Part A

The Driver team is planning to test a new incentive structure in which they will offer drivers an extra $5 per hour if they choose to drive during times of peak demand (4pm - 8pm) to increase the available supply. Drivers will be required to complete at least 5 trips during this window to qualify for the new incentive. As the data scientist on the team:

1. Propose and define the primary success metric of this test. In addition, propose and define 2 or 3 tracking metrics that will be important to monitor in addition to the success metric you have defined.

   Primary Success Metric:
   1. Average number of active drivers during 4pm - 8pm

   Additional Important Metric:
   1. Revenue per driver (for trips completed during 4pm - 8pm)
   2. Average # of completed trips per driver (for trips completed during 4pm - 8pm)
   3. Average passenger wait time (for trips completed during 4pm - 8pm)
   4. Average driver ratings (for trips completed during 4pm - 8pm)

2. Outline an experimentation plan to evaluate the effect of this incentive, according to the metrics you outlined.

a. What would be the rollout schedule, and how would you balance this with statistical rigor?

   Assuming that Uber drivers don't have much communication among each other and hence very little interference between the treated and untreated drivers (e.g. a driver's productivity might get negatively influenced after he/she knows someone else has the incentive while he/she doesn't), the experiment can be randomized at the driver level. Or else, it's possible to cluster the experiments by region but we will compromise sample size/statistical power and also have a potentially much larger between-sample variance.

   Ideally we will randomize at a driver level, if the budget is limited or we are worried about any potential negative impact on revenue, we can first conduct experiment in a few randomly selected cities (representing 5-10% of all uber drivers). Ideally we should conduct blocked random assignment, blocked at a city level, so we can control for the potential heterogeneous treatment effect caused by the inherent differences by different cities.

   If the learning from the initial group of drivers is that the incentives are working and increasing the driver supply which indirectly drives up the revenue, we can start gradually rolling out the incentives to more cities. It can be designed as a stepped-wedge experiment where treatment gradually rolls out to larger population (5%

-> 25% -> 50% -> 75% -> 100%). But it's important to monitor how metrics behave for different blocks (cities) as the incentives might not work in certain places (rural areas) for various reasons like different extents of surges during 4-8pm

b. What type of data analysis would you perform? Please explain why you chose that method over possible alternatives.

   For the first city where we don't have any blocked assignment and drivers are evenly split to control and variant group, we can just use a two-sample independent t-test to understand whether the metrics for control group are different from those of the variant group. I choose t-test over z-test since the variance for the population is unknown.

   As the experiment expand to more cities where we want to control for the city-level heterogeneous treatment effect, we can use a linear regression for the metrics we care about in the form like:

   # of active drivers between 4-8pm ~ incentive(1/0) + city(level)

   Although t-test would work, in this case we just need to do the OLS regression once instead of having to do a t-test for each block in the experiment.

Part B

A marketing team is planning a campaign to attract new riders to Uber in which they will put billboards up across a city, and they'll be up for several weeks. As the data scientist on the team,

1. what metrics would you be interested in when analyzing the impact of this campaign

   I would be mainly interested in:

   1) Weekly # of new rider sign-ups
   2) Weekly # of new driver sign-ups
   3) Total revenue (7 days, 14 days, 28 days)
   4) # of trips completed per rider (7 days, 14 days, 28 days)
   5) # of trips completed per driver (7 days, 14 days, 28 days)
   6) Revenue per rider (7 days, 14 days, 28 days)
   7) Revenue per driver (7 days, 14 days, 28 days)

2. how would you go about quantifying any effect on these?

   It's tricky to accurately quantify the impact of billboard ads impression since we can't track exactly who has seen the billboards and who hasn't. There are mainly two ways I'd go about measuring such impact:

1. Pre/Post Analysis: assuming that no other sorts of interventions occur during the billboards campaign (e.g. driver incentives, TV ads..), we can look at metrics with seasonality removed (like year-over-year metrics) before and after the campaign to approximate the impact of such campaign. Though we are making some strong assumptions that our ads is the only intervention during the timeframe.

2. Cluster-Level Experiment: we can also design an experiment clustered by region where we assume that the effect of the treatment (billboard ads) is contained within the same region and run an experiment where we randomly assign clusters to have/not have billboards ads. With that design, we can then do an A/B test to see the changes in metrics for the clusters in the variant groups vs. those in the control groups.