**Plantilla actividad**

**Máster en Big Data & Data Science**

**Módulo 1. Herramientas de Big Data**

## Actividad núm. 2

### Título de la actividad

Indicar un título representativo para la actividad propuesta.

HERRAMIENTAS DE ANALISIS: PROGRAMACIÓN EN R – MICROACTIVIDAD 2

## Tipo de actividad

Marcar si la actividad será individual o grupal.

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

## Instrucciones para el alumnado

### Descripción de la tarea

Describir de forma clara el enunciado de la actividad.

*Redactad en segunda persona del singular (trato de tú) para actividades individuales y para grupos segunda persona del plural (trato de vosotros).*

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

Para el ejercicio 2, utiliza, también, los datos 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) des de Githib.

## EJERCICIO 2.1.

Escribe un bucle for que calcule la media todas las columnas numéricas de millas.

# Solución:

## EJERCICIO 2.2.

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

# Solución:

## EJERCICIO 2.3.

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

# Solución:

## EJERCICIO 2.4.

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

Explica qué significan y pon un ejemplo.

# Solución:

## EJERCICIO 2.5.

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

# Solución:

## EJERCICIO 2.6.

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:

## EJERCICIO 2.7.

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:

## EJERCICIO 2.8.

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

# Solución:

## EJERCICIO 2.9.

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

# Solución:

## EJERCICIO 2.10.

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

Detallar los pasos que se seguirá en la elaboración de la actividad que servirán de pauta para el tutor/docente durante su impartición y para el alumno como guía para su realización.

Esta información será parte del enunciado de la actividad y por lo tanto visible para el alumnado.

|  |  |  |
| --- | --- | --- |
| 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**

Se deberá detallar:

* El tipo de documento que se espera (informe, artículo, vídeo, etc.), si debe tener una estructura determinada (se detallarán los apartados, los bloques, etc.), así como otros aspectos o consideraciones a tener en cuenta.
* Formato: si el documento a entregar será un word, pdf, gif, avi, etc.
* Nombre de archivo: Se indicará cómo deberá nombrarse el archivo. Como criterio general se detallará: “El archivo a entregar deberá nombrarse siguiendo la siguiente estructura (sin acentos): apellido1\_apellido 2\_moduloX\_actX”.
* Extensión: longitud mínima y máxima (en páginas o caracteres).
* Tipo de documento: Informe. El alumno 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\_temaX\_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)**

* Listado de los contenidos del temario que hace falta leer, consultar o trabajar para la elaboración de la tarea.
* Si la actividad requiere de algún material específico (consultar un informe, una web, un software, también se debe indicar aquí).
* Si aplica, adjuntar los documentos que se entregarán a los alumnos para realizar la actividad. Por ejemplo, herramientas, texto para trabajar, documentos modelo, tablas, organigramas, gráficos numéricos, esquemas…

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

¿Qué se pretende que el alumno consiga al finalizar esta 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

* Los criterios de evaluación serán públicos en el aula y visibles para el alumno.
* Se deben explicitar todos los criterios a tener en cuenta.
* Deben ser acordes con los objetivos de la actividad, y han de tener en cuenta no sólo el punto de vista de un profesor que evalúa sino también el punto de vista de quien recibiría la actividad en el lugar de trabajo (un responsable al dar feedback a un colaborador, o un cliente, o un compañero, o un proveedor) siguiendo la lógica y la cultura del sector. Responden por tanto a la pregunta: ¿Qué será lo que mirará y tendrá en cuenta el tutor y el profesional para evaluar la actividad?
* Se deberá cumplimentar el porcentaje de evaluación, es decir qué peso tendrá la actividad respecto del módulo.
* Indicar si la actividad es recuperable marcando la selección SÍ | NO
* En el caso de marcar SÍ indicar cuál va a ser la manera de recuperar la actividad propuesta por el tutor.
* 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 de clara y comprensible las dudas planteadas.

|  |
| --- |
| Porcentaje de evaluación de la actividad respecto al módulo: xx % |