

Autor: Paul Zambrano

1. Título del dataset. Poned un título que sea descriptivo.

Animales en peligro crítico de extinción.

2. Subtítulo del dataset. Agregad una descripción ágil de vuestro conjunto de datos por vuestro subtítulo.

Listado de todos los animales clasificados con riesgo crítico de extinción de acuerdo a la organización IUCN (International Union for Conservation of Nature)

3. Imagen. Agregad una imagen que identifique vuestro dataset visualmente



4. Contexto. ¿Cuál es la materia del conjunto de datos?

La organización IUCN mantiene un sitio web de lista roja de las especies de plantas y animales con diferentes grados de peligro de extinción. El conjunto de datos extraído será únicamente del reino Animal con las especies que se han considerado en el nivel más cercano a la extinción (CR-Critically Endangered).

5. Contenido. ¿Qué campos incluye? ¿Cuál es el periodo de tiempo de los datos y cómo se ha recogido?

name – nombre común

kingdom – reino

class – clase

phylum – filo

scientificName – nombre científico

family – familia

species – especie

order – orden

genus – género

yearPublished – año de publicación

populationTrend – Tendencia de la población

biogeographicalRealms – regiones biogeográficas

systems – sistemas donde habita

scopes – alcance

lastAssessed – última vez que se recibieron datos

estimatedPopulation – población estimada

threats – riesgos que afectan su existencia

habitats – los hábitats

Los datos son válidos desde 1996 hasta la fecha de ejecución del script en octubre 2018. El año está incluido en el campo “yearPublished”

6. Agradecimientos. ¿Quién es propietario del conjunto de datos? Incluid citas de investigación o análisis anteriores.

La organización IUCN es la propietaria del conjunto de datos. Según los términos y condiciones está permitido el uso de datos por motivos educativos.

7. Inspiración. ¿Por qué es interesante este conjunto de datos? ¿Qué preguntas le gustaría responder la comunidad?

La IUCN mantiene actualizada una lista global con los datos de todos los seres vivos. Esta misma organización tiene un sitio especializado llamado lista roja donde se encuentran los seres vivos con algún peligro de supervivencia. Los datos son interesantes para conocimiento general de especies, para tomar acciones preventivas con las especies en riesgo, para tener estadísticas de las malas prácticas humanas que afectan los ecosistemas. Este conjunto de datos va cambiando año tras año y se puede ver tendencias de aparición, supervivencia o desaparición de especies.

8. Licencia. Seleccionad una de estas licencias y decid porqué la habéis seleccionado:

Released Under CC BY-NC-SA 4.0 License

La licencia BY-NC-SA es necesaria porque los datos son para propósitos educativos únicamente según los términos de la IUCN. Non-Commercial es requerido en la definición de la licencia

9. Código: Hay que adjuntar el código con el que habéis generado el dataset, preferiblemente con R o Python, que os ha ayudado a generar el dataset

El código para obtener el dataset está en R. Para mejor entendimiento del código se utilizó R Markdown. Existe un archivo html generado a través de R Markdown con la explicación detallada del código.

10. Dataset: Dataset en formato CSV

La salida es un archivo llamado endangeredSpecies.csv