**Lab 3: Estandarización Terminológica con Glosarios**

Objetivo: Aplicar glosarios de referencia para estandarizar terminología inconsistente en campos de ecosistemas y métodos de muestreo.

Materiales: Archivo "ejercicio\_bloque2\_terminologia.xlsx" y "glosarios\_ambientales.xlsx"

Instrucciones Detalladas:

Abra el archivo de datos con problemas terminológicos y examine las columnas "Ecosistema" y "Metodo\_Muestreo" para identificar las variaciones presentes. Notará múltiples formas de referirse al mismo concepto, como "Manglar", "Manglares", "Bosque de mangle", y "MANGLAR" para el mismo tipo de ecosistema.

Importe las tablas de glosario desde "glosarios\_ambientales.xlsx", específicamente las hojas "Tipos\_Ecosistema" y "Metodos\_Muestreo". Examine la estructura de estas tablas, notando que incluyen códigos estándar, nombres estandarizados, y listas de variantes comunes separadas por comas.

Cree tablas auxiliares expandidas para las variantes comunes. Para la tabla de ecosistemas, duplique la tabla original y utilice "Dividir columna" en el campo "Variantes\_Comunes", especificando la coma como delimitador y seleccionando "Dividir en filas". Esto creará una tabla donde cada variante tiene su propia fila con el código estándar correspondiente.

Aplique las estandarizaciones mediante operaciones de merge. Para los ecosistemas, realice un merge entre su tabla de datos y la tabla de variantes expandida, utilizando la columna "Ecosistema" de los datos y "Variantes\_Comunes" de la tabla de referencia. Configure el merge como "Left Outer Join" para mantener todos los registros originales.

Después del merge, cree una columna condicional que utilice el código estandarizado cuando esté disponible y mantenga el valor original cuando no haya correspondencia. Esto le permitirá identificar términos que no están en el glosario y que requieren revisión manual o adición al glosario.

Implemente detección de errores ortográficos creando una tabla de frecuencias que muestre cuántas veces aparece cada término único. Términos que aparecen solo una vez pueden ser errores de escritura de términos más comunes. Utilice funciones de texto para normalizar la capitalización y detectar variaciones menores.

Repita el proceso para los métodos de muestreo, aplicando las mismas técnicas de estandarización y detección de errores. Documente todas las decisiones de estandarización en una tabla de log que registre el valor original, el valor estandarizado, y la justificación del cambio.

Criterios de Evaluación:

•Identificación correcta de todas las variaciones terminológicas (20%)

•Creación apropiada de tablas de correspondencia expandidas (25%)

•Aplicación exitosa de estandarizaciones mediante merges (30%)

•Detección e identificación de posibles errores ortográficos (15%)

•Documentación completa de las decisiones tomadas (10%)

Resultado Esperado: Datos con terminología completamente estandarizada, reducción significativa en el número de términos únicos, y documentación completa de todas las transformaciones aplicadas.