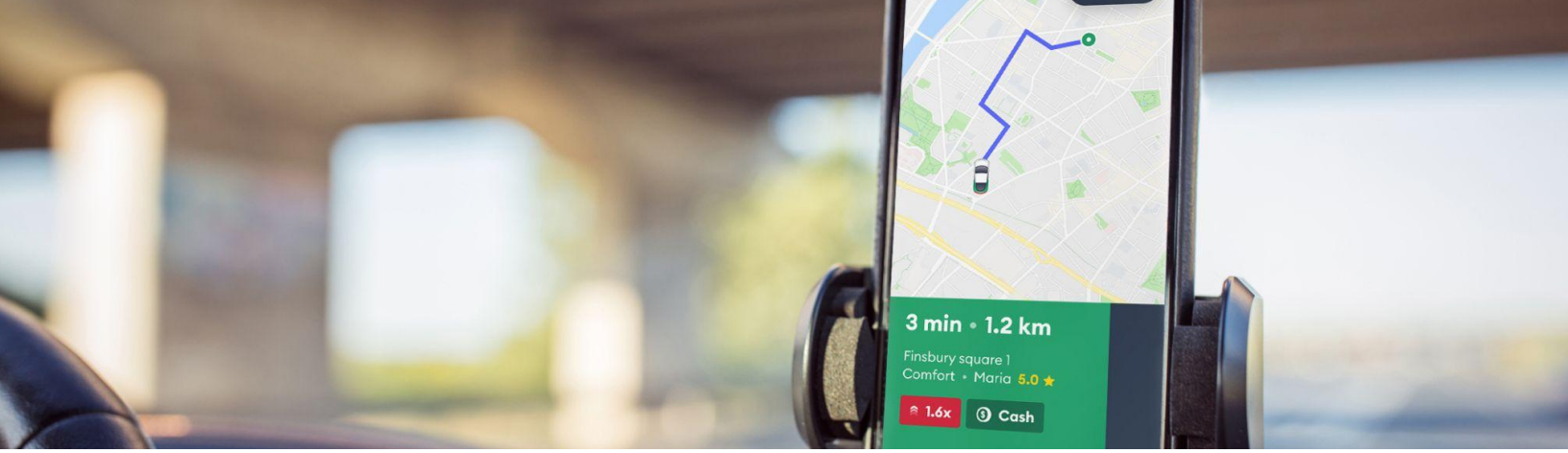

Senior Data Analyst (Central Operations) Home Task, Bolt

The key aspect of ride-hailing is upfront pricing, which works in the following way:

- ➔ Firstly, it predicts the price for a ride based on the predicted distance and time.
  - ◆ This price is what you see on the screen of the phone before ordering a ride.
- ➔ Secondly, if the metered price (based on actual distance and time) differs a lot from the predicted one, the upfront price switches to the metered price.
  - ◆ 'A lot' implies more than 20%.
  - ◆ For example, suppose you want to take a ride for which the upfront price is predicted as 5 euros. If the metered price is between 4 and 6 euros - the rider pays 5 euros, otherwise the metered price.

Task:

We would like to improve the upfront pricing precision. Please analyse the data and identify top opportunities for that. **Could you suggest a course of action?**

Bonus: Is there any other value you can extract from this data? (Feel free to get creative!)

Submission:

Assume that both business and technical people will check the results.

In a ZIP folder:

1. Please provide us with a PDF file no longer than 3 pages. *(Think of a business reader!)*

2. Please also submit a Jupyter notebook as a result.

We expect you to spend less than 8 hours on the task.

Data:

Sample data - test.csv

Variable	Description
order_id_new	ID of an order
order_try_id_new	ID of an order “attempt” (one order can be attempted on multiple drivers, until one accepts)
Metered_price, distance & duration	Actual price, distance and duration of a ride
upfront_price	Promised to the rider price, based on predicted duration (predicted_duration) and distance (predicted_distance)
distance	Ride distance
duration	Ride Duration
gps_confidence	Indicator for good GPS connection (1 - good one, 0 - bad one)
entered_by	Who entered the address
b_state	State of a ride (finished implies that the ride was actually done)
dest_change_number	Number of destination changes by a rider and a driver. It includes the original input of the destination by a rider. That is why the minimum value of it is 1
predicted_distance	Predicted duration of a ride based on the pickup and dropoff points entered by the rider requesting a car
predicted_duration	Predicted duration of a ride based on the pickup and dropoff points entered by the rider requesting a car
prediction_price_type	Internal variable for the type of prediction: *upfront, prediction* - prediction happened before the ride upfront_destination_changed - prediction happened after rider changed destination during the ride
change_reason_pricing	Indicates whose action triggered a change in the price prediction. If it is empty, it means that either nobody changed

	the destination or that the change has not affected the predicted price
ticket_id_new	ID for customer support ticket
device_token, device_token_new	ID for a device_token (empty for all the fields)
rider_app_version	App version of rider phone
driver_app_version	App version of driver phone
driver_device_uid_new	ID for UID of a phone device
device_name	The name of the phone
eu_indicator	Whether a ride happens in EU
overpaid_ride_ticket	Indicator for a rider complaining about the overpaid ride
fraud_score	Fraud score of a rider. The higher it is the more likely the rider will cheat.