

```

1 ## SCRIPT 7 - R ##
2 # Tratamiento y limpieza de datos inicial de las modificaciones
3 #
4 ######
5 # Carga de librerías -----
6 library(ggplot2)
7 library(dplyr)
8 library(openxlsx)
9 library(RODBC)
10 library(quanteda)
11 library(quanteda.textmodels)
12 library(quanteda.textstats)
13 library(readtext)
14 library(spacyr)
15 library(devtools)
16 library(zoo)
17 library(stringr) # Necesario para str_trim()
18 #
19 # Ruta de trabajo -----
20 # Usar una ruta relativa o configurar como proyecto en RStudio
21 setwd("C:/Users/guillermo.alonso/Desktop/Tesis Guillermo/_BD actualizada/Datos atom/PLACSP")
22 #
23 # Carga de archivos por año -----
24 años <- 2012:2025
25 lista_df <- list()
26 
27 for (año in años) {
28   archivo <- paste0("4.38_ContractModification_", año, ".RData")
29   load(archivo)
30   lista_df[[as.character(año)]] <- Proy
31 }
32 
33 # Unificación de datos -----
34 DF_all <- bind_rows(lista_df)
35 DF_all_u <- unique(DF_all) # Eliminamos duplicados
36 DF_all_u$entryID <- as.numeric(DF_all_u$entryID)
37 
38 # Filtro 1 (actualizado con limpieza de ID) -----
39 # Eliminamos contratos donde no hay modificación de coste ni valor final válido.
40 # También limpiamos espacios en blanco en `ContractID` y `entryID`.
41 
42 DF_all_u <- DF_all_u |>
43   mutate(
44     ContractID = str_trim(ContractID, side = "both")
45   ) |>
46   filter(
47     !is.na(ContractModificationLegalMonetaryTotal),
48     ContractModificationLegalMonetaryTotal != 0,
49     !is.na(FinalLegalMonetaryTotal),
50     FinalLegalMonetaryTotal != 0
51   )
52 
53 # Filtro 2 -----
54 # Para cada `ContractID`, seleccionamos las filas que tienen el valor máximo de
55 # `FinalLegalMonetaryTotal`. Esto nos deja potencialmente múltiples filas si hay
56 # empates.
57 
58 DF_max_final <- DF_all_u |>
59   group_by(ContractID) |>
60   filter(FinalLegalMonetaryTotal == max(FinalLegalMonetaryTotal)) |>
61   ungroup()
62 
63 # Filtro 3 -----
64 # De los casos empatados anteriores, nos quedamos con el más reciente según `updated`.
65 DF_Modis_FinalPrice <- DF_max_final |>
66   arrange(desc(updated)) |>
67   group_by(ContractID) |>
68   slice_head(n = 1) |>
69   ungroup()
70 
71

```

```
72 # Guardado de resultados -----
73 # Guardamos el dataframe como archivo .RData
74 save(DF_Modis_FinalPrice, file = "DF_Modis_FinalPrice.RData")
75 # Guardamos el dataframe como archivo .RData
76 save(DF_all_u, file = "DF_Modificaciones.RData")
77
78 # Exportamos como .csv con separador ";"
79 write.csv2(DF_Modis_FinalPrice, file = "DF_Modis_FinalPrice.csv", row.names = FALSE)
80
```