

Zadanie 1:pobieranie danych

```
""{r}
```

```
library(tidyverse) library(dplyr) library(tidyr) library(ggplot2)
```

```
url <- "http://mlr.cs.umass.edu/ml/machine-learning-databases/census-income/census-income.data" df <-  
read.table(url,sep=","na.strings = "?") head(df)
```

```
""
```

Nazywanie kolumn

```
""{r}
```

```
colnames(df) <- c("age","workclass","fnlwgt","education","education-num","marital-status","occupation","relationship","race",  
"gain","capital-loss","hour_per_week","native_country","class") head(df)
```

```
""
```

Zadanie 2 : analiza danych pod katem brakujacych wartosci

```
"" {r}
```

```
anyNA(df)
```

```
kompletne.dane <- complete.cases(df) df[kompletne.dane,][ ,1:10] is.na(df) sum(is.na(df))
```

```
colSums(is.na(df))
```

```
no.question.mark <- apply(df, 1, function(r) !any(r %in% ' ?')) no.question.mark clean <- df[no.question.mark,]  
clean head(clean)
```

```
""
```

Zadanie 3:

Wykres przedstawiajacy udzial, kobiet i mezczyzn w grupie

```
""{r}
```

```
head(clean)
```

```
table(cleansex)pie(table(cleansex), main="udzial, kobiet i mezczyzn w grupie")
```

```
""
```

Histogram przedstawiajacy wiek osob w grupie:

```
""{r} table(cleanage)plot(x = table(cleanage), main="Wiek osob w grupie",xlab="wiek",ylab="ilosc osob")
```

```
"" Zadanie 4
```

tabela przedstawiajaca procentowy udzial, natywnych amerykanow w grupie

```
""{r} a <- select(clean,native_country) head(a) liczebnosc <- table(a) head(liczebnosc) liczebnosc[-1, ] procenty  
<- (liczebnosc/sum(liczebnosc))*100 procenty round(procenty, digits = 2)
```

```
"" Zadanie 5 Poslugujac sie wykresami/tabelami pokazac jak ksztaltuje sie ilosc przepracowanych godzin w  
tygodniu ze wzgledu na sektor pracy (kolumna workclass) ""{r}
```

```
wclass <- select(clean,workclass,hour_per_week) head(wclass,10)
```

```
l <- arrange(wclass, workclass, hour_per_week) head(l,40)
```

```
levels(clean$workclass) summary(l)
```

```
tabela <- table(cleanhour_per_week,cleanworkclass) tabela[-1] tabela<-tabela[-1] tabela[-3] tabela <- tabela[-  
3] head(tabela, 10)
```

```

barplot(tabela, space = 0, xlab = 'workclass')
''' Zadanie 6 sredni czas pracy w sektorze prywatnym w grupach wyodrebnionych ze wzgledu na wykształcenie
wsrod osobb ponizej 30 roku zycia minimalny i maksymalny czas pracy w kazdej grupie '''{r} mum <-
select(clean, age, workclass, hour_per_week, education) head(mum) class(mum)

mum <- data.frame(lapply(mum, as.character), stringsAsFactors=FALSE)

mum %>% group_by(education) %>% filter(workclass == 'Private', age <= 30) %>% summarise(sredniCzas
= mean(hour_per_week), minCzas = min(hour_per_week), maxCzas = max(hour_per_week)) '''

Zadanie 7 Przedstawic jakis ciekawy wykres prezentujacy wczytane dane
'''{r}

head(clean) stat.sum <- summary(clean) stat.sum

install.packages("rmarkdown")
'''

```