Data Scientist Technical Test

Welcome to the Glovo Data Scientist recruitment pipeline. This is the second step of a 4-step recruitment process. In this step, you will be given some data sets and you will be asked to optimize a specific problem. If you have any questions, feel free to reach Koray (koray@glovoapp.com) for technical questions and Lindsay (lindsay.lefranc@glovoapp.com) for non-technical questions.

Glovo is an online delivery platform through which a customer can get items delivered to an address. Through an online app a customer places an order which is a request for an item to be delivered from a specific merchant. A glover or a rider then picks up the item from the merchant and delivers it to the customer's address. For effective running of this operation, it is essential to have the right number of riders in a city for any given day in advance in order that expected orders get delivered in a reasonable time. Furthermore, we wish to ensure that drivers are not idle for significant periods.

Task Description

In this task, we would like you to estimate the number of couriers in an entire day with a view to meeting an expected cancellation rate and utilization rate of Glovers (percentage of time that they are actively working and not idle. Essentially estimate how many riders or couriers should we assign to a city?

Formulate the mathematical problem that summarises the task at hand and implement it in a language of your choice (Python is preferred).

We need to minimize the cost and the cost function of the optimization is as following:

$$Cost = 10 * CR - UR$$

where CR stands for Cancellation rate and UR stands for Utilization rate and formulated as following:

$$CR = CancelledOrders/TotalOrders$$

$$UR = CumulativeDeliveryDuration/CumulativeWorkingTime$$

where Cumulative Delivery Duration and Cumulative Working Time is calculated at a daily resolution in the scale of the entire fleet for each city.

Some key rules govern our operations that should be factored into the problem description:

- The Glovers can book slots in every 1 hour interval and you can assume that 100% of the Glovers that book the slots will show up.
- If the assigned order to a Glover lasts more than the working hour of the Glover, the Glover does not leave the slot until he finishes the order.
- Glovers do not get any automatically assigned orders in the last 10 minutes of their last working slot.

- Each Glover can serve 1 order at a given time and each order can be assigned to at most 1 Glover.
- Each time orders assignment is delayed for 5 minutes, order will be cancelled with 2% possibility.

To implement this exercise you will be given the following data:

- estimated_delivery_times.csv
- last_slot_booked.csv
- orders.csv

Orders

In this csv, you will encounter 2 columns and the description of those are as follows:

- **time_frame:** Activation time: the time frame when an order is ready for assignment. For non-scheduled orders, it is the time an order has been placed and for the scheduled orders, it is the scheduled time minus the average delivery time. For example, if an order is scheduled at 12:00 and average delivery time is 20 minutes, activation time for this order is 11:40
- **orders_log:** Number of orders in the given time_frame until the next time_frame starts in **logarithmic** scale

Last Slot Booked

In this csv, you will encounter 2 columns and the description of those as follows:

- **hour:** Local hour of the day
- **last_slot_booked:** Probability that couriers booked this slot but not the next slot (indicates occupancy of next slot for the couriers that are coming from a previous slot)

Estimated Delivery Times

In this csv, you will encounter 3 columns and the description of those as follows:

- **day:** Day of the week where 0 is Monday and 6 is the Sunday
- **hour:** Local hour of the day
- **delivery_time:** Expected courier delivery time in this slot. Or in other words, time passed since courier starts the order until he delivers the order in **seconds**.

Further to the task description, we would like you to think about some proposed recommendations for us that will allow us to improve the rider assignment problem. For instance, do you think Glovo get some additional data? If so what datasets would you consider and why?

Would you change or explore the cost function? If so, on what basis and how would you change it?

Or could there be a change in the operational processes that can yield greater efficiencies? This is an open ended question and we are keen to get your suggestion on how we could improve our internal operating processes to best serve our customers.

Once you finish the task, please send it to Koray (koray@glovoapp.com) with Lindsay (lindsay.lefranc@glovoapp.com) in the copy. Good luck!