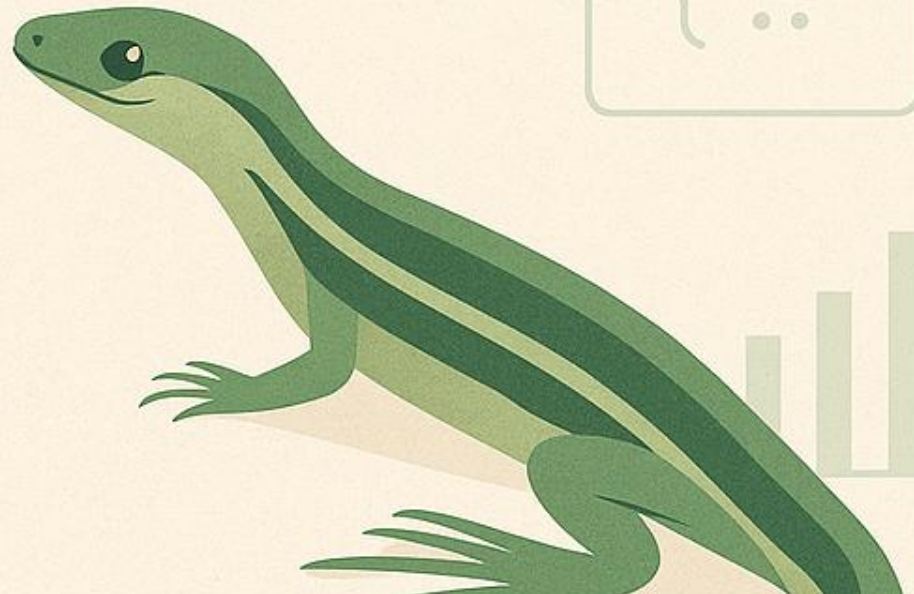


Análisis de Reptiles y Anfibios de México con Ciencia de Datos

Erik Manuel Barajas Torres
Diplomado en Ciencia de Datos – UNAM DGTIC
2025





Contexto e importancia

- México es uno de los países más biodiversos del mundo.
- Reptiles y anfibios son clave en redes tróficas y equilibrio ecológico.
- Amenazados por pérdida de hábitat, cambio climático y comercio ilegal.



Objetivo del estudio

- Analizar datos de CONABIO sobre herpetofauna en México.
- Detectar especies y zonas con mayor riesgo ecológico.
- Proyectar tendencias poblacionales y apoyar estrategias de conservación.

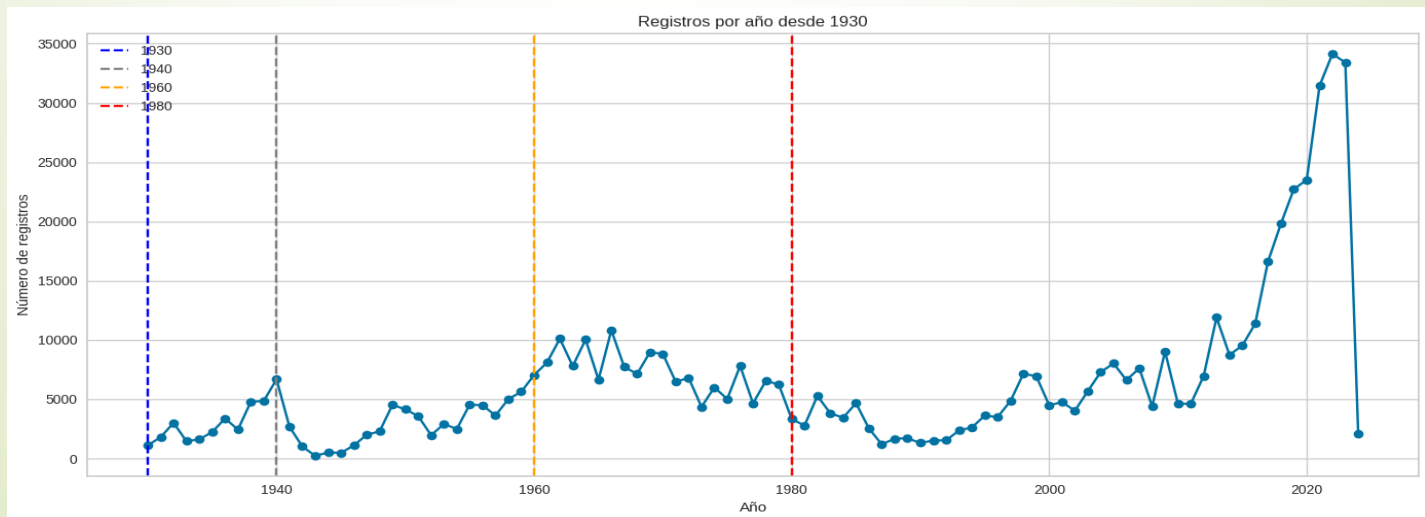


Flujo de trabajo del proyecto

- 1. Carga y limpieza de datos
- 2. Análisis exploratorio
- 3. Codificación y PCA
- 4. Clustering (KMeans, DBSCAN)
- 5. Series de tiempo (Prophet)
- 6. Modelos supervisados (clasificación y regresión)
- 7. Validación de hipótesis y análisis geográfico

Análisis exploratorio

- Filtrado desde 1950 para mayor consistencia.
- Se identificaron 30 especies con disminución y 10 en aumento.
- Visualizaciones mostraron declives por especie y por año.





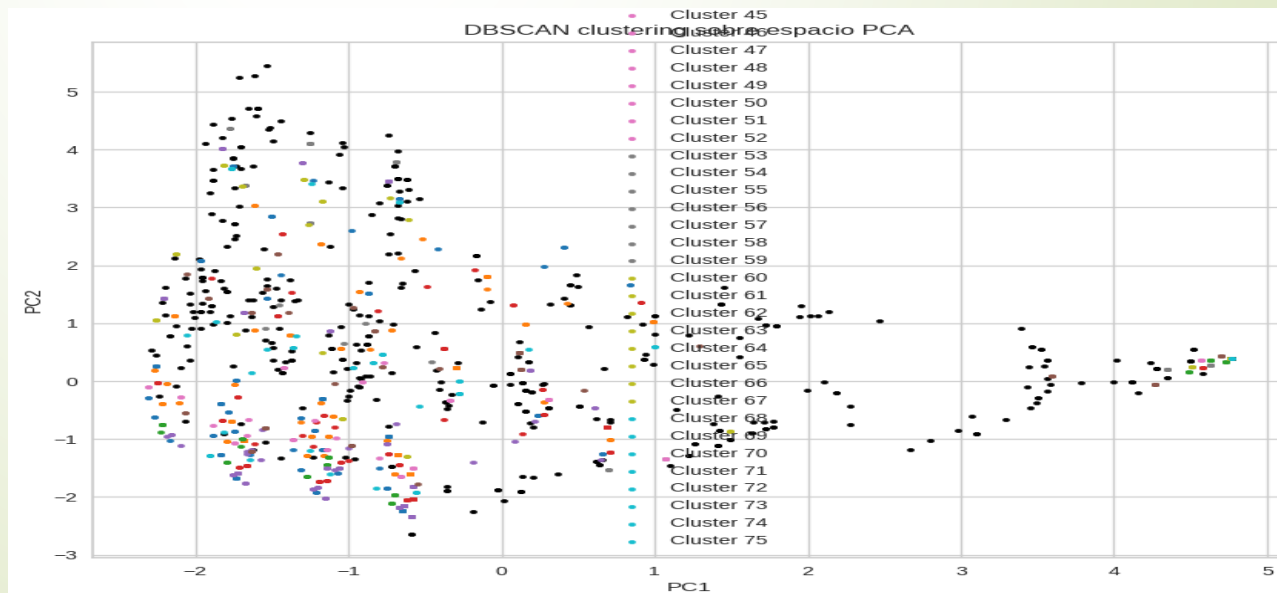
Clustering y reducción de dimensionalidad

- PCA explicó más del 70% de la varianza con 85 componentes.
- KMeans mostró mejor agrupamiento con $k=2$ (Silhouette Score: 0.42).
- DBSCAN identificó múltiples agrupaciones con ruido moderado.

Agrupación con K-means



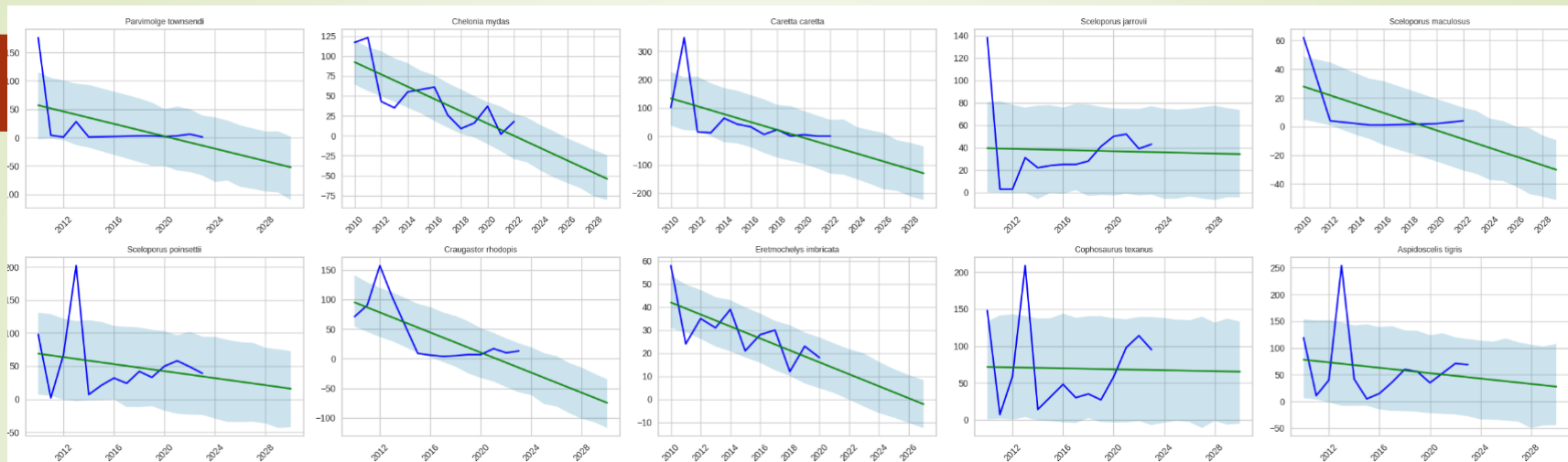
Agrupación con DBSCAN

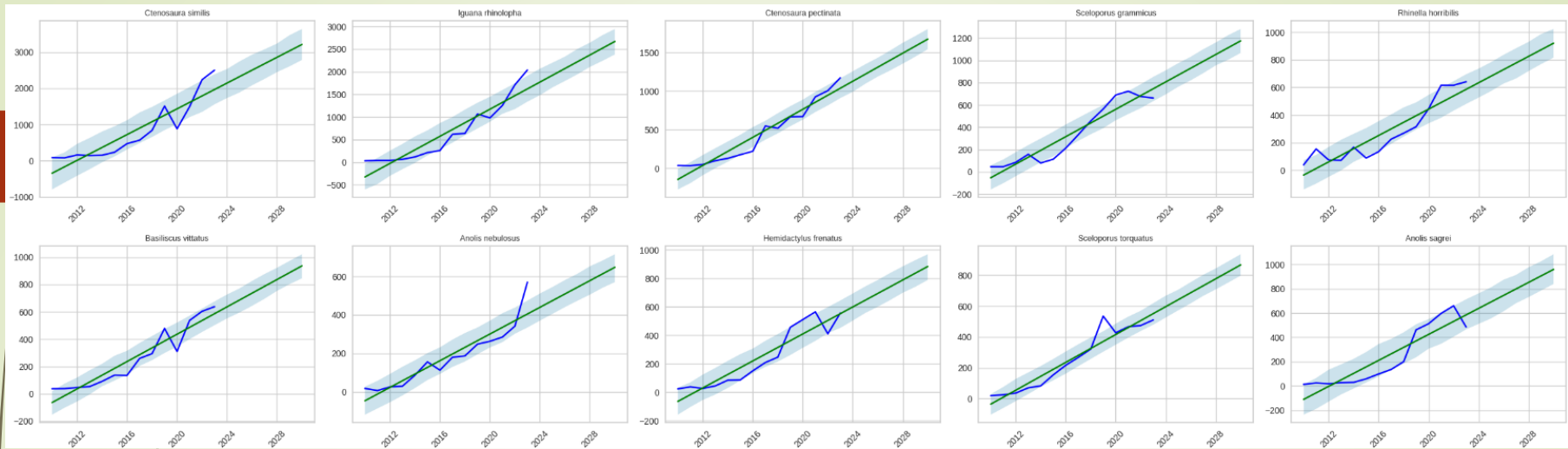




Series de tiempo con Prophet

- Se proyectaron registros 2025–2030 para 40 especies clave, 30 de ellas que está su población en declive y 10 de ellas que por el contrario, su población está en aumento.
- Se observaron tendencias descendentes persistentes.
- Permite anticipar posibles crisis de conservación.





Modelos supervisados

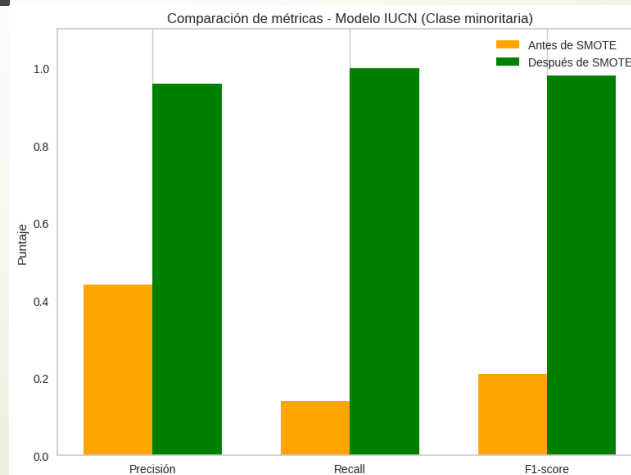
- Clasificación con Random Forest + SMOTE.
- F1-score mejoró notablemente en especies en riesgo tras balanceo.
- Regresión tuvo desempeño moderado ($R^2 < 0.4$, RMSE alto).

Antes de usar Smote

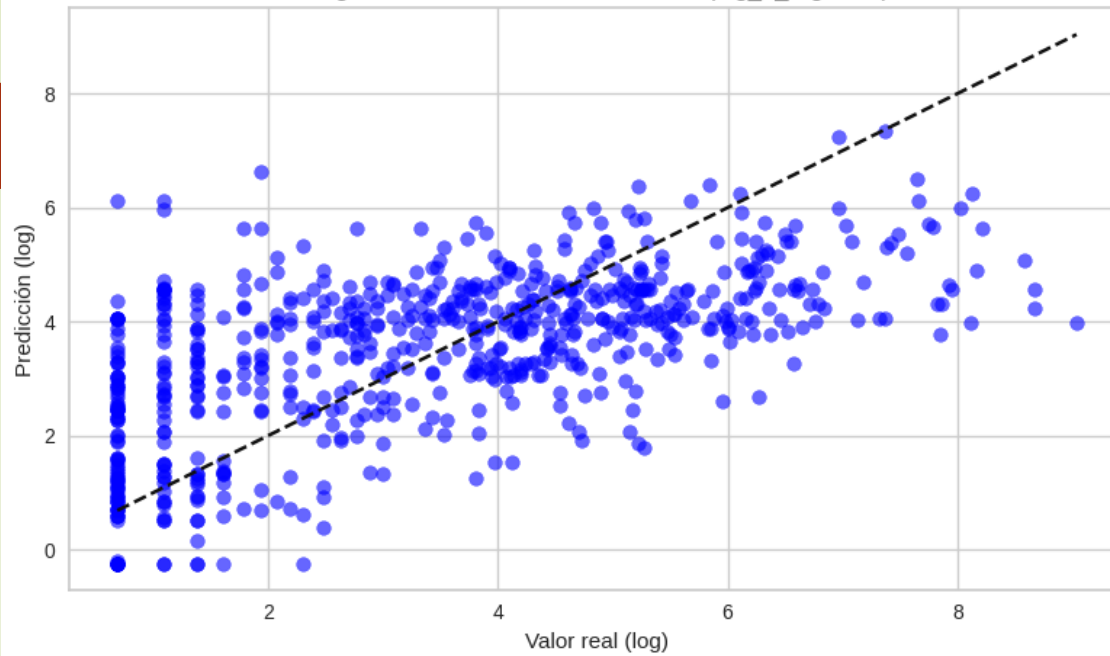
	Modelo	RMSE	MAE	R ²
0	Regresión Lineal	653.087972	311.224403	0.038570
1	Random Forest	666.538814	254.136669	-0.001441
2	XGBoost	662.106888	258.724762	0.011832

Después de usar smote

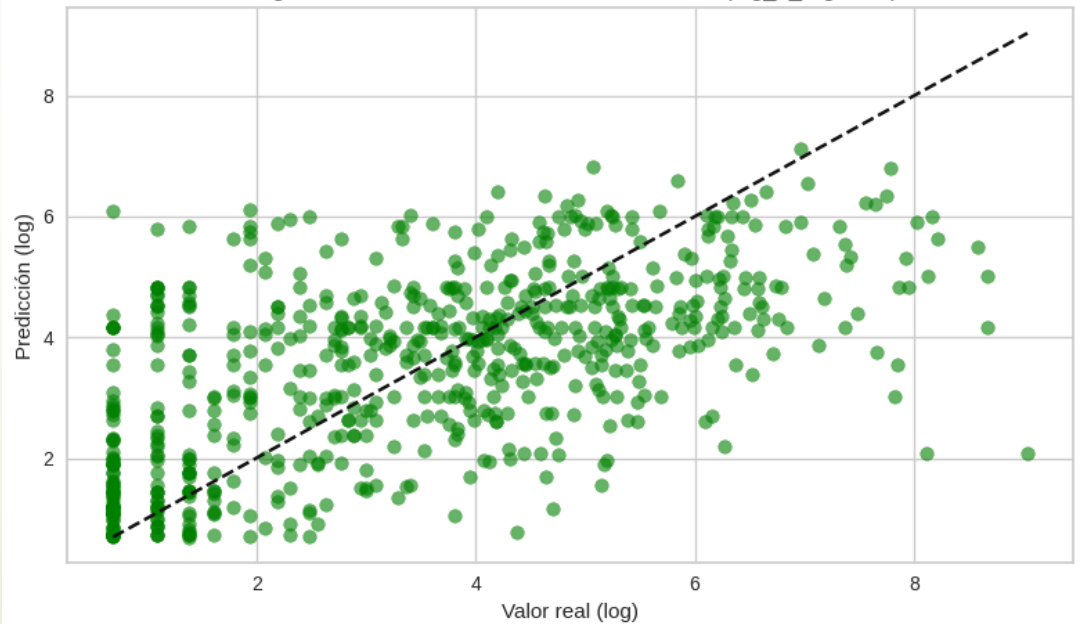
	Modelo	RMSE	MAE	R ²
0	Regresión Lineal (log)	1.621157	1.297752	0.341218
1	Random Forest (log)	1.649753	1.266191	0.317772



Regresión Lineal: Predicción vs Real (log_n_registros)



Regresión Random Forest: Predicción vs Real (log_n_registros)



Análisis geográfico e hipótesis

- Estados más críticos: Chiapas, Oaxaca, Veracruz y Guerrero.

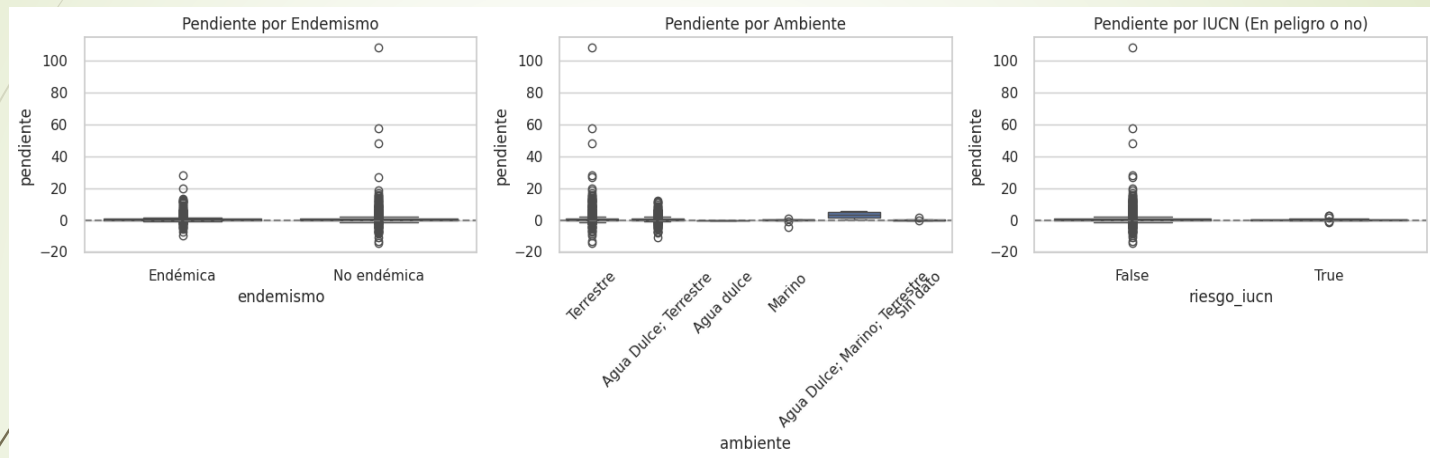
Estados con mayor diversidad

especievalida	estadomapa
OAXACA	626
CHIAPAS	506
VERACRUZ DE IGNACIO DE LA LLAVE	480
GUERRERO	383
PUEBLA	349
JALISCO	347
SONORA	328
MICHOACAN DE OCAMPO	312
CHIHUAHUA	309
SAN LUIS POTOSI	284

Estados más críticos

n_total	n_nom059	n_iucn	n_proyectadas	indice_prioridad	estadomapa
OAXACA	626	9.0	78.0	23.0	110.0
VERACRUZ DE IGNACIO DE LA LLAVE	480	17.0	52.0	23.0	92.0
CHIAPAS	506	8.0	38.0	15.0	61.0
PUEBLA	349	4.0	29.0	17.0	50.0
GUERRERO	383	4.0	27.0	16.0	47.0
MICHOACAN DE OCAMPO	312	5.0	18.0	16.0	39.0
SAN LUIS POTOSI	284	3.0	11.0	23.0	37.0
COAHUILA DE ZARAGOZA	260	5.0	11.0	20.0	36.0
JALISCO	347	7.0	12.0	16.0	35.0
HIDALGO	260	3.0	16.0	15.0	34.0

H2 validada: especies acuáticas tienen pendiente más negativa.



H1 - Endémica vs No endémica: $t = -1.3762, p = 0.1689$
H2 - Agua dulce vs otros: $t = -2.3045, p = 0.0213$
H3a - NOM-059 peligro vs otras: $t = -1.8419, p = 0.0656$
H3b - IUCN peligro/crítico vs otras: $t = -1.0906, p = 0.2756$



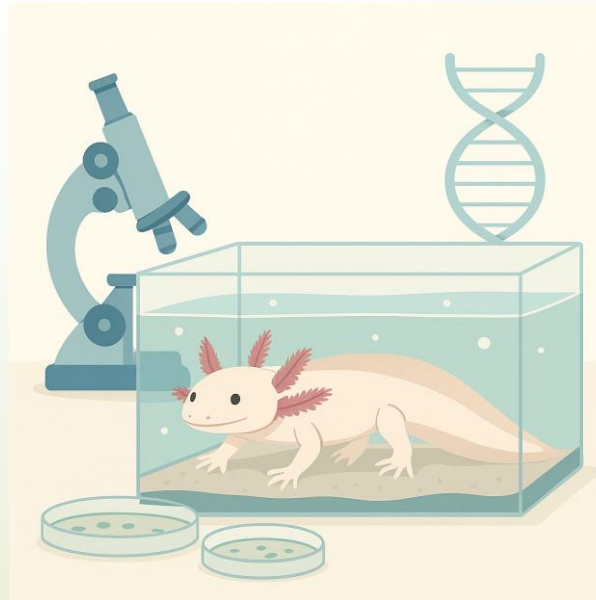
Se cruzaron especies en CITES con aquellas en declive poblacional.

	especievalida	cites	nom059	iucn
204286	Caretta caretta	I En peligro de extinción	Preocupación Menor	
20490	Chelonia mydas	I En peligro de extinción		Vulnerable
265900	Chelonia mydas	I En peligro de extinción		En peligro
209215	Eretmochelys imbricata	I En peligro de extinción		En peligro crítico
20699	Lepidochelys olivacea	I En peligro de extinción		Vulnerable



Ambystoma Mexicanum proyecto vs realidad (Ajolote)

- Aunque el *Ambystoma mexicanum* (ajolote) no mostró una tendencia de declive en nuestro análisis, esto puede deberse a la limitada cantidad de registros georreferenciados o a su reproducción en cautiverio. Sin embargo, múltiples fuentes coinciden en que esta especie enfrenta serios problemas de conservación en su hábitat natural, por lo que sigue siendo una prioridad ecológica."



Conclusiones

- Las especies endémicas y de agua dulce presentan mayor riesgo.
- El modelo Prophet y SMOTE aportaron proyecciones y balance.
- Este proyecto demuestra cómo la ciencia de datos puede impulsar la conservación.
- Se detectaron especies prioritarias para la conservación.

