

SUS⁵

1 Description of the Problem and Variables

The dataset consists of information from 9679 anonymized customers of Generali who have been involved in a car accident which resulted in an insurance claim for the company. For 6775 of these customers you know the cost of the insurance claim [Target_cost_euro], and you possess additional input variables describing several features of the customer, the vehicle and the accident. For the other 2904 customers, you have only information on the inputs and not on the cost of the insurance claim.

Your goal is to predict Target_cost_euro for the held-out 2904 customers.

There are 36 inputs and one response [Target_cost_euro, available only for the training set]. In the training set you will find also a variable [weight] which takes always value 1. This variable is only required to compute the score [see below], and hence is not necessary for the analysis. Missing values are denoted with NA. A detailed description of the variables can be found below.

Name of the Variable	Type of Variable	Short Description
Target_cost_euro	Numeric [Response]	Cost of the insurance claim [euro]
Date_claim_opening	Date [Input]	Date of claim opening
Date_of_accident	Date [Input]	Date of accident
Customer_code1	Qualitative [Input]	Type of customer [internal classification 1]
Customer_code3	Qualitative [Input]	Type of customer [internal classification 3]
Customer_code4	Qualitative [Input]	Type of customer [internal classification 4]
Customer_merit_class	Numeric [Input]	Merit class of the customer
Flag_virtuous_driver	Binary [Input]	Flag denoting virtuous drivers: N = No, S = Yes
Presence_of_black_box	Binary [Input]	Flag denoting the presence of a black-box in the car
Bareme_table_code_customer	Qualitative [Input]	See the specific description below
Bareme_table_code_other_driver	Qualitative [Input]	See the specific description below
Province_code	Qualitative [Input]	Province
Region	Qualitative [Input]	Region

Date_vehicle_immatriculation	Date [Input]	Date of vehicle registration
Vehicle_typology	Qualitative [Input]	Vehicle macro-typology
Vehicle_power_source	Qualitative [Input]	Vehicle power source
Vehicle_catalytic_converter	Qualitative [Input]	Type of the catalytic converter
Flag_hybrid_vehicle	Binary [Input]	Flag denoting hybrid vehicles: N = No, S = Yes
Vehicle_annual_distance_km	Numeric [Input]	Declared annual distance of the vehicle [km]
Flag_historical_vehicle	Binary [Input]	Flag denoting classic vehicles: N = No, S = Yes
Vehicle_engine_capacity	Numeric [Input]	Engine capacity of the vehicle
Vehicle_engine_rpm	Numeric [Input]	Speed of the engine in revolutions per minute [rpm]
Vehicle_torque_kgm	Numeric [Input]	Vehicle torque [kgm]
Vehicle_torque_rpm	Numeric [Input]	Vehicle torque [rpm]
Vehicle_n_valves_cyl	Numeric [Input]	Vehicle number of valves
Vehicle_fiscal_HP	Numeric [Input]	Vehicle fiscal horse-power
Vehicle_price_euro	Numeric [Input]	Price of the vehicle [euro]
Vehicle_max_speed_kmh	Numeric [Input]	Vehicle maximum speed [kmh]
Vehicle_acceleration_0_100_s	Numeric [Input]	Vehicle acceleration 0–100 [s]
Vehicle_length_m	Numeric [Input]	Vehicle length [m]
Vehicle_width_m	Numeric [Input]	Vehicle width [m]
Vehicle_height_m	Numeric [Input]	Vehicle height [m]
Vehicle_weight_kg	Numeric [Input]	Vehicle weight [kg]
Vehicle_weight_full_load_kg	Numeric [Input]	Vehicle weight in full load [kg]
Vehicle_n_doors	Numeric [Input]	Vehicle number of doors

Vehicle_n_seats	Numeric [Input]	Vehicle number of seats
Vehicle_code_supercharging	Qualitative [Input]	Vehicle type of supercharging

Two variables which require a more detailed description are **Bareme_table_code_customer** as well as **Bareme_table_code_other_driver**. The Italian regulations require that both the customer and the other driver describe the dynamic of the accident. In particular, **both drivers must declare what they were doing at the moment of the car accident by choosing among 18 options outlined below.**

Code	Description
0	No selection
1	Vehicle parked or stopped
2	Restarting after a break / opening a door
3	Parking
4	Coming out from a parking lot, a private place, or a local road
5	Entering in a parking lot, a private place, or a local road
6	Entering in a roundabout
7	Circulating in a roundabout
8	Proceeding in the same direction or line
9	Proceeding in the same direction but on a different line
10	Changing line
11	Passing
12	Turning right
13	Turning left
14	Going backward
15	Invading the roadway reserved for circulation in the opposite direction
16	Coming from the right

Given the declarations of the drivers, the **Barème table** [summarized below] allows to quickly identify if the customer was right or wrong. The possible outcomes are: **R** = customer is right, **W** = customer is wrong, **S** = shared responsibility of customer and other driver, **NV** = not verifiable.

Customer	Other driver																	
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
17	W	W	S	W	S	W	W	W	S	W	W	S	W	W	S	S	W	S
16	R	W	R	NV	R	R	R	R	R	NV	R	R	R	R	R	R	S	R
15	W	W	W	W	W	W	W	W	NV	W	W	W	W	W	S	S	W	S
14	W	W	W	W	S	W	W	W	W	W	S	W	W	W	S	S	W	S
13	NV	W	R	R	R	NV	NV	W	R	W	R	S	NV	S	R	R	W	R
12	NV	W	R	R	R	R	W	W	R	W	R	R	S	NV	R	R	W	R
11	W	W	R	NV	R	NV	NV	W	R	W	R	S	W	S	R	R	W	S
10	W	W	R	NV	R	NV	W	W	S	W	S	W	W	W	S	R	W	R
9	NV	W	R	NV	R	R	NV	S	NV	S	R	R	R	R	R	R	NV	R
8	W	W	S	W	S	W	W	W	NV	NV	S	W	W	W	R	NV	W	S
7	NV	W	W	R	R	R	R	S	R	S	R	R	R	R	R	R	W	R
6	W	W	W	R	R	R	S	W	R	NV	R	NV	R	NV	R	R	W	R
5	W	W	S	S	S	S	W	W	R	W	NV	NV	W	NV	R	R	W	R
4	W	W	S	S	S	S	W	W	S	W	W	W	W	W	S	R	W	S
3	W	W	S	S	S	S	W	W	R	NV	NV	NV	W	W	R	R	NV	R
2	W	W	S	S	S	S	R	R	S	W	W	W	W	W	R	R	W	S
1	R	NV	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
0	NV	W	R	R	R	R	R	NV	R	NV	R	R	NV	NV	R	R	W	R

2 Bocconi Data Science Challenges Platform

The submission of your predictions for the test set and of your report for the final evaluation is managed by the [Bocconi Data Science Challenges Platform](https://sus5-competition.unibocconi.eu/) [https://sus5-competition.unibocconi.eu/]. To access the page of the SUS⁵ challenge:

1. **Log-in** at <https://sus5-competition.unibocconi.eu/> using the guidelines and the credentials that have been sent to you by e-mail.
2. **Enter** the [Stats under the Stars 5](#) challenge.

The challenge page contains all the information regarding the data competition as well as a submission section where you can upload:

1. Your predictions for the test set [[upload submission](#)]
[UPLOAD GUIDELINES](#) for a description on how the predictions' file should be formatted.
2. Your report for the final evaluation [[upload workflow](#)].

NOTE: each upload overwrites the previous ones.

Once a prediction file has been submitted, the platform computes the score between the submitted predictions and the true values of [Target_cost_euro](#) [unknown to teams], showing in the [Leaderboard](#) the current ranking compared to the performance of the other teams.

During the challenge, only a [Temporary Leaderboard](#) is displayed. This temporary Leaderboard is based only on a subset of the test set, while the [Final Leaderboard](#)—available at the end of the hackathon—is based on all your submitted predictions for the test set. **In this challenge the score is calculated using the mean absolute error**

$$\frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|,$$

where $y_i, i = 1, \dots, n$ is the true value of [Target_cost_euro](#) for the i -th unit in the test set and $\hat{y}_i, i = 1, \dots, n$ is your submitted prediction for such a unit.

NOTE: the platform computes, more generally, weighted mean absolute errors $\sum_{i=1}^n w_i |\hat{y}_i - y_i| / \sum_{i=1}^n w_i$. However, in this case all the weights $w_i, i = 1, \dots, n$ are equal to 1. Hence **the formula used is the classical mean absolute error**.

The final report must be in pdf format and no longer than 5 pages. Good reports contain

- [i] A brief summary of the path you made to reach your final solution.
- [ii] The key guidelines to implement the code which led to your predictions.
- [iii] An accessible, but clear, description of the methods comprising your solution.
- [iv] Additional interpretations, comments and relevant data structures or patterns arising from your analysis which may be useful for the company.

REMEMBER to upload the report before the challenge ends.

3 Awards, Evaluation and Rules

All the detailed information regarding the awards, the evaluation committee and the rules which you agreed to follow during your registration can be found at <http://www.sus5.unibocconi.eu>. **Please, take a moment to re-read them carefully before starting the challenge.**