

Experiment Design

Part A

1. **Primary success metric:**

Average trips during the peak hours per driver.

The goal of this incentive is to encourage drivers to take on more trips and then to increase the supply so this metric should directly indicate whether the incentive is successful or not.

Additional metrics:

(1) *Average rating per trip during the peak hours.*

This metric will tell us whether the customer satisfaction will be affected when the available supply is possibly increased by the incentive.

(2) *Average revenue during the peak hours per driver.*

This metric will tell us whether the possible increase in available supply is also leading increase in revenue.

(3) *Average trips outside peak hours per driver.*

This metric will monitor whether the possible increase in the available supply during peak hours will affect the supply in other hours negatively, which could be a caveat worth attention.

2. **Primary null hypothesis:** The incentive has no effect on the primary metric.

Primary alternative hypothesis: The incentive has effect on the primary metric.

Additional null hypothesis: The incentive has no effect on the additional metrics.

Additional alternative hypothesis: The incentive has effect on at least one of the additional metrics.

Experiment objects: Drivers

- a. The rollout plan should be to limit the number of drivers exposed to this incentive to only a small portion of drivers in the cities that participate in the experiment.

Sampling method:

We should divide each participating city to multiple areas according to the extent to which their traffic is affected by the peak hours and do stratified sampling in each area for the same sample size for both the control group and treatment group, so that the effect of peak hours will be balanced in both control and treatment groups. The sample size should be decided by the power and the minimum desired lift.

Experiment process:

The incentive should only be provided to the drivers in the treatment group simultaneously before the incentive becomes effective and no more change should be added after the incentive becomes effective. The incentive should last for at least one week, to balance the different effect of peak hours on different days of week. Two weeks should be enough to collect adequate data. After the two weeks, we use the collected data of the treatment group and control group to do the analysis.

Decision Criteria:

If the test proves that there is a positive treatment effect exceeding the minimum desired lift, we will launch the incentive to all drivers. If we want to be more conservative, we will also consider the additional metrics and won't launch the incentive to all users if there is negative effect on at least one of the additional metrics.

- b. I will assume the two group has the same variation in the metrics and use z-test to compare the primary metric of the two groups, since the average trips should approximately follow a normal distribution when the sample size is large enough according to the law of large numbers. Z-test will give higher statistical power over non-parametric tests. T-test won't work here because we can't assume the population, number of trips of each individual driver, follows a normal distribution.

Part B

I will be interested in the metrics:

- 1) *Average number of new riders per day*, M1.

This metrics gives us the direct amount of lift in the goal of attracting more new riders.

- 2) *Average rides by the new riders per day*, M2.

This metrics allows us to know whether the riders attracted are active.

- 3) *Average revenue by new riders per day*, M3.

This metrics allows us to know whether the riders attracted are monetized.

To test the impact of this campaign, other campaign channels should stay the same and stable during the campaign period as before.

To better distinguish the impact of billboards from other campaign, I will advise that add a coupon code for new rides of new riders on those billboards that won't be available in any other channel. This will allow us to identity all new riders using that coupon as users stimulated by the billboards and get the above metrics for them.

We compare M1, M2, M3 in the billboard marketing period to M1, M2, M3 in the same length of time period before the billboard marketing and test whether there is any significant difference.