**Plantilla actividad**

**Máster en Big Data & Data Science**

**Módulo 1. Herramientas de Big Data**

## Actividad núm. 1

### Título de la actividad

Indicar un título representativo para la actividad propuesta.

HERRAMIENTAS DE ANALISIS: PROGRAMACIÓN EN R – MICROACTIVIDAD 1

## 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 1, 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) des de Githib.

## EJERCICIO 1.1.

A partir de los datos de millas, dibuja un gráfico de dispersión de puntos que muestre las millas recorridas en autopista por galón de combustible consumido (autopista) respecto a la cilindrada del motor de cada automóvil. No olvides añadir títulos al gráfico y a los ejes x e y.

# Solución:

## EJERCICIO 1.2.

A partir del gráfico del ejercicio 1.1., escoge una variable para cada uno de los siguientes parámetros estéticos: color, size y shape.

Truco: Observa que puedes seleccionar tanto variables continuas como categóricas. Si lo crees interesante, puedes repetir.

Comenta algún aspecto relevante que hayas descubierto sobre los coches a partir del gráfico.

# Solución:

## EJERCICIO 1.3.

Transforma el siguiente vector de tipo factor a tipo numeric de forma que el valor final mostrado sea exactamente el mismo en ambos vectores, pero con formato distinto. Para ello utiliza as.character() y as.numeric().

¿Que sucede si sólo utilizas as.numeric()?

vec <- factor(c("8","5","9","8","1","7"))  
print(vec) # valor mostrado

## [1] 8 5 9 8 1 7  
## Levels: 1 5 7 8 9

# Solución:

## EJERCICIO 1.4.

¿Es millas un objeto de la clase data.frame o matrix?

¿Y el siguiente objeto `obj`?

obj <- cbind(millas$cilindrada,millas$cilindros)

# solución

## EJERCICIO 1.5.

Crea una función que tome un vector como input y retorne una lista con los siguientes resultados:

1. El último valor del vector
2. Los elementos de las posiciones impares.
3. Todos los elementos excepto el primero.
4. Solo números impares (y no valores faltantes).

# solución

## EJERCICIO 1.6.

¿Por qué x[-which(x > 0)] no es lo mismo que x[x <= 0]?

Pon un ejemplo de vector x donde el resultado de ambas expresiones sea distinto.

# Solución:

## EJERCICIO 1.7.

Añade a millas una nueva columna llamada “fabr\_mod” que contenga la concatenación del nombre del fabricante, un guion “-” y el modelo del coche.

# Solución:

## EJERCICIO 1.8.

Selecciona todos los coches de millas que cumplan con todas las condiciones siguientes:

* La marca es distinta a “dodge”
* Tiene tracción en las cuatro puertas
* Han estado fabricados antes del 2008
* Las millas/galón en ciudad o en carretera no llega a 12 millas/galón.

Cuantos coches has encontrado?

# Solución:

## EJERCICIO 1.9.

Añade una nueva columna “vol\_por\_cil” a obj del ejercicio 1.4. que contenga el cociente de la cilindrada por el número de cilindros.

# Solución:

## EJERCICIO 1.10.

Modifica los valores de la columna “vol\_por\_cil” del objeto obj del ejercicio 1.9. asignando NA a los valores de la columna superiores a 0.7.

Presenta los datos con un summary del nuevo objeto obj.

# Solución:

### 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”:

* 1.3. Elementos básicos de R
* 1.4. Gráficos con GGPLOT2
* 1.5. Colecciones de objetos
  + 1.5.1. Definición de los objetos
  + 1.5.2. Filtros
  + 1.5.3. Modificación de valores

### Objetivo/s de la actividad

¿Qué se pretende que el alumno consiga al finalizar esta actividad?

* Desarrollar gráficos con todos los elementos que lo definen: titulo, título de los ejes, si aplica, leyenda, uso de elementos gráficos como la tipología, colores, tamaños y textos.
* 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.
* Programación de gráficos con todos los elementos solicitados, que sean visualmente limpios (con todos los elementos legibles), que utilicen toda la variedad de elementos gráficos disponibles.
* 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 % |