

# ÍNDICE

ÍNDICE	1
CONTEXTO	2
TÍTULO	2
DESCRIPCIÓN DEL DATASET	3
REPRESENTACIÓN GRÁFICA	3
CONTENIDO	4
AGRADECIMIENTOS	5
INSPIRACIÓN	5
LICENCIA	6
CÓDIGO	6
DATASET	6

# 1. CONTEXTO

La situación de la biodiversidad a nivel mundial ha sido reconocida como crítica en numerosos foros y publicaciones, y Argentina no es una excepción. La pérdida del área de distribución histórica debida a actividades humanas representa un paso previo a la pérdida de especies. Frente a esta situación, conocer el estado de conservación de las especies de fauna silvestre de un país es fundamental para diagramar planes de gestión y manejo que aseguren la persistencia de sus poblaciones y la coexistencia armónica con los humanos. Además, contar con los mapas de distribución actual de los mamíferos, facilita el trabajo de gestores, investigadores, etc., y alienta a la elaboración de modelos y análisis para entender patrones de distribución de la biodiversidad y su relación con las actividades humanas, entre otros factores.

En el año 2018 la Sociedad Argentina de Mastozoología, junto con la Secretaría Nacional de Ambiente y Desarrollo Sustentable pusieron en marcha el proceso de categorización de los mamíferos de Argentina que culminó con la publicación, en el año 2019, de un [portal web](#) que contenía la información recopilada durante el proceso. Si bien la idea original del portal era contener toda la información abierta para quien quisiera consultarla, los mapas de distribución de las especies no están disponibles para su descarga.

Este trabajo tiene como objetivo generar un pipeline que permita la descarga de la información geoespacial de la distribución de las 395 especies de mamíferos presentes en Argentina en un formato acorde que permita su procesamiento posterior. Pero además, nos propusimos recolectar información georeferenciada de las áreas protegidas presentes en Argentina que nos permita evaluar el nivel de protección efectiva de dichas áreas sobre las poblaciones de mamíferos. Para esto, recurrimos al portal [Protected Planet](#) que es la base de datos más actualizada respecto de las áreas protegidas a nivel global. Esta información, si bien está disponible para su descarga, se actualiza constantemente por lo que decidimos también aplicar técnicas de web scrapping que permita tener la información actualizada. Por último, nos pareció interesante complementar el dataset con datos a nivel provincial que sean indicadores socioeconómicos y que puedan utilizarse en modelos subsecuentes tendientes a evaluar el nivel de protección de las especies en cada provincia y que resulten en herramientas útiles para la gestión a nivel nacional.

## 2. TÍTULO

*Factores socioeconómicos que influyen en el nivel de protección a nivel provincia de los mamíferos de Argentina.*

## 3. DESCRIPCIÓN DEL DATASET

Este trabajo consistió en la recopilación de datos georeferenciados acerca de la distribución de las especies de mamíferos presentes en Argentina, combinando esta información a nivel provincial con una serie de indicadores socio-económicos y de protección ambiental. Como resultado del proyecto, se han generado tres conjuntos de datos diferenciados:

1. ***provincias\_argentina.csv***: Información demográfica y socioeconómica de las provincias de Argentina.
2. ***especies.json***: Información georeferenciada de la distribución de los mamíferos de Argentina, junto con el nombre científico y su categoría de conservación.
3. ***areas\_protegidas.zip***: Información georeferenciada de la extensión de las áreas protegidas de Argentina.

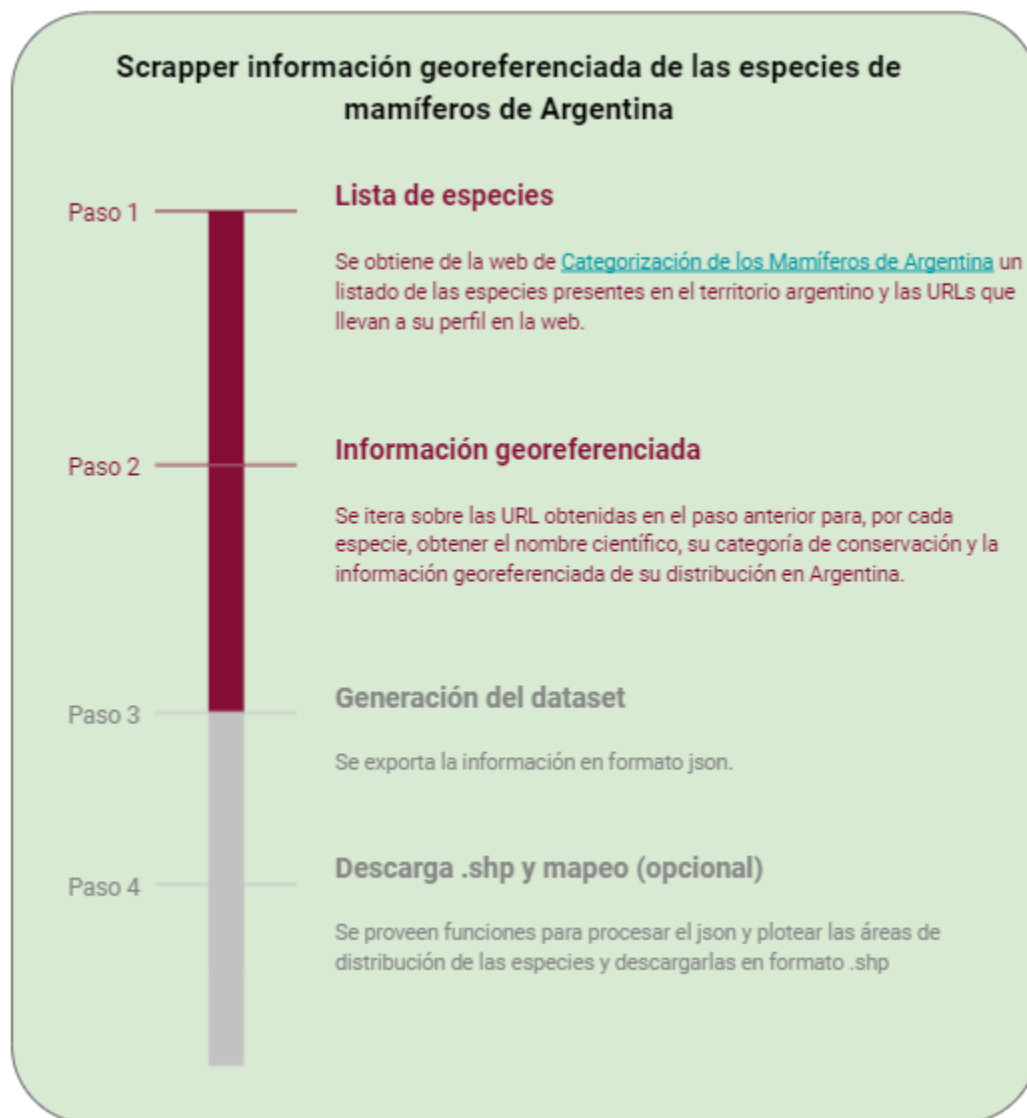
## 4. REPRESENTACIÓN GRÁFICA

Los siguientes diagramas muestran los pasos seguidos para los tres procesos de *web scraping* que se llevaron a cabo para obtener cada uno de los *datasets*.

- Para la generación del conjunto de datos con la información demográfica y socioeconómica de las provincias de Argentina (*provincias\_argentina.csv*) se han seguido los siguientes pasos:



- Para la generación del conjunto de datos con la información georeferenciada de las especies de mamíferos de Argentina (*especies.json*) se han seguido los siguientes pasos:



- Para la obtención de los datos con la información georeferenciada de las áreas protegidas de Argentina (*areas\_protegidas.zip*) se han seguido los siguientes pasos:



## 5. CONTENIDO

### especies.json

El conjunto de datos con la información georeferenciada de la distribución de las especies de mamíferos en Argentina está en formato json, el cual fue elegido ya que contiene la siguiente información:

- **nombre de la especie:**
- **categoría:**
- **información espacial:**

### provincias\_argentinas.csv

El conjunto de datos con la información demográfica y socioeconómica de las provincias de Argentina contiene la siguiente información:

- **provincia:** nombre de la provincia de Argentina
- **url:** enlace a wikipedia con la información de detalle para la provincia
- **poblacion:** número de habitantes

- **superficie:** extensión territorial de la provincia
- **densidad:** densidad de población
- **capital:** capital de la provincia
- **IDH:** Índice de Desarrollo Humano (ver detalle [aquí](#)) (año 2018)
- **analfabetismo:** tasa de analfabetismo (año 2010)
- **porc\_poblacion\_argentina:** porcentaje de población sobre el total de Argentina
- **autonomía:** fecha de declaración de la autonomía
- **altitud media:** altitud media sobre el nivel del mar
- **altitud maxima:** altitud máxima sobre el nivel del mar
- **altitud minima:** altitud mínima sobre el nivel del mar
- **coordenadas:** coordenadas geográficas (latitud y longitud)

Para la obtención de la información, se han seguido los pasos descritos en el punto anterior, obteniendo (mediante *web scraping* con la librería *Selenium*) la lista de provincias de la siguiente página de Wikipedia, la cual no está recogida en el fichero *robots.txt* como página cuya información no esté permitida obtener:

[https://es.wikipedia.org/wiki/Provincias\\_de\\_Argentina](https://es.wikipedia.org/wiki/Provincias_de_Argentina)

Posteriormente, se ha obtenido información detallada (también mediante la técnica de *web scraping*) de cada provincia de Argentina disponible en las respectivas páginas de Wikipedia para cada provincia. Cada atributo data de una fecha determinada, siendo por ejemplo el IDH del 2018 y la tasa de analfabetismo del 2010. Existe otra información que, por su naturaleza, resulta invariante en el tiempo (p.e. la declaración de Autonomía, altitudes o capital).

## 6. AGRADECIMIENTOS

La pregunta 6 hace referencia a un "mini estado del arte" indicando algunos estudios o análisis que se hayan realizado en base a los datos contenidos en el dataset/sitio web elegido, con el objetivo de justificar la utilidad de los datos extraídos. Dependiendo del sitio web elegido puede ser difícil encontrar dichos estudios o análisis, por eso, en el caso de no encontrarlos, se propone que incluyáis algunos estudios o análisis con datos similares.

## 7. INSPIRACIÓN

Explicar por qué es interesante este conjunto de datos y qué preguntas se pretenden responder. Es necesario comparar con los análisis anteriores presentados en el apartado 6.

Con la información proporcionada en los diferentes conjuntos de datos se puede dar respuesta a preguntas como las siguientes:

- ¿Existe alguna relación entre las especies en peligro de extinción y la tasa de analfabetismo de las provincias donde habitan?
- ¿Están las especies mejor protegidas en aquellas provincias con un IDH más alto?
- ¿Existe correlación entre las especies con menor extensión de hábitat (km<sup>2</sup>) y las provincias más pobladas?

Además por ejemplo, mediante la historificación de los datos de la distribución de las especies, también se podría dar respuesta a preguntas de la siguiente índole:

- ¿El hábitat de las especies de mamíferos en Argentina crece o por el contrario decrece con el paso de los años?
- ¿Cuál es la tasa de crecimiento interanual del hábitat de los mamíferos en Argentina?
- ¿Sobreviven mejor las especies que habitan en áreas protegidas?

## 8. LICENCIA

Los conjuntos de datos generados se encuentran bajo licencia [CC0: Public Domain License](#).

De esta forma, la información es completamente abierta al dominio público, permitiendo la copia, modificación y distribución de la obra sin pedir permiso a los autores. De esta manera se pretende fomentar el análisis de la información, a favor de la protección de las especies en Argentina. De igual forma, los autores no ofrecen ninguna garantía respecto a la obra y renuncian a cualquier responsabilidad por cualquier uso de la obra.

## 9. CÓDIGO

El código del proyecto se puede encontrar en el repositorio siguiente:

<https://github.com/ntaxus/WEB-SCRAPING>



Para la extracción de la información web (*scraping*) se ha utilizado la librería **selenium** con Python. Además, se han usado técnicas para la prevención del bloqueo de peticiones html (bloqueo por peticiones masivas desde una misma dirección IP) mediante el espaciado temporal de dichas peticiones.

## 10. DATASET

[Añadir enlace a Zenodo](#)