

DATA SCIENCE ASSIGNMENT

FOR JUNIOR DATA SCIENTIST

# Submission Deadline

Please submit your completed results within 7 days of receiving this assignment. Your results can be submitted to the HR representative who provided you the assignment.

# Policy

You are not allowed to discuss the assignment with anyone except (i) a Data Scientist at LOGIVAN and (ii) the HR person who sent you this assignment. You are welcome to consult available resources online.

# Assignment

The objective of this assignment is to validate the performance of a model. Model validation plays a very important role in modeling, since it provides essential information about the model.

At LOGIVAN, price of a shipment is auto-calculated by different regression models. Unfortunately, our data scientist has quitted the job to chase his childhood dream: becoming a progamer. We hope you can help us to validate the performance of our pricing models.

# Data

You are given four CSV (comma separated) data file:

* predicted\_value\_x.csv: The value of x is range from 1 to 3, stands for model’s ID. Each file contains the output of one pricing model.
* ground\_truth.csv: the price observed.

In these files, each row associates an *order\_id* with a *final\_price*.

# Deliverables

The following deliverables must be submitted to LOGIVAN:

* Your report in PDF format. The report should answer all below questions:
* Your code to calculate all performance metrics reported.

# Questions

1. What metrics did you choose to validate our pricing model? Why?
2. Based on the metrics, what do you think about our models? Which is the best and worst model? Why?
3. Beside the metrics used on the report, is there anything concerned you? Could you share with us your concerns during making the report?