**Plantilla actividad**

**Máster en Big Data & Data Science**

**Módulo 1. Herramientas de Big Data**

## Actividad número 2

### Título de la actividad

Indicar un título representativo para la actividad propuesta.

Ejercicios de programación en R: Control de flujo y gestión de datos.

## Tipo de actividad

|  |  |  |  |
| --- | --- | --- | --- |
| **Individual** | **X** | **Grupal** |  |

## Instrucciones para el alumnado

### Descripción de la tarea

**Important:** En aquest manual explicarem les macros programades que creiem són les més utilitzades.

Para este ejercicio, utiliza los datos “millas” que hay en el package “datos”.

Estos datos consisten en 238 filas y 11 columnas que describen el consumo de combustible de 38 modelos de coche populares.

Puedes consultar más sobre los datos en la ayuda “?millas”.

library(datos)  
suppressPackageStartupMessages(library(tidyverse))

?millas

Puedes descargar la [plantilla del ejercicio en Rmardown](https://github.com/griu/mbdds_fc20/blob/e0b645377f45789f663904b0173b3624d163b956/R/modulo1_tema4_R_Ejercicios.Rmd) desde Githib.

* Escribe un bucle “for” que calcule le media todas las columnas numéricas de “millas”.

# Solución:

* Haz lo mismo que en 2.1 pero utilizando “sapply()” en vez del bucle for.

# Solución:

* Explica la diferencia entre la función if() e ifelse(). Pon un ejemplo de uso de ambas.

# Solución:

* ¿Qué parámetros son necesarios para especificar la lectura de datos de ancho fijo mediante read.fwf()?

Explica qué significan y pon un ejemplo.

# Solución:

* Calcula la media de millas/galón en autopista para cada clase de coche de millas.

# Solución:

* Incorpora la media de calculada en 2.5. en el data frame “millas” como una nueva variable llamada “autopista\_clase”.

Utiliza la función “merge()” para juntar la table de 2.5 con millas.

# Solución:

* Utiliza las funciones del package dplyr: “group\_by()” y “mutate()” para realizar el mismo cálculo que en 2.5. y 2.6., sin necesidad de utilizar “merge()”. Llama a la nueva columna “autopista\_clase\_dplyr”

Haz un summary() para verificar que el resultado es el mismo que en 2.6.

# Solución:

* Analiza si “millas” tiene registros duplicado y, en caso afirmativo, crea un nuevo data frame que contenga una única copia de cada fila.

# Solución:

* Crea una función que tenga como input la fecha de tu nacimiento (en formato date) y devuelva tu edad en años.

# Solución:

* Explica por qué el resultado de “fechahora\_1” y “fechahora\_2” son distintos en la siguiente expresión:

library(lubridate)  
Sys.setlocale(locale="es\_ES.UTF-8")

## [1] ""

fechahora <- ymd\_hms("2020-03-28 15:11:23", tz = "Europe/Madrid")  
fechahora\_1 <- fechahora + dhours(24)  
fechahora\_2 <- fechahora + hours(24)  
  
print(fechahora\_1)

## [1] "2020-03-29 16:11:23 CEST"

print(fechahora\_2)

## [1] "2020-03-29 15:11:23 CEST"

### Pasos para la realización de la actividad

|  |  |  |
| --- | --- | --- |
| 1. | Presentación de la actividad. | Día XX/XX/2020. |
| 2. | Elaboración y entrega del informe/resultado. | Envío a través del espacio habilitado en el aula virtual (antes del XX/XX/2020). |
| 3. | Evaluación de la actividad por parte del tutor(a). | La nota se publicará en el aula virtual a partir del día XX/XX/2020. |

**Descripción del entregable**

* Tipo de documento: Informe. El alumno/a debe entregar un documento Rmarkdown, Rmd con el código R solicitado. Al ejecutar el Rmd se debe generar un documento (docx, pdf o html) que muestre el código R y las respuestas y los resultados solicitados.
* Formato: Rmd.
* Nombre de archivo: El archivo a entregar deberá nombrarse siguiendo la siguiente estructura (sin acentos): apellido1\_apellido 2\_moduloX\_bloque X\_apartadoX\_actX.
* Extensión: Libre.
* La entrega se deberá realizar en la correspondiente tarea del aula del curso.

**Materiales y/o recursos requeridos para la realización de la actividad (si procede)**

Se recomienda leer los siguientes apartados del tema “Programación en R”:

* 1.6. Control de flujo
  + 1.6.1. Funciones.
  + 1.6.2. Condicionales.
  + 1.6.3. Bucles.
* 1.7. Gestión de datos
  + 1.7.1 Importar y exportar datos.
  + 1.7.2 Cruce entre tablas.
  + 1.7.3 Resúmenes de agregados.
  + 1.7.4 Lubridate: gestión de fechas y horarios.

### Objetivo/s de la actividad

* Identificar la estructura de datos más adecuada a cada problema.
* Utilizar las funciones adecuadas de R para resolver el problema de la forma más simple posible.
* Presentar los resultados de forma limpia y organizada en el contenido y la forma.
* Redactar de forma clara y comprensible las dudas planteadas.

### Criterios que tendrá en cuenta el tutor para evaluar la actividad

* Precisión en los resultados obtenidos y explicaciones que se piden.
* Uso de funciones de R correctas y de la forma más simple posible.
* Presentación de los resultados de forma limpia y organizada en el contenido y en la forma.
* Redacción clara y comprensible las dudas planteadas.

|  |
| --- |
| Porcentaje de evaluación de la actividad respecto al módulo: 17,5 % |