

Curso: R Intermedio para CCSS y Gestión Pública

Docente: Christian Chiroque Ruiz

Asunto: Práctica Calificada N° 1

Fecha: 07 de marzo del 2023

INDICACIONES

- Fecha de entrega límite: lunes 13 de enero a las 23:59 horas. Subir a PAIDEIA
 - La PC se desarrollará de forma **INDIVIDUAL**.
 - Modo de entrega: SCRIPT con los códigos.
 - Recuerde que para comentar se debe utilizar el signo “#”.
 - Si presentan los códigos sin comentar (qué se está haciendo y para qué) no se le asignará puntaje.
 - Se considerará el orden del Script (uso de separadores o secciones /document outline) y la calidad de redacción en la calificación.
-

Pregunta 1 (10 PUNTOS)

- a) Busque y presente una base de datos que tenga las siguientes características **(3 puntos)**:
 - a. Fuente: Puede ser una base que Ud. haya utilizado en su centro de labores/estudios/investigación o una que pueda ser descargada de [Kaggle](#).
 - b. Número de casos: un mínimo de 100 casos y un máximo de 1000 (subset aleatorio, en caso la data encontrada sea más grande)
 - c. Variables: mínimo de 5 variables numéricas y un máximo de 10 variables numéricas. No se considera en el conteo aquellas variables de identificación de caso (Ej: el nombre de la región/país, el código del cliente/encuestado/IE/etc).
 - d. Tema: El tema de la data es libre.
 - e. Descripción: Realice una descripción muy directa sobre: de qué es la data, qué mide cada variable y cuál es la fuente utilizada.
- b) Aplique Análisis de Componentes Principales a la base de datos. **(2 puntos)**
- c) Presente una descripción del resultado, teniendo en cuenta que deseamos obtener los dos primeros componentes. **(4 puntos)**
- d) Muestre una visualización de los casos en dos dimensiones. **(1 puntos)**

Pregunta 2 (10 PUNTOS)

Con la misma data de la pregunta anterior, aplique lo siguiente:

- a) Explore la data. ¿Previamente puede identificar algún número de clústers potenciales? Confirme ello a través de la aplicación del algún algoritmo de recomendación de número de clúster. **(2 puntos)**
- b) Aplique al menos dos tipos de algoritmos de clasificación. Puede elegir entre jerárquico (AGNES, DIANA), partición (K-means) o basados en densidad (DBSCAN). Tome de referencia la recomendación obtenida en el punto anterior. **(2 puntos)**
- c) Compare los resultados de los dos métodos utilizados. ¿Cuál agrupó de mejor manera? **(2 puntos)**
- d) Presente una caracterización sencilla de cada grupo creado a partir de las variables originales. **(2 puntos)**

- e) Una vez seleccionado el mejor resultado, presente un gráfico de dispersión (puntos) utilizando ggplot2 y coloreando cada punto de acuerdo a los grupos formados. **(2 puntos)**