**Tarea práctica 3**

Cargue el archivo de datos llamado *diabetes* en su entorno de trabajo. Este conjunto de datos contiene información de pacientes relacionada con factores de riesgo de diabetes y una etiqueta indicando si el paciente tiene o no diabetes.

Revise el conjunto de datos en busca de registros duplicados. En caso de encontrar duplicados, elimínelos del conjunto de datos, y documente la cantidad de duplicados encontrados y eliminados.

Identifique las columnas que contienen valores faltantes y analice su proporción en el conjunto de datos. Anote los porcentajes de datos faltantes en cada columna.

En caso de encontrar datos faltantes, utilice imputación avanzada para los valores ausentes.

Revise la distribución de las clases en la variable objetivo (si el paciente tiene o no diabetes) y determine si existe un desbalance de clases significativo.

Ejecute una estrategia para manejar el desbalance justifique su selección.