Data scientist challenges

Introduction

As context, here's a 10,000-foot view of the Thumbtack product:

- A consumer posts a **request** for a service needed. Every request is in some**category** (e.g., Catering, Personal Training, Interior Design) and some **location**(e.g., New York, San Francisco).
- We match the request up with appropriate service providers and send each of those providers an **invite** to quote on the request.
- Providers view the invite and some choose to send a **quote** to the consumer expressing interest.

For the following question, please be as specific and thorough as possible in your answers, quantify your statements as much as you can, and explain your computations. Include code you used where appropriate. You're free to use any software you like; it's OK if we can't run the analysis ourselves (but you should still include your code). Your response should not skim over technical details and should not be written for general readers.

Quote rate trending

A critical long-term question for our business is: *Are service providers becoming more or less inclined to quote over time?* Measuring the **invite-to-quote rate**, the proportion of invites that result in a quote, is a key metric for monitoring this. One thing that makes this analysis tricky (along with just about every analysis we do) is that characteristics of our ecosystem can vary dramatically by category, location and other factors. The problem of confounding is ubiquitous.

Your task is to analyze a (fictional) dataset containing requests, invites and quotes over a two month period. You should examine the trends in quote rates over time. **Is there evidence that product changes over the last two months have caused site-wide**

shifts in quoting behavior?You may focus on this question exclusively and ignore irrelevant aspects of the dataset.

Please provide both numerical and graphical characterizations as well as statistical evidence of the significance and precision of your conclusions.

You can download the dataset here in the form of a gzipped SQLitedatabase. Here's a quick summary of the schema, though it should be largely self-explanatory:

- `categories` and `locations` store the names of all categories and locations
- `users` holds the user ID and email for all users, both consumers and service providers
- `requests` has a row for each request posted by a consumer
  - `user_id` references the consumer who posted the request
  - `category_id` and `location_id` reference the `categories` and `locations`tables for the category and location of the request
  - `creation_time` is when the consumer posted the request
- `invites` has a row for each invite sent to a service provider
  - `request_id` references the request that this invite was sent for
  - `user_id` references the service provider that this invite was sent to
  - `sent_time` is when the invite was sent to the service provider
- `quotes` has a row for each quote a service provider sent in response to an invite
  - `invite_id` references the invite that this quote was sent in response to
  - `sent_time` is when the service provider sent the quote