# Data Science Challenge at Cabify

An experiment design for measuring
impact of professional pictures on Airbnb revenue

Irene Ovejero Gimeno

October 24, 2021

## 1 The Problem

As stated in the given problem definition, throughout the years 2011-2016 Airbnb has realized that providing verified and better quality pictures rises trust among guests and helps hosts to make more gains from their property listings. The cost of providing this service for free to hosts has raised issues and is being questioned. The value of the service in terms of trust and demand rising is already proven here. That paper contains very valuable ideas that I share and agree with, and thus several concepts and proposals included in this design document are based on its content or try to enhance it.

The aim is to find if Airbnb should stop or continue funding this service, by checking how strong the relationship is between the revenue for the company and the value provided by this service.

As materialistic as it sounds, one could also think that profits made by hosts could be covering this service that has proven to be of value to them. However this is an additional question.

It is important to find a proper and rigorous solution in terms of statistical evidence but also to understand the context and user experience from a business point of view.

## 2 Questions that arise

These questions are related to the problem and were wondered out of curiosity of the context. Some of them might be answered or partially answered later but they do not necessarily have to:

**Q1** What is a *good* picture?

**Q2** Does every photographer value his/her work the same? Are their rates related to geo-location or even seasons of the year?

**Q3** Does every property need super high quality pictures to be better considered? Is that related to geo-location? Is that related to property type?

**Q4** What do guests value more: aesthetics or "official" pictures (verified)?

**Q5** What is the impact observed in a property's demand if it has verified/good pictures?

**Q6** How much more profit a host is making after having professional pictures taken and verified?

## 3 Proposed process

The key aspects to keep in mind in order to design an experiment to solve this matter are: business context, data and statistics. Therefore the experiment is proposed as follows:

1. Context, definitions and hypothesis
2. Gathering and structuring data
3. Modeling to prove or revoke hypothesis
4. Reading results

## 3.1 Context, definitions and hypothesis

Please note that none of the following statements are necessarily official definitions from Airbnb. Among the whole Airbnb database, there are many **sub-regions** where hosts offer properties, and guests can book them each month, as a chosen **unit of time period**. These two bold concepts mean that there is seasonality at a zipcode-date[1] level.

Each property has a certain **set of characteristics**, static as number of rooms or dynamic as price per night or pictures. These pictures can be made and uploaded by the host (**non-verified**), or taken by a professional photographer and verified for uploading (**verified**).

Each month, a property is offered during a certain number of days and is booked during some of those offered days. So the demand for property $i$ during a month $t$ can be defined as

$$demand_{it} = \frac{\#booked\_days_{it}}{\#offered\_days_{it}}$$

The price for each property $i$ during a month $t$ is calculated as the average of the prices per night during all nights offered in that month: $avg\_price\_per\_night_{it}$.

The revenue for Airbnb, out of one property $i$ during a month $t$ is defined as

$$revenue_{it} = Airbnb\_fee * \#booked\_days_{it} * avg\_price\_per\_night_{it} - associated\_costs_{it}$$

where $associated\_costs_{it}$ includes possible photography services rates calculated and fixed as $avg\_pic\_rate_{rt}$ (taking into account zipcode and date too).

For further thinking, the profit for a host, made by one property $i$ during a month $t$ is defined as

$$profit_{it} = (1 - Airbnb\_fee) * \#booked\_days_{it} * avg\_price\_per\_night_{it}$$

### 3.1.1 Hypothesis to verify

The goal then is to check

(i) if the revenue for Airbnb increases/decreases related to verified

(ii) and better quality pictures,

(iii) and whether it does so enough to cover/not cover the average photographer rate.

An additional analysis would be checking if the profits for hosts are enhanced by professional pictures in a way that Airbnb can justify transferring the cost to hosts.

## 3.2 Gathering and structuring data

It is necessary to define a time period for data extraction as close as possible to present date and large enough to cover at least one cycle of seasonality (maybe not as long as 5 years like economic cycles, but at least 2 years long), also taking into account macro events like COVID-19 situation.

The period must include dates where there are enough properties that:

- begin with non verified pictures and at some point decide to include verified pictures from the service funded by Airbnb.
- begin with non verified pictures and never go for the professional service.

### 3.2.1 Interesting variables

It would be desirable to have, for each property and date, static and dynamic information about its location, type, size, capacity, rooms, equipment, services, availability, price, bookings, reviews (positive and negative), number of pictures, the pictures themselves, etc. Information about the host is also interesting to add. These variables will serve as independent items in the modeling and need to be processed to build dummies, categorical or continuous terms. Besides, new variables made from the initial ones can be calculated.

Given the period and sample mentioned before, a key variable $serviceind_{it}$ will be built as follows:

[1] a date is a year and a month

$$\begin{cases} 0 & \text{if property } i \text{ does not have verified pictures at moment } t \\ 1 & \text{if property } i \text{ does have verified pictures at moment } t \end{cases}$$

This variable is the one to be studied and faced against $revenue_{it}$.

Derived from this definition, it is clear that the sample has two subsections: properties with professional pictures and properties with normal pictures. That means the sample has been divided by a definition, and not by a random sampling. This can be an issue in terms of comparability. At least, it is necessary to check if the means of gathered variables in the two subsections are statistically indifferent by applying a T-test analysis for instance. The $H_0$ would be "the samples are comparable because the means are indifferent", and once calculated t's and p-values we could not reject $H_0$ if p-values are greater than a certain level of significance $\alpha$. If it is not the case, it would be necessary to over-sample or under-sample the subsections to reach a balance.

Please note that the quality of pictures is not yet considered in the experiment.

### 3.2.2 Descriptive analysis

After gathering data, a table representing the time period of at least two years is available containing the columns related to independent information about properties and its hosts, and a column containing the service indicator binary flag. The revenue, profit and demand are separate columns containing dependent information calculated as defined before. The main unit for modeling (a row) is a property in a certain date.

It is necessary to carry out a descriptive analysis of the table in terms of rows, null values and how to filter them, means and standard deviation to observe how variables behave. A treatment for outliers (using the first and third quartile to calculate the inter-quartile range) is also recommended either by eliminating them or applying a threshold. Correlations between variables are also verified to eliminate collinearity.

It would also be interesting to compare this sample to old samples, from the aforementioned paper for example. A KS analysis (maximum distance between the two samples using the accumulated distribution over levels of demand) or PSI (Population Stability Index) analysis would tell how much the population has changed over time and will give a hint about the validity of past results.

## 3.3 Modeling to prove or revoke hypothesis

In this step, the aim is to find a relationship between key variables of the problem in order to be able to extract conclusions. The relationship can be found with different types of experiments or modeling, here two options are discussed:

(a) An A/B test is implemented through the hosts front: two random lists of hosts are separated, one sample is offered to pay for the professional photography service and the other is offered the same service for free. Then after a while, the two samples are compared in terms of revenue for Airbnb.

(b) A regression is forced between independent and dependent variables: this experiment is opaque to hosts since is based on already observed data.

Option (a) does not seem to fit for the present purpose since it implies to change the host experience all of a sudden. The change is huge and shocking, to the point that a marketing strategy would have to be put in place to please the hosts selected to pay for the service (and that is sort of a bias too). It does not measure the impact of a new feature in the product interface, but tries to measure a whole new experience. Moreover, even reverting the test has an impact on users (what are they going to say if the service stays free of charge at the end). For all of the above, this approach is hereby discarded. Option (b) is the preferred approach, and it consists of building a linear regression model written as:

$$revenue_{it} = a_0 + \alpha * serviceind_{it} + \beta * \{property\_variables\}_{it}$$

where $a_0$ is the intercept, $\alpha$ is the coefficient subject to analysis, and $\{property\_variables\}_{it}$ is a vector of the bunch of variables related to property static and dynamic characteristics and the host, taking into account zip codes and dates seasonality.

Once the model is adjusted, a table with the coefficients, for each variable, is obtained along with a level of significance in terms of p-values. The $R^2$ and corrected $R^2$ coefficient, AIC coefficient or SBC coefficient can give an idea of the robustness of the model itself (the higher the better).

## 3.4 Reading results

The key is to observe coefficient $\alpha$ at the end of the modeling:

1. First of all, the p-value associated with it must be small enough (e.g. less than 0.05) to consider it significant (meaning that the hypothesis of all model coefficients being equal to zero can be rejected). If this is the case, it is possible to extract conclusions of the relationship between Airbnb revenue and the professional photography service.

2. If $\alpha$ is positive and high, that means how many times the revenue itself (including costs derived from professional photographers fees) is related to the fact that hosts choose the professional photography service. That would indicate that is advisable to continue funding the service because it does not affect the revenue for Airbnb in a negative way.

3. On the other hand, if $\alpha$ is negative, that would mean that the professional service is making a negative impact in the revenue, so assuming the cost of it should stop.

### 3.4.1 Additional conclusions

- If Airbnb concludes that they should no longer fund the service, an updated version of the model estimated in the previously mentioned paper could be build to inform the hosts of the impact of the professional service on their profits. Instead of modeling against the demand, the model would be taking $profit_{it}$ as the dependent variable, and if the coefficient $\alpha$ is significant and enough positive, it could serve as justification of "with this improvement in your profit due to verified pictures you actually can cover the cost of the service in less than X bookings and still make money".

- Given the fact that Airbnb is already charging hosts for the professional service, it would be interesting to build a scoring for quality picture based on computer vision (using convolutional neural networks and tagging good/bad images). For hosts, it would fair if Airbnb offered an initial service of "evaluating your pictures" for free, making a quality score available and thus informing the hosts about the aesthetic and quality of his/her own homemade pictures. That would add an aditional value to the professional photographer budget offered to hosts, meaning "you can improve your X% quality pictures to a granted Y% quality pictures from this photographer".

- Airbnb is using a pricing suggestion feature to help the hosts define the price per night for their properties. This pricing model could include variables related to the professional photography service. However, this is a risky move since it has a direct impact on guests and business in general, and would need of further experiments to confirm viability.

# 4 References


- Zhang, Shunyuan and Lee, Dokyun and Singh, Param Vir and Srinivasan, Kannan, *What Makes a Good Image? Airbnb Demand Analytics Leveraging Interpretable Image Features* (May 25, 2017). Available here or here

- Zhang, Shunyuan and Lee, Dokyun and Singh, Param Vir and Srinivasan, Kannan, *How Much Is An Image Worth? An Empirical Analysis of Property's Image Aesthetic Quality on Demand at AirBNB* (Dec 11, 2016). Available here