

Activitat 4

Marc Cervera Rosell

21-06-2024

```
setRepositories(ind=2) # Per descarregar paquets de CRAN
```

1 Preprocessament

1.1 Variables Income i Year__Birth

```
tryCatch({  
  data <- read.csv("marketing.csv", header = TRUE)  
  print("Fitxer llegit correctament")  
}, error = function(e){  
  cat("ERROR en la lectura del fitxer:",conditionMessage(e),"\n")  
})
```

```
## [1] "Fitxer llegit correctament"
```

```
columns <- names(data)  
type <- sapply(data, class)  
for (i in seq_along(columns)) {  
  cat("La columna", columns[i], "és de tipus", type[i], "\n")  
}
```

```
## La columna ID és de tipus integer  
## La columna Year_Birth és de tipus integer  
## La columna Education és de tipus character  
## La columna Marital_Status és de tipus character  
## La columna Income és de tipus integer  
## La columna Kidhome és de tipus integer  
## La columna Teenhome és de tipus integer  
## La columna Dt_Customer és de tipus character  
## La columna Recency és de tipus integer  
## La columna MntWines és de tipus integer  
## La columna MntFruits és de tipus integer  
## La columna MntMeatProducts és de tipus integer  
## La columna MntFishProducts és de tipus integer  
## La columna MntSweetProducts és de tipus integer  
## La columna MntGoldProds és de tipus integer  
## La columna NumDealsPurchases és de tipus integer  
## La columna NumWebPurchases és de tipus integer  
## La columna NumCatalogPurchases és de tipus integer  
## La columna NumStorePurchases és de tipus integer  
## La columna NumWebVisitsMonth és de tipus integer  
## La columna AcceptedCmp3 és de tipus integer  
## La columna AcceptedCmp4 és de tipus integer
```

```
## La columna AcceptedCmp5 és de tipus integer
## La columna AcceptedCmp1 és de tipus integer
## La columna AcceptedCmp2 és de tipus integer
## La columna Complain és de tipus integer
## La columna Z_CostContact és de tipus integer
## La columna Z_Revenue és de tipus integer
## La columna Response és de tipus integer
```

Després d'observar els tipus de les variables del conjunt de dades, solament es procedirà a fer un canvi de tipus. Aquest canvi es produirà en la variable *Dt_Customer* que passarà de ser de tipus *character* a tipus *date*.

```
data_transformed <- transform(data,
                               Dt_Customer = as.Date(Dt_Customer))

columns <- names(data_transformed)
type <- sapply(data_transformed, class)
for (i in seq_along(columns)) {
  cat("La columna", columns[i], "és de tipus", type[i], "\n")
}
```

```
## La columna ID és de tipus integer
## La columna Year_Birth és de tipus integer
## La columna Education és de tipus character
## La columna Marital_Status és de tipus character
## La columna Income és de tipus integer
## La columna Kidhome és de tipus integer
## La columna Teenhome és de tipus integer
## La columna Dt_Customer és de tipus Date
## La columna Recency és de tipus integer
## La columna MntWines és de tipus integer
## La columna MntFruits és de tipus integer
## La columna MntMeatProducts és de tipus integer
## La columna MntFishProducts és de tipus integer
## La columna MntSweetProducts és de tipus integer
## La columna MntGoldProds és de tipus integer
## La columna NumDealsPurchases és de tipus integer
## La columna NumWebPurchases és de tipus integer
## La columna NumCatalogPurchases és de tipus integer
## La columna NumStorePurchases és de tipus integer
## La columna NumWebVisitsMonth és de tipus integer
## La columna AcceptedCmp3 és de tipus integer
## La columna AcceptedCmp4 és de tipus integer
## La columna AcceptedCmp5 és de tipus integer
## La columna AcceptedCmp1 és de tipus integer
## La columna AcceptedCmp2 és de tipus integer
## La columna Complain és de tipus integer
## La columna Z_CostContact és de tipus integer
## La columna Z_Revenue és de tipus integer
## La columna Response és de tipus integer
```

S'observa que després de l'aplicació de la funció *transform()* el tipus de la variable *Dt_Customer* queda modificat.

Finalment, cal excloure les variables *Z_CostContact* i *Z_Revenue*, atès que són variables de control i que així s'indica a l'enunciat de l'activitat.

```

columns_to_exclude <- c("Z_CostContact", "Z_Revenue")
data_with_no_control_variables <- data_transformed[, !(names(data_transformed) %in%
                                                    columns_to_exclude)]
print(names(data_with_no_control_variables))

```

```

## [1] "ID"                "Year_Birth"        "Education"
## [4] "Marital_Status"    "Income"            "Kidhome"
## [7] "Teenhome"          "Dt_Customer"       "Recency"
## [10] "MntWines"          "MntFruits"         "MntMeatProducts"
## [13] "MntFishProducts"   "MntSweetProducts"  "MntGoldProds"
## [16] "NumDealsPurchases" "NumWebPurchases"    "NumCatalogPurchases"
## [19] "NumStorePurchases" "NumWebVisitsMonth"  "AcceptedCmp3"
## [22] "AcceptedCmp4"       "AcceptedCmp5"       "AcceptedCmp1"
## [25] "AcceptedCmp2"       "Complain"           "Response"

```

S'observa que en treure per pantalla les columnes del *dataset* `data_with_no_control_variables` les variables de control indicades anteriorment ja no hi són. Per tant, aquest conjunt final queda completament operatiu per a poder treballar.