Predicting Bad Deliveries

# The Problem

Delivery Hero is a market leader in food delivery across many countries. Every delivery, every customer, and every rider are different, and we and the models that we build have to be sensitive to that.

With tens of millions of orders a day, not every delivery goes smoothly. In this task, your job will be to predict which deliveries will be delivered poorly.

When a delivery rider arrives at the customer-location that the customer entered for the order, the rider logs in our rider-app that they have arrived at the customer. However, because of any number of reasons (buildings blocking GPS signal, poor data entry by the customer, a rider’s newness, etc) the customer’s *actual* location might not match the customer’s entered location. As a result, the rider then starts the journey of finding their way from the customer’s entered location to the customer’s actual location. We track this traversal distance as **dropoff\_distance**.

In an ideal world, all dropoff\_distance’s are zero. However, this is rarely the case, and so we’d like to be able to predict what an order’s dropoff-distance will be, so that our logistics team can plan accordingly. Your job, then, is to ***create a model that predicts the dropoff\_distance of an order, given some features about the order. A model like this could then be fed into the logistics rider-scheduling algorithms.***

# Data

The data for this task is all orders in Singapore for the month of March 2021. A zip-compressed JSON file with the order data can be downloaded from [here](https://drive.google.com/file/d/1AljKMYLoGjQSwWQdhHoJU0FEl7dc8mLB/view?usp=sharing).

Note that there are some boolean fields **has\_X\_instruction**. Where does this come from, you might ask? For every order, the customer has a field “delivery instructions to the rider”, where they can type any message (e.g. “call me when you arrive 555-555-5555”, “leave it at the door”, “please knock when you arrive”, etc). With this, the boolean field **has\_bell\_instruction** means “the delivery instruction has the word ‘bell’”, etc.

All fields are described in the following data schema--

## Data schema:

* **order\_id (str)**
* **customer\_id (str)**
* **order\_items\_count (int) →** number of items in the order
* **order\_value (float) →** total Euro value of the order
* **created\_timestamp\_local (timestamp) →** timestamp of the order
* **delivery\_geohash\_precision8 (str) →** geohash of the order, at precision-level 8
* **delivery\_postal\_code (int)**
* **logistics\_dropoff\_distance (float) →** the target variable, in meters
* **has\_bell\_instruction (boolean) →** delivery\_instructions contains “bell”
* **has\_call\_instruction (boolean)**
* **has\_gate\_instruction (boolean)**
* **has\_instruction (boolean)**
* **has\_knock\_instruction (boolean)**
* **has\_leave\_instruction (boolean)**
* **has\_lift\_instruction (boolean)**
* **has\_lobby\_instruction (boolean)**
* **has\_phone\_number (boolean) →** delivery\_instructions contains a phone #

# Acceptance Criteria

Please solve this exercise in a single Jupyter Notebook, using the Python data ecosystem libraries and tools (e.g. pandas, numpy, scikit-learn, environment-management, etc). Some specifications for the submission:

* Jupyter notebook should have the following structure:

1. Data cleanup and validation
2. Feature exploration and any feature engineering
3. Models you tried, and the one you kept.
4. Ideas about improvements that can be done if you had more time

* Create a GitHub repo (please don't use the company name) for the project and provide us a link.

# Final Notes

The whole exercise should take around 2-3h of your time.

**We do not value the accuracy of the results very highly,** because we understand that it is an open-ended and time-consuming task. We will instead focus on your approach, process, data science intuition, and creativity.

We hope you'll find this exercise interesting and challenging. We are happy to answer any questions to clarify the problem. Please feel free to send us emails as needed.

Thank you very much for your time!