

# Modelowanie Statystyczne w Zarządzaniu Wierzytelnościami Masowymi

## Laboratorium 6

### Wymagane biblioteki

```
library(data.table)
```

### Zadanie 1

Przeanalizuj wartości NA na zmiennej `LoanAmount`. Czy wszystkie wartości NA w przypadku tej zmiennej oznaczają brak danych?

### Zadanie 2

Podaj biznesowy (ekspercki) sposób uzupełnienia wartości NA dla zmiennej `Other`.

### Zadanie 3

Uzupełnij braki danych na zmiennej `Land` wykorzystując rozkład empiryczny. Jak wykorzystać nowo pozyskane informacje w uzupełnieniu zmiennych `GDPPerCapita` i `MeanSalary`?

### Zadanie 4

Zweryfikuj dokładność uzupełniania braków danych dla zmiennej `TOA` poprzez modele lasów losowych i najbliższych sąsiadów (Wsk. Braki danych w `TOA` należy zasymulować).

### Zadanie 5

Zweryfikuj różnice pomiędzy wartościami średnich oraz przeciętnych dla rozkładów poszczególnych zmiennych opisujących sprawy. Oceń jaki wpływ na różnice mają wartości skrajne.

### Zadanie 6

Posługując się wykresami typu boxplot zidentyfikuj wartości odstające (jaka reguła jest przyjęta w funkcji `boxplot`) na poszczególnych zmiennych opisujących sprawy. Usuń przypadki z wartościami odstającymi, a następnie wykonaj wykres ponownie. Czy nadal możesz zaobserwować wartości odstające?