Visualization Assignment

Horizon Company Business Introduction:

Horizon is a company which provides car services like Uber or lyft. It matches passengers with drivers, and every ride has one driver and one or more passengers. Drivers are paid out - based on time and distance driven, regardless of how many passengers are present in the ride. For the purpose of this assignment, we are going to be mainly focusing on the driver side of the business.

Problem Statement

A Product Owner is attempting to gauge the overall health of our business and driver engagement for Horizon drivers. Some ways to measure driver engagement are:

1. Number of rides

2. Time on trips

3. Number of days that the driver was active on

You have been tasked with helping the PO by constructing a data visualization interface that will allow the easy exploration of factors that affect key metrics.

Using the provided data, please:

1) Construct an interactive visual exploration of the data.

2) Share any code and a method of viewing what you have built. You may submit any supporting material to demonstrate functionality or explain your design choices, focusing on any important relationships you discover in the provided data.

When deciding what features to support in your interface, it may be useful to think about some of the questions that a PO might be trying to answer using your tool:

• What metrics are the most important to determine business health and driver engagement?

• How many new drivers are we onboarding every week?

• How many rides do different drivers give per unit time? Choose a time unit that makes sense for analysis.

• What is the average projected lifetime rides for a driver? Do all drivers act alike? Are there specific segments of drivers (segmented by onboarding date and region) that generate more value for Lyft than the average driver?

• Which drivers should we offer retention incentives to?

• Which locations have better or worse performing driver populations, and where should we offer geospatially targeted incentives during onboarding?

• Are there any predictive indicators for driver churn?

Feel free to come up with reasonable metric definitions or assumptions that enable you to answer these questions. We are looking for visualizations that make it easy to interpret the aggregated data from these datasets.

Data Overview

visualization\_data/driver\_ids.csv

driver\_id Unique identifier for a driver

driver\_onboard\_date Date on which driver was on-boarded

visualization\_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

visualization\_data/ride\_locations.csv

ride\_id Unique identifier for a ride

start\_geohash Pickup location of the passenger, encoded as a geohash

visualization\_data/ride\_timestamps.csv

ride\_id Unique identifier for a ride

event event describes the type of event (see below)

timestamp Time of event

Here’s an overview of the event types:

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