Ultimate Challenge 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?

The key measure of success for this experiment in encouraging driver partners to serve both cities would be the percentage increase in driver partners completing trips in both Gotham and Metropolis. This metric captures the experiment's core objective, which is to increase the number of drivers serving both cities.

This metric directly measures the desired change in driver behavior, providing a clear measure of the experiment's success.

Quantifiable and Comparable: The percentage increase is a numerical value, allowing for easy quantification and comparison between the baseline period (before the toll reimbursement policy) and the intervention period (after the policy implementation). This facilitates objective evaluation of the policy's effectiveness.

We can compare the percentage increase between:

- The baseline period (before the toll reimbursement policy) and the intervention period (after the policy implementation).
- Weekdays and weekends to understand if the policy has different impacts on different days.

3. Actionable Insights:

- The percentage increase provides clear insights for the city operations team.
- A significant increase indicates that the toll reimbursement policy is effective and worth considering for permanent implementation.
- A minimal or non-existent increase suggests that the policy needs refinement or might not be the best solution.

4. Focuses on Drivers Serving Both Cities:

- Unlike other metrics, such as total number of trips, this metric specifically focuses on drivers completing trips in both cities.
- This ensures that we are measuring the desired outcome of the experiment: cross-city service.

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**

The total duration of the experiment would be 16 weeks.This schedule allows for a sufficient baseline period to capture typical driver behavior before the intervention, a long enough intervention period to observe the policy's impact, and adequate time for data analysis and interpretation.

Phase 1: Baseline Data Collection (4 weeks)

1. Collect data on driver partners trip patterns, focusing on trips completed within each city and trips crossing between cities.
2. Record the number of trips, trip completion time, distance traveled, and driver earnings.
3. Conduct a survey among driver partners to understand their current attitudes towards cross-city service and potential barriers.

Phase 2: Intervention Period (8 weeks)

1. Implement the toll reimbursement policy.
2. Continue collecting data on driver partners' trip patterns and any changes in their behavior.

Monitor changes in key metrics, including:(8 weeks)

3. Percentage increase in drivers completing trips in both cities.
4. Total number of cross-city trips.
5. Average trip completion time.
6. Average driver earnings.
7. Driver feedback on the toll reimbursement policy

Phase 3: Data Analysis and Interpretation (4 weeks)

1. Compare data from the baseline and intervention periods to assess the impact of the toll reimbursement policy.

2. Conduct appropriate statistical tests to determine the significance of observed changes.
3. Analyze driver feedback to understand their perceptions and experiences with the policy.

- **b) what statistical test(s) you will conduct to verify the significance of the observation**

  Paired t-test or Wilcoxon signed-rank test: Comparing the percentage of cross-city trips by the same driver partners before and after the intervention. This assesses the individual-level effect of the policy.

  Additional Tests:

  Independent t-test or Mann-Whitney U test: Compare the percentage of cross-city trips for all drivers between the baseline and intervention periods. This assesses the overall effect of the policy.

  Repeated measures ANOVA: If data is normally distributed, analyze the interaction between the intervention period (baseline vs. intervention) and day of the week (weekday vs. weekend) to understand potential variations in policy effectiveness.

- **c) how you would interpret the results and provide recommendations to the city operations team along with any caveats.**

Interpret Results:

- Analyze the results of the statistical tests to determine if the observed increase in cross-city service is statistically significant.
- Consider the magnitude of the increase and compare it to the expected impact.
- Analyze the trends in trip patterns, completion times, and driver earnings to understand the broader implications of the policy.
- Consider driver feedback to identify any concerns or suggestions for improvement.

Recommendations to City Operations:

- Based on the analysis, recommend whether to implement the toll reimbursement policy permanently.

If implementing permanently, consider:

- Adjusting the reimbursement amount based on data insights.
- Implementing targeted incentives for specific time periods or driver demographics.
- Communicating the policy's benefits and impact to drivers through targeted campaigns.

- If not implemented permanently, suggest alternative strategies to encourage cross-city service.

Caveats:

- Acknowledge potential limitations of the experiment, such as the duration, sample size, and generalizability of the findings.
- Recommend further research or data collection to address these limitations and strengthen the evidence base.
- Emphasize the need for continuous monitoring and evaluation of the policy if implemented permanently.