Driver Lifetime Value

Download Datasets

This data project has been used as a take-home assignment in the recruitment process for the data science positions at Lyft.

Assignment

After exploring and analyzing the data, please:

1. Recommend a Driver's Lifetime Value (i.e., the value of a driver to Lyft over the entire projected lifetime of a driver).
2. Please answer the following questions:
    - What are the main factors that affect a driver's lifetime value?
    - What is the average projected lifetime of a driver? That is, once a driver is onboarded, how long do they typically continue driving with Lyft?
    - Do all drivers act alike? Are there specific segments of drivers that generate more value for Lyft than the average driver?
    - What actionable recommendations are there for the business?
3. Prepare and submit a writeup of your findings for consumption by a cross-functional audience.

You can make the following assumptions about the Lyft rate card:

- Base Fare $2.00
- Cost per Mile $1.15
- Cost per Minute $0.22
- Service Fee $1.75
- Minimum Fare $5.00
- Maximum Fare $400.00

Data Description

You'll find three CSV files attached with the following data:

driver_ids.csv

- `driver_id` Unique identifier for a driver
- `driver_onboard_date` Date on which driver was on-boarded

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 duration in seconds
- `ride_prime_time` Prime Time applied on the ride

ride_timestamps.csv

- `ride_id` Unique identifier for a ride
- `event` describes the type of event; this variable takes the following values:
  - *requested_at* - passenger requested a ride
  - *accepted_at* - driver accepted a passenger request
  - *arrived_at* - driver arrived at pickup point
  - *picked_up_at* - driver picked up the passenger
  - *dropped_off_at* - driver dropped off a passenger at destination
- `timestamp` Time of event

You can assume that:

- All rides in the data set occurred in San Francisco
- All timestamps in the data set are in UTC

Practicalities

Please work on the questions in the displayed order. Make sure that the solution reflects your entire thought process - it is more important how the code is structured rather than the final answers. You are expected to spend no more than 2 hours solving this project.