

Análisis Exploratorio de Datos para la Biodiversidad Mexicana

Caso: Familia Leporidae (GBIF)

Santiago Flórez Suárez, Andrés Iván Rodríguez Hernández
Centro de Investigación en Matemáticas Aplicadas

Universidad Autónoma de Coahuila

18 de noviembre de 2025

Agenda

1 Introducción

2 Objetivo

3 Datos y Descripción

4 Implementación

5 Resultados

- México es un país que cuenta con una gran diversidad.
- Plataformas como **GBIF** permiten acceder a millones de observaciones biológicas.
- Los reportes tradicionales suelen ser:
 - Estáticos
 - Poco interactivos
 - Difíciles de replicar

Como lo es: <https://redbioma.org/proyectos/2024-04-python-ciencia-datos/diversidad-abejas-familia-halictidae-mexico.html>, que nos ofrece una gran información pero se vuelve complejo a la lectura y el uso.

Limitaciones de los reportes tradicionales:

- Figuras estáticas.
- Difícil exploración por especie, región o tiempo.
- Requiere conocimientos avanzados de programación.
- Poco accesibles para biólogos o tomadores de decisiones.

Limitaciones de los reportes tradicionales:

- Figuras estáticas.
- Difícil exploración por especie, región o tiempo.
- Requiere conocimientos avanzados de programación.
- Poco accesibles para biólogos o tomadores de decisiones.

Se observa la necesidad de: una herramienta interactiva accesible, dinámica y reproducible.

Objetivo General

Desarrollar una plataforma interactiva (Streamlit) para realizar un **Análisis Exploratorio de Datos (EDA)** sobre registros biológicos de la familia **Leporidae** en México.

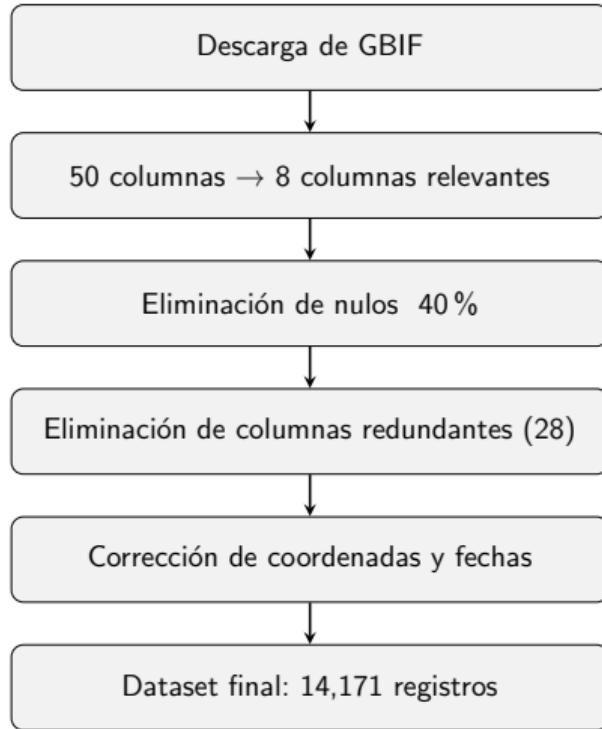
Estructura Taxonómica

GBIF provee jerarquía completa desde Reino → Género.

Reino	
Animalia.....	890,355
Filو	
Chordata.....	26,432,409
Clase	
Mammalia.....	890,355
Orden	
Lagomorpha.....	23,786
Familia	
Leporidae.....	23,672
Géneros	
<i>Sylvilagus</i>	14,803
<i>Lepus</i>	8,136
<i>Romerolagus</i>	451
<i>Oryctolagus</i>	61
<i>Hypolagus</i>	50
<i>Notolagus</i>	16
<i>Aztlanolagus</i>	10
<i>Paranotolagus</i>	9
<i>Pewelagus</i>	6
<i>Aluralagus</i>	4
<i>Pratilepus</i>	4
<i>Pronotolagus</i>	2
<i>Unknown genus</i>	148
<i>Unknown family</i>	86

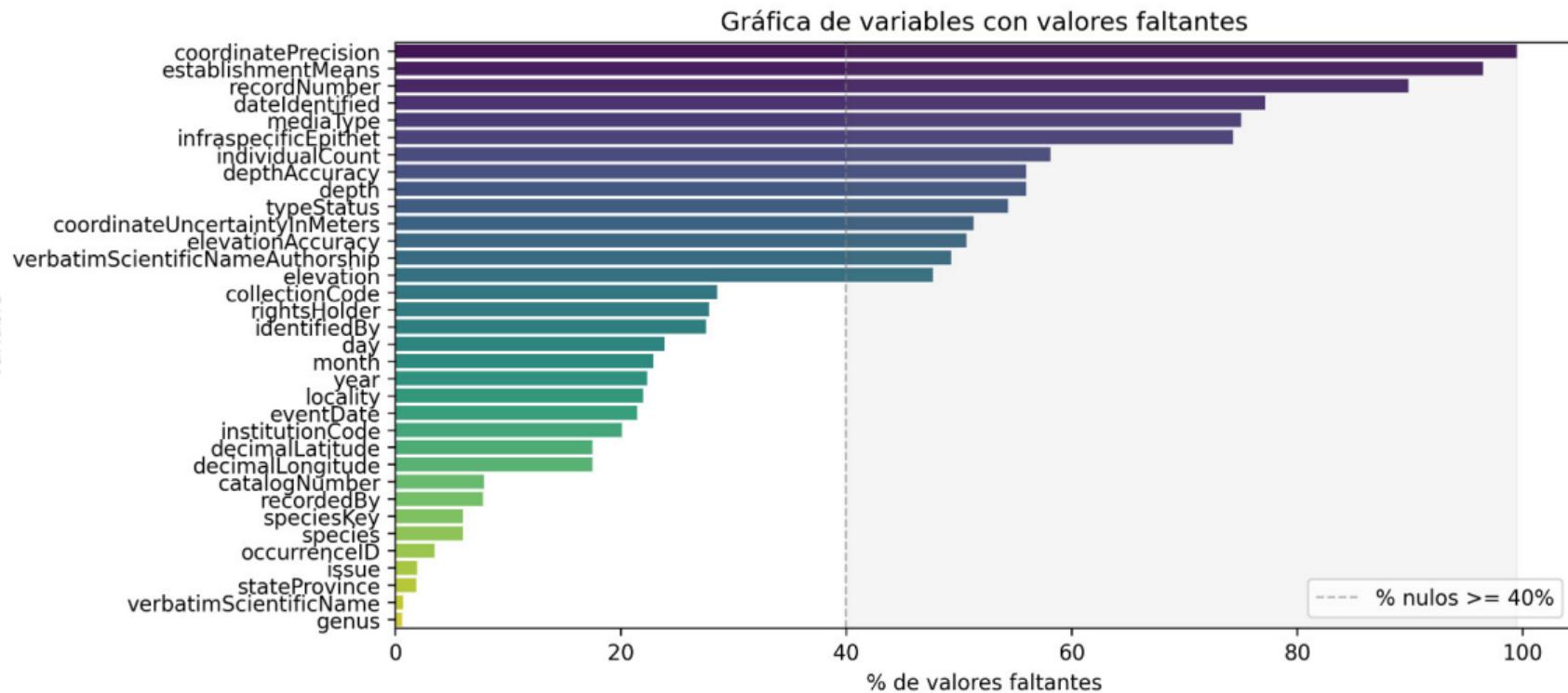
Figura: Estructura taxonómica (Leporidae).

Esquema de limpieza aplicado



Variables con altos porcentajes de nulos

Se identificaron 14 columnas con 40 % de valores faltantes.



Columnas eliminadas

50 columnas → 8 finales.

Eliminación motivada por:

- Redundancia
- Constancia sin valor analítico
- Metadatos administrativos

Columna	Justificación
gbifID	Identificador técnico
genus	Redundante
family	Constante
speciesKey	No útil

Variables finales utilizadas

Columna	Descripción
verbatimScientificName	Nombre científico
locality	Localidad
stateProvince	Estado
decimalLatitude / Longitude	Coordenadas
eventDate	Fecha
basisOfRecord	Tipo de evidencia
institutionCode	Institución

Carga y Filtros:

Carga optimizada con `@st.cache_data`

Filtros dinámicos por:

- Especie
- Año
- Estado

Visualizaciones

- Histogramas
- Series temporales
- Mapas geoespaciales (PyDeck)

Mapa Geoespacial

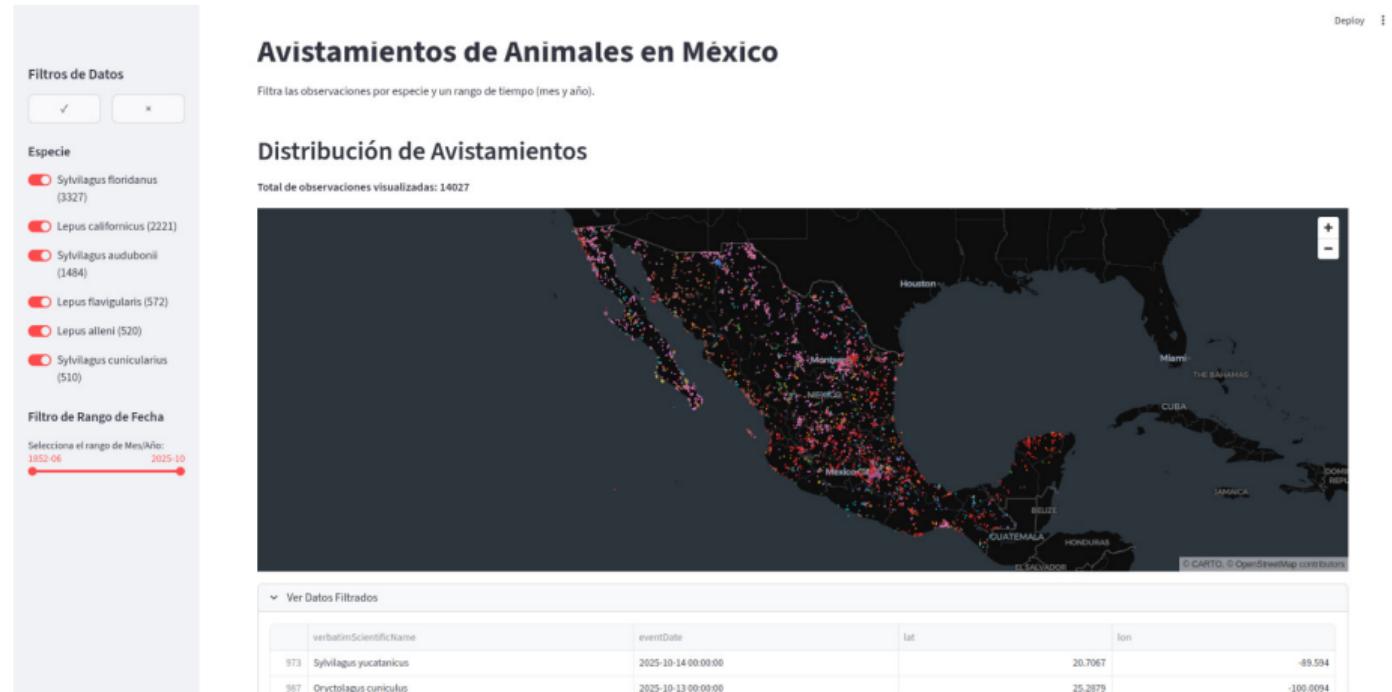
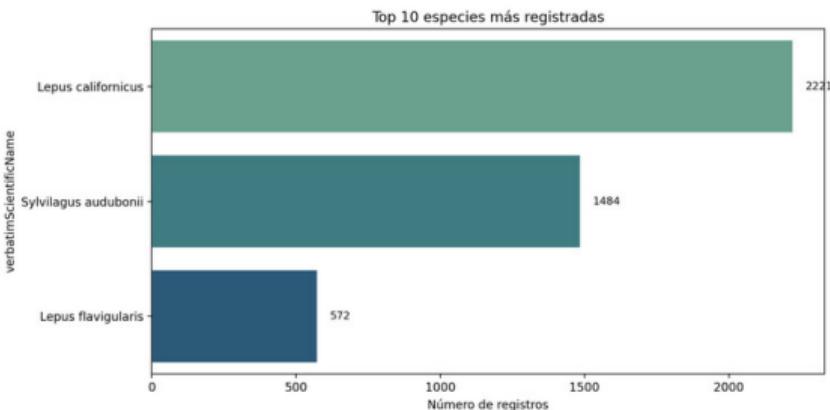


Figura: Mapa Interactivo.

Gráficas de barras

Análisis de Distribución: Top 10 Especies y Provincias

Top 10 especies más registradas



Top 10 provincias con más registros

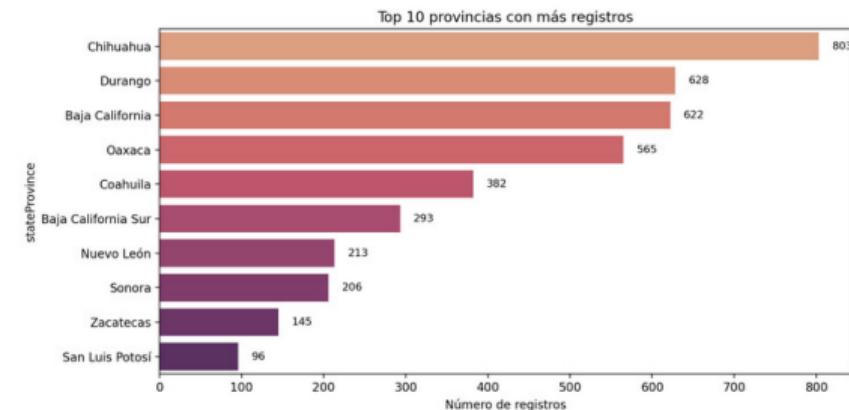


Figura: Distribución de registros por especie y estado.

Distribución temporal

Tendencia Temporal: Histograma de Observaciones por Año (Top 5 Especies)

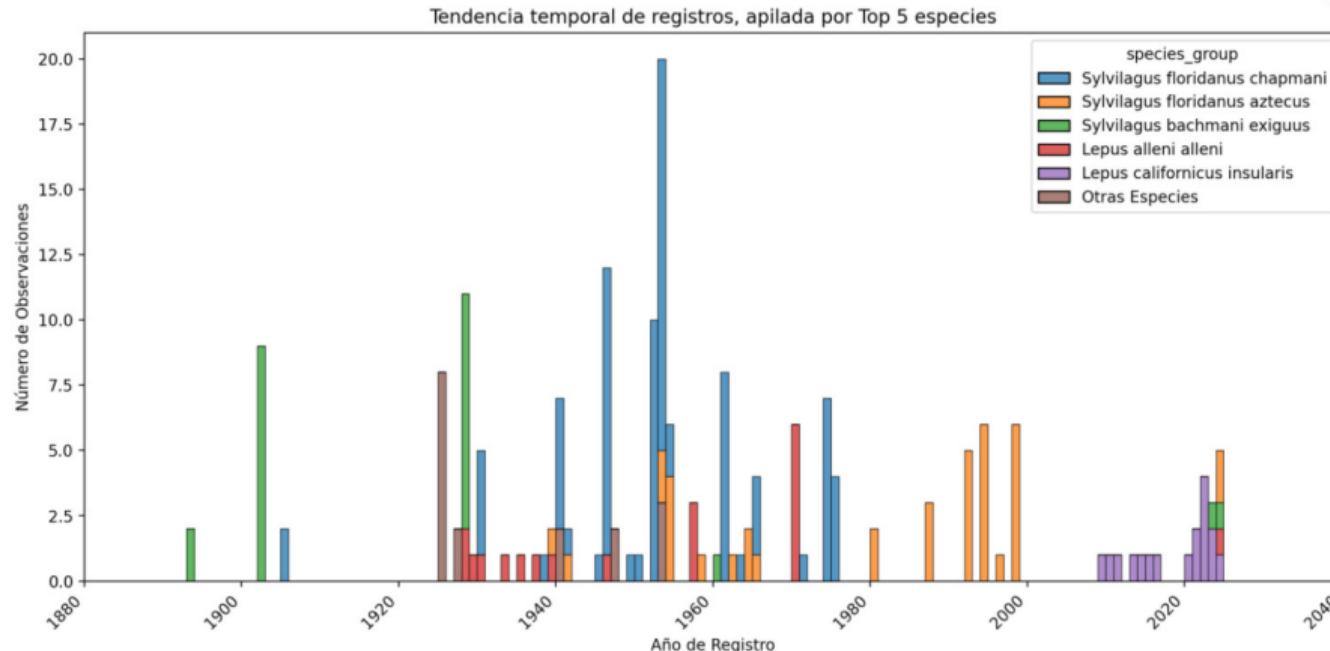


Figura: Cantidad de registros por año.

Conclusiones

- La limpieza permitió reducir ruido y mejorar la calidad analítica del dataset.
- Streamlit facilita un entorno accesible para especialistas sin experiencia en programación.
- PyDeck habilita una comprensión geoespacial profunda de los datos biológicos.
- La herramienta permite análisis reproducibles, dinámicos y actualizables.

¡Gracias!