Caso práctico: Lenguaje R

Descripción de la actividad

Deberás resolver los siguientes apartados utilizando el lenguaje de programación R. El entregable será un solo script con el código de R correspondiente que resuelve todos los apartados.

Dado el dataset flights.csv, y utilizando solo funciones del paquete básico de R, se pide:

* Leerlo con R para convertirlo en un data.frame
* Seleccionar aquellas filas cuyo retraso es negativo, y guardarlas en otro data.frame llamado *puntuales.df*. ¿Cuántas son?
* Sobre el data.frame *puntuales.df*, calcular el número de ocurrencias de cada combinación (origen, dest). Es decir, el número de vuelos que salen de un origen y llegan a un destino, para todas las posibles combinaciones de aeropuerto de origen y de destino.
* Reemplazar en el data.frame *puntuales.df* la columna *month* por una nueva, de tipo factor, en la cual los meses del año estén expresados como texto y no como número. Sugerencia: crear un vector constante de 12 elementos con los 12 meses del año como cadenas de caracteres, y usar la columna month para indexar sobre él. No deben usarse bucles bajo ningún concepto.
* Extraer a un nuevo data.frame las columnas year, month y day del data.frame original (completo, no puntuales.df).
* Escribir una *función* en R capaz de recibir un data.frame de tres columnas year, month, day (como números enteros) y calcular, para cada fila, el **día de la semana correspondiente a esa fecha.** La función que se pide debe tener **exactamente** esta forma:

diaSemanaDesdeDF <- function(df) {

apply(df, MARGIN = 1, FUN = diaSemanaDesdeFila)

}

diaSemanaDesdeFila <- function(fila){

… # procesar la fila, que es un vector de 3 elementos

# Devolver un string “L”, “M”, “X”, “J”, “V”, “S” o “D”

}

* La función debe utilizar por dentro la función *apply* con MARGIN = 1 aplicada al data.frame de entrada de la función. La función interna con la que se llamará a *apply* debe ser otra función llamada *diaSemanaDesdeFila* que recibe como argumento un vector de tres elementos [año, mes, dia] y calcula el día de la semana de esa fecha.

Las reglas para calcular el día de la semana a partir de una fecha dada por el número de año, número de mes y número de día del mes puedes encontrarlas en el siguiente enlace (asumimos calendario gregoriano):

<https://artofmemory.com/blog/how-to-calculate-the-day-of-the-week-4203.html>

* Usar la función diaSemanaDesdeDF para crear una nueva columna, de tipo factor, y añadirla al data.frame original de vuelos, llamada “diaSemana”.
* Mostrar un diagrama de caja (boxplot) de la variable arr\_delay (retraso del vuelo al llegar, medido en minutos) agrupado por el día de la semana. El diagrama debería contener, por tanto, siete cajas con sus bigotes.
* Mostrar dos histogramas uno al lado de otro (usar mfrow con 1 fila y 2 columnas). Ambos deben mostrar la distribución de la variable *distance*, en un caso para los meses 6 a 9 (verano) y en otro, para los meses 1 a 5 (invierno). ¿Hay alguna diferencia?
* Ningún ejercicio requiere usar bucles *for* ni *while*. Su uso será penalizado.
* **Entregable**: fichero *apellidos\_nombre.R* con el script que resuelve los apartados anteriores. Se recomienda incluir también el nombre y apellidos del alumno dentro del script, como comentario en la primera línea. El enunciado de cada apartado debe estar copiado como comentario justo antes de su resolución. Se entregará también un documento (muy breve) que incluya solamente las imágenes de los dos últimos apartados y un comentario acerca de su interpretación: ¿qué conclusión se obtiene gracias a dichos gráficos?

**Objetivos**

Con esta actividad pondrás en práctica los conocimientos estudiados en las clases de teoría del lenguaje de programación R. Volveremos a analizar el dataset de vuelos para hacer operaciones simples de limpieza, resumen y exploración de datos.

**Extensión máxima de la documentación**: una página, fuente utilizada Georgia 11 e interlineado, 1,5.

**NOTA**: cuidado con los valores perdidos (NA) en algunas columnas. Eliminar esas filas si fuese necesario.