TAKE-HOME CHALLENGE: Data Science - Analytics

Airbnb is a two sided marketplace which matches guests to hosts. The booking flow at Airbnb is as follows: a guest finds an available room (listing) that they like, and then they contact the host. Once the guest finds a listing they are interested in, there are three ways to send the host an inquiry: 'contact_me', 'book_it', or 'instant_book' (detailed at the bottom of this document). Upon receiving the inquiry, the host can then decide whether or not to accept the request (for 'contact_me' and 'book_it' methods -- `instant_book` is auto-accepted). One of our goals at Airbnb is to help maximize the likelihood of a successful guest-host match on our platform.

Prompt:

We would like to better understand what causes guests on our platform to end up with a successful booking. **Please identify, analyze, and visualize the opportunities to increase successful guest-host matching in Rio de Janeiro using the artificial datasets provided.**

1. Prepare these findings to present to the market manager of Rio and your fellow data team.
   a. In particular, questions you should address include:
      - What key metrics would you propose to monitor 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 in addition to a visualization.
      - What opportunities exist to increase the number of successful bookings in Rio de Janeiro? What segments are doing well and what could be improved?
      - What are 2-3 specific recommendations that could address these opportunities? Demonstrate rationale behind each recommendation AND prioritize your recommendations in order of their estimated impact. At Airbnb, one way we base decisions is around the absolute number of additional bookings that are expected to happen over a year due to a product improvement.

2. Next, imagine that your analysis was particularly effective, and has garnered interest from the broader company. Summarize your findings in a 3-5 slide presentation intended for the Head of Product and VP of Operations (who leads market managers and is not technical).
   a. Include a broader framing of the challenge of matching supply and demand when you present.
   b. Think outside of the data set you were given. What other research, experiments, or approaches could help the company get more clarity on the problem?

Further Instructions:

1) Create a PDF of your presentation of your results, with relevant tables / graphs / visualizations to explain what you have learned about the market, first for 1) and then adding slides for 2).
2) Append all code you use to analyze results to the above PDF, including code used for data exploration. We typically see data processed in SQL/R/Python and a presentation with results made in Keynote/Google slides/Powerpoint. But you are welcome to use any software you feel comfortable with. If you use Excel, please describe your operations to process the data and append your spreadsheet.
3) Please do NOT include your name or email address on this PDF.
4) You will have 48 hours to complete the assignment.

Suggestions:

1) We are looking for candidates who go beyond basic data summary to make clearly reasoned recommendations that are backed by data (and clarify assumptions).
2) We suggest that you browse the Airbnb website and look at listings to see the different ways that you can message a host.
3) Aim to have a writeup that is possible to share in a presentation that would last ~10-15 minutes (feel free to include more detailed slides in your appendix as needed).

Grading:

Your assignment will be judged according to:
1) The analytical approach and rigor used to produce your graphs, tables, visualizations,
2) The data decisions you made and the ease of reading and reproducibility of the analysis,
3) Clarity of recommendations, prioritizations, and rationale behind them,
4) The narrative of your presentation,
5) How well you followed the directions.

Data Provided:

Contacts - contains a row for every time that an user makes an inquiry for a stay at a listing in Rio de Janeiro.
- id_guest_anon - id of the guest making the inquiry.
- id_host_anon - id of the host of the listing to which the inquiry is made.
- id_listing_anon - id of the listing to which the inquiry is made.
- ts_interaction_first - UTC timestamp of the moment the inquiry is made.
- ts_reply_at_first - UTC timestamp of the moment the host replies to the inquiry, if so.
- ts_accepted_at_first - UTC timestamp of the moment the host accepts the inquiry, if so.
- ts_booking_at - UTC timestamp of the moment the booking is made, if so.
- ds_checkin_first - Date stamp of the check-in date of the inquiry.
- ds_checkout_first - Date stamp of the check-out date of the inquiry.
- m_guests - The number of guests the inquiry is for.
- m_interactions - The total number of messages sent by both the guest and host (at the time the data was pulled).
- m_first_message_length_in_characters - Number of characters in the first message sent by the guest, if a message was sent
- contact_channel_first - The contact channel through which the inquiry was made. One of {contact_me, book_it, instant_book}. *See bottom of page for more detail*
- guest_user_stage_first - Indicates whether the user has made a booking before sending the inquiry ("past booker"). If the user has not booked before, then the user is a new user.

Listings - contains data for every listing
- id_listing_anon - anonymized id of the listing to which the inquiry is made.
- room_type - indicates whether the room is an entire home, private room, or shared room
- listing_neighborhood - the neighborhood of the listing
- total_reviews - the total number of reviews of the listing (at the time the data was pulled).

Users - contains data for every user
- id_user_anon - anonymized id of user

- words_in_user_profile - the number of words in the "about me" section of the user's Airbnb profile (at the time of contact)
- country - origin country of the user

Further Information:

There are three ways to book a listing on Airbnb:

1) **contact_me** - The guests writes a message to the host to inquire about the listing. The host has the option to 1) pre-approve the guest to book their place, or 2) they can reject, or 3) they can write a free text message with no explicit acceptance or rejection. If the host pre-approves, the guest can then go ahead and click to make the booking (but is not obligated to).

2) **book_it** - The guest puts money down to book the place directly, but the host has to accept the reservation request. If the host accepts, the booking happens automatically. If you have used Airbnb before, this shows up as a button labeled "Request to book".

3) **instant_book** - The guest books the listing directly, without any need for the host to accept or reject actively (it is auto-accepted by the host). This shows up as a button labeled "Book".

Note: A host can opt-in to the `instant_book` feature. If a host does so, a guest can use the `contact_me` or `instant_book` channels for booking that particular listing, but cannot use the `book_it` functionality. Alternatively, if a host does not opt in, a guest can use the `contact_me` or `book_it` channels only.