

EXAMEN

Instrucciones.

1. La Evaluación es Individual.
2. Las respuestas escritas realícelas en el Script con la enumeración correcta a la pregunta a la que pertenece.
3. No se aceptará como respondida aquellas preguntas que el profesor reciba en formato manuscrito, a mano, escaneado o fotografiado.
4. Escriba su nombre, Apellido y RUN al comienzo de su Script.R
5. Recuerde que el profesor debe ejecutar con éxito su sintaxis línea por línea. Si la pregunta X depende de la ejecución exitosa de la línea anterior y esta no se ejecuta correctamente, su respuesta X estará errónea.
6. Asuma que el profesor es una máquina o interfaz al correr su script. Sólo ejecutará "Ctrl+Enter" por cada línea de código. No corregirá código para que sea ejecutado exitosamente.

No asuma que el profesor posee los paquetes y librerías ya instalados(as) y ejecutados(as) previamente para la ejecución exitosa de su línea de código. Debe enunciarlos en el script para habilitar por primera vez la función que desea utilizar. Si su desarrollo es correcto, pero arroja error dado que usted no habilitó la librería correspondiente, el desarrollo no tendrá puntaje.

7. El Script debe ser enviado a través de un solo correo electrónico (no dos, tres, etc.). En el caso que el alumno envíe más de un correo electrónico dentro del horario consignado para el desarrollo y entrega de la evaluación, con elementos adjuntos repetidos en un correo anterior, el profesor seleccionará aquellos elementos del último correo electrónico recibido dentro del horario de entrega.

No se aceptarán enmendaciones de las entregas enviadas fuera de la hora indicada.

8. En dicho correo electrónico se debe adjuntar su Script con el desarrollo del Script solicitado.
9. Enviar sólo por U-Cursos hasta las 23.59 hrs. del día miércoles 08 de febrero No se recibirán evaluación ni elementos desde las 00.00 hrs. o 24 hrs., es decir, al segundo de comenzar el día jueves 09 de febrero.

Observación 1: Revise que su Script se haya adjuntado correctamente al enviar la evaluación. Hágalo 25 minutos antes de la hora de cierre.

Observación 2: Observe la asignación de puntaje por cada pregunta.

Observación 3: Revise la Rúbrica adjunta.

Desarrolle y Responda.

1. Desde su Script cargue o llame el set de datos correspondiente a su identificación que se consigna en la columna NOMBRE del archivo.xlsx. titulado como “Set_de_datos_a_Utilizar_”.

- a. Para llamar a su set de datos, ocupe la sintaxis de la columna “Llamada”.

Para mayor información del data set, consulte la Fuente: <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/>

Tenga en cuenta que los sets de datos de la planilla o archivo citado fueron distribuidos de manera aleatoria por el algoritmo enseñado de aleatoriedad en el módulo 3.

(0,104 pts.)

2. Titule su set de datos que se le asignó, como “Set_Original”. O dicho de otra forma genere un objeto nuevo de tipo dataframe a partir de su set de datos cargado, denominado “Set_Original”.

Su nuevo “Set_Original” debe cumplir con las siguientes 5 condiciones:

- a. Si su set de datos posee menos de 50 rows, anide o duplique su set de datos (hacia abajo) con la misma información, para así poseer 50 o más rows. **(0,104 pts.)**
 - b. Si el nombre de las variables de su set de datos es compuesto, por ejemplo, separados ambos nombres por un punto, coma, dos puntos u otro guarismo, renombre esas variables por el nombre que usted desee no compuesto y sin separaciones. **(0,104 pts.)**
 - c. Recuerde poseer valores numéricos. **(0,104 pts.)**
 - d. Si su set de Datos posee NA, reemplácelos por el segundo dígito de su run de izquierda a derecha (si este es 0 o letra, avance uno hacia la derecha y escoja dicho número). **(0,104 pts.)**
 - e. Si su set de datos posee observaciones que toman el valor 0, reemplácelos por el tercer dígito de su run de izquierda a derecha (si este es 0 o letra, avance uno hacia la derecha y escoja dicho número y reemplace). **(0,104 pts.)**
 - f. Genere un ID correlativo e indéxelo. Recuerde que esta variable es un ID. Un identificador. No entrega información del set de datos procesable. **(0,104 pts.)**
3. Responda sí o no, si en su “Set_Original” obtenidos en el punto f, se encuentran variables fuertemente correlacionadas.

Para responder a lo anterior, demuestre su respuesta en su Script con lo enseñado en clases. Para esto considere un criterio de correlación igual o mayor a 0.8. **(0,165 pts.)**

4. Responda en un máximo de 2 líneas, ¿qué correspondería hacer si tuviera variables correlacionadas fuertemente antes de generar los clústers? ¿por qué? **(0,165 pts.)**
5. Responda sí o no, si su set de datos posee outliers. Para esto, realice una demostración a través de la **aproximación visual** más completa enseñada en clases para responder si su set de datos posee outliers o no. **(0,165 pts.)**
6. Responda si para la demostración de la pregunta anterior usted debió escalar los datos o no. Justifique en no más de 4 líneas por qué hacerlo o no. **(0,165 pts.)**
7. Responda en un máximo de 3 líneas, ¿Qué debe hacer si su modelo de datos contiene outliers para resolver un problema real? **(0,104 pts.)**
8. Responda en un máximo de 3 líneas ¿cómo influye en la generación de clústers la presencia de outliers en mi modelo de datos? **(0,165 pts.)**

Para las siguientes preguntas puede considerar la presencia de outliers.

9. Genere un nuevo set de datos titulado “Nuevo_Set” a partir del “Set_Original” no escalado obtenido en el punto 2 el cual deberá agregar las siguientes condiciones:
 - a. Añada un vector nuevo denominado “nuevo1” obtenido de la suma de su segundo número de su run con el cuarto número de su run (si este es 0 o letra, avance uno hacia la derecha y escoja dicho número). **(0,104 pts.)**
 - b. Añada un vector nuevo denominado “nuevo2” obtenido de la multiplicación de 2 unidades al vector 2 de su “Nuevo_Set” obtenido en el punto anterior. **(0,104 pts.)**
 - c. Añada al set de datos obtenido en el punto anterior un vector nuevo denominado “nuevo3” obtenido de la multiplicación de 5 unidades al cuarto número de su run (si este es 0 o letra, avance uno hacia la derecha y escoja dicho número). **(0,104 pts.)**
10. Al set de datos obtenido en el punto anterior, elimine las variables correlacionadas bajo el criterio de correlación visto en clases. **(0,104 pts.)**
11. Responda si, para la generación de clústers, usted debe escalar los datos o no del set de datos obtenidos en la pregunta anterior. Justifique en no más de 4 líneas por qué hacerlo o no. **(0,116 pts.)**
12. Aplique la respuesta anterior dada por usted, a su “Nuevo_Set” obtenido en la pregunta 10. **(0,104 pts.)**

13. Al set de datos obtenido en el punto anterior y según las técnicas vistas en clases, proponga un número óptimo de K y con qué algoritmo decidirá generar Clústers. Para esto:
- Se requiere la justificación técnica de los parámetros APN, AD, ADM, FOM, Connectivity, Dunn, y Silhouette. Responda como influye en sus clústers la lectura obtenida por cada uno de estos parámetros. **(1 pt.)**
 - Ocupe una segunda técnica distinta a la anterior para seleccionar el óptimo. Y explíquelo a negocio, qué métrica aritmética está ocupando dicha técnica, que significa y como interpretarla. **(0,25 pts.)**
14. Genere los clústers con el mejor número de clústers (óptimo) de K y algoritmo identificado en la pregunta anterior. **(0.5 pt.)**
15. Traiga los Clústers generados a su set de datos con el que generó la clusterización y plotee en 3D los clústers generados. **(0,145 pt.)**
16. Traiga los Clústers generados a su "Set_Original" obtenido en la letra f de la pregunta 2. **(0.458 pt.)**
17. Del set de datos anterior, su equipo de analítica le pide que entregue las medidas de tendencia central de 3 variables escogidas por usted y las explique. **(0,5 pts.)**
- Déjese apoyar por el significado de las variables de su set de datos que se encuentran en <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/>
18. Su equipo de analítica le pide que prepare con sus palabras para negocio o gerencia, una explicación entendible, traducible y/o cualitativa de las lecturas de las variables del punto 17. Y por qué escogió usted estas para explicar los clústers generados.

Asuma que el ID, significa un país, animal, cosa, persona, u objeto. Pero asuma algo.

Déjese apoyar por el significado de las variables de su set de datos que se encuentran en <https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/>

(0.7 pt.)