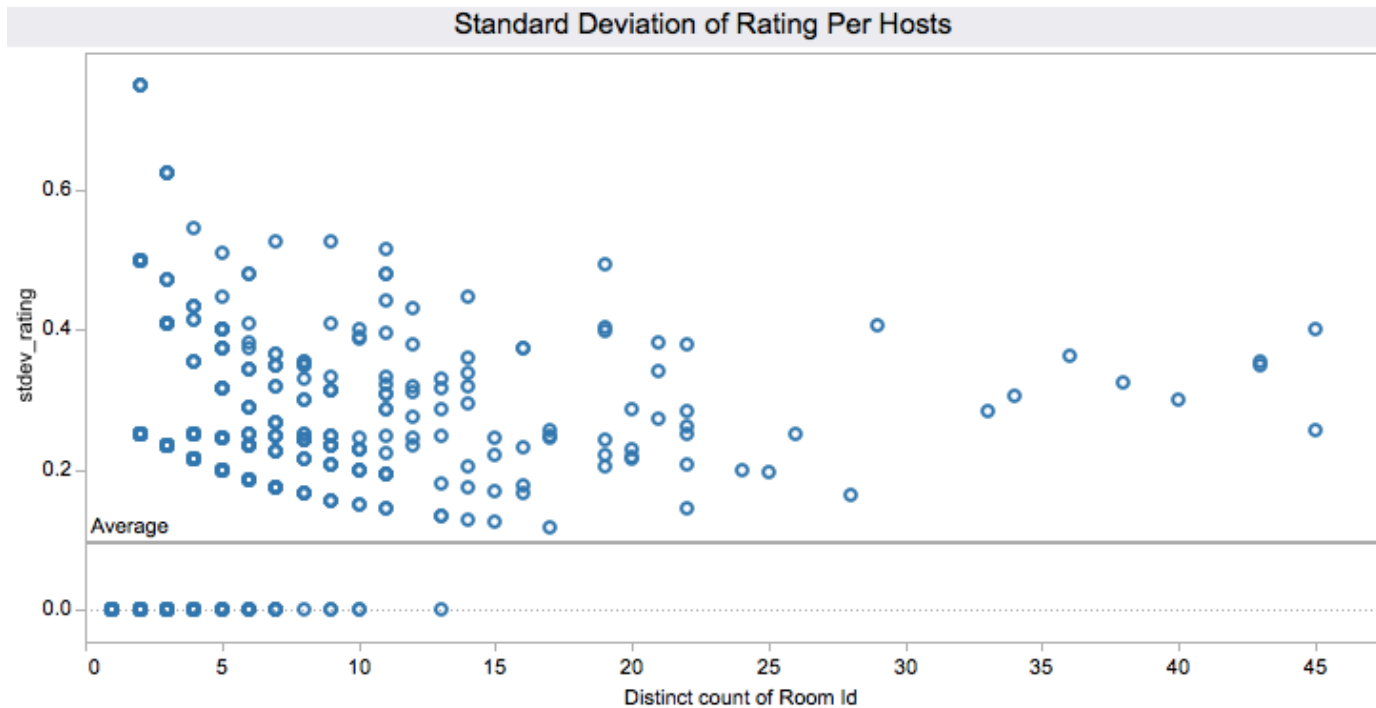
Insight: Room type does not correlate with ratings

Also No Pattern Between Price and Overall Satisfaction



Next, Looking At The Standard Deviation To See How Dispersed The Overall Satisfaction Are

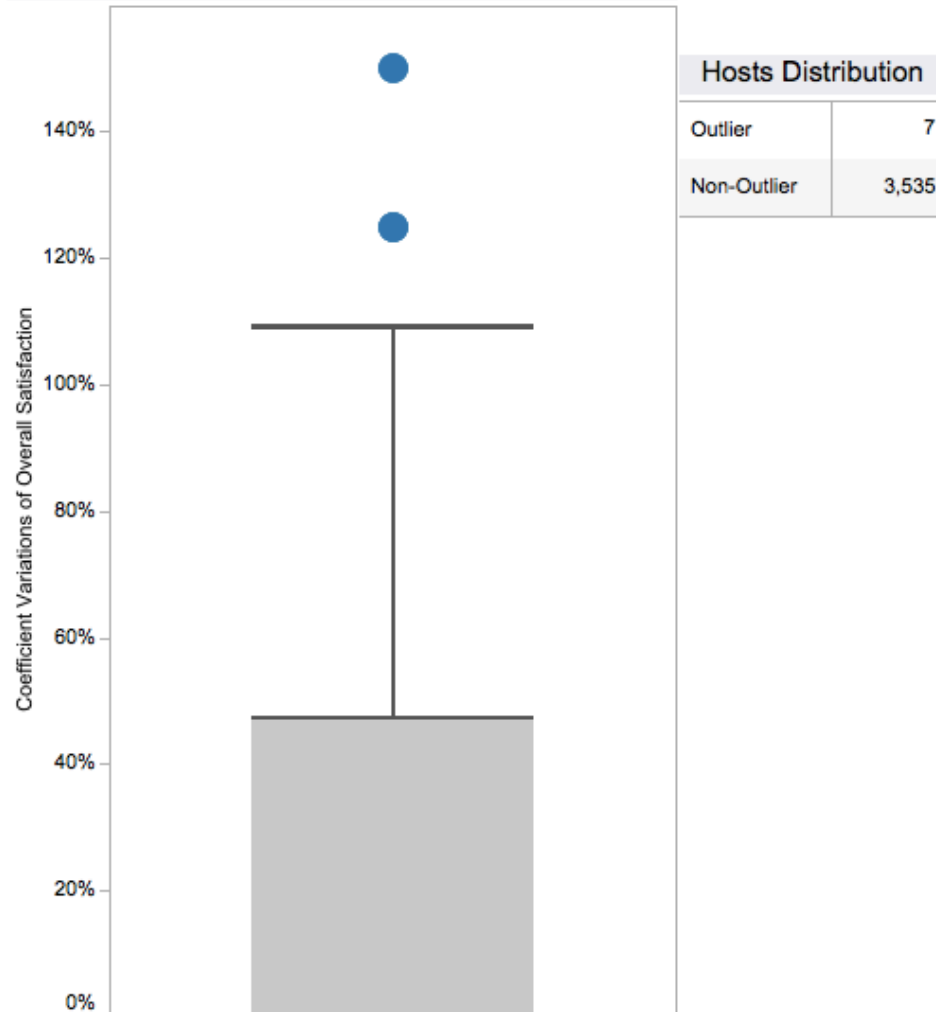
Now that we know there is no association between room overall satisfactions and types/no. of listings owned/price, we can start mapping the standard deviation.



But we want to identify and compare the consistency of overall satisfactions for each multi-host, not from the average rating of all properties. Hence, using coefficient of variations will be more suitable

Conclusion: Multi-Property Hosts Are Able To Maintain Consistent Ratings Across Their Listings

Coefficient of Variations of Ratings Across Listings Owned



- In order to determine whether to change the superhost criteria or not, we came up with the questions: **Do multi-property hosts got consistent reviews across all of their listings?**
- Looking at the Coefficient of Variance (CV) of the ratings received by multi-property hosts, we found that only 0.2% of hosts are outliers, in this case means their CV is higher than most hosts.
- This means 99.8% hosts have consistent ratings across their listing, which pretty much show almost all hosts with multi-property can maintain the visitors' satisfaction in all of their properties.
- **Thus, we don't need to change the superhost evaluation criteria**

Closer Look Into Some of the Outliers

- By putting boxplot into our CV distribution (refer to final visualization on slide 4), we can identify the outliers.
- Using simple SQL Query, we can check our conclusion and be convinced that there is no extreme differences between ratings of listings owned by multi-property Hosts. Here we'll see the details of two outliers, with host_id '19857936' and '9722879'.

```
1 select host_id, room_id, reviews, overall_satisfaction
2 from rated_host where host_id=19857936 or host_id=9722879
3 order by host_id
```

⚙️ Query Favorites ▾ Query History ▾			
host_id	room_id	reviews	overall_satisfaction
19857936	7212020	8	5
19857936	13756458	9	3.5
9722879	7067657	7	3
9722879	4782010	34	4.5
9722879	5866612	22	4

**End of
Document**