One of our team's goals is to drive revenue on vacasa.com by encouraging guests to book their next vacation home rental through us. One of the levers we have to do this is email marketing, and one of our most important emails is our abandoned cart email, which gets sent to guests who have started the checkout process but who do not complete a reservation.

It looks like this:

Hello Vacasa Customer,

Good news! There's still time to book your stay at Villa Exotica.

Finish up your reservation to secure your spot. If you'd prefer booking over the phone, call 503.345.9399 to speak to our team. Don't wait! Your reservation cannot be guaranteed until you complete your booking.

Complete Your Reservation

Check in: Thursday, May 30th
Check out: Thursday, June 6th

Total Cost: $1,103.35

Sincerely,
Vacasa
503.345.9399

As an improvement to this email, our digital team would like to include a section at the bottom that features three units that are similar to the current unit, so that users who did not book the first unit have some encouragement to choose additional options.

However, we do not currently have a programmatic way of determining the "most similar" units. As a data scientist, you are asked to develop an approach. Your objective is to maximize conversion / revenue from the email. This same model might be used on vacasa.com as a "similar units" module on the unit page later.

Note: We realize that to develop a *complete* solution here would take longer than you probably have to work on this. We are looking to see some analysis of the data and an attempt at an

initial solution. You need not use all the datasets but they are there if you want them.

About the data:
You have been given four datasets:

- “official_reservations” is a record of all reservations for Vacasa units made from both vacasa.com and our channel partners like AirBnB. What you are getting is a subset of our full table.
- “Units” contains descriptive fields about our units. You are seeing a subset of all of our units that excludes lat and lng for privacy reasons.
- “Segment page views” contains a list of all page views of unit pages on the website (these are when a user clicks through to an individual unit). The key here is an anonymous id, which is an id assigned to a particular browser that visits our web page. Theoretically, whenever that same browser comes back, it should be assigned the same anonymous id through browser cookies.
- “segment_reservations” contains a list of the browser sessions that ended in a reservation. It is joinable to the unit pages viewed table on the anonymous id and joinable to the official reservations table on the reservation id. You are seeing a masked subset of our full dataset.