

# A9: Creación de la visualización y entrega del proyecto (PRA2). Procesamiento de los datos.

Pedro Uceda Martínez

14/05/2021

## 1. Introducción

En el presente documento se procederá a unir los distintos *dataset*, cada uno con los datos anuales desde el 2017 a 2021, que se encuentran en el siguiente enlace: <https://datos.madrid.es/portal/site/egob/menuitem.c05c1f754a33a9fbe4b2e4b284f1a5a0/?vgnextoid=710e760a581a7610VgnVCM1000001d4a900aRCRD&vgnextchannel=374512b9ace9f310VgnVCM100000171f5a0aRCRD&vgnextfmt=default>

```
library(stringi)

#Carga del csv
data_2017 <- read.csv('agencia_empleo_inscritos_2017.csv',stringsAsFactors = FALSE, sep = ';')
data_2018 <- read.csv('agencia_empleo_inscritos_2018.csv',stringsAsFactors = FALSE, sep = ';')
data_2019 <- read.csv('agencia_empleo_inscritos_2019.csv',stringsAsFactors = FALSE, sep = ';')
data_2020 <- read.csv('agencia_empleo_inscritos_2020.csv',stringsAsFactors = FALSE, sep = ';')
data_2021 <- read.csv('agencia_empleo_inscritos_2021.csv',stringsAsFactors = FALSE, sep = ';')

months_list <- list("ene" = "01", "feb" = "02", "mar" = "03", "abr" = "04", "may" = "05", "jun" = "06",
                    "jul" = "07", "ago" = "08", "sep" = "09", "oct" = "10", "nov"="11", "dic"="12",
                    "enero" = "01", "febrero" = "02", "marzo" = "03", "abril" = "04", "mayo" = "05",
                    "junio" = "06", "julio" = "07", "agosto" = "08", "septiembre" = "09", "octubre" = "10",
                    "noviembre"="11", "diciembre"="12")

#Normalización de las fechas
transform_date <- function(string_date){
  month <- substr(string_date, 1, 3)
  year_beginning <- nchar(string_date)-1
  year_end <- nchar(string_date)

  year <- substr(string_date, year_beginning, year_end)

  complete_date <- paste0("01-",months_list[month],"-", "20", year)

  complete_date
}

convert_to_not_accent <- function(value){
  stri_trans_general(str=value, id = "Latin-ASCII")
}
```

```

all_data <- rbind(data_2017, data_2018, data_2019, data_2020, data_2021)

all_data[, 'Fecha.de.Inscripción.Formateada'] <- sapply(all_data[, 'Fecha.de.Inscripción'], transform_data)

colnames(all_data) <- c('fechaInscripcion', 'genero', 'distrito', 'tramosDeEdad', 'nacionalidad',
                        'codigoObjetivoProfesional1', 'objetivoProfesional1', 'codigoObjetivoProfesional2',
                        'objetivoProfesional2', 'codigoObjetivoProfesional3', 'objetivoProfesional3',

all_data[, 'genero'] <- sapply(all_data[, 'genero'], convert_to_not_accent)
all_data[, 'distrito'] <- sapply(all_data[, 'distrito'], convert_to_not_accent)
all_data[, 'tramosDeEdad'] <- sapply(all_data[, 'tramosDeEdad'], convert_to_not_accent)
all_data[, 'nacionalidad'] <- sapply(all_data[, 'nacionalidad'], convert_to_not_accent)
all_data[, 'codigoObjetivoProfesional1'] <- sapply(all_data[, 'codigoObjetivoProfesional1'], convert_to_not_accent)
all_data[, 'objetivoProfesional1'] <- sapply(all_data[, 'objetivoProfesional1'], convert_to_not_accent)
all_data[, 'codigoObjetivoProfesional2'] <- sapply(all_data[, 'codigoObjetivoProfesional2'], convert_to_not_accent)
all_data[, 'objetivoProfesional2'] <- sapply(all_data[, 'objetivoProfesional2'], convert_to_not_accent)
all_data[, 'codigoObjetivoProfesional3'] <- sapply(all_data[, 'codigoObjetivoProfesional3'], convert_to_not_accent)
all_data[, 'objetivoProfesional3'] <- sapply(all_data[, 'objetivoProfesional3'], convert_to_not_accent)

write.csv(all_data, 'agencia_empleo_inscritos_agrupados.csv', row.names = FALSE)

```