

# Data Science for Non-Life Insurance

Academic Year 2022-2023

KU Leuven



## Instructions for the Assignments

You should provide answers in English. You get points for the methods you use, for your clear explanation and discussion of the model assumptions, strengths and limitations, and for the overall quality of your report, not just for the final answer. You should use either **R** or **Python** to solve this assignment.

Happy learning!

## Deliverables for the Assignments

Please hand in on or before May 22, 2023 via TOLEDO:

1. A report (format: pdf file or html) or notebook (e.g. in RMarkdown or Colab) in which you answer the *Assignment Questions* stated below and which contains a selection of relevant figures. The figures should be labeled properly and integrated in your report.
2. If the code is not integrated in your report or notebook, you should provide a separate **R** or **Python** script that allows to reproduce your calculations. Your code should be well-organized and easy to read.

Please mention the names and student numbers of your team members on both items. It is allowed to work in teams (with three students maximum); it suffices to submit one solution per team.

Each team of students will deliver an (online) pitch presentation in the final week of the course (schedule to be determined). This allows the teaching team to give feedback on the report, the models constructed and the presentation.

## Assignment Questions

You analyze the data set (in `.csv`) that is available on TOLEDO. This data set contains observations on the variables listed in the table printed below. Your report should document the following steps:

1. An exploratory data analysis.
2. The construction of a (technical) tariff plan for a motor insurance product. Hereto you analyze both the frequency and severity information in the data with (at least) two of the methods/algorithms discussed in the lectures (GLM, GAM, regression tree, bagging, random forest, gradient boosting, neural networks). You combine frequency and severity models into a technical tariff. You compare the performance of the constructed models, based on your own defined set of criteria. You discuss the resulting (technical) tariff structure and provide insights in the designed tariff plan.
3. As an extra step you will discuss and demonstrate the calculation of a safety (or risk) loading on top of the technical pure premium. You explore the literature on insurance pricing and propose a suitable strategy. You implement your proposed strategy and demonstrate your findings on the data set.

There is no need to answer the above questions separately (question by question) in your report. A well structured text that covers the above items is preferred. Be creative and rigorous!

---

<code>ageph</code>	age of the policyholder
<code>CODPOSS</code>	postal code in Belgium
<code>duree</code>	exposure, fraction of the year the insured is covered
<code>lnexpo</code>	log of exposure
<code>nbrtotc</code>	total number of claims during period of exposure
<code>chargetot</code>	total claim amount
<code>agecar</code>	age of the car: 0 – 1, 2 – 5, 6 – 10, > 10
<code>sexp</code>	sex of the policyholder: male or female
<code>fuelc</code>	type of fuel: petrol or gasoil
<code>split</code>	split of the premium: monthly, once, twice, three times per year
<code>usec</code>	use of the car: private or professional
<code>fleetc</code>	car belonging to a fleet: yes or no
<code>sportc</code>	sport car: yes or no
<code>coverp</code>	coverage: MTPL, MTPL+, MTPL+++
<code>powerc</code>	power of the car: < 66, 66-110, >110

---