




Reto 1: Procesamiento de Cosechas

 **Dataset:** Cosechas_Cosechas_2023b.csv

 **Variable de análisis:** Toneladas Cosechadas

 **Objetivo:** Preprocesar la variable de producción para prepararla para modelos de análisis o predicción.

Pasos detallados:

1. Cargar los datos:

- o Leer el archivo CSV.
- o Ignorar la primera fila si contiene títulos repetidos o vacíos.
- o Verifica que las columnas sean: Empresa, Especie, Toneladas Cosechadas, Periodo Información.

2. Limpiar y convertir:

- o Asegúrate de que la columna Toneladas Cosechadas esté en formato numérico.
- o Reemplaza comas por puntos si es necesario.
- o Elimina filas con valores nulos o faltantes en columnas clave.

3. Codificación de variables categóricas:

- o Aplica One-Hot Encoding a las columnas:
 - Empresa
 - Especie
 - Periodo Información
- o Conserva las columnas codificadas y descarta las originales.

4. Escalamiento de la variable numérica:

- o Escala Toneladas Cosechadas utilizando:
 - StandardScaler: centra los datos en media 0, std=1.
 - MinMaxScaler: convierte los datos al rango [0, 1].
 - RobustScaler: usa mediana e IQR, útil si hay outliers.

5. Transformaciones matemáticas:

- o Aplica transformaciones sobre Toneladas Cosechadas:
 - np.log1p(): para reducir sesgo positivo.
 - np.sqrt(): para reducir varianza en datos grandes.
 - PowerTransformer(method='yeo-johnson'): mejora normalidad.

6. Normalización:

- o Usa Normalizer para escalar cada fila a una norma de 1.

7. Visualización KDE:

- o Genera gráficos de densidad (sns.kdeplot) para comparar:
 - StandardScaler, MinMaxScaler, RobustScaler.
- o Guarda la figura como .png.


8. Resumen estadístico:


- o Calcula y guarda la media y desviación estándar de cada técnica.


9. Interpretación automática en HTML:

- o Crea una tabla en HTML que explique:
 - Cómo cambió la distribución.
 - Qué transformación normalizó mejor.
 - Cuáles técnicas serían más útiles para un modelo de regresión.

◆ Reto 2: Agrupación de Producción por Empresa

 **Dataset:** Cosechas_Cosechas_2023b.csv

 **Variable de análisis:** Toneladas agregadas por Empresa


 **Objetivo:** Agrupar datos y analizar la producción total por empresa, aplicando transformaciones.


Pasos detallados:

1. **Cargar y preparar los datos:**
 - Leer el CSV.
 - Convertir Toneladas Cosechadas a tipo numérico.
2. **Agrupar los datos:**
 - Agrupa por Empresa.
 - Calcula la **suma total** de Toneladas Cosechadas por empresa.
3. **Codificación:**
 - Aplica OneHotEncoding a la columna Empresa.
4. **Escalamiento:**
 - Aplica:
 - StandardScaler
 - MinMaxScaler
 - RobustScaler
 - Escala el total de toneladas por empresa.
5. **Transformaciones:**
 - Aplica log1p, sqrt, y Box-Cox a la producción total.
6. **Normalización:**
 - Normaliza los valores agregados.
7. **Visualización KDE:**
 - Grafica la distribución de las producciones por empresa escaladas.
8. **Resumen estadístico e interpretación:**
 - Tabla con medias y std de cada transformación.
 - HTML explicativo sobre:
 - Qué empresa tiene producciones extremas.
 - Qué técnica destaca mejor las diferencias.

◆ Reto 3: Simulación de Análisis de Escapes

 **Dataset:** escapes.csv (simulado)

 **Variable de análisis:** Escapes


 **Objetivo:** Simular un análisis de centros con escapes masivos y evaluar qué técnicas detectan mejor estos casos.

Pasos detallados:

1. **Simular o cargar datos con columnas:**
 - Centro
 - Especie
 - Escapes
 - Año
2. **Verificar tipos:**
 - Asegúrate que Escapes sea numérica.
3. **Codificación:**


- o OneHotEncoder para:
 - Especie
 - Año
 - 4. **Escalamiento:**
 - o Escala Escapes con los tres métodos vistos.
 - 5. **Transformaciones:**
 - o Aplica log1p, sqrt, Box-Cox.
 - 6. **Normalización:**
 - o Aplica Normalizer a Escapes.
 - 7. **Visualización:**
 - o KDE para comparar la forma de las distribuciones escaladas.
 - 8. **Interpretación:**
 - o Genera HTML explicando:
 - Qué técnica reduce mejor los valores extremos.
 - Cómo se identifican centros con escapes masivos.
-

◆ Reto 4: Índice de Riesgo Acuícola

 **Dataset:** Consolidado (simulado)

 **Variable de análisis:**

$\text{Índice de Riesgo} = (\text{Escapes} + \text{Mortalidad}) / \text{Toneladas Cosechadas}$

 **Objetivo:** Crear una métrica compuesta y evaluar qué técnicas permiten distinguir centros con alto riesgo.

 **Pasos detallados:**

1. **Simular o preparar datos con:**
 - o Centro
 - o Escapes
 - o Mortalidad
 - o Toneladas Cosechadas
2. **Crear columna calculada:**
 - o $\text{Indice_Riesgo} = (\text{Escapes} + \text{Mortalidad}) / \text{Toneladas Cosechadas}$
3. **Verificar valores:**
 - o Reemplaza ceros o nulos para evitar divisiones por cero.
4. **Escalamiento del índice:**
 - o StandardScaler, MinMaxScaler, RobustScaler
5. **Transformaciones:**
 - o Aplica log1p, sqrt, Box-Cox sobre Indice_Riesgo
6. **Normalización:**
 - o Aplica Normalizer
7. **Visualización:**
 - o KDE para comparar cómo cambia el índice con cada técnica
8. **Resumen e interpretación:**
 - o Calcula media y std.
 - o Crea HTML explicando:
 - Qué técnica acentúa mejor los centros en riesgo.
 - Qué centro podría necesitar intervención.