1. What key metrics would you propose to monitor over time the success of the team's efforts in improving the guest host matching process and why? Clearly define your metric(s) and explain how each is computed.

2. What areas should we invest in to increase the number of successful bookings in Rio de Janeiro? What segments are doing well and what could be improved? Propose 2-3 specific recommendations (business initiatives and product changes) that could address these opportunities. Demonstrate rationale behind each recommendation AND prioritize your recommendations in order of their estimated impact.

3. There is also interest from executives at Airbnb about the work you are doing, and a desire to understand the broader framing of the challenge of matching supply and demand, thinking beyond the data provided. What other research, experiments, or approaches could help the company get more clarity on the problem?

<https://www.lucidchart.com/documents/edit/48054bf7-f43d-404e-9b72-514e56867e8a/1>

------------------------------------------------------------------------------------------------------------------------------------------

# Goal: One of our goals at Airbnb is to increase bookings on our platform.

* To increase the number of bookings…
  + Increase number of accepted regular booking
    - Increase number of booking requested from travelers
    - Increase number of accepted booking requests from hosts
  + Increase number of instant booking
* Thing to consider
  + Cancel rate (from both host and traveler sides)
    - Might be different for instant booking from regular booking
      * So we need a balance of instant booking and regular booking
  + Hosts can also become guest
  + Hosts might have multiple listings
* Users table
  + id\_user\_anon is comprised of both guest and host

# Clustering

## Guest-level attributes

* Country
* Words\_in\_user\_profile
* Number of distinct booking requests made
* Number of distinct instant-bookings made
* Number of guests inquiry made for

## Host-level attributes

* Country
* Words\_in\_user\_profile
* Number of distinct booking requests received
  + Number of distinct booking requests replied (ts\_reply\_at\_first)

# Regression and Classification

## Listing-level attributes

* Total\_reviews
* Listing\_neighborhood
* Room\_type
* Number of days or % booked <- target

## Inquiry (Contact/Booking/Instant) level variables

*Split inquiry by first contact\_channel\_first*

*Run for* Rio de Janeiro separately and at the overall level as well

* **Target variable**
  + Whether or not inquiry has turned into legit booking (1,0)
* **Explanatory variable**
  + **Reservation related**
    - Duration of the request stay
    - Number of people requested to stay (M\_guest)
    - How far ahead reservation was requested (ds\_checkin\_first - ts\_interaction\_first)
    - Whether or not requested checkin date is weekday or weekend (weekly periodicity)
    - Week of the year of requested checkin date (yearly seasonality)
    - *Which holiday overlaps with requested stay (Not here, as there are different countries for the scope of the assessment)*
  + **Guest’s attributes**
    - Whether or not a guest has previously stayed at the listing requested guest\_user\_stage\_first
    - country
    - words\_in\_user\_profile
  + **Host’s attributes**
    - country
    - words\_in\_user\_profile
  + **Listing’s attribute**
    - Total review counts
    - Room type
    - Listing neighborhood
  + **Host and guest Interaction attributes**
    - Time between first interaction (ts\_interaction\_first) and host’s first reply (ts\_reply\_at\_first) as this tells us how long guest had to wait to get reply from host
    - Time between first reply from host (ts\_reply\_at\_first) and when host approves guests to stay (ts\_accepted\_at\_first) as this tells us how quickly host approved guest after they had replied initially
    - Length of the first message from guest to host (m\_first\_message\_length\_in\_characters)

Misc

* While ‘Contact host’ this is needed, this is not the best channel, as it takes time and imposes messaging time required for both hosts and guests. Rather, think of different programs, features, tutorials, guides to both parties to specify most discussed topics to be explicitly written or easy to understand to minimize the use of this channel, and its much lower overall booking rate (7%) than ‘Book It’ channel (48%).
* Instantbook
  + They are the best
  + Look at past guest vs new guest ratio?

What action can we take

## Booking-level attributes

* How long it takes to accept