

Análisis y Limpieza de Datos de Incidentes de Ataques de Tiburones

Informe gerencial

Solana Navarro, Ines Suarez y Lourdes Wittmund Montero

13 de Noviembre del 2024

1 Introducción

El dataset en cuestión, proporcionado por la página Shark Attack File, incluye información sobre ataques de tiburones, con detalles como el tipo de incidente, país y actividad de la víctima. El desafío principal fue garantizar la consistencia de los registros y eliminar datos incompletos o mal clasificados.

2 Objetivos

Buscamos obtener una representación fiel de los incidentes en cuanto a su clasificación por tipo de ataque y distribución geográfica.

3 Análisis y Resultados

Se observó que un 8.1% de los datos no correspondían a ninguna de las categorías de tipos de incidentes principales, lo que representaba una fuente de posibles sesgos. Los registros con errores tipográficos, formatos inadecuados y entradas ambiguas afectaban la calidad del dataset. Un análisis detallado de las columnas "Type" y "Country" reveló la necesidad de una estandarización en la escritura de los nombres de los países y una clasificación más precisa de los tipos de incidentes.

4 Principales decisiones

- Eliminación de filas con datos vacíos o etiquetas incorrectas.
- Reclasificación de incidentes etiquetados como "Invalid" a categorías más apropiadas como "Questionable" cuando era posible.
- Estandarización de los nombres de países y eliminación de símbolos o términos ambiguos como "Ocean" o "Sea".

5 Hallazgos Principales

El análisis de los datos limpios reveló que los ataques “Unprovoked” son la categoría más común, especialmente en países con extensas costas y alta actividad humana en el agua, como Estados Unidos, Australia y Sudáfrica. Otros países como Nueva Zelanda, Brasil, México, Papúa Nueva Guinea y Fiji también mostraron incidencias destacadas, lo cual es coherente con sus condiciones geográficas y climáticas, que facilitan los encuentros entre tiburones y personas. Los ataques “Questionable” por otro lado son la categoría con menor número de ataques registrados.

6 Conclusiones

La limpieza de datos realizada permitió mejorar significativamente la coherencia y calidad del dataset. Con los registros depurados y estandarizados, el análisis reflejó de manera más precisa la distribución y los tipos de incidentes de ataques de tiburones, destacando la prevalencia de ataques no provocados en países como Estados Unidos y Australia.