Modelowanie Statystyczne w ZarzÄ…dzaniu WierzytelnoĹ›ciami Masowymi

## Laboratorium 2.

Biblioteki:

## Wymagane pakiety:   
  
library(data.table)  
library(corrgram)  
library(corrplot)  
  
## Opcjonalne pakiety:  
  
library(ggplot2) # wizualizacja  
library(discretization) # dyskretyzacja zmiennych ciągłych  
library(arules) # dyskretyzacja zmiennych ciągłych  
library(arulesCBA) # dyskretyzacja zmienncyh ciągłych

### Zadanie 1.

Wyznacz skuteczność SR (suma wpłat przez wartość zadłułenia) w różnych horyzontach czasu (np. 3M, 6M, 12M) i wyznacz jej statystyki opisowe (kwantyle) w podziale na:

* Gender
* ExtrernalAgency
* Bailiff
* ClosedExecution
* M\_LastPaymentToImportDate (zaproponuj podział wg tej zmiennej)
* DPD (zaproponuj podział wg tej zmiennej)
* Age (zaproponuj podział wg tej zmiennej)
* TOA (zaproponuj podział wg tej zmiennej)

### Zadanie 2.

Wyniki zaprezentuj również na wykresie. Które zmienne najlepiej różnicują skuteczność (co rozumiesz poprzez “różnicują”“)?

### Zadanie 3.

Wyznacz korelacje pomiędzy zmienną skuteczności 12M a innymi zmiennymi.

Czy w zbiorze danych istnieją inne zmienne (od wskazanych) z dużą wartością korelacji?

### Zadanie 4.

Wyznacz dotarcie per sprawa (czy był kontakt w sprawie telefoniczny, lub bezpośredni) w do 6M obsługi i wyznacz w statystyki opisowe (kwantyle) w podziale na wybrane zmienne (np. zmienne z zadania 1, lub zmienne, które różnicują dotarcie).

### Zadanie 5.

Czy istnieje zaleśność pomiędzy ilością wykonywanych telefonów, wizyt, lub wysyłanych listów, a zmiennymi opisującymi sprawę (zmienne w cases).

### Zadanie 6.

Dla wybranych zmiennych dokonaj przekształceń i zapisz jako nowe zmienne:

* standaryzowane (o średniej zero i warinacji 1)
* normalizowane (przekształcenie wartości zmiennej na odcinek [0, 1]).
* logarytmowane
* pierwiatskowane

Wyznacz korelację dla zmiennych orginalnych oraz korelację ich przekształconych odpowiedników. Co można zauważyć?

### Zadanie 7.

Wyznacz wykres liniowy pokazujący skumuluwaną skuteczność SR w kolejnych miesiącach obsługi dla następujących typów spraw:

* SR w sprawach bez kontaktu (zarówno telefoniczny jak i wizyta)
* SR w sprawach z kontaktem
* SR w sprawach z ugód?
* SR w sprawach przekazanych do sądu.

Powyższe zdarzenia narzucają hierarchię procesu, tzn. jeżli na sprawie był kontakt w 3M to sprawa ta jest uważana za sprawę z kontaktem do 12M (do końca). Jeżli w sprawie był kontakt w 2M oraz w tej sprawie została podpisana ugoda w 2M, to zaliczamy tą sprawę (jej skuteczność) do kategorii spraw z ugodą w 2M.