



UCZELNIA LUDZI CIEKAWYCH

Studia podyplomowe Big Data — Analytics & Society

IMIĘ NAZWISKO

---

## Wykorzystanie metod Big Data w prognozowaniu zapotrzebowania na nauczycieli akademickich

Utilitlization of Big Data methods in demand  
forecasting for academic teachers

---

Praca dyplomowa

Promotor: dr Łukasz Wawrowski

Poznań 2023

# Spis treści

<b>Wstęp</b>	<b>4</b>
<b>1. Opis badanego zjawiska</b>	<b>5</b>
<b>2. Wybrana metoda statystyczna</b>	<b>7</b>
2.1. Metoda 1 . . . . .	7
2.2. Metoda 2 . . . . .	7
<b>3. Przeprowadzona analiza</b>	<b>9</b>
3.1. Przygotowanie danych . . . . .	9
3.2. Zastosowanie algorytmu . . . . .	10
3.3. Rezultaty . . . . .	10
<b>4. Przeprowadzona analiza</b>	<b>12</b>
4.1. Przygotowanie danych . . . . .	12
4.2. Zastosowanie algorytmu . . . . .	13
4.3. Rezultaty . . . . .	13
<b>Podsumowanie</b>	<b>15</b>
<b>Bibliografia</b>	<b>16</b>
<b>Załączniki</b>	<b>16</b>
<b>A. Kod wykorzystany w analizie</b>	<b>17</b>

# Spis rysunków

1.1. Logo uczelni CDV . . . . .	6
3.1. Porównanie bieżącego i początkowego wynagrodzenia pracowników . . . . .	11

## Spis tablic

1.1. Pewne losowe liczby w tabeli . . . . .	5
3.1. Statystyki samochodów . . . . .	9
4.1. Statystyki samochodów . . . . .	12

# Wstęp

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

# 1. Opis badanego zjawiska

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

Przedstawiony problem został opisany w pracy (Knuth, 1984). Bieżący stan wiedzy na ten temat przedstawia Tabela 1.1. Referencja do tabeli musi się zaczynać od **tbl-**, a do rysunku od **fig-**.

Tablica 1.1.: Pewne losowe liczby w tabeli

Col1	Col2	Col3
A	B	C
E	F	G
A	G	G

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit

eleifend ullamcorper.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.



Rysunek 1.1.: Logo uczelni CDV

Logo uczelni przedstawia Rysunek 1.1.

Więcej na temat wstawiania obiektów jest na stronie <https://quarto.org/docs/authoring/cross-references.html>

## 2. Wybrana metoda statystyczna

### 2.1. Metoda 1

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Wzory w tekście wstawiamy z wykorzystaniem pojedynczego znaku dolara:  $y = ax + b$ , natomiast jak chcemy go wycentrować to wtedy używany jest podwójny.

$$c^2 = a^2 + b^2 \tag{2.1}$$

Równanie 2.1 stanowi podstawę współczesnej matematyki.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

### 2.2. Metoda 2

W podrozdziale 2.1 przedstawiono pierwszą wersję, natomiast w tym podrozdziale znajdzie się opis jej ulepszenia. Jeżeli nie chcemy wstawiać prefiksu razem z odsyłaczem to poprzedzamy go znakiem minusa.



Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

## 3. Przeprowadzona analiza

### 3.1. Przygotowanie danych

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Tablica 3.1.: Statystyki samochodów

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

## 3.2. Zastosowanie algorytmu

W tym miejscu zastosujemy metodę opisaną w podrozdziale [2.1](#).

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

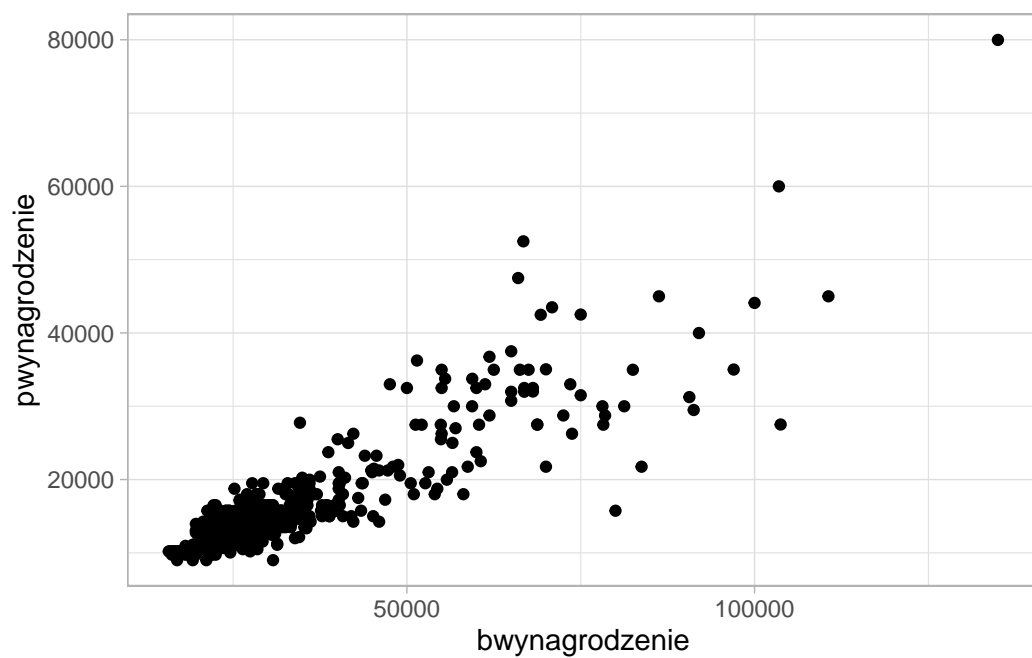
Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

## 3.3. Rezultaty

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Wykres rozrzutu przedstawiający wynagrodzenie przedstawia Rysunek [3.1](#).

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.



Rysunek 3.1.: Porównanie bieżącego i początkowego wynagrodzenia pracowników

## 4. Przeprowadzona analiza

Żeby używać bibliotek pythona trzeba je najpierw zainstalować w zakładce Terminal poleceniem `pip install pandas` itp.

```
import pandas as pd

a = 20

dane = pd.read_csv("data/klienci.csv")
```

### 4.1. Przygotowanie danych

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Tablica 4.1.: Statystyki samochodów

speed	dist
4	2
4	10
7	4
7	22
8	16
9	10

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend.

Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

## 4.2. Zastosowanie algorytmu

W tym miejscu zastosujemy metodę opisaną w podrozdziale [2.1](#).

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

## 4.3. Rezultaty

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend.

Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.

## Podsumowanie

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Nam condimentum odio sit amet nisl dictum, condimentum porta libero placerat. Proin ipsum lacus, condimentum quis tortor vel, malesuada scelerisque lacus. Curabitur ut metus id quam dapibus commodo. Vestibulum eget pharetra nibh. Maecenas pretium faucibus euismod. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Proin suscipit eleifend ullamcorper.

Sed varius fermentum leo ac tempus. Nunc tortor justo, euismod eget scelerisque in, consequat sed neque. Aliquam vulputate nibh ac tellus fringilla, et interdum sem eleifend. Nulla eleifend id dolor at egestas. Morbi magna tellus, sagittis egestas efficitur at, mollis id metus. Aliquam eget felis eget velit hendrerit vehicula nec efficitur nulla. Nam vitae nibh ut velit feugiat consectetur non sit amet justo. Vestibulum ultrices mollis odio, eget dapibus ex. Praesent posuere nisi eu ante aliquam vulputate. Praesent sed risus ex. In tincidunt tempor dolor, quis efficitur nisl interdum et. Vestibulum rutrum arcu at libero consectetur, sed porta lorem imperdiet. Duis sit amet turpis in nulla mattis volutpat. Mauris in tristique neque.



## Bibliografia

Knuth, D. E. (1984). Literate Programming. *Comput. J.*, 27(2), 97–111. <https://doi.org/10.1093/comjnl/27.2.97>

## A. Kod wykorzystany w analizie

```
# generowanie danych

rnorm(10)

dane <- data.frame(x1=rnorm(10), x2=rnorm(10), x3=rnorm(10), x4=rnorm(10))

dane$w1 <- (min(dane$x1)-max(dane$x1))/mean(dane$x1)
dane$w2 <- (min(dane$x2)-max(dane$x2))/mean(dane$x2)
dane$w3 <- (min(dane$x3)-max(dane$x3))/mean(dane$x3)
dane$w4 <- (min(dane$x4)-max(dane$x4))/mean(dane$x4)

# funkcje

oblicz_wskaznik <- function(x){

  wskaznik <- (min(x)-max(x))/median(x)

  return(wskaznik)

}

dane$w1 <- oblicz_wskaznik(dane$x1)
dane$w2 <- oblicz_wskaznik(dane$x2)
dane$w3 <- oblicz_wskaznik(dane$x3)
dane$w4 <- oblicz_wskaznik(dane$x4)

przeskaluj01 <- function(x){

  z <- (x - min(x))/(max(x)-min(x))

  return(z)
```

```

}

przeskaluj01 <- function(x){

  (x - min(x))/(max(x)-min(x))

}

dane$z1 <- przeskaluj01(dane$x1)
dane$z2 <- przeskaluj01(dane$x2)
dane$z3 <- przeskaluj01(dane$x3)
dane$z4 <- przeskaluj01(dane$x4)

plot(dane$x1, dane$z1)

# warunki

sprawdz_liczbe <- function(liczba){

  if(is.character(liczba)){
    stop("Argument jest tekstem")
  } else if(is.na(liczba)){
    stop("Argument jest brakiem danych")
  }

  if(liczba > 0){
    print("Liczba większa od 0")
  } else if(liczba < 0){
    print("Liczba mniejsza od 0")
  } else{
    print("Liczba równa 0")
  }

}

sprawdz_liczbe(9)
sprawdz_liczbe(0)
sprawdz_liczbe("abc")
sprawdz_liczbe(NA)

```

```

# ifelse v1

liczba <- 10

tekst <- ifelse(test = liczba > 0, yes = "Liczba większa od 0", no = "Liczba <= 0")

if(liczba > 0){
  tekst <- "Liczba większa od 0"
} else{
  tekst <- "Liczba <= 0"
}

# ifelse v2

tekst <- ifelse(test = liczba > 0, yes = "Liczba większa od 0",
               no = ifelse(liczba < 0, "Liczba mniejsza od 0", "Liczba równa 0"))

if(liczba > 0){
  tekst <- "Liczba większa od 0"
} else if(liczba < 0){
  tekst <- "Liczba mniejsza od 0"
} else{
  tekst <- "Liczba równa 0"
}

# przykład na danych

dane$tekst <- ifelse(test = dane$x1 > 0, yes = "Liczba większa od 0",
                   no = ifelse(dane$x1 < 0, "Liczba mniejsza od 0", "Liczba równa 0"))

# zadanie

czy_podzielna35 <- function(liczba){

  if(liczba %% 3 == 0 && liczba %% 5 == 0){
    print("Liczba podzielna przez 3 i 5")
  } else if(liczba %% 5 == 0){
    print("Liczba podzielna przez 5")
  } else if(liczba %% 3 == 0){
    print("Liczba podzielna przez 3")
  }
}

```

```

    } else {
      print("Liczba niepodzielna ani przez 3 ani przez 5")
    }
  }

  czy_podzielna35(15)
  czy_podzielna35(17)
  czy_podzielna35(12)
  czy_podzielna35(10)

  for(month in 1:length(month.name)){
    print(month.name[month])
  }

  seq_along(month.name)
  1:length(month.name)

  for(month in seq_along(month.name)){
    print(month.name[month])
  }

  for(month in month.name){
    print(month)
  }

  # rzut kostką

  wynik_kostka <- numeric(1000000)

  for(rzut in seq_along(wynik_kostka)){
    wynik_kostka[rzut] <- sample(x = 1:6, size = 1)
  }

```

```

}

wynik_kostka
mean(wynik_kostka)
table(wynik_kostka)

# zadanie

for(i in seq(99,1,-1)){
  if(i == 1){
    print('1 bottle of beer on the wall, 1 bottle of beer! So take it down, pass it around,')
  } else if(i == 2){
    print('2 more bottles of beer on the wall, 2 more bottles of beer! So take one down, pa')
  } else {
    print(paste0(i, ' bottles of beer on the wall, ', i, ' bottles of beer! So take it down,')
  }
}

```