Airbnb Data Science Challenge - Take Home (Entry Level)

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 she likes, and she then contacts the host. Upon receiving the inquiry, the host can then decide whether or not to say yes to the request (accept the request). However, the host may not accept a guest for many reasons. Some might be logistical, e.g. dates do not work in the calendar, and some may be more personal, e.g. the guest seems risky. Our goal is to help our guests maximize their likelihood of being accepted by the hosts they contact.

Suppose we run an experiment where we require the guest to write a message that is at least 140 characters long to explain why he or she is interested in staying with the host, and we run this as a 50-50 experiment (50% in treatment, 50% in control). We then look at data on the contacts and bookings of users in the treatment group compared to the control group. We are interested in what happens to the experience of contacting and booking a place on Airbnb when the guest is required to write a message like this. We are also looking for suggestions for evaluating the future of this change. Should we launch the experiment to everyone or stop it? How would you explain the results and the decision to a non-technical person, who is smart but not a statistician?

Using the (fabricated) experiment assignment and the contact and booking data attached, please provide an analysis and write up to answer these questions.

Further instructions:

- Please attach both your write-up and your code, in its entirety (please pdf into **one** document).
- At Airbnb we value our core value of “Simplify.” Keep this in mind in your write-up, and aim to have a write-up that is easy to understand in 10-15 minutes. One way you might help your reader is by including a summary, perhaps the length of a tweet**,** at the top of your report that could be sent to the entire Airbnb product team.
- Please include at least one **visualization** to communicate your results.
- You are free to use whatever tools you are most comfortable with to work through the analysis.
- Please document the different steps of your analysis in your write-up. Make sure to show that you checked for inconsistencies in the data.

Data provided:

Experiment Assignments - contains a row for every time that a user gets assigned to an treatment group.
* id_user - random id of the user.
* ab - The experimental group the user is assigned to.

Contacts - contains a row for every time that a user makes an inquiry
* id_guest - random id of the guest (user) making the inquiry. Can be linked to id_user.
* id_host - random id of the host (user) of the listing to which the inquiry is made.
* id_listing - random id of the listing to which the inquiry is made.
* ts_interaction_first - UTC timestamp of the moment the inquiry is made.
* ts_accepted_at_first - UTC timestamp of the moment the host accepts the inquiry, if so.
* m_first_message_length - length of the message the guest sent the host, in characters. If missing then there was no message.