Part I: Data Analysis

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

First, it predicts the price for a ride based on predicted distance and time. This price is what you see on the screen of the phone before ordering a ride. Second, 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' means more than 20%. For example, suppose you want to make a ride that upfront price predicts a cost of 5 euros. If the metered price is between 4 and 6 euros, you will end up paying 5 euros. You will end up paying something else if the metered price is anything less than 4 or more than 6 euros.

No customer likes surprises (especially when it comes to money!), that's why we always strive to improve our upfront pricing precision for our customers' smooth journeys.

In the attached file, you will find a sample data test.csv. **Your task is to analyze this data and identify the top 1 to 2 opportunities that can help us improve upfront pricing precision**. The expected output is

- A business PDF report (2 pages maximum). Assume that both business and technical people will read your results.
- A Jupyter notebook (using Python or SQL) as a result.

Variables in the file:

- order_id_new, order_try_id_new - id of an order
- **calc_created- time when the order was created**
- 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.

Part II: Business Research

Please create an assessment for food delivery (courier delivery of food from restaurants) launch in a city of your choice. The output should be a spreadsheet including the following:

1. A top-down estimation of market size
2. Unit economics with profitability per order
3. What conclusions can you come up with? Should we launch or not?
4. How would you measure success or failure if the launch happens? What KPI’s can you come up with?