

# **STATYSTYKA - PROJEKT**

Marcel Miłosz, Data Science,

Niestacjonarne

Nr. Albumu: 25403

## **Informacje do projektu:**

Plik główny - Main.r Baza danych 10. towarów - Wszystko.csv

! UWAGA ! Do odpalenia programu wystarczy odpalić dwie funkcje, które znajdują się na samym dole programu (linia: 127 i 128) Do zmiennej 'lista\_Pkt' przypisujemy funkcję 'Oblicz(nazwa\_bazyDanych)' DO funkcji 'Wykres(lista\_Pkt)'

W Wyniku otrzymujemy gotowy wykres pokazujący jak zmieniała się cena w każdym roku dla każdego województwa ! Linia czerwona (pogrubiona) to średnia !

Zdjęcia przykładowych wykresów:

- Cukier.png
- Wizyty.png

Plik 'Cukier\_Analiza.pdf' zawiera przykładową analizę dla towaru "Cukier 1kg"

===== KOD =====

```
library(ggplot2)

library(tidyverse)


# Dane z pliku

# W pliku głównym zamieniłem ',' -> '.'

dataExc = read.csv(file = 'Wszystko.csv', sep = ';', dec = ',', encoding = 'UTF-8')

dataExc = dataExc[-nrow(dataExc),]


# Zapisanie nazw kolumn bardziej normalnie

produkt = c("Cukier", "Mleko", "Olej", "Podzielowanie", "Rajstopy", "Ryz", "Strzyzenie", "Szynk", "Wizyta", "Woda") # 10

miesiac = c("Styczeń", "Luty", "Marzec", "Kwiecien", "Maj", "Czerwiec", "Lipiec", "Sierpień", "Wrzesień", "Październik",
"Listopad", "Grudzien") # 12

rok = c("2006":"2019") # 14


ind = 3


for (r in 1:10) {
  for (m in 1:12) {
    for (p in 1:14) {
      names(dataExc)[ind] = paste0(miesiac[m], "_", produkt[r], "_", rok[p])

      ind = ind + 1
    }
  }
}


# Oblicza średnia dla wszystkich lat dla konkretnego towaru

Oblicz = function(database) {

  Srednie_Wojewodztw = rep(0, each = 16)

  wojewodztwa = c('Dolno', 'Kujaw', 'Lubel', 'Lubus', 'Łódzk', 'Małop', 'Mazow', 'Opols', 'Podka', 'Podla', 'Pomor', 'Śląsk',
'Święt', 'Warmi', 'Wielk', 'Zacho')

  lata = c("2006", "2007", "2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019")

  lista_punktow = list()
```

```

for (woj in 2:17) {

  wiersz_wojewodztwa = database[woj, 1:168]    # Tutaj wybieramy konkrety wiersz / województwo


  emptyV = rep(0, each = 168)                  # Pusty wektor który później wypełnimy liczbami dla konkretnego wojewodztwa

  emptyV_Srednia_wRoku = rep(0, each = 14)      # Pusty wektor który później wypełnimy średnimi dla każdego roku oraz
konkretnego wojewodztwa


  for(i in 1: 168) {

    emptyV[i] = as.numeric(as.vector(wiersz_wojewodztwa[[i]]))

  }


  for(rr in 1:14) {

    val = rr

    suma = 0

    for(i in 1:12) {

      suma = suma + emptyV[val]

      val = val + 14

    }

    emptyV_Srednia_wRoku[rr] = (suma / 12)

  }


  Srednie_Wojewodztw[woj - 1] = mean(emptyV_Srednia_wRoku)

  # print(emptyV_Srednia_wRoku)

  lista_punktow[woj - 1] = list(emptyV_Srednia_wRoku)

}


jaki_Produkt = substr(names(database[1]), unlist(gregexpr(pattern = "_", names(database[1])))[1] + 1,
unlist(gregexpr(pattern = "_", names(database[1])))[2] - 1)


return(lista_punktow)

}


# Funkcja do stworzenia wykresu

Wykres = function(x) {

  finalna_baza = data.frame(x)

```

```
daty = c("2006", "2007", "2008", "2009", "2010", "2011", "2012", "2013", "2014", "2015", "2016", "2017", "2018", "2019")
```

```
# (OBLICZENIE średniej)
```

```
srednia = rep(0, each = 14)
```

```
for(i in 1:14) {  
  suma = 0  
  pointer = i  
  for(j in 1:224) {  
    if(j == pointer) {  
      suma = suma + unlist(finalna_baza)[j]  
      pointer = pointer + 14  
    }  
  }  
  srednia[i] = suma / 16  
}
```

```
ggplot(finalna_baza) +
```

```
  geom_line(aes(x = daty, y = srednia, colour = "- ŚREDNIA -"), group = 1, size = 1.7) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[1]))), colour = 'Dolnośląskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[2]))), colour = 'Kujawsko-Pomorskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[3]))), colour = 'Lubelskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[4]))), colour = 'Lubuskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[5]))), colour = 'Łódzkie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[6]))), colour = 'Małopolskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[7]))), colour = 'Mazowieckie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[8]))), colour = 'Opolskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[9]))), colour = 'Podkarpackie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[10]))), colour = 'Podlaskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[11]))), colour = 'Pomorskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[12]))), colour = 'Śląskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[13]))), colour = 'Świętokrzyskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[14]))), colour = 'Warmińsko-Mazurskie', group = 1)) +  
  +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[15]))), colour = 'Wielkopolskie', group = 1)) +  
  geom_line(aes(x = daty, y = as.numeric(as.vector(unlist(finalna_baza[16]))), colour = 'Zachodniopomorskie', group = 1)) +
```

```

labs(title = "Wykres zmiany cen dla konkretnego towaru, dla wszystkich województw na przestrzeni lat 2006 - 2019") +
xlab("Lata 2006 -> 2019") +
ylab("Średnie ceny dla konkretnego województwa")

}

# Stworzenie konkretnych data.frame'ow dla każdego towaru

Cukier = data.frame(dataExc[3:170])          # data frame dla cukru
Mleko = data.frame(dataExc[171:338])         # data frame dla mleka
Olej = data.frame(dataExc[339:506])          # data frame dla oleju
Podzielowanie = data.frame(dataExc[507:674]) # data frame dla podzielowania
Rajstopy = data.frame(dataExc[675:842])      # data frame dla rajstopy
Ryz = data.frame(dataExc[843:1010])          # data frame dla ryżu
Strzyżenie = data.frame(dataExc[1011:1178])  # data frame dla strzyżenia
Szynka = data.frame(dataExc[1179:1346])      # data frame dla szynki
Wizyta = data.frame(dataExc[1347:1514])      # data frame dla wizyty
Woda = data.frame(dataExc[1515:1682])        # data frame dla wody

# ! TUTAJ JEST SERCE PROGRAMU !

# Do zmiennej 'lista_Pkt' przypisujemy funkcję 'Oblicz' a w niej parametr bazy (podane wyżej) np. Cukier, Mleko, Olej itd.
# Funkcja zwróci nam potrzebne dane do stworzenia wykresu
# Wystarczy zaznaczyć linię nr. 111 i 112 i odpalić program (dostaniemy wykres dla konkretnego towaru)

# Czerwona linia (pogrubiona) to średnia
lista_Pkt = Oblicz(Wizyta)
Wykres(lista_Pkt)

```