«Car Market

In this challenge, you'll work for CarMarket, a marketplace where thousands of Merchants can sell used vehicles (*any similarity to actual companies, living or dead, is purely coincidental*).

CarMarket's algorithm aims to select the vehicles with the best quality / price ratio for our customers. It is therefore crucial to estimate in advance the **defective rate** of the cars sold in the marketplace by our Merchants.

What is your challenge ?

Build a model that predicts the defective rate at 30 days (that is the probability that a given car sold on our platform by a given merchant has an issue in the 30 days following the order).

You should keep in mind that your model will be used in production, and therefore any feature or characteristic should be feasible in that setting.

We want to be mindful of your time and we expect you to spend around 4 hours on this challenge.

We understand that this is pretty short and the goal of this exercise is for us to see how you tackle a real-life machine learning problems, with the constraints that it involves.

This means that this is not a Kaggle problem, and you will be judged more **on your approach (how rigourous you are and how you justify the different steps you take) than on the actual technical performance of your model**. We will reward approaches that correctly identify the challenges specific to this problem.

Deliverable

You are free to use any way to present your results, although a Jupyter Notebook and (if necessary) scripts will be more than enough.

We appreciate having a detailed overview of:

- the approach and assumptions (training set, feature selection, etc)
- any data analysis (if necessary)

- model choice
- any improvements from a ML perspective if you had had more time (additional features, optimisation, etc)
- how you would evaluate performance from a technical and business perspective

---

Dataset

Every line in the dataset is one, and only one, order at CarMarket.

- `DEFECTIVE`: whether there was a Technical customer service request linked with the order
- `DATE_ORDER`: date of the order
- `CONTACT_DATE`: date the customer contacted customer service. Is NaN if no contact has been made
- `CONTACT_TYPE_DETAILS`: details about the problem that was reported (e.g. *'Steering Wheel Shaking'*, *'Faulty Ignition Coil'*)
- `PRODUCT_CATEGORY`: product category (e.g. car, motorcycle, ...)
- `BRAND`: product brand (e.g. *'Ford'*)
- `MODEL`: product model (e.g. *'Mustang diesel'*).
- `STATE`: aesthetic grade of the car, A, B, C and D (where A is the closest to a new product, and D the worst)
- `PRICE`: price the merchant is selling the vehicle
- `PRICE_NEW`: MSRP of the vehicle
- `MERCHANT_ID`: unique ID for each merchant at CarMarket
- `CUSTOMER_COUNTRY`: customer's country
- `MERCHANT_COUNTRY`: merchant's headquarters' country
- `PRODUCT_RELEASE_DATE`: year the model was released (e.g. 1982 for Nissan Micra), note that this is not the date the vehicle was built (which would be between the release date and today)

Good luck ! Feel free to reach out to us should you have any additional question

The *BackBox* team @BackMarket