**Consigna**

**Modelo de expresión genética**

Contexto

Trabajas en un laboratorio biomédico que recopila datos de expresión génica de pacientes con diferentes condiciones de salud. El objetivo es identificar patrones subyacentes en la expresión génica y agrupar a los pacientes en clusters que compartan características genéticas similares. Esto podría ayudar en la identificación de subtipos de enfermedades o en la personalización de tratamientos.

Datos

El conjunto de datos incluye las siguientes columnas:

* Variables de Expresión Génica: Cada columna representa la expresión de un gen específico. Existen 10 genes para cada muestra.
* Edad: Edad del paciente.
* Género: Género del paciente.
* Condición de Salud: La condición de salud del paciente:
  + 0 significa salud normal
  + 1 al 9 especifica un tipo de enfermedad.

Objetivo

Utilizar técnicas de aprendizaje no supervisado, específicamente PCA y K-Means, para agrupar a los pacientes en clusters basados en sus perfiles de expresión génica, edad y género.

Documentar cada paso y realizar comentarios explicativos del código.

Pasos a realizar

* Exploración de Datos
* Preprocesamiento de Datos
* Análisis de Componentes Principales (PCA):
* Selección de Características:
* Clustering con K-Means:
* Evaluación del Modelo
* Interpretación y Visualización

Recomendaciones Biomédicas

Proporcionar recomendaciones basadas en los clusters identificados. ¿Se pueden inferir subtipos de enfermedades? ¿Cómo podría esto informar sobre enfoques de tratamiento personalizado?