**Lab 4: Normalización de Nombres de Especies**

Objetivo: Homologar nombres de especies utilizando un catálogo taxonómico de referencia y resolver ambigüedades nomenclaturales.

Materiales: Archivo "ejercicio\_bloque3\_taxonomia.xlsx" y "catalogo\_taxonomico\_referencia.xlsx"

Instrucciones Detalladas:

Examine el archivo de datos taxonómicos problemáticos, notando la presencia de nombres científicos completos, abreviados, con autoridades, nombres vernáculos en español e inglés, y campos de familia incompletos. Su tarea es normalizar toda esta información utilizando el catálogo de referencia proporcionado.

Importe el catálogo taxonómico de referencia y examine su estructura. Note que incluye nombres científicos válidos, autoridades taxonómicas, jerarquía taxonómica completa, nombres comunes en múltiples idiomas, e identificadores únicos de GBIF.

Cree una tabla de especies únicas desde sus datos principales utilizando "Quitar duplicados" en la columna "Nombre\_Especie". Esta tabla contendrá aproximadamente 25 variaciones diferentes que necesitan ser homologadas con los 6 nombres válidos en el catálogo de referencia.

Limpie los nombres de especies eliminando espacios adicionales con Text.Trim, normalizando la capitalización con Text.Proper, y separando autoridades taxonómicas cuando estén presentes. Para nombres como "Lutjanus griseus (Linnaeus, 1758)", extraiga solo "Lutjanus griseus" para la comparación con el catálogo.

Realice un merge entre la tabla de especies únicas y el catálogo taxonómico utilizando el nombre científico como clave de unión. Configure como "Left Outer Join" para mantener todas las especies originales. Para especies con correspondencia exacta, la homologación será directa.

Para especies sin correspondencia exacta, implemente búsquedas a nivel de género. Cree una columna que extraiga solo el género de cada nombre científico y realice un segundo merge a este nivel. Esto permitirá identificar especies que pertenecen a géneros válidos pero cuyo epíteto específico no está en el catálogo.

Desarrolle una tabla de correspondencias para nombres vernáculos. Mapee nombres como "Pargo", "Pargo gris", y "Gray snapper" al nombre científico válido "Lutjanus griseus". Incluya información sobre el nivel de confianza de cada correspondencia, especialmente para nombres vernáculos que podrían referirse a múltiples especies.

Cree columnas de clasificación que categoricen cada especie según el resultado de la homologación: "Correspondencia\_Exacta" para nombres que coinciden perfectamente, "Correspondencia\_Genero" para especies donde solo el género está validado, "Nombre\_Vernaculo" para nombres comunes que fueron homologados, y "Sin\_Correspondencia" para nombres que no aparecen en el catálogo.

Enriquezca sus datos originales con la jerarquía taxonómica completa del catálogo, incluyendo Reino, Filo, Clase, Orden, Familia, Género, y Especie. Implemente validaciones cruzadas que verifiquen la consistencia de esta jerarquía.

Criterios de Evaluación:

•Limpieza apropiada de nombres científicos y separación de autoridades (20%)

•Homologación exitosa de nombres con correspondencia exacta (25%)

•Implementación de búsquedas a nivel de género para especies no encontradas (20%)

•Desarrollo de correspondencias para nombres vernáculos (20%)

•Enriquecimiento con jerarquía taxonómica completa y validaciones (15%)

Resultado Esperado: Todos los nombres de especies normalizados a nomenclatura científica válida, jerarquía taxonómica completa para cada registro, y documentación de todas las decisiones de homologación.