



## Claim frequency estimation

Competition challenge: Calculate the expected number of claims for testing portfolio

We are presenting you with a typical insurance pricing challenge: to predict claim frequency which is the main driver for determining the price of a product. We have selected motor third party liability insurance since it can provide a variety of potential pricing factors which have impact on claim frequency. The prevalence of this insurance product on the market provides consistent market information for numerous risks.

The provided database of claims is split into two parts. In the first part, the contenders will be provided with exposure and number of claims for some combinations of potential factors at certain levels. The exposure represents a number of insured units in that group. In the second part, contenders will receive just exposure for a set of combinations of pricing factors. Some factors are categorical, some are ordinal. The main goal is to estimate the number of claims for each row in the output file as accurately as possible. Solutions are real numbers. As a measure weighted square of difference between realized and estimated number of claims will be used. The lower you get, the better the result.

**Technical instructions:** All the data can be found on a cloud:

Quickconnect.to

ID: aclogatec

User: MChallenge

Password: Portoroz.2021

Folder: Math Challenge/Challenge 4

The data is available in the folder Math Challenge/Challenge 4. Learning data is available in file L-Group.csv. The competitors must complete a file T-Group.csv with best estimates of number of claims for each combination of premium factors in the rightmost column, which is currently empty so the T-Group file would have the same structure as L-Group file. The data is based on insurance company experience, so it includes also some noise. Besides T-Group.csv file the competitors must provide also a description of the solution.