



CASCADE CUP 2022

Round 3

Problem Statement



Welcome to the third and final round of the *Cascade Cup 2022*.

Task

Your job is to compile a detailed Data analysis report based on the dataset. Use graphs to make as many significant inferences and relationships as possible. Make sure the report isn't more than ten pages (including the introduction and thank you pages).

Deadline

15th Feb EOD (i.e 11:59 PM)

Background

At Shadowfax, part of our business includes delivering food orders from clients such as Swiggy and Zomato to customers. The typical order flow goes something like this:

- Client creates the order in our system
- The order gets allocated to a rider
- The rider accepts the order
- The rider goes to the pickup location
- The rider picks up the order
- The rider goes to the delivery location and delivers the order

The rider also has the option to get the order cancelled before delivery by calling the client's call centre. We would like to predict this kind of cancellation before it happens so that we can try and reassign the order to another rider before it gets cancelled.

Data

The sample data consists of two files:

1. `train_data.csv`: This file contains one row for each order created in a specific time-frame. The columns are:
 - a. `order_id` : unique id for each order
 - b. `order_time`: time of the creation of order by the client
 - c. `order_date`: date of the order
 - d. `allot_time`: time of allocation of order to the rider
 - e. `accept_time`: time of acceptance of order by the rider (if available)

- f. pickup_time: time of pickup of the order (if available)
 - g. delivered_time: time of delivery of order (if available)
 - h. cancelled_time: time of cancellation of order (if the order was cancelled)
 - i. cancelled: whether the order was cancelled
 - j. rider_id: unique id for each rider
 - k. first_mile_distance: road distance from rider's location to the pickup location
 - l. last_mile_distance: road distance from pickup location to delivery location
 - m. allotted_orders: total number of orders allotted to the rider in the 30 days before (not including) order_date
 - n. delivered_orders: total number of orders delivered by the rider in the 30 days before (not including) order_date
 - o. undelivered_orders: total number of orders allotted to but not delivered by the rider (i.e. cancelled) in the 30 days before (not including) order_date
 - p. lifetime_order_count: total number of orders delivered by the rider at any time before order_date
 - q. reassigned_order: whether the order was reassigned to this rider
 - r. reassignment_method: if the order was reassigned, whether the reassignment was done manually (by the ops team) or automatically
 - s. reassignment_reason: more detailed reason for the reassignment
 - t. session_time: total time the rider had been online on order_date before order_time
2. call_data.csv. This file contains one row for each call placed by the rider (whether to the customer or to our call centre) for the mentioned order_id. It contains the following columns:
- a. order_id: the order id for which the call was placed
 - b. rider_id: the rider who placed the call
 - c. user_type: whether the call was placed to the customer or to the call centre
 - d. reason_text: the reason given, if any, by the rider for the call

Regards

Team CnA, IIT Guwahati