**INTRODUCTION:**

StayZilla is a Bengaluru-based Indian homestay network founded in 2005 as “Inasra” and rebranded as “StayZilla” in 2010 which acted as a marketplace for homestays and alternate stays in India, with around 55,000 stay options across 4,500 towns in the country. Although it had first-mover advantage in the space, its growth amped up only after it was rebranded. Its founder and CEO, Yogendra accepted that this was achieved at quite a high-cost and had a lot of pitfalls. He laid out a couple of reasons which led to the failure as follows: Supply-Demand Mismatch, Creating A Market and High Costs, Low Revenues.

**Supply-Demand Mismatch:**

There was supply – demand mismatch as the travel market in India does not experience a network effect. When they get a demand for the booking, they experienced bad supply for the need on the homestay**.**

**Creating A Market:**

The system of homestays is something very new in a country like India. For Stayzilla, there was no ready market to sell the product and, thus, it required investing in educating the market about the concept of homestay market place, how to use the product and even on how to use Internet to the users.

**High Costs, Low Revenues:**

In an industry where offers and promotions were the norm, Stayzilla’s team was focusing on marketing ROI and getting bookings without any discounts. Discounting-based growth rampant in the travel industry was another reason what led to the fall of Stayzilla. They Forced to match prices, but could not even recoup what they put in, necessitating a very large capital requirement, simply to sustain growth.

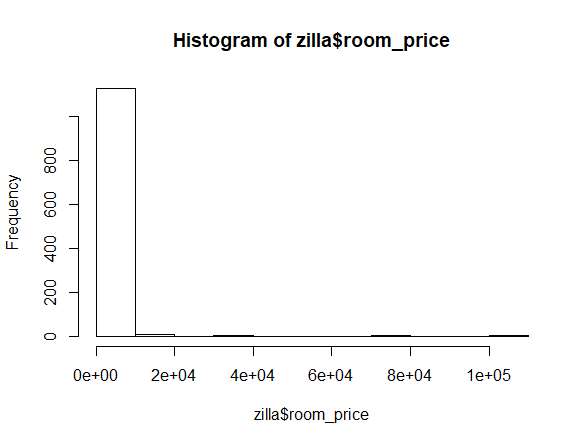
[Reference: https://inc42.com/buzz/stayzilla-shutdown/](https://inc42.com/buzz/stayzilla-shutdown/)

**Analysis based on Explanatory Data Analysis:**

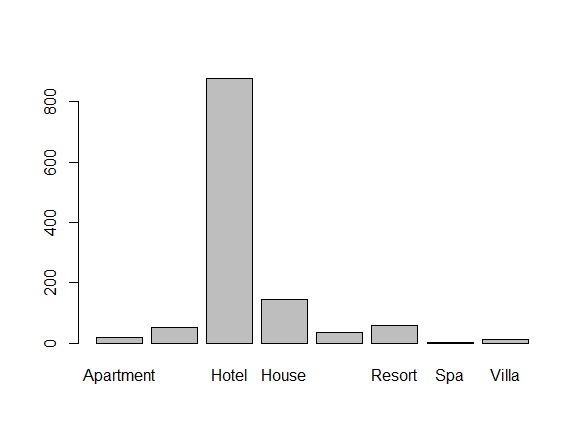
**On room price:**

Room prices are heavily right-skewed to a handful of extravagantly prized luxury listings





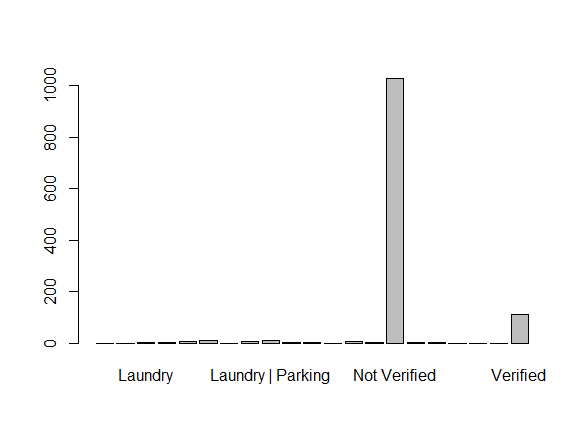
**On Property type:**



The vast majority of the properties on the site are not homes but (self-described) hotels. This is a prominent difference between StayZilla and the US equivalent, AirBnB, which does not allow (explicit) hotels in its listings. Hotels are subject to a raft of legal oversights that listing on a homestay platform sidesteps, a fact which has led to huge and very public fights between AirBnB and many of the US cities it operates in; evidently this is much less of or not even a concern in India.

**On Service value:**





The vast majority of properties on the StayZilla platform seems to be unverified which is not trust worthy for the customers. **We can recommend them to make the properties verified to gain the trust based relationship from the customers.**

**Poisson Regression:**

To predict the number of occupants and the stay period of the customers per booking from the customer end, with the given features, we created a regression model. The reason for selecting poisson regression is because the response variables “No.ofoccupants” and “StayDays” are count variables. Count variables often follow a Poisson distribution, and can therefore be used in Poisson Regression Models to model the counts of event occurrence.

**Steps followed:**

1. Removing the unwanted columns from the dataset and changing the factor variables into its appropriate data types.

2. Create a column “No.ofoccupants” with the existing “occupancy” column by writing a custom function in R to extract the numbers and sum it, so we get the total number of occupants per booking.

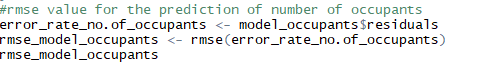
3. Create a column “StayDays” with the difference between the check-in and checkout date attributes.

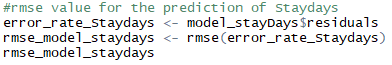
4. Clean the room\_price column by extracting only the price from it

5. Handling the missing values of the significant attributes to create the model for predicting the “No.of Occupants” per booking and “No.of StayDays” of customer per booking.

6. Two Poisson regression models on the response variable “No.ofOccupants and “StayDays” based on other significant features are created.

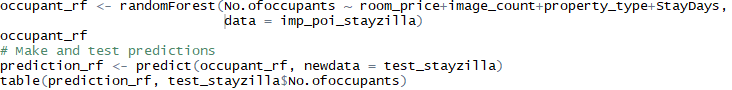
7. To check the performance metrics of the models created, the root mean squared error for both the models are calculated.





The root mean squared error for the model created on the response variable “No.ofoccupants” is 0.2862205 and the rmse for the response variable “StayDays” is 5.377056e-17. Both the values seemed to be very decent to trust the models. The lower the rmse, the better the model will perform. So, **these models can be used to predict the total number of occupants per bookings and the duration of the customer stay per booking.**

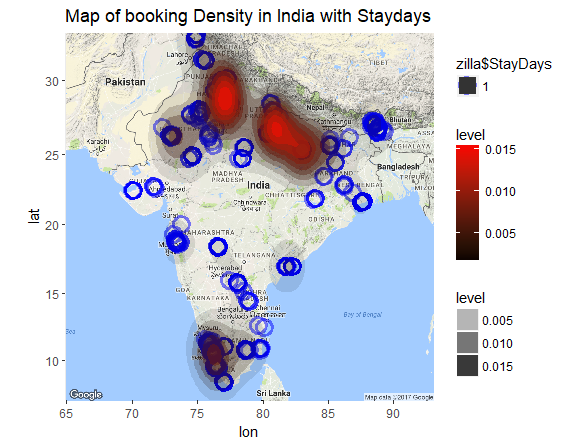
8. To back up the model, we have also created a random forest model to predict the total number of occupancy per booking.

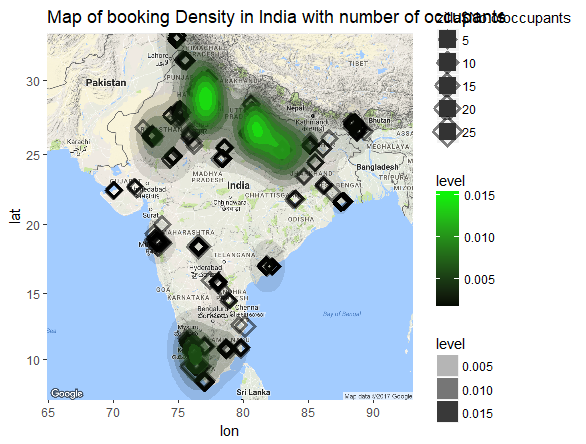


9. But the accuracy of the model turned to be only 12% which is a very poor model. This may be due to several reasons like it is not appropriate technique to be applied on count variable, less data etc.,

**Visualization based on the model analysis:**

The geocoding API and ggplot helped us to plot the most common areas where the properties are getting booked and shows the actual duration of stay period and the total number of occupants per booking respectively.





We can notice, that most of the properties are located in northern side of India, which appears to be where the platform is most popular. As we have seen in the introduction, they have spent lot of investment on advertising their market which is one of the major reason which led to the failure. So, with our analysis, **we can recommend the company to invest on creating awareness /marketing campaign about their product on the other regions as well.**