**Rideshare Case Study**

Lyft ridesharing is a two-sided marketplace with drivers and passengers. Every day new drivers

join the platform and existing drivers either drive or they do not. Suppose you are working as a

Data Scientist on the Driver Retention team whose primary goal is to reduce the rate of churn of

activated drivers (a driver becomes ‘activated’ once they complete their first ride).

The team would like to understand churn better. Explore the data to provide the team with a

deeper understanding of churn at Lyft. Your summary should include:

● The definition (with justification) for a driver to be considered churned.

● An assessment on the current business impact of churn to Lyft.

● Insights on factors affecting churn.

● Insights on segments of drivers more likely to churn.

Next, the team would like to size the opportunity of reducing churn in order to prioritize their

roadmap. The team is considering the following two hypotheses:

i. Doubling the number of rides in an activated driver’s first week.

ii. Another hypothesis you recommend.

Using the data, help the team prioritize these two hypotheses. You should cover:

● How big the opportunities are.

● What might be the longer-term consequences on the marketplace of each hypothesis.

● Which segments of drivers are most likely affected by each hypothesis.

● Which hypothesis you have more confidence in.

Data Provided

data/driver\_ids.csv

driver\_id Unique identifier for a driver

driver\_onboard\_date Date on which driver was onboarded

data/ride\_ids.csv

driver\_id Unique identifier for a driver

ride\_id Unique identifier for a ride that was completed by the driver

ride\_distance Ride distance in meters

ride\_duration Ride durations in seconds

ride\_prime\_time PrimeTime applied on the ride

data/ride\_timestamps.csv

ride\_id Unique identifier for a ride that was completed by the driver

ride\_picked\_up\_at Timestamp for when driver picked up the passenger