

Project Instructions:

Data Analysis Interview Challenge

This is your chance to wow us with creative and rigorous solutions! Please include your code at the end of your submission, or in a separate file. We also accept incomplete solutions.

Part 2 - Experiment and metrics design

The neighboring cities of Gotham and Metropolis have complementary circadian rhythms: on weekdays, Ultimate Gotham is most active at night, and Ultimate Metropolis is most active during the day. On weekends, there is reasonable activity in both cities.

However, a toll bridge, with a two way toll, between the two cities causes driver partners to tend to be exclusive to each city. The Ultimate managers of city operations for the two cities have proposed an experiment to encourage driver partners to be available in both cities, by reimbursing all toll costs.

- 1) What would you choose as the key measure of success of this experiment in encouraging driver partners to serve both cities, and why would you choose this metric?
- 2) Describe a practical experiment you would design to compare the effectiveness of the proposed change in relation to the key measure of success. Please provide details on:
 - a) how you will implement the experiment
 - b) what statistical test(s) you will conduct to verify the significance of the observation
 - c) how you would interpret the results and provide recommendations to the city operations team along with any caveats.

Note: The two cities of Gotham and Metropolis are not in the provided dataset; however, you do not need this information to answer Part 2.

Project submission (Leo Evancie):

Question 1: What would you choose as the key measure of success of this experiment in encouraging driver partners to serve both cities, and why would you choose this metric?

Key measure of success: Mean customer booking latency

Metric choice rationale: The status quo is that drivers limit themselves to their respective city. Let's say there are n drivers in Gotham, and m drivers in Metropolis. A Gotham customer requests a ride. In theory, if there were no barrier between cities, the driver pool available to that customer would be $m + n$. In practice, due to the toll, there are only n available drivers. The lower the number of available drivers, the longer it takes for a customer ride-request to match with an available driver (i.e., customer booking latency).

In other words: If each customer has $m + n$ drivers at their disposal, which is to say that the toll reimbursement does in fact induce drivers to accept rides from/to/between both cities, customers will not wait as long to match with a driver.

Why not choose other metrics? It may seem unorthodox to use a customer-side behavioral metric to measure the success of an experiment ostensibly designed to change driver behavior. But the experiment in question, as should be the case for all strategy decisions, should be understood as an attempt to increase value. And the best way to increase value is to increase the adoption and usage of the product by our customers. When customers feel there's a consistently long wait, their sense of product reliability erodes.

For the sake of argument, we could for example measure the rate of ride-requests accepted per driver. But this would not necessarily reflect success in the experiment. Riders may increase or decrease their activity within their original city for reasons unrelated to the question at hand.

We could also use map data to directly assess whether drivers increase the rate at which they accept rides to/from/between both cities. But it's not enough for drivers to simply cross the bridge more frequently. It's possible for drivers to cover both cities, but for value not to increase, if they are simply fulfilling the same number of rides (or fewer!) across both cities.

Project submission (Leo Evancie):

Question 2: Describe a practical experiment you would design to compare the effectiveness of the proposed change in relation to the key measure of success.

Experiment implementation: Before initiating the toll reimbursement, we must establish baseline levels for our key metric: Customer booking latency (CBL). We will calculate a 7-day rolling averages (to smooth out latency variations across weekdays/weekends) for CBL across all customers, recording the customer IDs used in the calculation.

Then, we push emails and in-app alerts to drivers informing them of a toll-reimbursement pilot. We reinforce the message by adding a reminder pane to the driver's home screens, so that as they peruse available rides, they see the reassurance that bridge tolls are no longer a barrier.

Since our more seasoned drivers have built firm habits about where to position themselves and which kinds of fares to accept, we must give adequate time for drivers to adapt to the new condition. We should also ensure we are investigating a change in behavior among the *same set of customers*, so that any significant change can be attributed to the pilot, rather than being polluted by behaviors of new, unrelated customers.

And so, 30 days after launching the pilot, we will reassess 7-day rolling CBL averages *for that same group of customers*, making note of customer churn. Now that we have means from two samples where individuals in each sample can be paired, we conduct a paired t-test and check for significance of $p < 0.05$.

If the latter CBL is significantly lower than the former, we can support the claim that our toll-reimbursement pilot was associated with a decrease in latency between riders initiating ride-requests and drivers accepting said requests.

Having said that, we should *also* note the amount of customer churn from our sample and assess whether that differs significantly from typical churn. If we see a decrease in CBL, but an increase in churn, we will have enjoyed success in one sense and failure in another.

Assuming we see no churn problems, a significant result would lead to a recommendation to city operations teams that the pilot is working. Reimbursing drivers for bridge tolls showed initial success in speeding up ride-request fulfillment due to increased driver availability.