# Dokumentacja Kodu R - Zadanie 2

# Biblioteka datasets  
library(datasets)  
  
# Ustawienie ścieżki dostępu do danych  
path = "C:\Users\petit\Desktop\repos\UO\rok 3\Wprowadzenie do eksploracji danych\lista5\"  
setwd(path)  
  
# -------------- Zadanie 2a: Wczytanie i identyfikacja punktów oddalonych -------------- #  
# Wczytanie danych  
wine <- read.csv('wine\wine.data', header = FALSE)  
  
# Funkcja do identyfikacji punktów oddalonych  
identify\_outliers <- function(data) {  
 outliers <- list()  
 for(i in 1:ncol(data)) {  
 Q1 <- quantile(data[,i], probs = 0.25)  
 Q3 <- quantile(data[,i], probs = 0.75)  
 IQR <- Q3 - Q1  
 outliers[[names(data)[i]]] <- which(data[,i] < Q1 - 1.5 \* IQR | data[,i] > Q3 + 1.5 \* IQR)  
 }  
 return(outliers)  
}  
  
# Zastosowanie funkcji do danych wine  
outliers <- identify\_outliers(wine)  
  
# -------------- Zadanie 2b: Wizualizacja punktów oddalonych -------------- #  
# Wizualizacja punktów oddalonych dla pierwszych dwóch zmiennych  
plot(wine[,1], wine[,2], main = "Punkty oddalone w pierwszych dwóch zmiennych", xlab = "Zmienna 1", ylab = "Zmienna 2")  
points(wine[outliers[[1]],1], wine[outliers[[1]],2], col = 'red', pch = 19)  
points(wine[outliers[[2]],1], wine[outliers[[2]],2], col = 'blue', pch = 19)  
  
# -------------- Zadanie 2c: Usuwanie punktów oddalonych -------------- #  
# Usunięcie punktów oddalonych z danych  
clean\_wine <- wine  
for(i in 1:length(outliers)) {  
 clean\_wine <- clean\_wine[-unique(outliers[[i]]), ]  
}  
  
# -------------- Zadanie 2d: Wizualizacja po usunięciu punktów oddalonych -------------- #  
# (Kod wizualizacji po usunięciu punktów oddalonych powinien być tutaj)  
  
# -------------- Zadanie 2e: Zapisanie zmodyfikowanego zbioru danych -------------- #  
# Zapisanie oczyszczonych danych do pliku CSV  
write.csv(clean\_wine, "wine\_cleaned.csv", row.names = FALSE)