



A la deriva: análisis exploratorio de los registros de cetáceos en aguas españolas (2000–2024)

Autoría: Cecilia Vanessa Mendoza Peña

Bootcamp de Data Science – The Bridge

Proyecto de Análisis Exploratorio de Datos (EDA)

Entrega: mayo de 2025

*Imagen de portada: Delfín común (*D. delphis*). Fotografía propia*

Índice

Índice

1. Resumen ejecutivo
2. Contexto y motivación
 - 2.1 El valor ecológico de los cetáceos
 - 2.2 Ciencia ciudadana, datos abiertos y retos de interpretación
 - 2.3 Comparación de especies: más que números
 - 2.4 ¿Por qué comunicar esto?
3. Preguntas de investigación e hipótesis
 - 3.1 Hipótesis inicial
 - Hipótesis contrastadas
 - 3.2 Preguntas de investigación
4. Descripción del dataset y proceso de limpieza
 - 4.1 Proceso de limpieza y transformación
5. Análisis Exploratorio
 - 5.1 Análisis univariante
 - 5.1.1 Año del evento (`anio`)
 - 5.1.2 Mes del evento (`mes`)
 - 5.1.3 Número de individuos por registro (`individualCount`)
 - 5.1.4 Especie observada (`nombre_cientifico`)
 - 5.1.5 Comunidad Autónoma (`comunidad_autonoma`)
 - 5.1.6 Nivel taxonómico (`taxonrank`)
 - 5.1.7 Tipo de incidencia (`issue`)
 - 5.1.8 Coordenadas geográficas (`decimallatitude` y `decimallongitude`)
 - 5.2 Análisis bivariante
 - 5.2.1 Comparativa de abundancia por especie
Distribución de individuos por especie
 - 5.2.2 Distribución del número de individuos por comunidad autónoma
 - 5.2.3 Relación entre latitud y número de individuos observados
 - 5.2.4 Distribución conjunta de especies y comunidades autónomas
Nota sobre los tests estadísticos
 - 5.3 Análisis multivariante: correlaciones entre variables numéricas
 - 5.3.1 Correlación entre variables (Spearman)
 - 5.3.2 Distribución espacial por especie (mapa multivariado)
6. Comparativa entre especies
 - 6.1 Introducción
 - 6.2 Evolución temporal de registros
 - 6.3 Distribución geográfica
 - 6.4 Estacionalidad
 - 6.5 Comparación entre especies por períodos
 - Períodos 2000-2005 y 2006-2011
 - Período 2012-2017
 - Período 2018-2024
 - 6.6 Síntesis comparativa
 - 6.7 Interpretación general de los cambios espacio-temporales en la distribución
7. Conclusiones y próximos pasos
 - 7.1 Evaluación de las hipótesis de trabajo
 - Hipótesis contrastadas
 - 7.2 Conclusiones generales
 - 7.3 Próximos pasos y oportunidades de mejora
8. Sobre la autora

1. Resumen ejecutivo

Este proyecto realiza un análisis exploratorio de los registros de cetáceos observados en aguas españolas entre los años 2000 y 2024. A partir de datos abiertos extraídos del portal [GBIF](#), se ha llevado a cabo una limpieza y transformación del conjunto de datos para explorar la distribución espacio-temporal de las principales especies registradas, con especial atención a *Delphinus delphis* (**delfín común**) y *Stenella coeruleoalba* (**delfín listado**).



Delfín común (*D. delphis*) | Src: propia



Delfín listado (*S. coeruleoalba*) | Src: [Dolphin Watching Mallorca](#)

El análisis se estructura en distintas fases: un estudio univariante de variables clave, análisis bivariantes y multivariantes para detectar patrones y correlaciones, así como una visualización comparativa centrada en las dos especies dominantes. Se han empleado herramientas de Python como pandas, matplotlib, seaborn y folium para construir las visualizaciones y extraer conclusiones.

Entre los principales hallazgos, se identifican **diferencias en la distribución geográfica y temporal entre las especies analizadas**, con una concentración destacada de observaciones en áreas específicas del **litoral mediterráneo y el Atlántico sur**. Estos patrones podrían estar vinculados a variables ecológicas y a la variabilidad en el esfuerzo de muestreo.

Este trabajo pone de relieve la **utilidad de los datos abiertos** en la investigación ambiental, y abre la puerta a futuros estudios que integren variables ambientales o registros de esfuerzo para una mejor interpretación. La metodología seguida es replicable y representa un ejemplo claro de cómo el análisis de datos puede contribuir a la conservación marina.

2. Contexto y motivación

2.1 El valor ecológico de los cetáceos

Los cetáceos son especies clave en los ecosistemas marinos. Su papel como **depredadores tope** influye en el equilibrio de las redes tróficas, y su comportamiento de migración, alimentación y reproducción está estrechamente ligado a la salud del océano.

Además, muchas especies, como los delfines pelágicos, **contribuyen al reciclaje de nutrientes** en aguas abiertas y funcionan como **indicadores ecológicos** del estado del medio marino. Cambios en su distribución o abundancia pueden alertar de alteraciones en temperatura, disponibilidad de alimento, contaminación acústica o presión pesquera.

2.2 Ciencia ciudadana, datos abiertos y retos de interpretación

El presente análisis se basa en datos recopilados por diversas fuentes a través de la plataforma [GBIF](#), que integra tanto observaciones científicas como registros provenientes de **ciencia ciudadana**. Esta combinación permite obtener una imagen amplia y dinámica de la biodiversidad marina, pero también requiere una interpretación crítica.

Algunos puntos clave a considerar:

- **Esfuerzo de muestreo desigual:** no todas las zonas ni todos los años tienen el mismo número de observaciones.
- **Detectabilidad variable:** especies más carismáticas o activas pueden estar **sobrerrepresentadas**.
- **Sesgos espaciales y temporales:** registros más abundantes cerca de puertos, rutas turísticas o en períodos vacacionales.

Aun así, estos datos son **fundamentales para comprender tendencias a gran escala**, identificar áreas de especial interés y formular hipótesis científicas validadas por patrones observables.

2.3 Comparación de especies: más que números

La comparación entre *Stenella coeruleoalba* y *Delphinus delphis* va más allá del conteo de individuos. Permite abrir una reflexión sobre:

- **Preferencias ecológicas distintas:** una especie más asociada a plataformas continentales (*Delphinus*), otra a aguas oceánicas profundas (*Stenella*).
- **Diferencias en accesibilidad para su estudio:** zonas costeras vs. alta mar.
- **Impacto de actividades humanas:** tráfico marítimo, pesca incidental, contaminación acústica.

Entender estas diferencias ayuda a:

- **Diseñar estrategias de conservación adaptadas a cada especie y hábitat.**
- **Priorizar zonas para monitoreo y protección.**
- **Involucrar a la ciudadanía con información veraz y contextualizada.**

2.4 ¿Por qué comunicar esto?

En el contexto del cambio global y la pérdida de biodiversidad, comunicar de forma efectiva lo que los datos nos revelan es más urgente que nunca. Este análisis no solo muestra números, sino que contribuye a:

- **Dar visibilidad a especies que no siempre reciben atención mediática.**
- **Reforzar el uso de datos abiertos como herramienta de conservación.**
- **Fomentar una ciudadanía informada y participativa.**

En resumen, el conocimiento científico debe estar al servicio de la sostenibilidad. Este trabajo aspira a ser una pequeña contribución en esa dirección.

3. Preguntas de investigación e hipótesis

Este proyecto parte de una hipótesis inicial basada en la percepción común y la literatura ecológica, la cual fue contrastada mediante el análisis exploratorio y reformulada durante el proceso.

3.1 Hipótesis inicial

Delphinus delphis (delfín común) es la especie con mayor número de individuos observados en las costas españolas.

Esta hipótesis se apoyaba en:

- Su amplia distribución en aguas templadas y subtropicales.
- Su familiaridad para científicos y ciudadanía.

- Su comportamiento gregario, que facilita su detección en superficie.

Sin embargo, el análisis reveló lo siguiente:

Stenella coeruleoalba (delfín listado) es la especie con mayor número total de individuos observados entre 2000 y 2024.

Este resultado dio lugar a una **comparación específica entre ambas especies**, reformulando el enfoque del análisis comunicativo en torno a tres hipótesis revisadas:

Hipótesis contrastadas

H1: *Delphinus delphis* es la especie con mayor número de individuos observados.

H2: Galicia, Andalucía y Canarias concentran los registros con mayor abundancia y diversidad.

H3: *Delphinus delphis* ha modificado su distribución espacial, desplazándose hacia el norte peninsular o el Mediterráneo.

3.2 Preguntas de investigación

- ¿Qué especies dominan el registro de observaciones de cetáceos en aguas españolas?
- ¿Cómo varían espacial y temporalmente las observaciones de *Delphinus delphis* y *Stenella coeruleoalba*?
- ¿Existen patrones estacionales o geográficos que permitan caracterizar su distribución?
- ¿Qué implicaciones tienen estas diferencias para la conservación marina?

Estas hipótesis y preguntas estructuran el análisis exploratorio y la narrativa visual del proyecto.

4. Descripción del dataset y proceso de limpieza

Los datos utilizados en este análisis proceden del portal [GBIF](#), una infraestructura internacional que proporciona acceso abierto a datos sobre biodiversidad global. Se aplicaron los siguientes filtros:

- **Taxón:** Cetacea
- **Región:** España
- **Tipo de registro:** Observación humana y observación por máquina
- **Estado del individuo:** Presente
- **Coordenadas válidas:** Sí
- **Problemas geoespaciales:** `false`

El archivo descargado fue un `.csv` almacenado en [data/raw/gbif_cetaceos_espana.csv](#), con un total de **18.169** registros.

4.1 Proceso de limpieza y transformación

El proceso de limpieza abordó varias incidencias comunes en registros abiertos:

- Error de tokenización por delimitadores inconsistentes → se resolvió con `on_bad_lines="skip"` y `low_memory=False`.

- Columnas con fechas múltiples (`eventDate`, `year`, `month`) → se unificaron extrayendo `anio` y `mes`.
- Registros sin coordenadas o sin nombre científico a nivel de especie → fueron eliminados.
- Estandarización y normalización textual de columnas categóricas (`taxonrank`, `stateprovince`, `locality`, etc.).
- Conversión de nombres científicos completos a forma binomial (`nombre_científico`).
- Identificación de valores extremos en `individualCount`, manteniéndolos para inspección posterior.
- Agrupación y corrección de nombres geográficos para minimizar la pérdida de datos (por ejemplo, "illes balears" y "islas baleares").

Se eligieron las variables más relevantes para el análisis y los datos fueron guardados en un nuevo archivo limpio: [data/clean/df_cetaceos_limpio.csv](#), con un total de **15.783** registros.

Este conjunto final sirvió de base para todos los análisis exploratorios posteriores.

Variable	Descripción
<code>gbifid</code>	Identificador único del registro en la base de datos GBIF
<code>anio</code>	Año del evento de observación o muestreo
<code>mes</code>	Nombre científico completo (incluye autor y año)
<code>taxonrank</code>	Nombre científico completo (incluye autor y año)
<code>scientificname</code>	Nombre científico binomial (género y especie) extraído de <code>scientificname</code>
<code>nombre_científico</code>	Latitud geográfica del evento, en formato decimal
<code>decimallatitude</code>	Longitud geográfica del evento, en formato decimal
<code>decimallongitude</code>	Fecha completa del evento, si está disponible
<code>eventdate</code>	Fecha completa del evento, si está disponible
<code>individualcount</code>	Número de individuos registrados en ese evento
<code>locality</code>	Descripción local específica del lugar de observación
<code>stateprovince</code>	Provincia o región geográfica declarada en el registro original
<code>issue</code>	Posibles incidencias o problemas detectados en el registro
<code>comunidad_autonoma</code>	Comunidad autónoma derivada del campo <code>stateprovince</code>

5. Análisis Exploratorio

5.1 Análisis univariante

El análisis univariante permite examinar la distribución individual de cada variable relevante del conjunto de datos. Su objetivo es identificar tendencias, valores atípicos, sesgos de muestreo o patrones destacados que puedan influir en el análisis posterior.

En este proyecto, se ha aplicado el análisis univariante sobre variables temporales, espaciales, taxonómicas y de abundancia. Para cada variable, se construyeron visualizaciones que permiten evaluar su comportamiento y su posible utilidad explicativa.

A continuación, se presentan los principales resultados organizados por tipo de variable.

5.1.1 Año del evento (`anio`)

Esta variable representa el año en el que se registró cada observación de cetáceos en aguas españolas. Su análisis permite evaluar la cobertura temporal del conjunto de datos, detectar sesgos cronológicos y validar la

correcta construcción del campo de fechas.

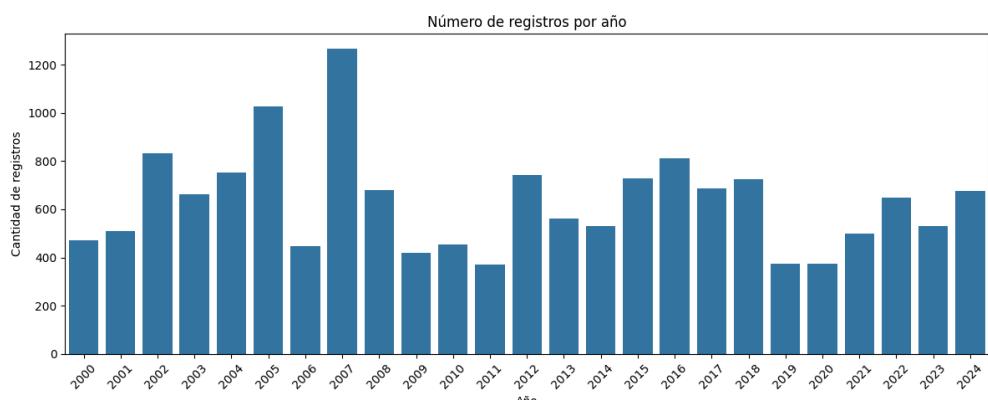


Figura 1. Distribución de registros por año (2000–2024). Se observa una fuerte variabilidad interanual, con un máximo en 2007 y descensos marcados en períodos como 2009–2011 y 2019–2021. Estas fluctuaciones pueden reflejar cambios en el esfuerzo de muestreo, la digitalización de datos o, potencialmente, variaciones ecológicas en la distribución o abundancia de cetáceos.

El conjunto de datos abarca desde el año 2000 hasta 2024, aunque con una distribución marcadamente irregular. El año con mayor número de registros fue 2007, seguido de otros picos relevantes como 2005 y 2016. En cambio, los años 2006, 2009–2011 y 2019–2021 presentan volúmenes considerablemente menores.

Estas fluctuaciones pueden deberse a múltiples factores: variaciones en el esfuerzo de observación (número de campañas, participación ciudadana), diferencias en la digitalización o carga de datos en GBIF, e incluso posibles cambios reales en la distribución o abundancia de cetáceos. Por ejemplo, el descenso sostenido entre 2019 y 2021 podría estar influido por la pandemia de COVID-19, pero también podría reflejar una menor presencia o detectabilidad de las especies objetivo.

5.1.2 Mes del evento (mes)

La variable `mes` permite evaluar la estacionalidad de las observaciones de cetáceos a lo largo del año. Esta dimensión temporal es fundamental para identificar posibles patrones migratorios, concentraciones estacionales o sesgos ligados al esfuerzo de observación en determinadas épocas.

El análisis muestra un patrón claro: los registros aumentan progresivamente desde la primavera y alcanzan un pico muy pronunciado en julio (mes 7), con más de 3.400 observaciones. Agosto y septiembre mantienen también una frecuencia elevada. A partir de octubre, la cantidad de registros desciende de forma marcada hasta alcanzar su mínimo en diciembre.

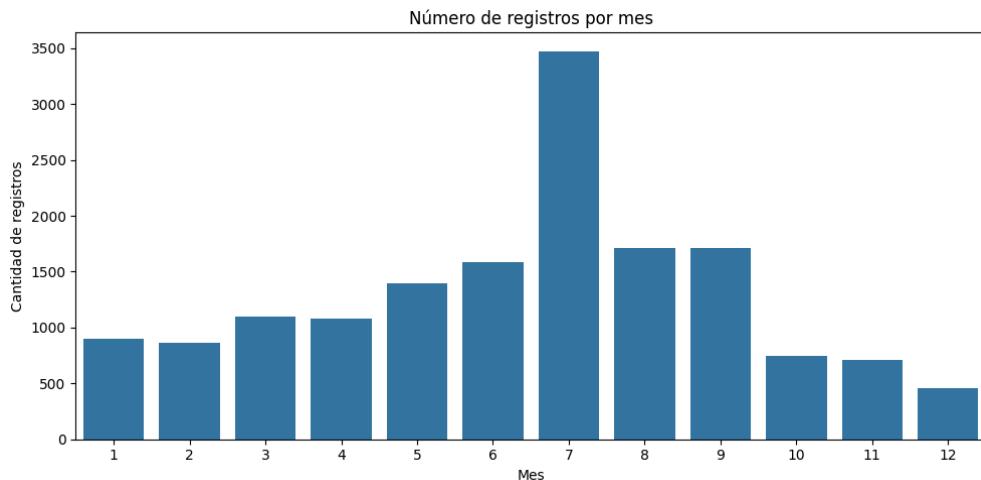


Figura 2. Distribución de registros por mes (1 = enero, 12 = diciembre). Se observa un claro patrón estacional, con un pico muy pronunciado en julio y una disminución progresiva a partir de octubre. Este comportamiento puede reflejar tanto patrones ecológicos reales como una mayor actividad de observación durante los meses de verano.

Este patrón puede estar influido por dos factores principales:

- Una mayor presencia o detectabilidad de cetáceos en verano, por cuestiones ecológicas (migraciones, reproducción, alimentación).
- Un esfuerzo de observación más intenso en los meses estivales, tanto por campañas científicas como por turismo marino o ciencia ciudadana.

Dado que el dataset no incluye información sobre esfuerzo de muestreo, es necesario interpretar estos patrones con cautela. Aun así, las variables `año` y `mes` ofrecen una base sólida para evaluar tendencias temporales.

5.1.3 Número de individuos por registro (`individualCount`)

Esta variable indica cuántos individuos fueron observados en cada evento registrado. Su análisis ayuda a entender la dinámica de avistamientos (individuales o grupales) y la posible calidad o precisión del dato.

El análisis muestra una distribución fuertemente sesgada hacia valores bajos: más del 75 % de los registros contienen entre 1 y 10 individuos. Sin embargo, también se detectan picos inusuales en valores como 10, 20, 30 o 50, lo que sugiere que algunos registros podrían haber sido **estimados o redondeados** manualmente por el observador, más que contados con precisión.

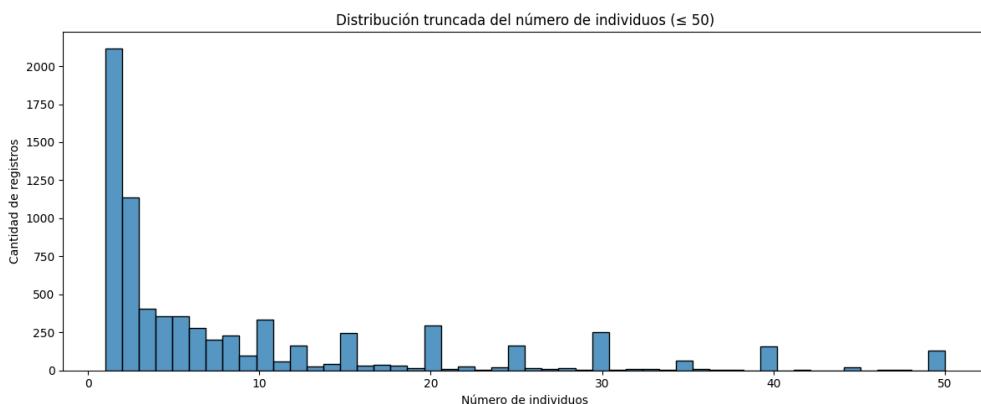


Figura 3. Distribución del número de individuos por registro (≤ 50). La mayoría de los registros agrupan entre 1 y 10 individuos. Se observan picos en múltiplos de 10 (10, 20, 30), lo cual podría deberse a estimaciones redondeadas durante el muestreo.

Desde el punto de vista ecológico, este patrón es coherente con el comportamiento grupal de muchas especies de cetáceos. Es común observar a individuos formando **pods o grupos sociales**, cuyo tamaño varía según la especie, la época del año o el contexto ecológico. En algunas ocasiones, especialmente en zonas ricas en alimento o durante migraciones, pueden formarse **macro pods** que superan los 100 individuos, lo que explicaría parte de los valores extremos registrados.

Aunque existen valores extremos por encima de 100, estos no son frecuentes. Por ello, se optó por una visualización **truncada a valores ≤ 50** , que permite apreciar mejor la distribución real de la mayoría de registros.

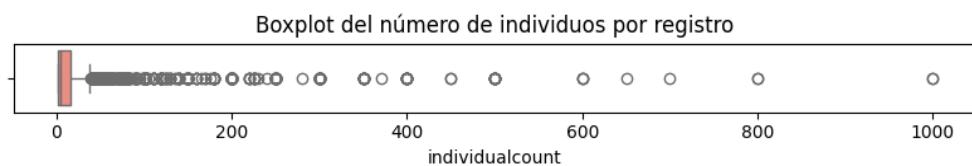


Figura 4. Diagrama de caja del número de individuos por registro. La gran mayoría de observaciones se agrupan en valores bajos, pero se identifican múltiples valores atípicos por encima de 200, lo que confirma una dispersión elevada.

5.1.4 Especie observada (**nombre_científico**)

Permite identificar la especie registrada en cada observación. Su análisis revela qué especies dominan el conjunto de datos en términos de frecuencia de registro, una métrica relevante para conocer qué especies son más detectadas, reportadas o presentes en zonas accesibles.

El análisis muestra que **Tursiops truncatus** (delfín mular) es la especie **más registrada** en aguas españolas entre 2000 y 2024, seguida por **Stenella coeruleoalba** (delfín listado) y **Delphinus delphis** (delfín común). Esta frecuencia puede reflejar una combinación de factores como la abundancia real, la facilidad de detección visual, el comportamiento costero o la familiaridad del observador.

Esta información es útil como punto de partida, aunque no debe confundirse con abundancia real: para ello se requiere considerar el número total de individuos registrados, lo cual se explora en el siguiente bloque bivariante.

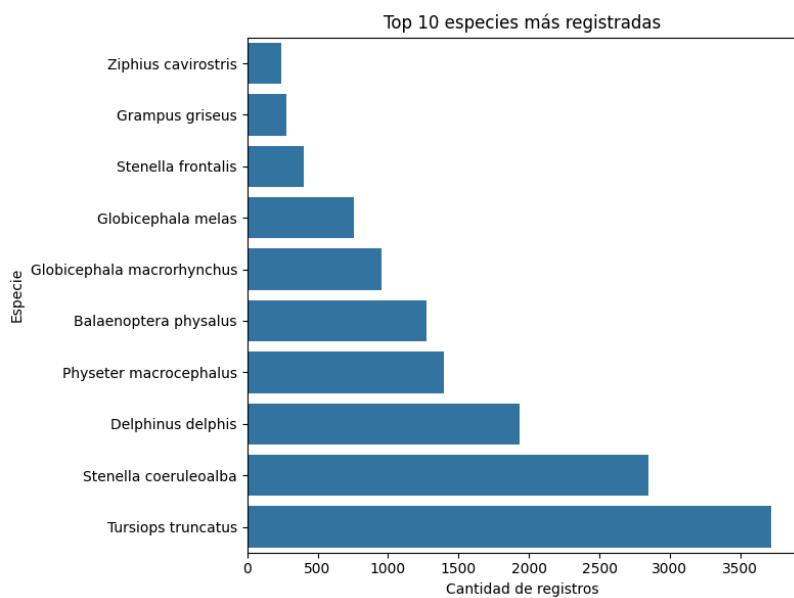


Figura 5. Especies con mayor número de registros. *Tursiops truncatus*, *Stenella coeruleoalba* y *Delphinus delphis* son las especies más frecuentemente observadas en el conjunto de datos. Esta métrica refleja la frecuencia de registro, no necesariamente su abundancia poblacional.

5.1.5 Comunidad Autónoma ([comunidad_autonoma](#))

Esta variable geográfica deriva del campo [stateprovince](#) y agrupa los registros por comunidad autónoma española, permitiendo una primera evaluación de la distribución espacial de los avistamientos. Su análisis revela tanto la cobertura geográfica como posibles sesgos espaciales en la recolección de datos.

El primer gráfico muestra el top 10 de comunidades autónomas considerando todos los registros. Destaca que aproximadamente **el 76 % de los registros no pudieron ser asignados a una comunidad autónoma concreta**, ya que correspondían a valores como [NaN](#), [Mar](#), [Northern Atlantic Ocean](#) u otros puntos sin normalizar. Estos registros fueron agrupados en la categoría "**no asignado**".

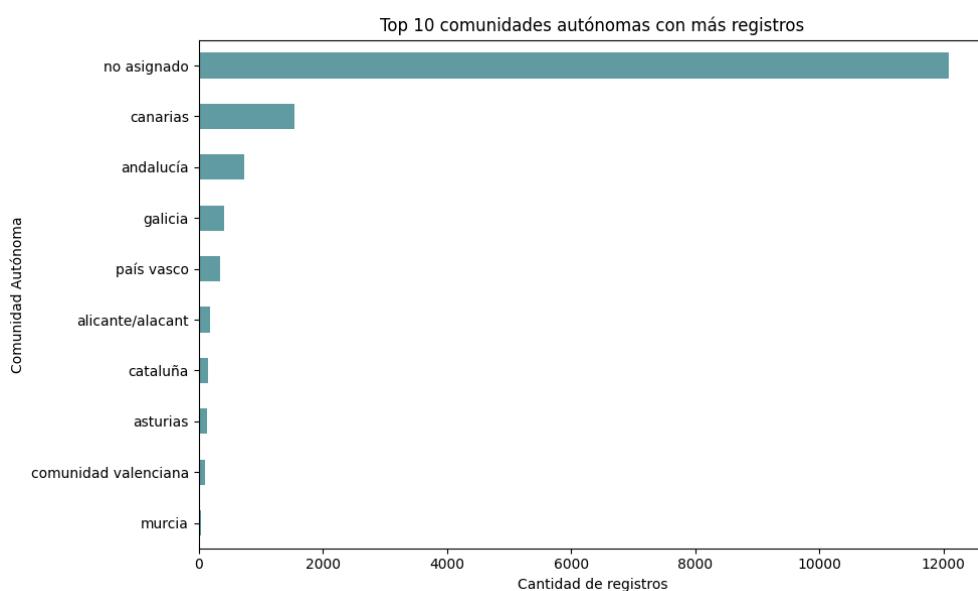


Figura 7. Top 10 comunidades autónomas con más registros (incluyendo valores no asignados). Más del 75 % de los registros no están geográficamente asignados a una comunidad autónoma específica, lo que limita parcialmente el análisis regional.

Aunque este grupo aporta información sobre observaciones fuera del litoral directo, su peso dominante distorsiona la lectura del resto del conjunto. Por ello, se realizó un segundo análisis excluyendo los registros no asignados.

Sin esos valores, se observa una concentración de registros en **Canarias, Andalucía, Galicia y el País Vasco**, lo que podría reflejar tanto patrones ecológicos (áreas de tránsito o residencia de cetáceos) como zonas con mayor esfuerzo de observación o proyectos de monitoreo activos. Este patrón se explora con mayor profundidad en la sección de análisis espacial.

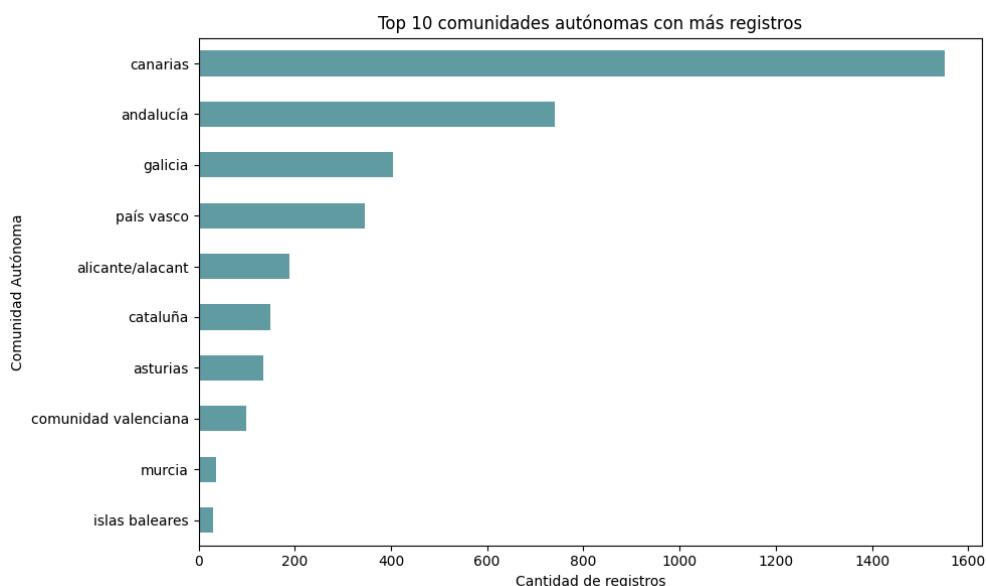


Figura 8. Registros por comunidad autónoma (sin "no asignado"). Canarias, Andalucía, Galicia y País Vasco encabezan el listado, lo que sugiere una combinación de alta actividad de observación y presencia de cetáceos en esas regiones.

5.1.6 Nivel taxonómico (**taxonrank**)

La variable **taxonrank** indica el nivel de clasificación taxonómica al que corresponde cada registro (por ejemplo, especie, género o familia). Su análisis permite validar la **resolución biológica del dataset** y definir con precisión el alcance de los análisis posteriores.

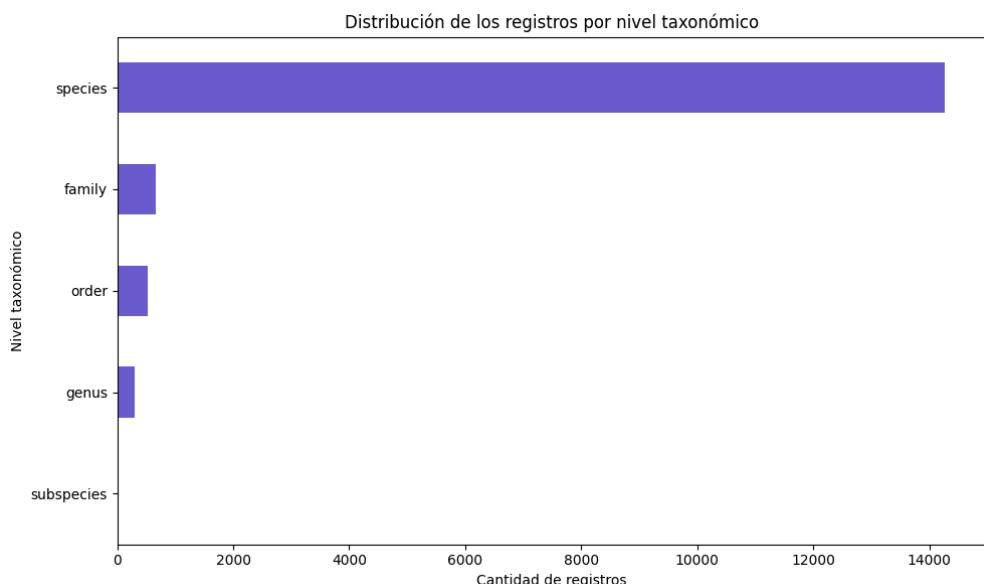


Figura 9. Nivel taxonómico de los registros. La mayoría de los registros están identificados a nivel de especie, lo que permite análisis ecológicos y comparativos con una resolución adecuada. Los registros más generales fueron conservados, pero no se usaron en los análisis por especie.

El gráfico muestra que más del 90 % de los registros se encuentran identificados a nivel de **especie**, lo cual representa una fortaleza importante del conjunto de datos. Esta resolución permite trabajar con variables como `nombre_científico` o `individualCount` sin necesidad de generalizaciones excesivas.

Una minoría de registros está clasificada a nivel de familia, orden o género, lo cual podría deberse a:

- Observaciones donde la identificación precisa no fue posible.
- Cargas automáticas incompletas.
- Especies con taxonomía aún debatida o sin descripción formal.

Estos registros fueron conservados en el dataset limpio por completitud, aunque los análisis enfocados por especie se restringen exclusivamente a aquellos con identificación completa (`taxonrank = species`).

5.1.7 Tipo de incidencia (`issue`)

La variable `issue` identifica posibles incidencias detectadas en cada registro, como errores de coordenadas, fechas inválidas o problemas taxonómicos. Su análisis es crucial para evaluar la **calidad del conjunto de datos** y justificar el proceso de limpieza realizado.

El gráfico muestra las 10 incidencias más frecuentes. Entre ellas destacan:

- **geodetic_datum_assumed_wgs84**: implica que el sistema de referencia espacial se asumió por defecto, lo que puede introducir errores en la localización.
- **country_derived_from_coordinates** y **continent_coordinate_mismatch**: reflejan discrepancias entre la ubicación geográfica declarada y la inferida a partir de las coordenadas.
- **geodetic_datum_invalid**: indica errores en el sistema de coordenadas.
- **recorded_date_invalid** y **coordinate_rounded**: apuntan a problemas en la precisión temporal o espacial del dato.

Estas incidencias no siempre invalidan el registro, pero deben considerarse al interpretar los resultados.

Este análisis pone de manifiesto la importancia de revisar los metadatos asociados a los conjuntos abiertos, especialmente cuando se utilizan con fines científicos o de conservación.

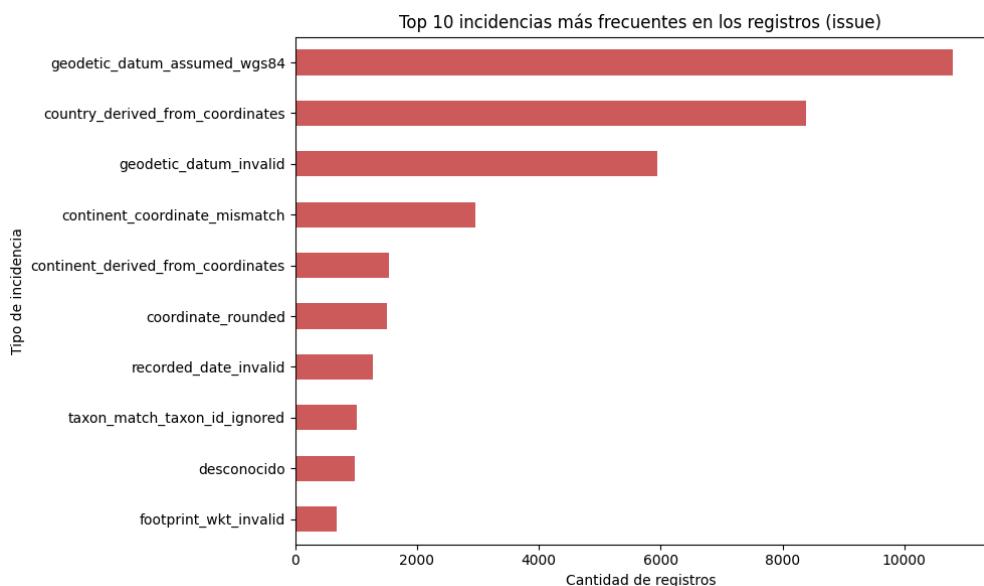


Figura 10. Incidencias más frecuentes en los registros. Los problemas más comunes afectan al sistema de coordenadas y la concordancia geográfica. Aunque muchos no comprometen el análisis, su revisión es esencial para asegurar la calidad del dataset.

5.1.8 Coordenadas geográficas (`decimallatitude` y `decimallongitude`)

Las variables `decimallatitude` y `decimallongitude` representan la ubicación geográfica decimal de cada observación, y permiten evaluar la distribución espacial de los registros. Aunque suelen analizarse conjuntamente en mapas, su estudio por separado revela patrones relevantes sobre las zonas de observación y posibles anomalías.

El histograma de latitud muestra **tres núcleos principales de concentración:** uno en torno a los **28°** (archipiélago canario), otro cerca de los **36°** (Andalucía y litoral sur) y un tercero entre **42°–44°** (norte peninsular). Esta multimodalidad refleja la diversidad geográfica de los registros y sugiere que los avistamientos se distribuyen en varias regiones bien diferenciadas.

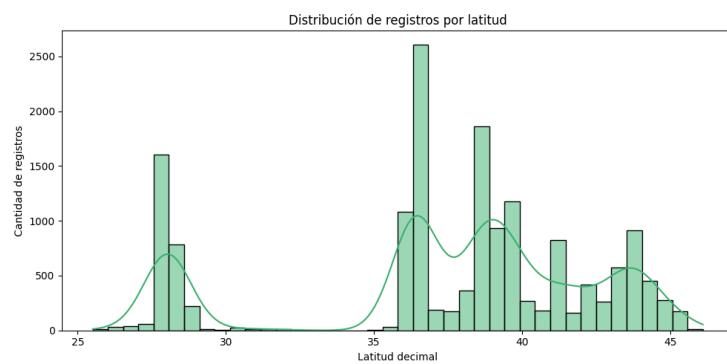


Figura 11. Distribución de registros por latitud decimal. Se observan tres núcleos de observación principales: Canarias (~28°), Andalucía (~36°) y la cornisa norte (~43°), reflejando la amplitud latitudinal del territorio español.

En cuanto a la longitud, se observa una mayor densidad de registros entre los **-18° y 5°**, lo que abarca desde las islas Canarias hasta el Mediterráneo occidental. También se identifican picos secundarios, probablemente relacionados con zonas de observación costera activa o campañas específicas.

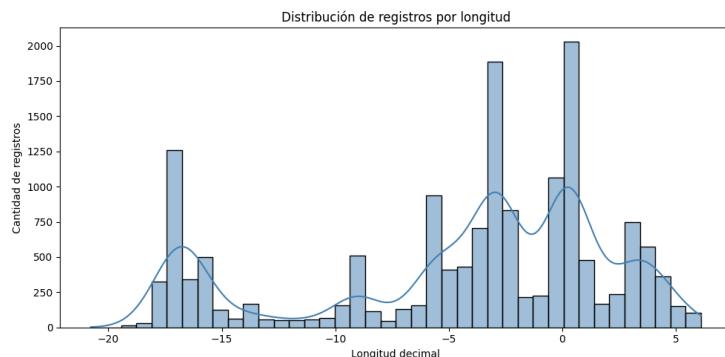


Figura 12. Distribución de registros por longitud decimal. Los registros se concentran entre los -18° y $+5^{\circ}$, cubriendo desde el Atlántico hasta el Mediterráneo. Los picos reflejan áreas costeras de observación frecuente.

Para la detección de outliers se ajustaron los umbrales de latitud y longitud conforme a los límites geográficos oficiales del territorio español, incluyendo Canarias, Baleares, Ceuta y Melilla, según datos del Instituto Geográfico Nacional. Los registros fuera de estos rangos fueron etiquetados como posibles outliers geográficos. Sin embargo, **no se eliminan en esta etapa**, ya que **podrían corresponder a ubicaciones válidas dentro de las aguas jurisdiccionales de España**, como su Zona Económica Exclusiva (ZEE), especialmente en regiones como el suroeste de Canarias.

Se recomienda una revisión más precisa si se desea excluir estos registros en análisis particularmente sensibles a la ubicación geográfica.

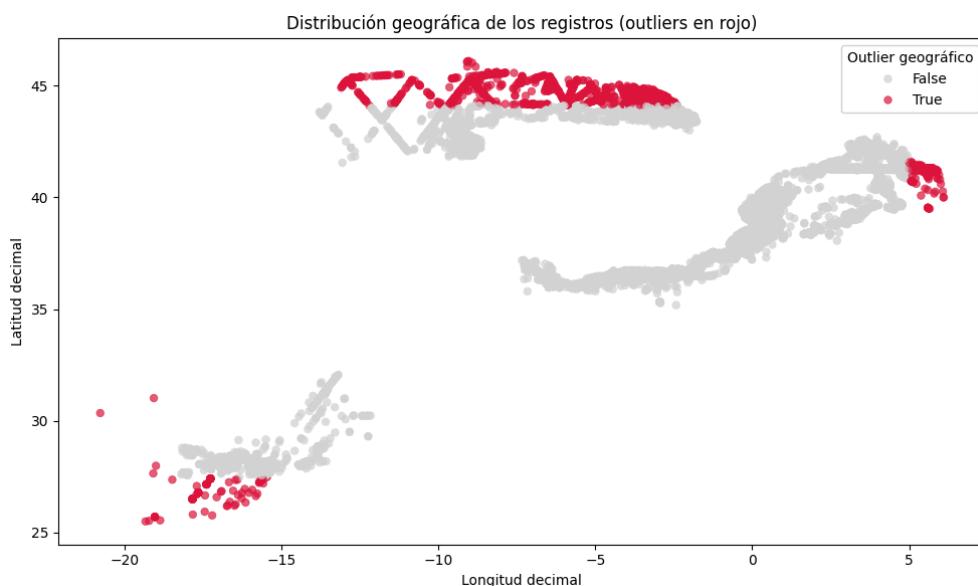


Figura 13. Registros geográficos con detección de “outliers”. Los puntos rojos indican registros geográficamente anómalos según los límites establecidos para el territorio español. Su distribución sugiere valores alejados de los núcleos de observación, como los litorales canario, atlántico o mediterráneo.

5.2 Análisis bivariante

El análisis bivariante permite explorar relaciones entre pares de variables para detectar patrones conjuntos que no son evidentes en los análisis univariantes. En este proyecto, se aplica este enfoque principalmente para

identificar diferencias entre especies, comparar distribuciones geográficas y evaluar posibles asociaciones entre variables ecológicas y temporales.

Iniciamos esta sección examinando la relación entre el número de individuos observados (`individualcount`) y la especie registrada (`nombre_cientifico`). Este análisis permite evaluar no solo la frecuencia de observaciones, sino también **la abundancia relativa y el comportamiento grupal típico** de las especies más comunes.

5.2.1 Comparativa de abundancia por especie

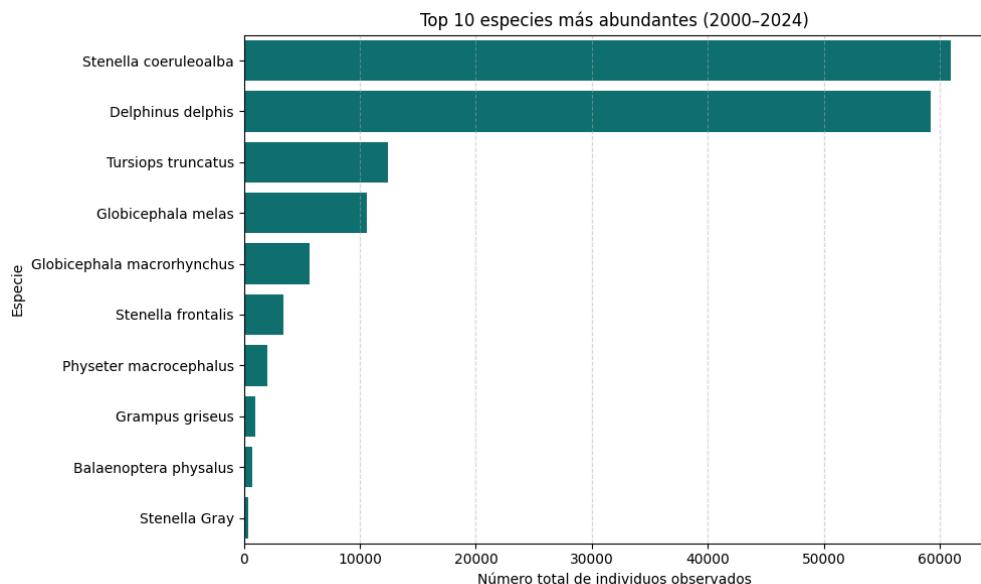


Figura 14. Top 10 de especies más abundantes (2000–2024).

La especie *Stenella coeruleoalba* encabeza el listado con más de 60.000 individuos observados, seguida de *Delphinus delphis*. Estas cifras reflejan no solo su frecuencia en los registros, sino también su tendencia a agruparse en manadas grandes.

Por el contrario, especies como *Balaenoptera physalus* y *Stenella Gray* presentan valores totales considerablemente más bajos, lo que puede deberse tanto a una menor frecuencia de observación como a sus patrones de comportamiento más solitarios o difíciles de detectar.

Distribución de individuos por especie

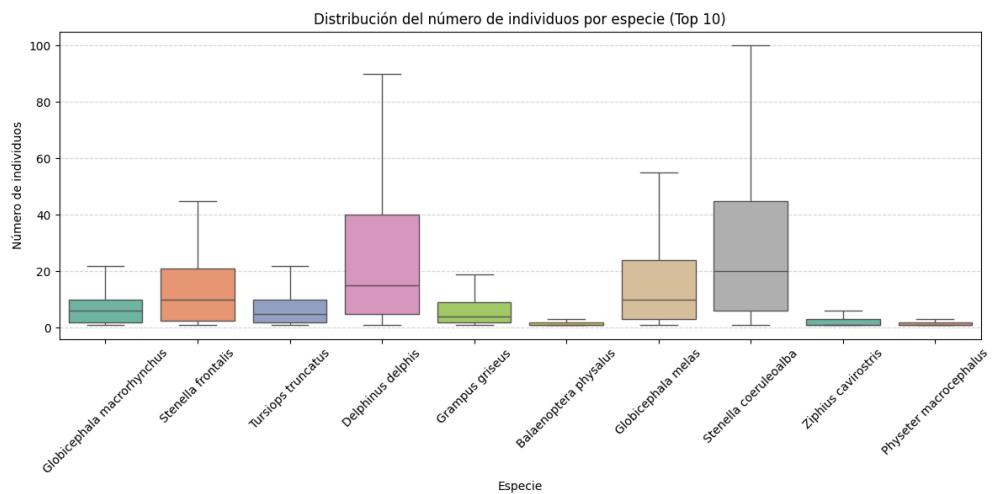


Figura 15. Distribución del número de individuos por registro en las 10 especies más comunes.

El boxplot revela diferencias claras en la dispersión y rango de tamaño de grupo entre especies. *Stenella coeruleoalba* y *Delphinus delphis* muestran una **alta variabilidad y valores máximos elevados**, lo que sugiere la presencia frecuente de **macropods** o agrupaciones masivas. Otras especies como *Globicephala melas* y *Tursiops truncatus* también presentan registros con grupos numerosos, aunque con menor dispersión. En contraste, especies como *Balaenoptera physalus* y *Ziphius cavirostris* tienden a ser observadas en grupos pequeños, consistentes con su biología y hábitos más solitarios.

Este análisis refuerza la importancia de combinar la frecuencia de observaciones con el número de individuos registrados para obtener una visión más precisa del comportamiento y abundancia relativa de cada especie.

5.2.2 Distribución del número de individuos por comunidad autónoma

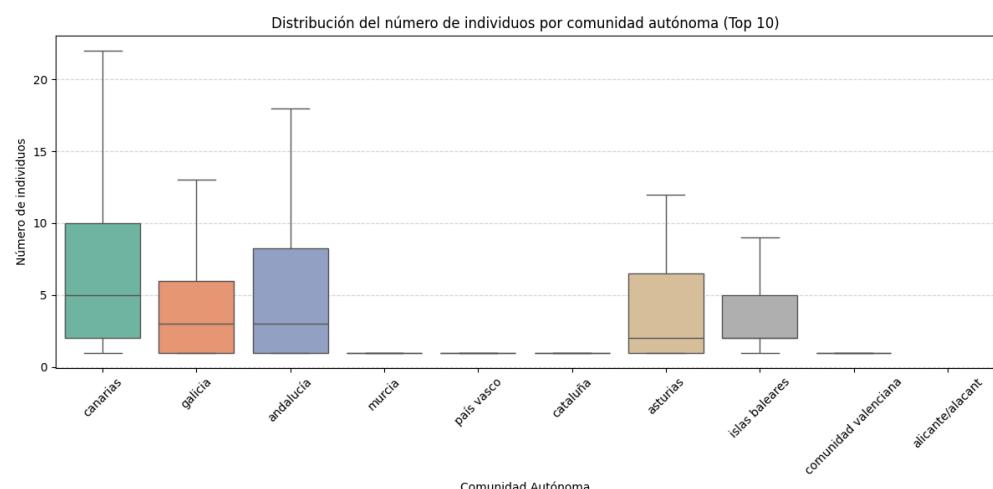


Figura 16. Distribución del número de individuos por comunidad autónoma (Top 10).

Este boxplot revela diferencias en el número de individuos observados por registro según la comunidad autónoma. Destacan **Canarias, Andalucía y Asturias**, que muestran una mayor mediana y dispersión, lo cual podría estar vinculado a una mayor presencia de especies gregarias, una visibilidad más alta de los grupos o un mayor esfuerzo de observación en zonas clave.

En contraste, comunidades como **Murcia, País Vasco o Comunidad Valenciana** presentan distribuciones planas, con valores bajos y sin variabilidad aparente. Esto puede deberse a factores como baja actividad de monitoreo, limitaciones de cobertura o una baja abundancia real en esas regiones.

5.2.3 Relación entre latitud y número de individuos observados

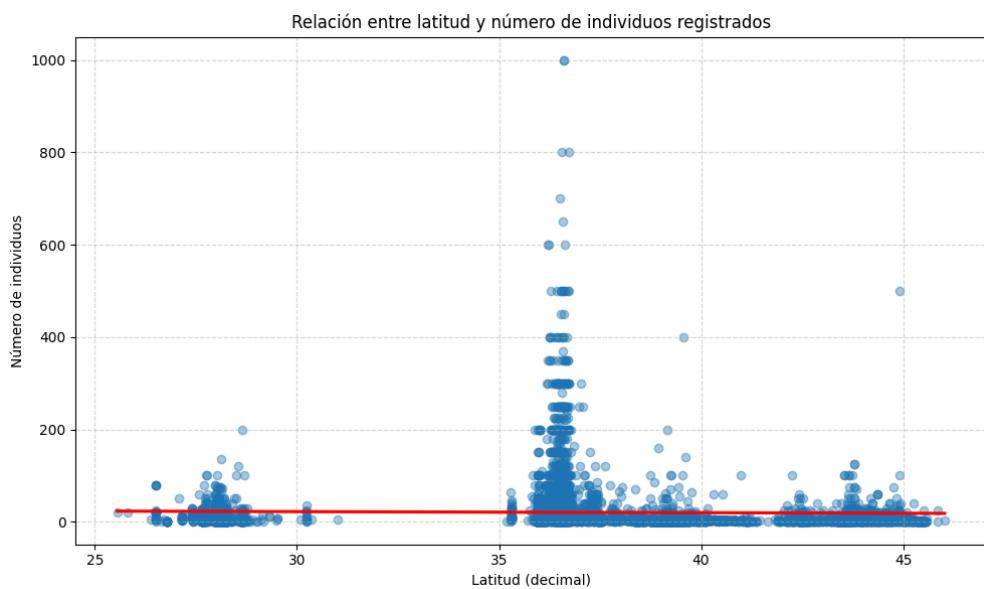


Figura 17. Relación entre la latitud decimal y la cantidad de individuos registrados por evento. La línea roja representa el ajuste lineal por mínimos cuadrados.

A pesar de la alta variabilidad observada a lo largo del eje de latitud, se detecta una **tendencia ligeramente decreciente** en el número de individuos conforme aumenta la latitud. Esta relación negativa, aunque débil, **resultó estadísticamente significativa** en los tests aplicados ($p < 0.001$), lo que sugiere que existen diferencias en la abundancia promedio a lo largo del gradiente latitudinal.

Esta distribución podría estar influenciada por múltiples factores, como el **tipo de especies predominantes en cada región**, el **esfuerzo de muestreo** o las **condiciones oceanográficas** propias de cada zona (p. ej., afloramientos en Canarias o la complejidad batimétrica del Mar de Alborán). No obstante, es importante interpretar esta relación con cautela, dado que la mayoría de los registros concentran valores bajos de `individualCount` y existen outliers marcados (p. ej., eventos con cientos de individuos).

5.2.4 Distribución conjunta de especies y comunidades autónomas

El siguiente gráfico muestra una **matriz de frecuencias** entre las 10 especies más observadas y las 12 comunidades autónomas con más registros (incluyendo la categoría "no asignado").

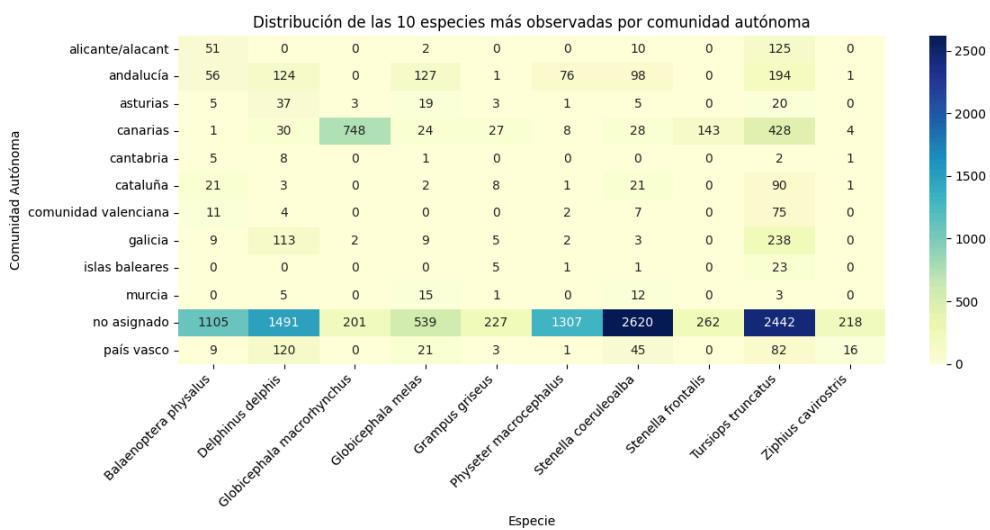


Figura 17. Mapa de calor que representa la frecuencia de registros de las 10 especies más comunes por comunidad autónoma. La escala de color refleja el número de observaciones (cuenta absoluta).

Esta visualización permite identificar asociaciones geográficas claras entre especies y zonas de observación. Destaca la presencia masiva de registros sin comunidad autónoma asignada, que afecta a todas las especies y refleja una **limitación importante del dataset**. Aun así, pueden observarse patrones como:

- **Globicephala macrorhynchus** aparece concentrado en **Canarias**, en línea con su distribución tropical-subtropical.
- **Stenella coeruleoalba** y **Delphinus delphis** presentan un amplio rango, con registros destacados en Andalucía, Galicia y la categoría no asignada.
- La distribución es más homogénea en especies costeras como **Tursiops truncatus**, observada en múltiples comunidades.

Este cruce ilustra bien la variabilidad regional en los datos y subraya la importancia de mejorar la georreferenciación administrativa para futuros análisis espaciales más precisos.

Nota sobre los tests estadísticos

Para complementar el análisis bivariante, se aplicaron **pruebas estadísticas**:

Relación	Tipo	Test aplicado	Resultado
comunidad_autonoma × individualcount	Categórica vs Numérica	Kruskal-Wallis	p < 0.0001 → diferencias significativas
scientificname × individualcount	Categórica vs Numérica	Kruskal-Wallis	p < 0.0001 → diferencias significativas
decimallatitude × individualcount	Numérica vs Numérica	Correlación de Spearman	p ≈ 0.0000 rho = -0.23 → correlación débil negativa
comunidad_autonoma × nombre_científico	Categórica vs Categórica	Chi-cuadrado	p < 0.0001 → asociación significativa

5.3 Análisis multivariante: correlaciones entre variables numéricas

5.3.1 Correlación entre variables (Spearman)

En conjunto, las correlaciones cuantitativas son en su mayoría bajas o débiles, lo que indica que no hay dependencias lineales fuertes entre las variables consideradas. Esto refuerza la utilidad de otros enfoques, como los análisis categóricos y espaciales, para comprender la dinámica de los registros.

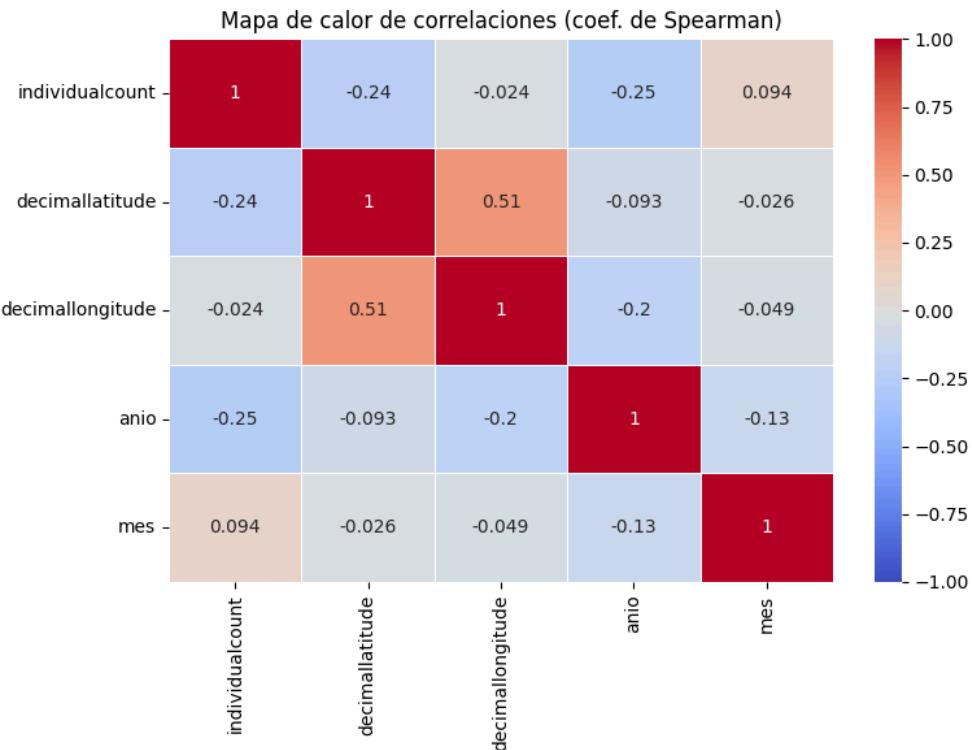


Figura 18. Mapa de calor de correlaciones entre variables numéricas, calculadas mediante el coeficiente de Spearman. Los valores van de -1 (correlación negativa perfecta) a 1 (positiva perfecta), siendo 0 indicativo de ausencia de asociación.

- La única correlación moderada destacable es la observada entre `decimalLatitude` y `decimalLongitude` ($p = 0.51$), lo que refleja una posible **relación geográfica entre ambas variables**, coherente con la distribución espacial de los registros.
- Se detecta una **débil correlación negativa entre el número de individuos registrados (`individualCount`) y el año (`anio`)**, lo que sugiere una leve tendencia a observar grupos más pequeños en años recientes.
- También se observa una **correlación negativa baja entre `individualCount` y la latitud**, lo cual podría reflejar patrones ecológicos o sesgos en la cobertura geográfica de los muestreos.
- La variable `mes` muestra correlaciones prácticamente nulas con el resto, a excepción de un leve valor positivo con `individualCount`, lo cual podría estar relacionado con la estacionalidad del esfuerzo de observación.

5.3.2 Distribución espacial por especie (mapa multivariado)

Para cerrar la exploración multivariante, se construyó un mapa de dispersión que integra simultáneamente tres variables clave: latitud, longitud y número de individuos registrados (`individualCount`), desglosadas por especie. Este gráfico permite visualizar cómo se distribuyen espacialmente las distintas especies de cetáceos en aguas españolas, así como su abundancia relativa por observación.

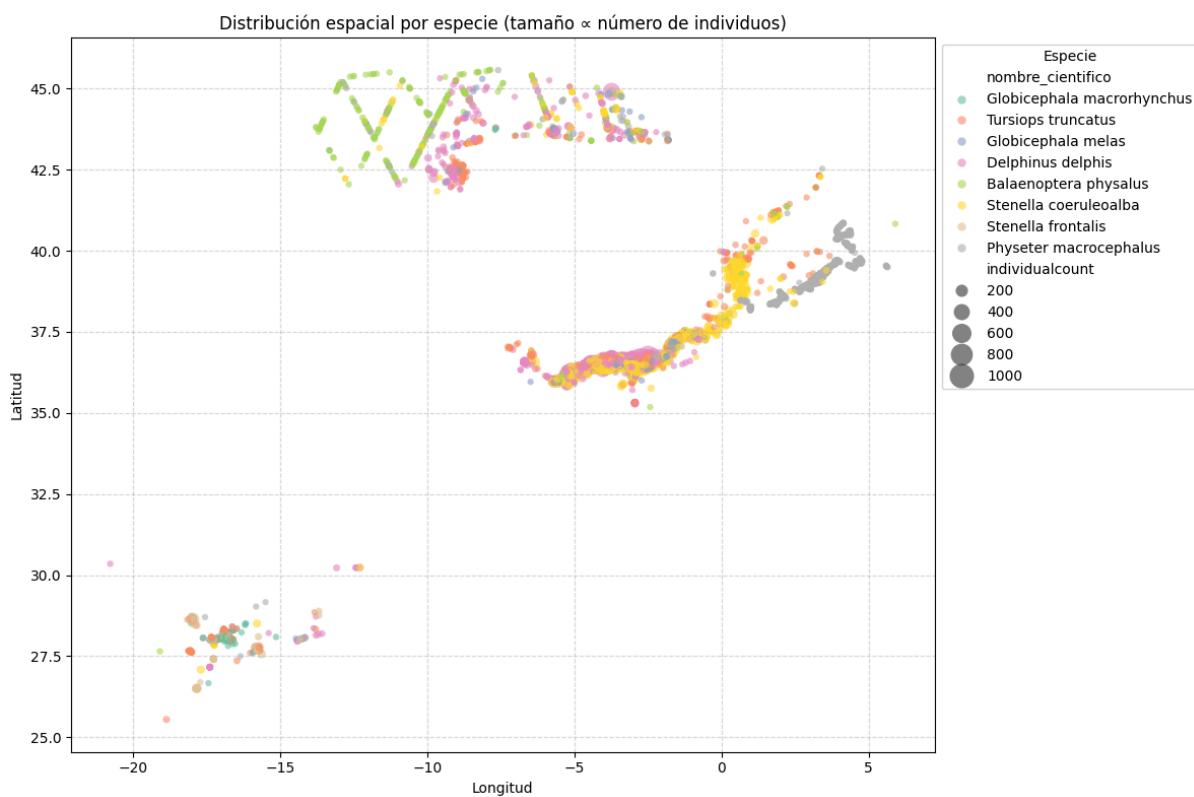


Figura 19. Mapa multivariado de los registros de cetáceos entre 2000 y 2024. Cada punto representa un registro georreferenciado, coloreado por especie y con tamaño proporcional al número de individuos observados. Se muestran las diez especies más frecuentes en el conjunto de datos.

- Se identifican tres núcleos espaciales principales de actividad: la **cornisa cantábrica**, el **mar de Alborán** y las **Islas Canarias**, lo que coincide con los patrones descritos en análisis anteriores.
- **Stenella coeruleoalba** y **Delphinus delphis**, además de ser las especies más abundantes, presentan una distribución espacial muy amplia, con registros en todas las zonas, especialmente en la fachada atlántica y mar de Alborán.
- Algunas especies muestran un **patrón espacial más restringido**: por ejemplo, *Globicephala macrorhynchus* se concentra casi exclusivamente en Canarias, mientras que *Grampus griseus* o *Physeter macrocephalus* parecen presentar agrupamientos más localizados.

6. Comparativa entre especies

6.1 Introducción

La comparación entre *Delphinus delphis* (delfín común) y *Stenella coeruleoalba* (delfín listado) permite profundizar en cómo se diferencian —y solapan— las distribuciones espacio-temporales de dos de los cetáceos más frecuentemente registrados en aguas españolas. Aunque comparten rasgos ecológicos (ambos son delfínidos gregarios y ampliamente distribuidos), sus preferencias de hábitat y patrones de registro difieren notablemente.

Esta sección integra información espacial, estacional y anual para comprender mejor las dinámicas de ambas especies entre 2000 y 2024.

6.2 Evolución temporal de registros

Ambas especies mostraron una alta frecuencia de registros en los primeros años del siglo, especialmente entre 2000 y 2008. Desde entonces, se observa una caída abrupta y un posterior repunte leve y oscilante. En los últimos años, los registros son más estables, aunque en menor cantidad que en la primera década.

Esto podría reflejar cambios en el esfuerzo de muestreo, la digitalización de datos históricos o el auge de la ciencia ciudadana a partir de 2010, más que variaciones reales en las poblaciones. *D. delphis* presenta una mayor regularidad en los años recientes, mientras que *S. coeruleoalba* muestra más picos aislados.

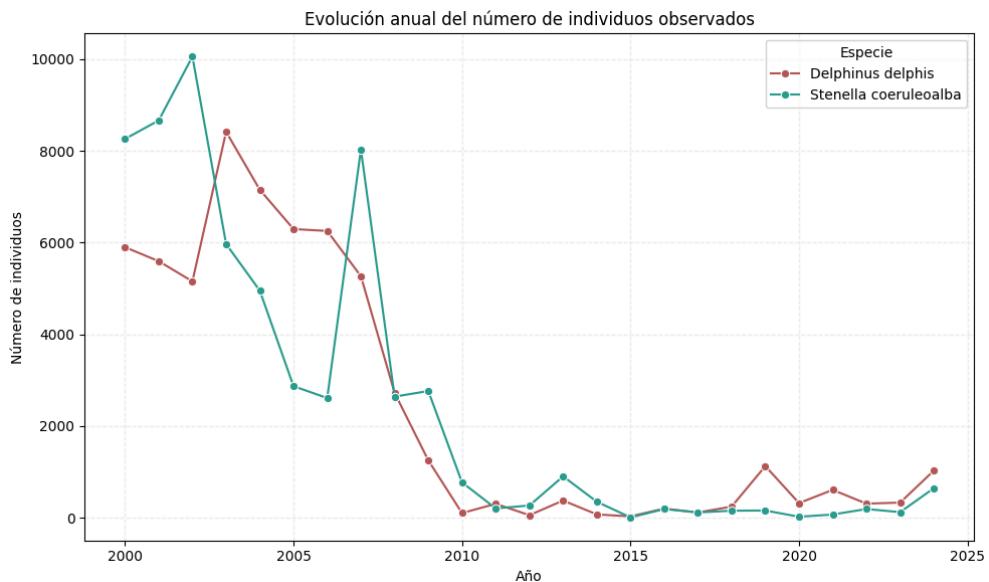


Figura 20. Mapa de distribución geográfica de los registros de *Stenella coeruleoalba* (verde) y *Delphinus delphis* (rojo) entre 2000 y 2024. Se observa una alta superposición espacial en el Mediterráneo occidental y el Cantábrico, con algunas diferencias en Canarias.

6.3 Distribución geográfica

El análisis conjunto de distribución espacial revela un patrón de solapamiento en el Mediterráneo occidental (especialmente mar de Alborán) y archipiélago canario. Sin embargo, existen diferencias clave:

- *D. delphis* tiene una distribución más amplia en la fachada atlántica, con presencia notable en Galicia y mar Cantábrico.
- *S. coeruleoalba* se concentra más en aguas del Mediterráneo, en zonas como Cataluña, Comunidad Valenciana y Baleares.

Estas diferencias responden a sus preferencias ecológicas: el delfín listado es más pelágico y vinculado a aguas profundas, mientras que el delfín común utiliza tanto plataformas continentales como aguas costeras.

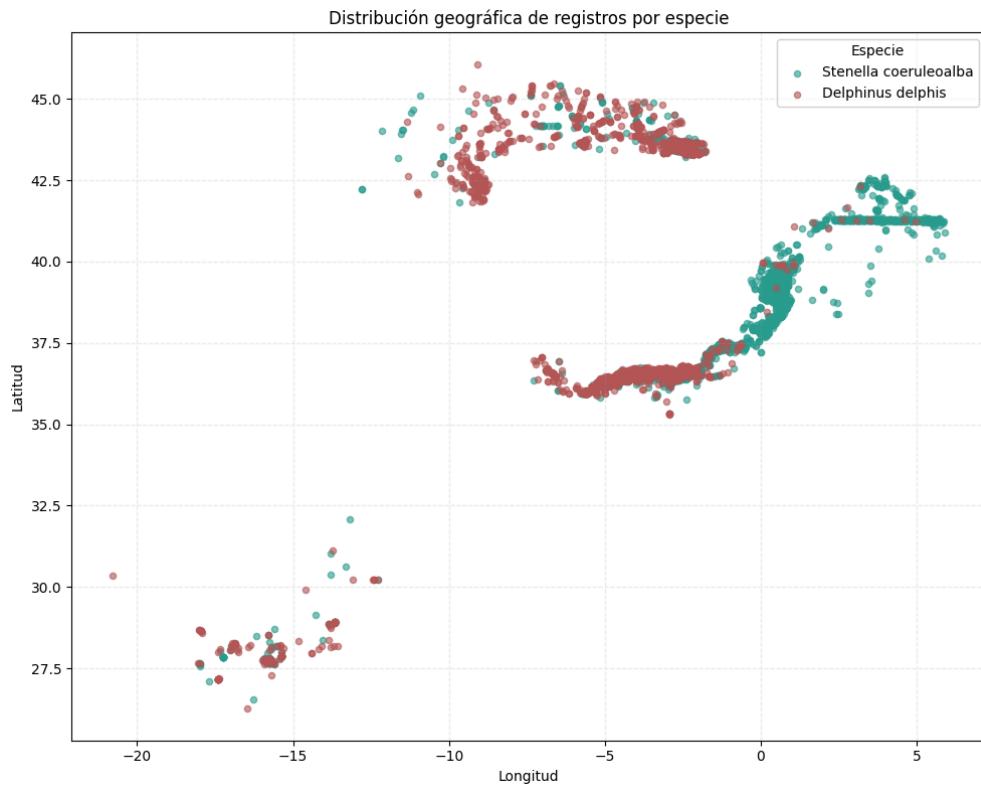


Figura 21. Distribución geográfica de las dos especies objetivo.

6.4 Estacionalidad

Ambas especies muestran una fuerte concentración de registros entre junio y septiembre. Este patrón estacional se repite anualmente, con picos más marcados en julio y agosto. Es probable que este aumento responda a un mayor esfuerzo de observación durante el verano (cruceros científicos, avistamientos turísticos), más que a una verdadera estacionalidad en su presencia.

No obstante, se detectan ligeras diferencias:

- *D. delphis* alcanza su máximo en julio, con un descenso progresivo posterior.
- *S. coeruleoalba* muestra un segundo pico en septiembre, lo que podría sugerir una variación en patrones de uso del hábitat o migración intraespecífica.

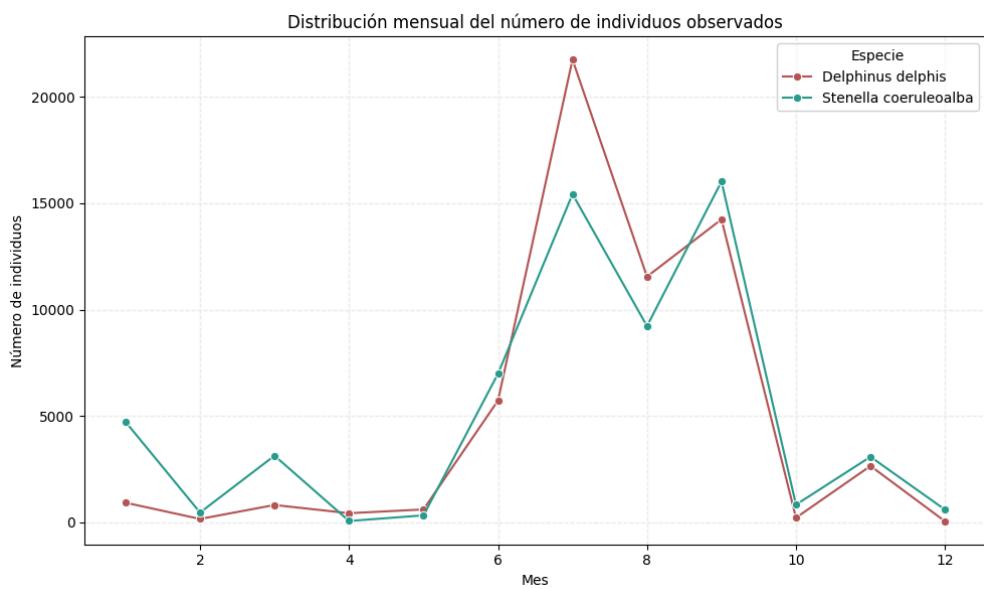


Figura 22. Distribución mensual del número de individuos observados para ambas especies. Ambas presentan un claro pico de registros durante los meses de verano (julio-agosto).

6.5 Comparación entre especies por períodos

Este bloque muestra de forma paralela cómo varía la distribución espacial de *Stenella coeruleoalba* y *Delphinus delphis* dentro de cada uno de los cuatro períodos definidos (2000–2005, 2006–2011, 2012–2017, 2018–2024). En cada figura se presentan dos paneles —uno por especie—, permitiendo detectar con mayor claridad posibles diferencias o solapamientos geográficos asociados al momento temporal específico.

Períodos 2000-2005 y 2006-2011

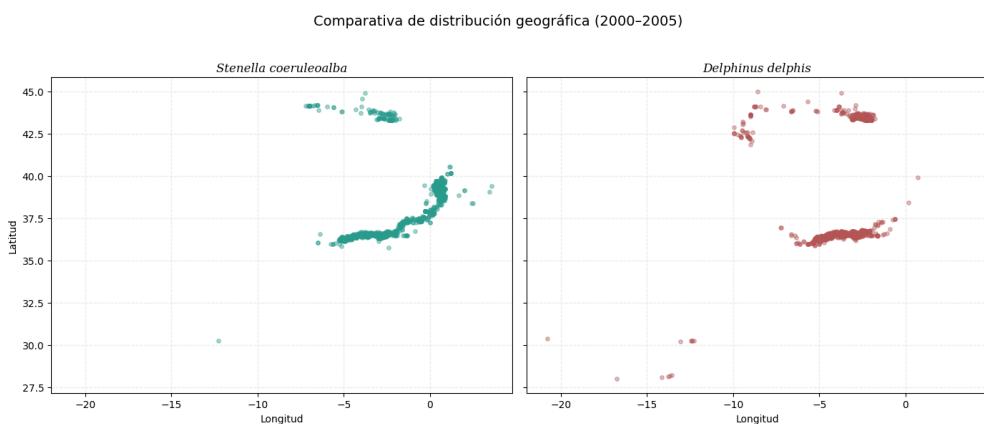


Figura 23. Comparativa de distribución geográfica (2000–2005) entre *Stenella coeruleoalba* y *Delphinus delphis*.

Comparativa de distribución geográfica (2006–2011)

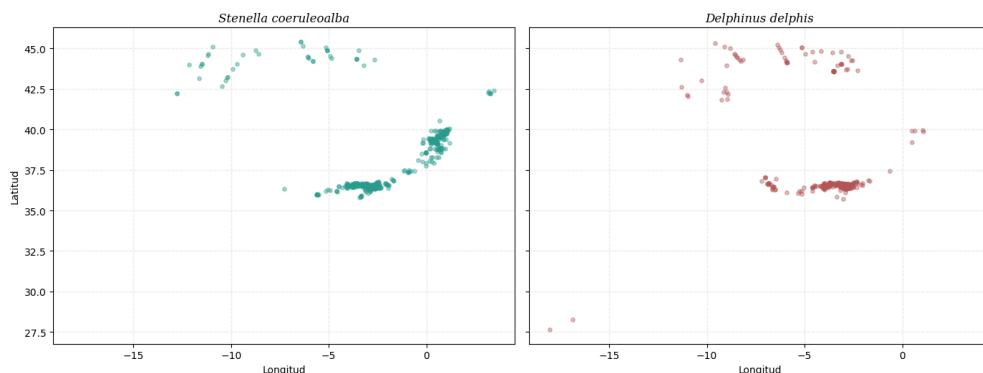


Figura 24. Comparativa de distribución geográfica (2006–2012) entre *Stenella coeruleoalba* y *Delphinus delphis*.

-Delfín listado

2000–2005 y 2006–2011: La especie muestra una marcada afinidad por el Mediterráneo, con una alta densidad de registros a lo largo de Cataluña, Baleares y el Levante español. Esta distribución se corresponde con su preferencia por aguas más profundas y alejadas de la costa, así como con su carácter más oceánico.

Delfín común

2000–2005 y 2006–2011: El núcleo de observaciones se concentra en el Mediterráneo occidental, especialmente en el mar de Alborán, la costa levantina y Cataluña, así como en algunas áreas del Golfo de Cádiz. Esta distribución está en línea con la preferencia del delfín común por aguas costeras y zonas de transición batimétrica con alta productividad.

Periodo 2012–2017

Comparativa de distribución geográfica (2012–2017)

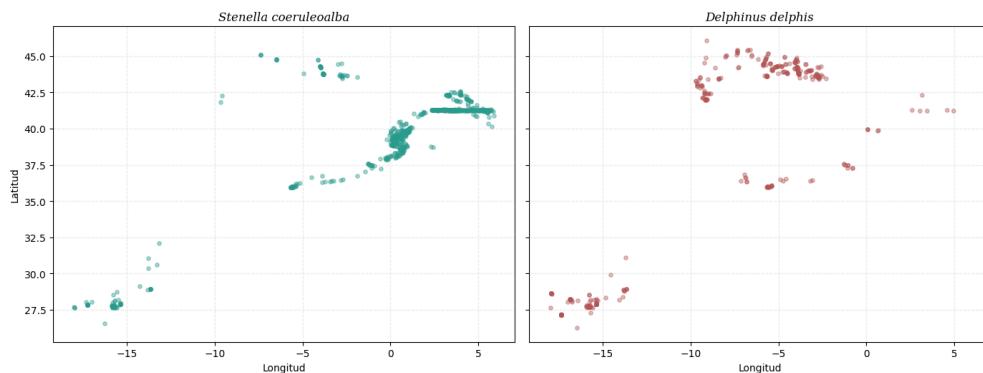


Figura 25. Comparativa de distribución geográfica (2012–2017) entre *Stenella coeruleoalba* y *Delphinus delphis*.

Delfín listado

2012–2017: Se mantiene la dominancia en el Mediterráneo, pero con una expansión visible hacia el sur, con nuevos registros en el mar de Alborán y las Islas Canarias, donde su presencia era escasa o inexistente en períodos previos. Esto puede sugerir una flexibilidad ecológica en respuesta a cambios térmicos o corrientes.

Delfín común

2012–2017: Se observa una expansión evidente hacia el Atlántico. Aumentan los registros en la costa gallega y el Cantábrico, así como en el archipiélago canario, con una presencia especialmente notoria en el norte de Tenerife y Gran Canaria. Este patrón sugiere una mayor cobertura de muestreo o una redistribución estacional potencialmente vinculada a factores oceanográficos o cambios en la disponibilidad trófica.

Periodo 2018-2024

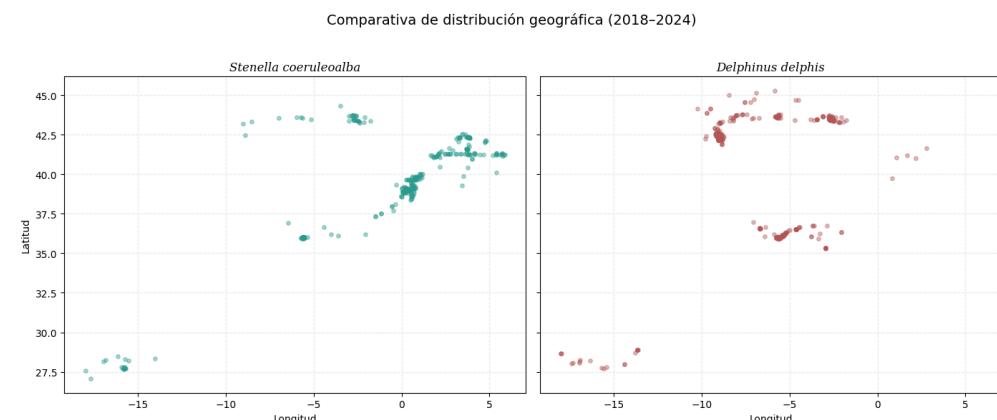


Figura 26. Comparativa de distribución geográfica (2018–2024) entre *Stenella coeruleoalba* y *Delphinus delphis*.

Delfín listado

2018–2024: La especie mantiene una distribución amplia y consistente, consolidando su presencia tanto en el Mediterráneo como en zonas insulares atlánticas. A diferencia de *D. delphis*, sus registros en el norte peninsular siguen siendo menos frecuentes, lo que refuerza la hipótesis de una diferenciación ecológica en el uso del espacio marino.

Delfín común

2018–2024: La especie mantiene su amplia distribución, pero con menos concentración en el Mediterráneo oriental, y una presencia consolidada en el noroeste peninsular y Canarias. Esta tendencia podría reflejar una redistribución latitudinal moderada o la estabilización de poblaciones en zonas previamente menos representadas.

6.6 Síntesis comparativa

- Aunque ambas especies muestran **solapamiento espacial** en áreas clave como el mar de Alborán o el archipiélago canario, la evolución temporal sugiere que *D. delphis* ha tenido una mayor expansión hacia el Atlántico norte y aguas costeras occidentales, mientras que *S. coeruleoalba* ha mantenido una preferencia más marcada por zonas de influencia mediterránea y aguas oceánicas más profundas.
- Estas diferencias podrían estar mediadas por estrategias de forrajeo diferenciadas, distribución del esfuerzo de muestreo, o incluso por cambios en la productividad primaria y circulación marina derivados del cambio climático, como el calentamiento de las aguas superficiales o modificaciones en el sistema de afloramiento.
- Resulta interesante que la expansión espacial de ambas especies hacia áreas nuevas sea más evidente a partir de 2012, lo que podría coincidir con una mayor integración de datos de ciencia ciudadana, aumento del esfuerzo de monitoreo, o una mayor movilidad estacional de las poblaciones.

6.7 Interpretación general de los cambios espacio-temporales en la distribución

El análisis de distribución por períodos revela **patrones espaciales distintivos y cambios relevantes a lo largo del tiempo** para ambas especies, que permiten plantear hipótesis sobre sus preferencias ecológicas y posibles respuestas a presiones ambientales.

En el caso de *Stenella coeruleoalba* (delfín listado), se observa una **fuerte presencia en el Mediterráneo occidental**, particularmente en la franja costera de **Cataluña, las Islas Baleares y el Levante español**. Aunque esta especie se considera típicamente **oceánica** y asociada a **aguas profundas**, esta aparente contradicción se resuelve si consideramos la **naturaleza batimétrica del Mediterráneo español**, donde la **plataforma continental es muy estrecha y da paso rápidamente a fondos de gran profundidad**. Zonas como la **fosa de Valencia**, los

cañones submarinos del Cabo de Palos o el Golfo de León permiten que *S. coeruleoalba* acceda a hábitats pelágicos sin alejarse significativamente de la costa.

Sin embargo, resulta especialmente llamativo el **descenso en el número de registros del listado en los últimos seis años**, lo cual puede interpretarse de varias formas, no necesariamente excluyentes:

- Podría reflejar un **declive real de la población** o un **cambio en su distribución**, motivado por factores como la **disminución de presas pelágicas**, el **aumento del ruido submarino**, o **cambios térmicos** en la columna de agua.
- También es plausible que se deba a una **redistribución hacia zonas menos accesibles para los programas de monitoreo actuales**, especialmente si están cambiando las rutas de migración, los patrones de residencia o la disponibilidad trófica.
- No se puede descartar un **cambio metodológico o en el esfuerzo de muestreo**, ya que la creciente participación ciudadana en campañas de avistamiento no garantiza un monitoreo homogéneo entre especies ni regiones.

De forma complementaria, es interesante observar que *Delphinus delphis* muestra un **aumento sostenido de registros en la costa norte peninsular**, especialmente en Galicia y el Cantábrico. Esta tendencia podría estar indicando un **desplazamiento latitudinal progresivo**, posiblemente vinculado a **cambios térmicos en el Atlántico**, una **redistribución de presas pelágicas** hacia aguas más frías. Esta observación invita a nuevas preguntas:

- ¿Está el delfín común colonizando nuevas áreas por desplazamiento climático?
- ¿Existe una mayor productividad en el Cantábrico que favorezca su presencia?
- ¿Qué papel juega la conectividad entre poblaciones atlánticas y mediterráneas en estos movimientos?

Estas preguntas podrían orientar estudios futuros centrados en el seguimiento de movimientos poblacionales en escenarios de cambio global.

En conjunto, los resultados apuntan a la necesidad de **integrar datos ecológicos, oceanográficos y de esfuerzo de muestreo** para interpretar con precisión estos cambios. La comparación espacio-temporal permite generar preguntas clave:

- ¿Está el calentamiento del Mediterráneo desplazando a *Stenella coeruleoalba* hacia otras zonas?
- ¿Mantiene *Delphinus delphis* su estatus de especie generalista con alta plasticidad ecológica?
- ¿Cómo pueden influir las actividades humanas (pesca, turismo, tráfico marítimo) en estas tendencias?

Este tipo de análisis exploratorio constituye un primer paso fundamental hacia la comprensión de **dinámicas poblacionales complejas**, y refuerza la importancia de disponer de **datos geográficamente representativos y temporalmente consistentes** para la conservación efectiva de los cetáceos.

7. Conclusiones y próximos pasos

7.1 Evaluación de las hipótesis de trabajo

A lo largo de este análisis comparativo se han contrastado las hipótesis planteadas inicialmente, con los siguientes resultados:

Hipótesis contrastadas

H1: *Delphinus delphis* es la especie con mayor número de individuos observados.

 **Refutada:** *Stenella coeruleoalba* registra más individuos.

H2: Galicia, Andalucía y Canarias concentran los registros con mayor abundancia y diversidad.

- ✓ **Parcialmente confirmada:** Andalucía y Canarias destacan, pero el área de mayor concentración fue el mar de Alborán y la cuenca mediterránea occidental.

H3: *Delphinus delphis* ha modificado su distribución espacial, desplazándose hacia el norte peninsular o el Mediterráneo.

- ✓ **Parcialmente confirmada:** Se observa un aumento sostenido en Galicia y el Cantábrico, lo que sugiere un posible desplazamiento latitudinal.

7.2 Conclusiones generales

- *Stenella coeruleoalba* se consolida como la especie más registrada en cuanto a individuos observados, con una marcada presencia en el Mediterráneo occidental.
- *Delphinus delphis* mantiene una distribución más costera y dispersa, destacando en el Golfo de Cádiz, el mar de Alborán y progresivamente en el norte peninsular.
- Se identifican patrones espacio-temporales diferenciados entre ambas especies, tanto en la distribución general como en su evolución durante las últimas dos décadas.
- Las diferencias observadas parecen responder a una combinación de factores ecológicos (batimetría, distribución de presas, comportamiento social), oceanográficos (temperatura, productividad) y metodológicos (esfuerzo de muestreo y ciencia ciudadana).
- El análisis evidencia el valor de los datos abiertos y la visualización como herramienta para detectar tendencias, generar hipótesis y apoyar la conservación marina.

7.3 Próximos pasos y oportunidades de mejora

Este análisis exploratorio ofrece una base sólida para futuros estudios, pero también revela limitaciones inherentes a los datos utilizados. A partir de ello, se proponen los siguientes pasos:

- **Integrar variables oceanográficas:** cruzar los registros con datos de temperatura superficial del mar (SST), salinidad o concentración de clorofila podría ayudar a explicar patrones de presencia y cambios en la distribución.
- **Incluir capas batimétricas y geomorfológicas:** para relacionar presencia de especies con gradientes de profundidad o estructuras submarinas (cañones, taludes).
- **Incorporar el calendario de campañas científicas:** para analizar la estacionalidad del esfuerzo de muestreo y minimizar sesgos en la interpretación.
- **Cruzar con eventos climáticos extremos** (ej. olas de calor marinas, anomalías de El Niño): identificar posibles impactos sobre la distribución y abundancia de cetáceos.
- **Ampliar el enfoque multiespecífico:** explorar patrones de coexistencia, solapamiento o reemplazo entre especies, especialmente en áreas críticas como el estrecho de Gibraltar o las islas Canarias.
- **Evaluar dinámicas poblacionales:** si se dispone de datos temporales más robustos o metodologías de marcaje, estudiar la fidelidad al sitio, movimientos migratorios y posibles cambios poblacionales.
- **Fomentar la estandarización y correcta introducción de datos:** una parte importante de los registros presentaba problemas en los campos administrativos (como las localidades de registro), lo que habría limitado seriamente un análisis regional más detallado. Esto pone de relieve la importancia de contar con datos bien estructurados y validados para maximizar su valor analítico y de apoyo a la toma de decisiones.

Este tipo de análisis, replicable y basado en datos abiertos, constituye una herramienta clave para la **gestión adaptativa de la biodiversidad marina**, así como para el diseño de estrategias de conservación más informadas,

proactivas y basadas en evidencia.

8. Sobre la autora

¡Hola! 

Mi nombre es Ceci y soy bióloga marina especializada en análisis de datos. Tras varios años dedicada a la investigación científica, actualmente estoy en proceso de transición hacia el ámbito de la ciencia de datos aplicada a la conservación.

Este análisis surge de mi interés por los cetáceos y por explorar cómo los datos abiertos pueden ayudarnos a entender mejor la biodiversidad marina. Combina mis conocimientos en ecología con nuevas herramientas de visualización, programación y análisis exploratorio adquiridas en el bootcamp de Data Science con The Bridge.

Mi objetivo es aplicar este enfoque para contribuir a la toma de decisiones informadas en gestión ambiental, monitoreo de fauna y conservación marina, tanto desde entidades científicas como desde organizaciones dedicadas a la sostenibilidad.