DATA RESEARCH LAB

Welcome to the next step in the DATA Initiative Research Lab application process. Based on your background and credentials, we have identified you as a potential preliminary fit for the Research Lab. However, to further assess your fit, we would like you to complete a short assignment. Here is the problem that you must solve:

Attached to the email is a dataset that contains information on one of the local Northeastern University shuttle services. The information contains all requests for on-demand, drop-off to door-step transportation service throughout the year 2020. The shuttle currently operates by picking up riders near the Cabot Physical Education Center and dropping them off at their requested destination. The dispatch manager for the shuttle would like to predict when and how often riders will cancel their request. More specifically, they would like to have a machine learning model draw a prediction about whether or not a rider will cancel their request using the information about the request. Your task is to design a few possible models that can predict whether or not a rider will cancel their request. You are free to do this in any software (R, Python, Excel, etc). You are free to consider any model type as well as any prediction performance metric or method to use for model specification, estimation, and selection.

- Unit of Analysis: Ride Request from Single Rider
- Properties:
  - Request Creation Date – The date the request was created.
  - Request Creation Time – The time of the date the request was created.
  - Request ID – A unique id that uniquely identifies the ride request.
  - Request Status – The ending status of the request.
  - Rider ID – The unique identifier of the rider.
  - Booking Method – The method the request was placed.
  - Destination Lat – The latitude of the request's destination.
  - Destination Lng – The longitude of the request's destination.
  - Original Planned Pickup Time – The projected planned time the rider will be picked up.
  - Original Planned Dropoff Time – The projected planned time the rider will be dropped off.
  - Time from Request Creation to Planned Pickup – The projected length of time of the ride.
  - Proposed Pickup Walk Distance – The distance a rider needs to walk to to arrive to their pickup location.
  - Ride ID
  - Cancellation Time – The time the ride was cancelled, if it was cancelled.
  - Actual Pickup Time – The actual time the shuttle picked up the rider.
  - Actual Dropoff Time – The actual time the shuttle dropped off the rider.
  - Ride Distance – The overall distance of the ride from pickup to dropoff.
  - Ride Duration – The actual length of time of the ride.
  - Ride Rating – The rating the ride on a scale from 1 – 5.

Deliverables (DUE June 15th):
- Provide code, additional data files, picture files, and any and all other material that you used or generated to answer the problem at hand.
- A 5-minute recording (a Zoom recording link WITH PASSWORD) of yourself that presents on the result. You do not need to use a powerpoint presentation, but it would be helpful in presenting your results. Please do NOT submit a video file, or use any other platform!!!