**Zadanie rekrutacyjne**

Do wykonania poniższych zadań należy wykorzystać załączony zbiór modelowanie\_pricing\_EH.csv. Jest to publicznie dostępny zbiór dotyczący francuskiego rynku polis komunikacyjnych, jednak przygotowaliśmy go w taki sposób, żeby sprawiał problemy przy wczytywaniu – naprawienie pliku, tak aby dało się go wczytać do pythona, jest częścią zadania.

W okresie ubezpieczenia klient może mieć wiele szkód, dlatego nie mówimy o prawdopodobieństwie szkody, a o częstości szkód – jest to liczba szkód podzielona przez liczbę lat ubezpieczenia (ekspozycję). Dla polis krótszych niż rok ekspozycja jest poniżej 1.

1. Przedstawić dane, tak aby słuchacz, który nie widział tych danych wiedział co się w nich znajduje.
2. Zbudować jak najlepszy model prognozujący częstość szkód, w tym przynajmniej jeden model klasy GLM oraz przynajmniej jeden model ML/DL.
3. Porównać wyniki zbudowanych modeli i wybrać najlepszy z nich.
4. Wskazać segmenty klientów, dla których predykcje dwóch zbudowanych modeli GLM i ML/DL najbardziej się różnią.
5. Dla modelu GLM wskazać przykładowych klientów o możliwie najniższym ryzyku oraz o możliwie najwyższym ryzyku i policzyć dla nich prognozowaną częstość szkód.
6. Dla modelu GLM wskazać zmienną, która najbardziej różnicuje prognozowaną częstość szkód.