

# Data Analytics 2024

# Project overview

## Base de datos original

- ★ **Origen de los datos:** fuente pública con 6.944 registros.
- ★ **Estructura de la base de datos:** 23 columnas con información sobre las ubicaciones, fechas, actividades, perfil de las personas que fueron atacadas, etc.
- ★ **Tamaño de la base de datos:** 6.944 filas y 23 columnas.
- ★ **Periodo de tiempo cubierto:** desde 1845 hasta marzo de 2024.
- ★ **Limitaciones:** la base de datos estaba muy desordenada, tenía muchos valores nulos y carecía de formato.

## Nuestra hipótesis

El número de ataques de tiburones está relacionado con la mayor presencia de tiburones. Por tanto, esperábamos encontrar patrones que permitieran identificar momentos y lugares con mayor actividad de tiburones de gran tamaño, lo que facilitaría la recomendación de ubicaciones para el avistamiento.

# Project overview

## Estructura, proceso de limpieza y análisis de datos

- ★ **Limpieza:** identificación y manejo de valores nulos, estandarización de formatos de fecha, país y ubicación, eliminación de registros irrelevantes y corrección de outliers.
- ★ **Formateo:** organización de los datos en tablas estructuradas, normalización de variables para facilitar la comparación, conversión de datos categóricos a numéricos.
  - Algunas técnicas empleadas: dropna para eliminar valores nulos, reseteo del index o grouping.
- ★ **Visualización de datos:** creación de gráficos y mapas para representar patrones y tendencias, identificación de puntos calientes de avistamientos de tiburones, exploración de la distribución temporal y geográfica de los ataques.
- ★ **Análisis de datos:** aplicación de técnicas estadísticas para identificar correlaciones y relaciones, evaluación de la hipótesis sobre la relación entre la frecuencia de ataques y la presencia de tiburones, generación de insights para recomendar ubicación óptimas para actividades de buceo.

# Data Wrangling and cleaning

## Retos de limpieza superados

Los datos faltantes y los problemas de formateo pueden afectar a la calidad de los resultados, la confianza en las conclusiones, la eficiencia del análisis y la relevancia de las decisiones.

La base de datos tenía registros desde hace más de 2 siglos por lo que limpiarla y formatearla ha sido todo un reto que hemos superado aplicando las técnicas de limpieza aprendidas y enfrentando cada problema paso a paso.

Además, finalmente se decidió limitar el análisis a los últimos 10 años (con 946 registros), ya que los registros anteriores eran los que estaban especialmente sucios y tampoco los considerábamos relevantes para extraer conclusiones que pudieran ser de utilidad para nuestro proyecto.



# Exploratory data analysis

## Métodos de análisis de datos exploratorio

- ★ Estadísticas descriptivas
- ★ Gráficos de barras y de dispersión
- ★ Mapas de calor
- ★ Análisis temporal
- ★ Limitaciones

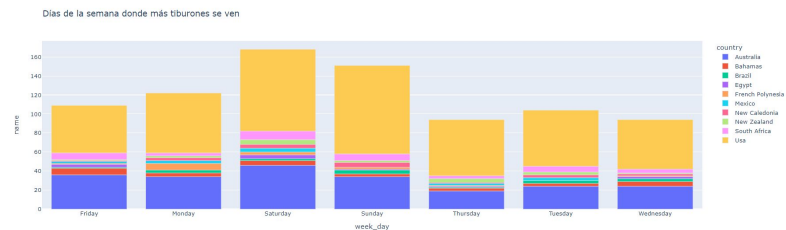
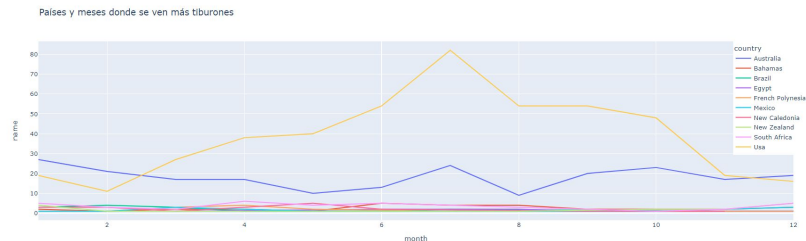


En los últimos 10 años ha habido un total de 946 avistamientos de tiburones.

EE.UU. (465) y Australia (221) lideran el nº de avistamientos con un 49.2% y 23.3% respectivamente. Seguidos de Sudáfrica (39) y Bahamas (31) a bastante distancia.

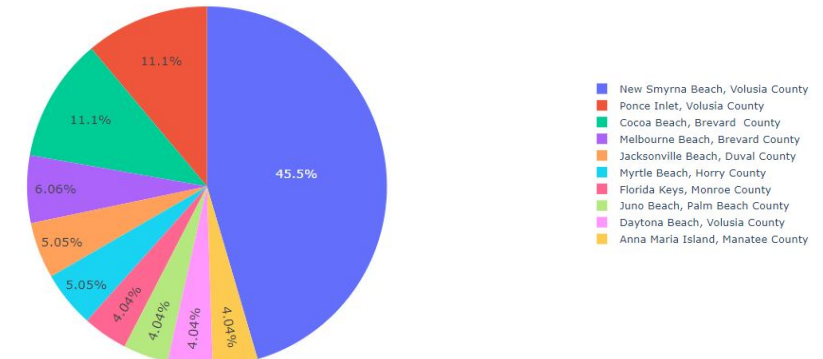
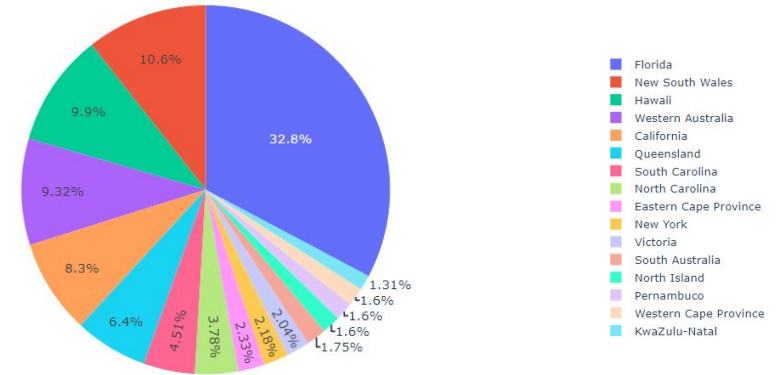
Los meses de verano (junio, julio y agosto) representan aproximadamente el 36% del total de avistamientos, siendo julio el que condensa mayor número con 135 durante el periodo analizado.

Los días de fin de semana (sábado y domingo) representan aproximadamente el 36% del total.



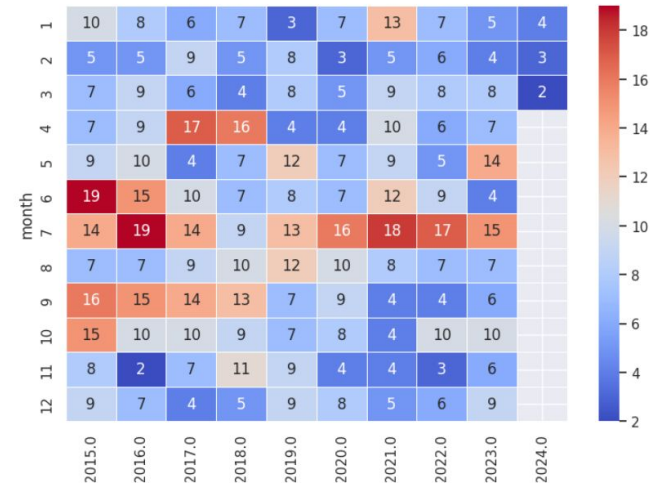
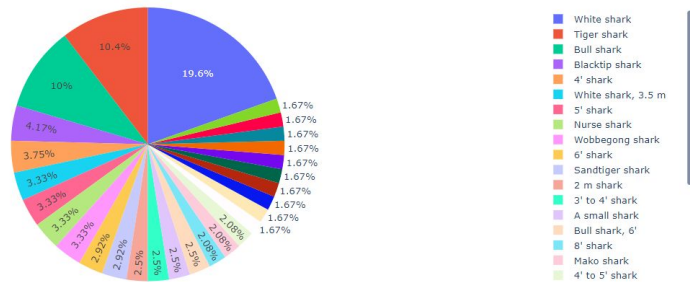
El 72% de los avistamientos tienen lugar en EE.UU. y Australia, por lo que las ciudades y localizaciones en las que más tiburones se ven se concentran en estos países y localizaciones:

- Florida (EE.UU.), New South Wales (Australia) y Hawaii (EE.UU.) representan aproximadamente el 35% del total de avistamientos de tiburones durante los últimos 10 años.
- De igual forma, New Smyrna encabeza la lista con 45 avistamientos, seguida de Ponce Inlet con 11, lo cual representa más de un 50% de los avistamientos.



Los tipos de tiburones que los clientes de ‘Dark Blue’ pueden esperar avistar, por ser los más vistos en los últimos 10 años, son el tiburón blanco (47), el tiburón tigre (25) y el tiburón toro (24).

Por último, parece que la evolución de avistamientos a lo largo de los últimos 10 años ha permanecido bastante estable y, como podemos ver en el mapa de calor y en los gráficos anteriores, se concentra en los meses centrales del año.





# Major obstacle

Una vez realizada la limpieza preliminar de los datos nos dimos cuenta de que implicaba la eliminación de muchos registros.

Por tanto, tuvimos que decidir si preferíamos priorizar la limpieza o el volumen de datos.

Finalmente, se decidió que priorizar el volumen era mejor para el análisis, ya que esos valores no eran esenciales para el objetivo del proyecto y contábamos con un tiempo muy limitado para su desarrollo.

Hemos aprendido a priorizar, teniendo siempre en cuenta el objetivo principal del proyecto de análisis de datos.

Creemos que esta decisión ha repercutido favorablemente en el análisis y hemos podido avanzar a la vez que presentar los datos más importantes para el proyecto.

# Insights

Hemos podido comprobar que, efectivamente, nuestra hipótesis se cumple: existen zonas calientes para el avistamiento de tiburones, concentrándose los ataques -y por tanto avistamientos- en dos países principalmente: Estados Unidos y Australia.

Por otro lado, el avistamiento se aglutina de manera clara en los meses centrales del año.

Gracias a este análisis, la empresa de buceo 'Dark Blue' puede ir optimizando su servicio focalizando el catálogo de actividades en estos destinos y fechas.

Quedaría pendiente complementar el análisis con otros datos para concluir si este mayor número de avistamientos en épocas concretas se debe a que coincide con épocas estivales.

Por último, en una siguiente fase del proyecto, habría que tener en cuenta variables como la temperatura del agua o las condiciones climáticas, además del desarrollo de gráficos más visuales, interactivos o visualizaciones 3D y establecer modelos predictivos que ayuden a la empresa de buceo a ofrecer un mejor servicio.

# Quest: Shark Attacks

¡Gracias!

Andrea Wahnnon  
Alicia Pérez