



Informe análisis Ovnis

1906-2014

Integrantes

- Iraitz Garmendia Alberdi (estudiante analítica de datos Upgrade-Hub)
- Jaime Vecina Montesinos (estudiante analítica de datos Upgrade-Hub)
- Javier Muñoz Ballesteros (estudiante analítica de datos Upgrade-Hub)

Presentación y objetivos

El análisis realizado se ha llevado a cabo sobre un conjunto de datos de NUFORC el cuál contenía, de manera inicial, más de 80.000 filas y 16 columnas con diferentes tipos de datos y relativos a cosas como coordenadas, época del año, fechas, etc.

El objetivo que se ha perseguido en nuestro proyecto ha consistido en ver los diferentes tipos de Ovnis, avistamientos por época del año, número de avistamiento por años y saber cuándo sería más fácil un avistamiento en base a la época, mes, día y hora.

Preparación y preprocesamiento de los datos

Una vez hemos descargado nuestros datos, procedemos a visualizar el tipo de datos, datos faltantes y su distribución para evaluar qué hacer con ellos

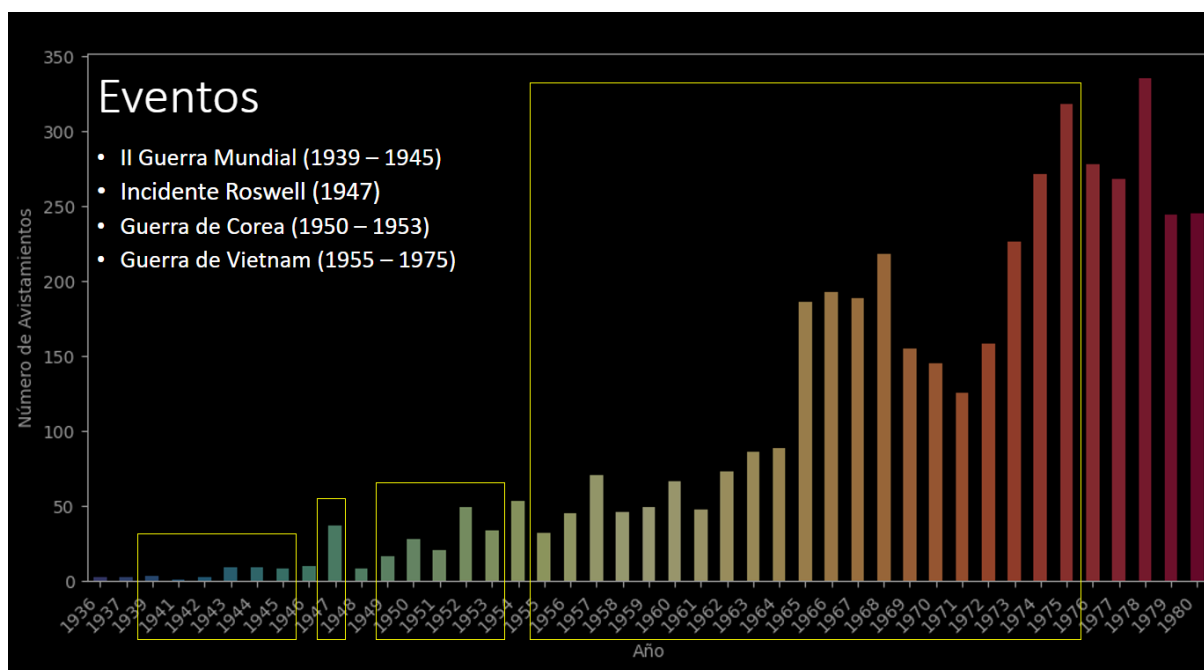
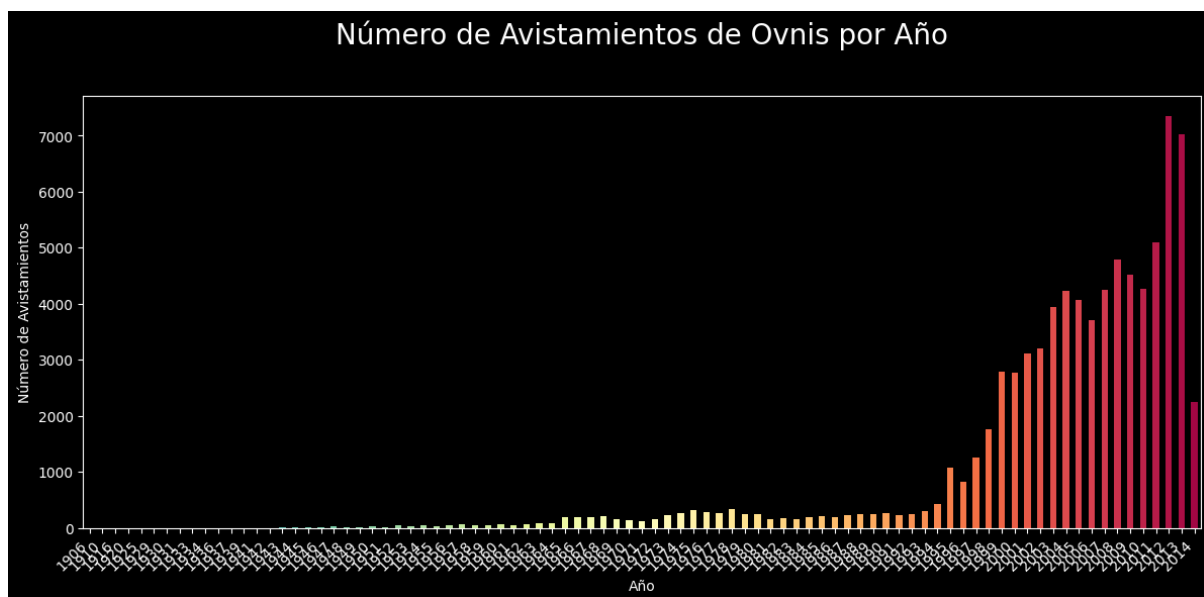
Una vez realizado lo anterior, observamos como los valores nulos que tenemos dentro de nuestro conjunto no son muchos e intentamos reparar la gran mayoría y siendo otros eliminados al no ser un porcentaje significativo dentro de nuestro conjunto.

Además, creamos columnas en base a las necesidades que nos sugieren los datos y eliminamos aquellas que no se van a utilizar para tener los datos disponibles de la mejor manera posible.

Análisis exploratorio

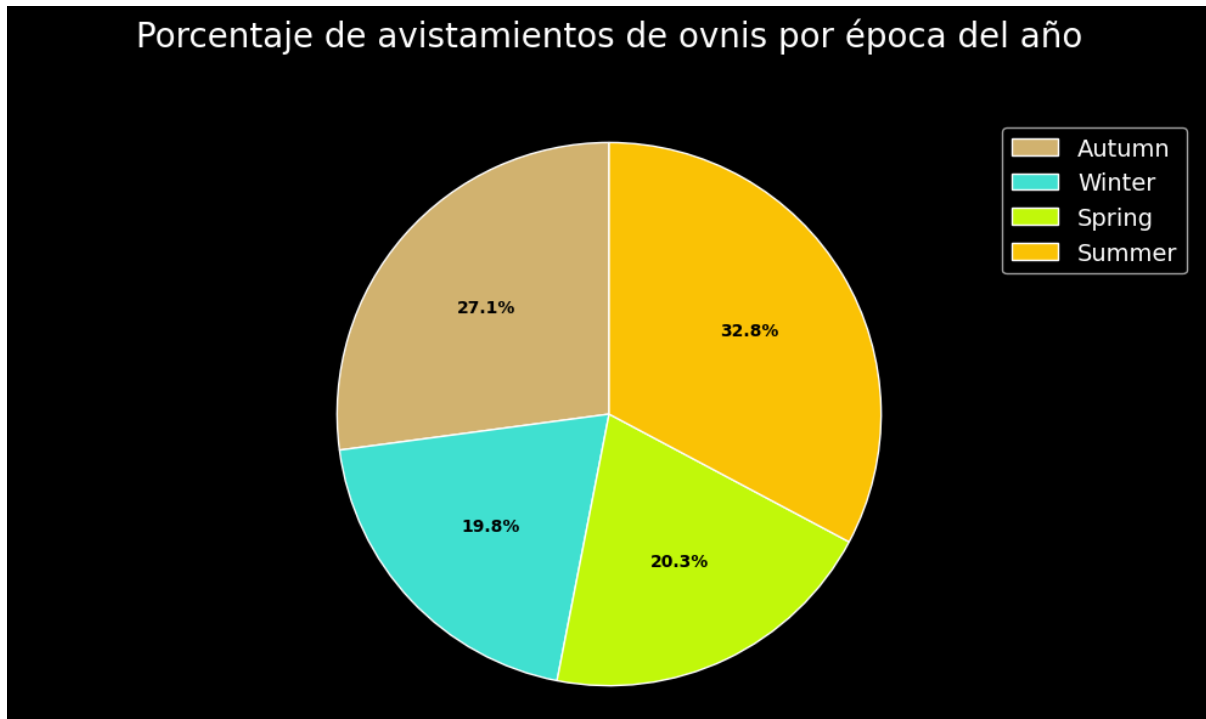
Evaluación de avistamientos por años

Como podemos observar en la gráfica, los avistamientos han ido en ascenso a lo largo de los años, con alguna ligera variación dependiendo de la época y los eventos acontecidos a lo largo de los años.



Avistamiento por estación del año

En la siguiente gráfica podemos ver como la mejor época para el avistamiento de ovnis es el verano, seguido de cerca por el otoño. Podemos deducir que es debido a que se disfruta de un cielo más claro y con menos nubes.

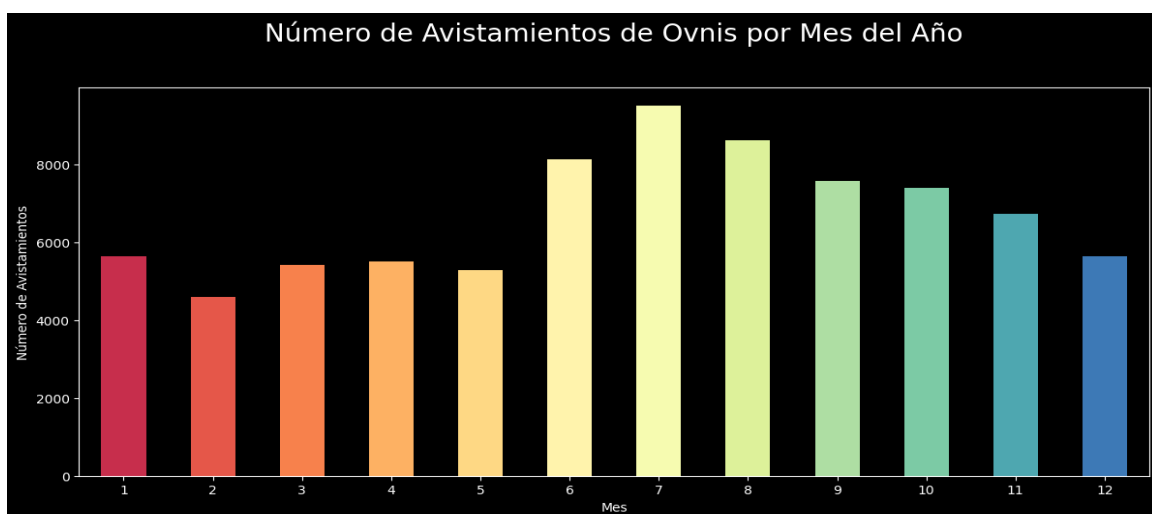


Avistamiento por mes, día y hora

En este apartado vamos a mostrar los mejores momentos para que se den este tipo de fenómenos en base a datos relativos al tiempo.

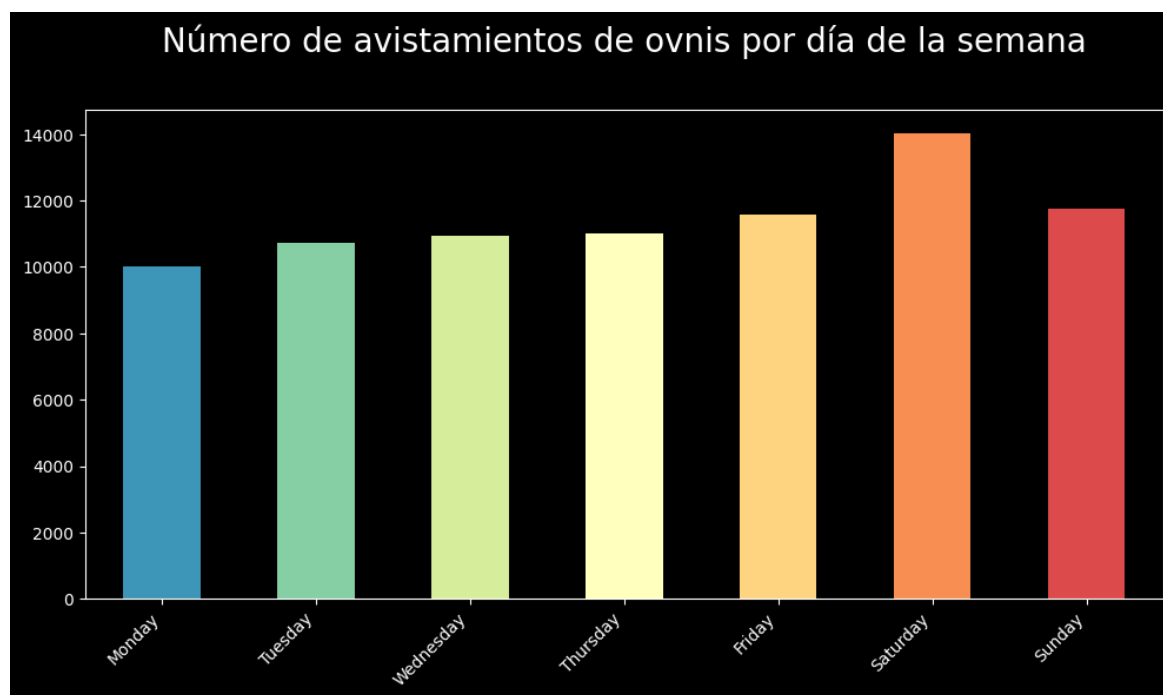
Mes

Según muestra nuestra gráfica, y como ya hemos visto anteriormente con las estaciones, los mejores meses para poder observar este tipo de fenómenos son los de verano, junio, julio y agosto, siendo los de primavera los menos habituales.



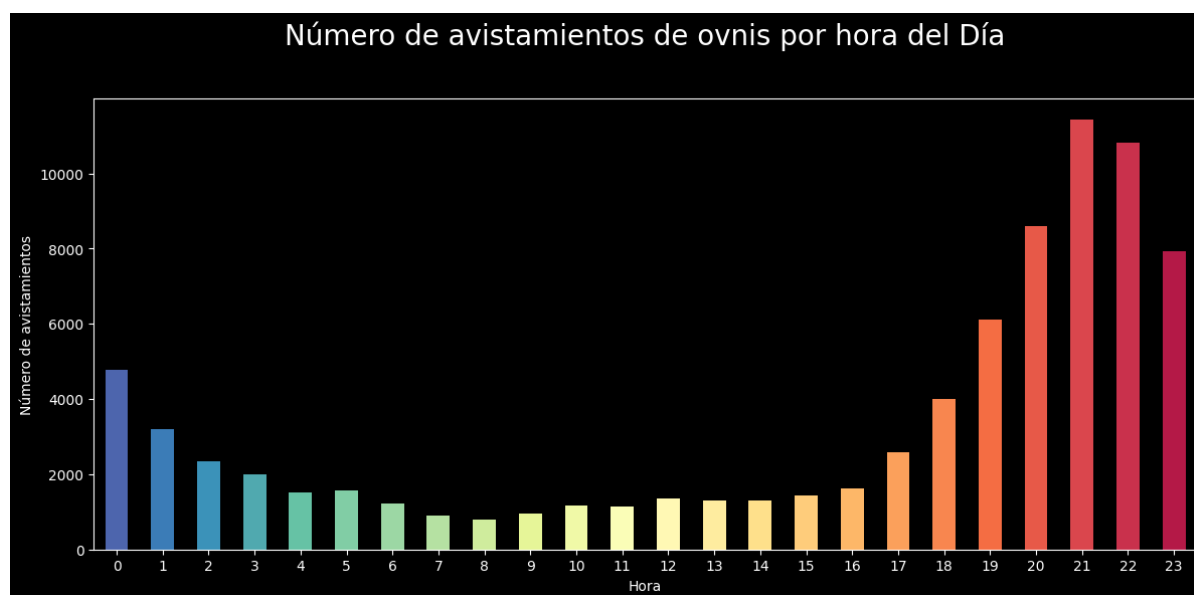
Día

El día que nos indican los datos que más avistamientos se han registrado son los correspondientes al fin de semana, viernes, sábado y domingo, siendo el sábado el que mayor número registra con diferencia.



Hora

Si ya analizamos las horas del día donde se ven con mayor frecuencia, vemos que las horas sufren una evolución, con alguna bajada sueva, desde las 10h hasta las 21h donde ya comienza una caída progresiva.

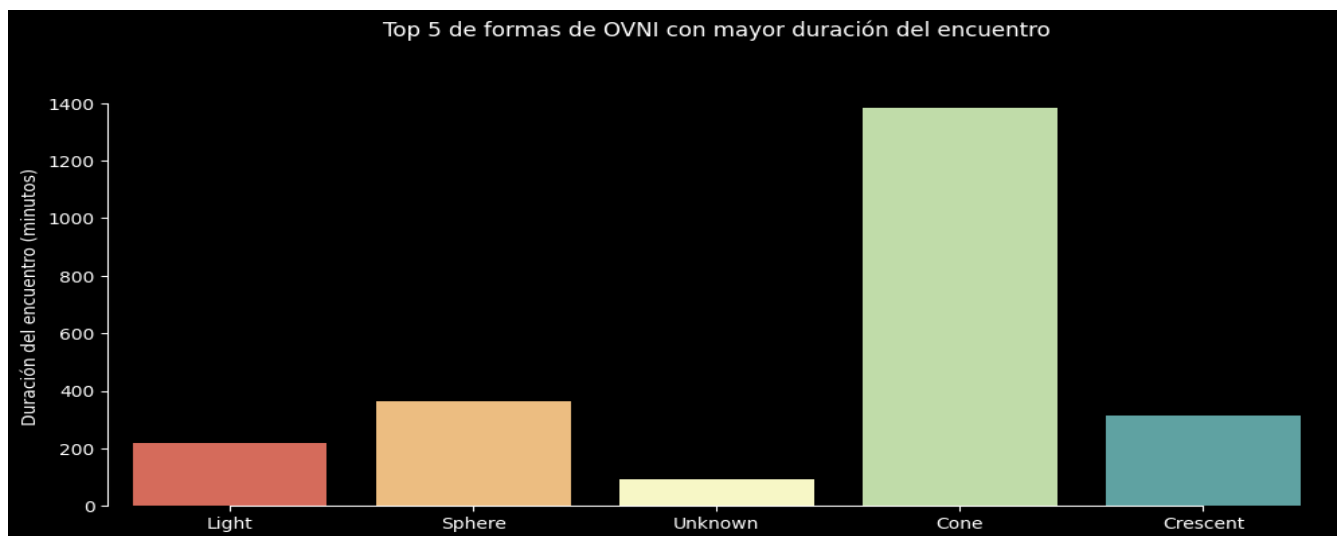


Tipo de ovni y su relación con el tiempo del encuentro

Vemos como en la distribución hay una gran diferencia en entre las luces (lo que algunos usuarios denominan como luces que se encienden y se apagan) con respecto al resto de fenómenos, siendo la forma triangular y circular las siguientes más habituales y que se llevan poca diferencia entre sí.



A continuación, tenemos el top 5 de las formas con respecto a la duración de los encuentros. Como se puede observar, la forma de cono predomina sobre el resto de las formas con mucha diferencia.



Avistamiento por países

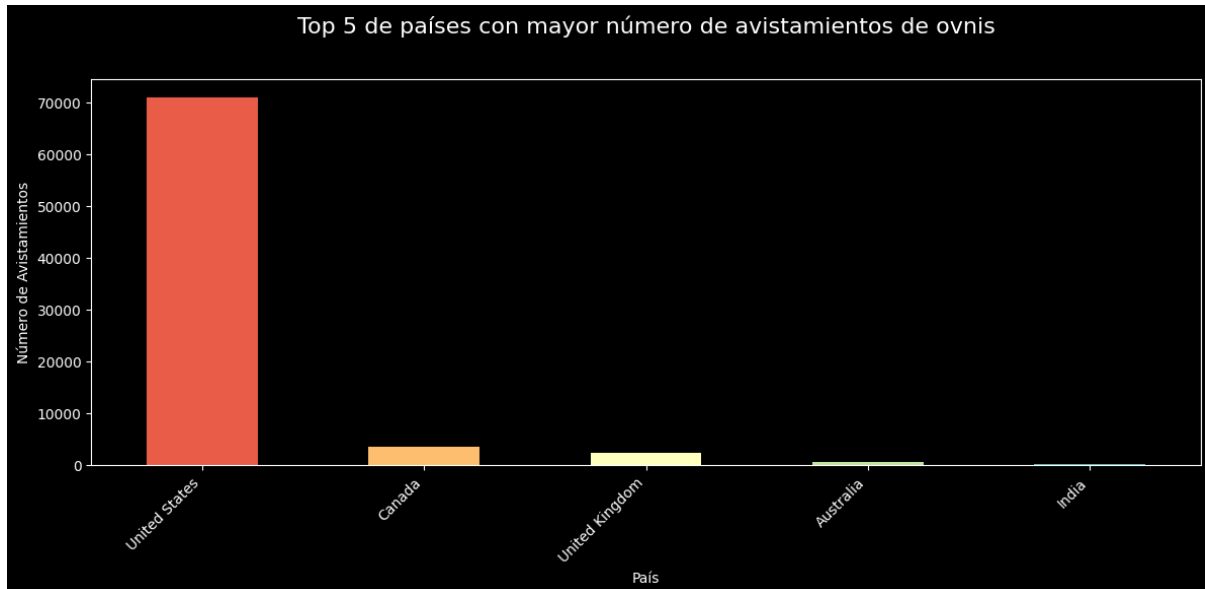
En este apartado vamos a ver como varían los datos de avistamientos en relación a los países en los que se han registrado estos fenómenos. Vamos a graficar el top 5 con EEUU entre ellos y posteriormente el top 15 con contar con EEUU debido a los porcentajes que nos arrojan los datos con respecto a países como España.

Porcentaje de avistamientos en Estados Unidos: 88.52%

Porcentaje de avistamientos en España: 0.095%

Top 5

Como comentábamos anteriormente, la diferencia con el siguiente es muy grande, por lo que se decidió separar y analizar sin esos datos.



Top 15 sin EEUU

Aquí ya tenemos una relación de los 15 países con mayor registro de avistamientos sin contar con Estados Unidos. Vemos como hay una gran diferencia entre los dos primeros con el resto, siendo también una diferencia muy amplia la que hay entre Canadá y Reino Unido.



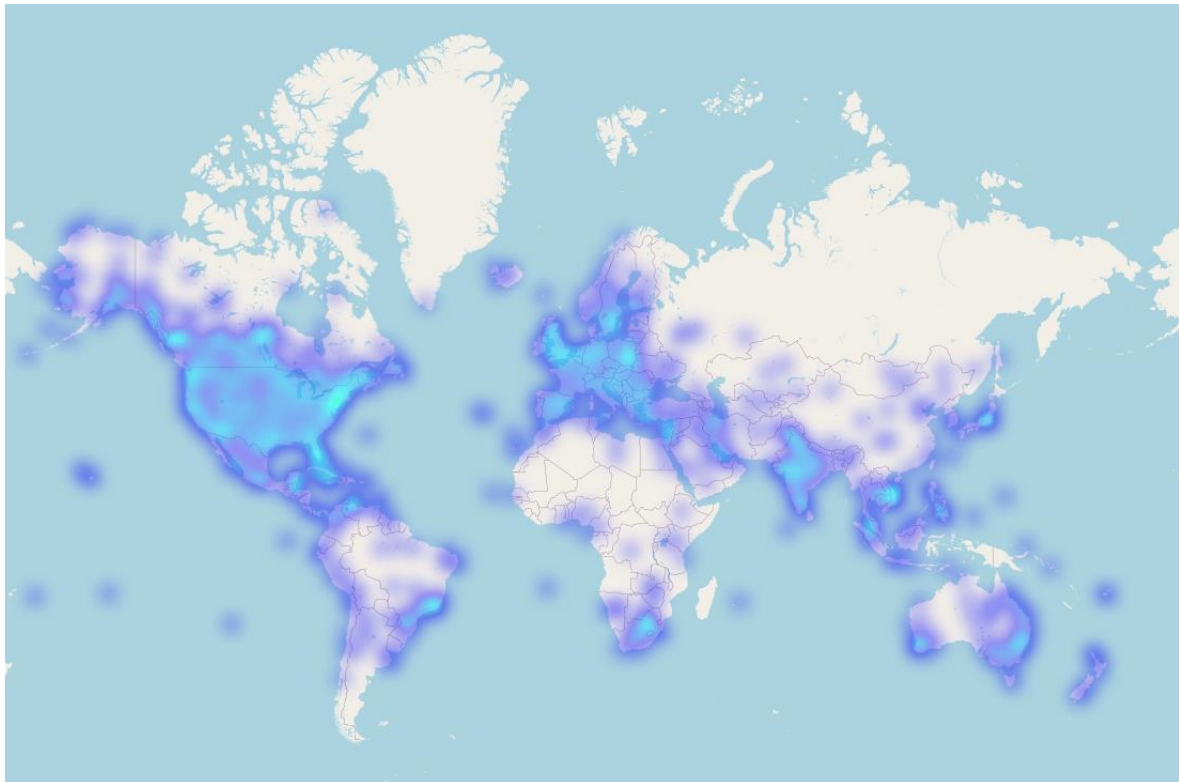
A la vista de ambas gráficas, y ya en la primera se puede observar, la mejor zona para el avistamiento de ovnis sería Norte América, debido a que es donde se encuentra el mayor número de registros hasta 2014, fecha hasta la que son nuestros datos.

Mapas

A continuación, vamos a ver la distribución que siguen basándonos en el uso de mapas de diferentes tipos y objetivos de los mismos.

Mapa de calor

En este mapa podemos observar cómo se distribuyen los avistamientos, llamando la atención como en todo Rusia se han reportado casi ningún avistamiento sobre estos fenómenos.



Avistamientos en Europa

En el siguiente mapa analizamos el número de avistamientos en Europa, con y sin Reino Unido, el cual hemos visto que despuntaba con respecto al resto de países.

Avistamientos en Europa

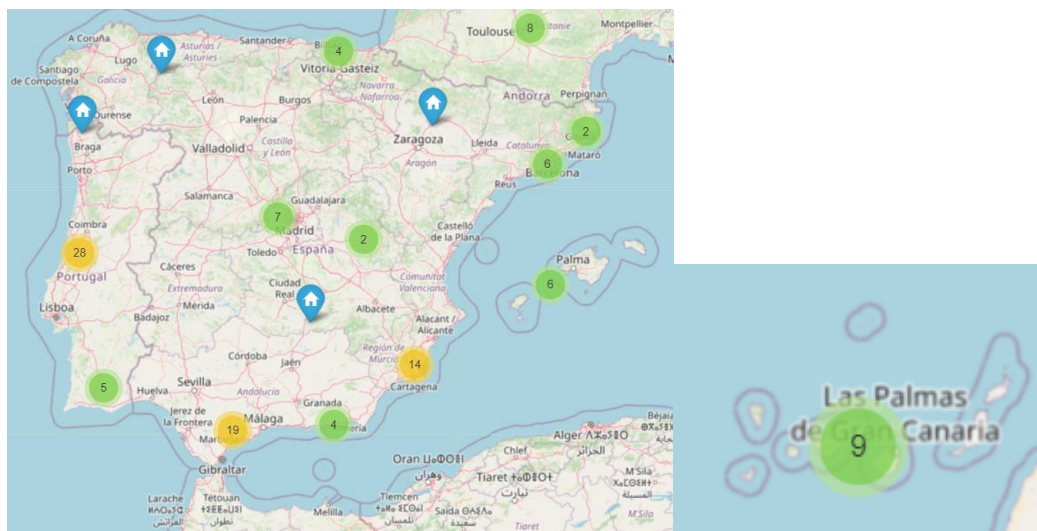


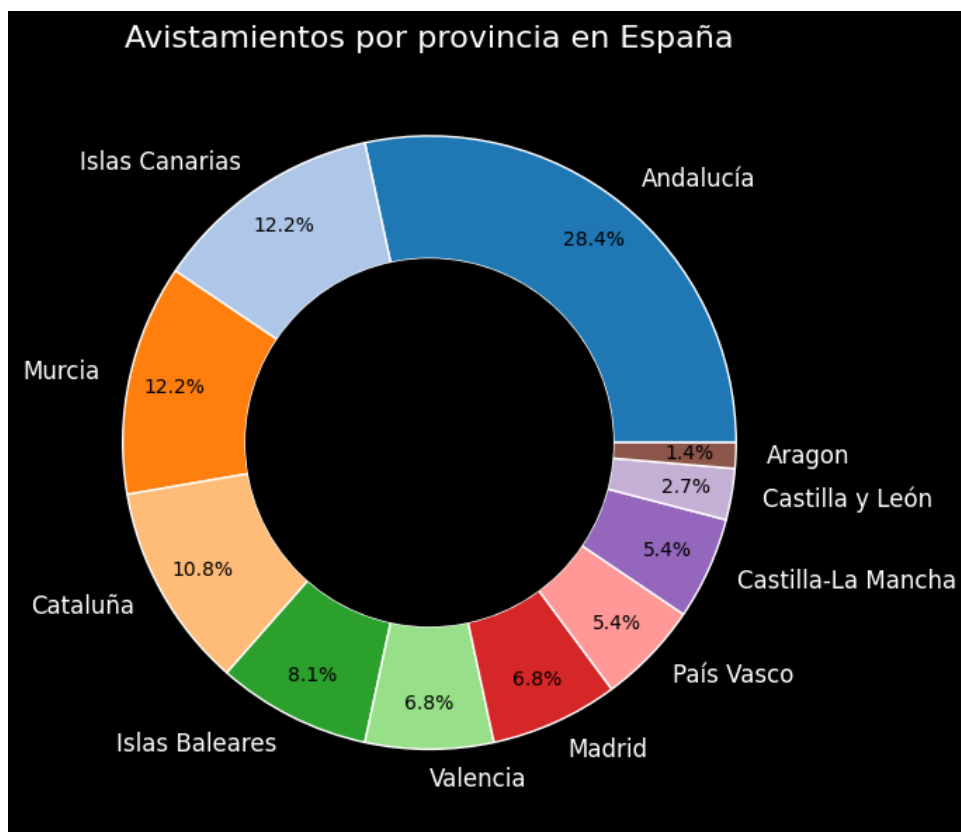
Avistamientos en Europa sin Reino Unido



Avistamientos en España

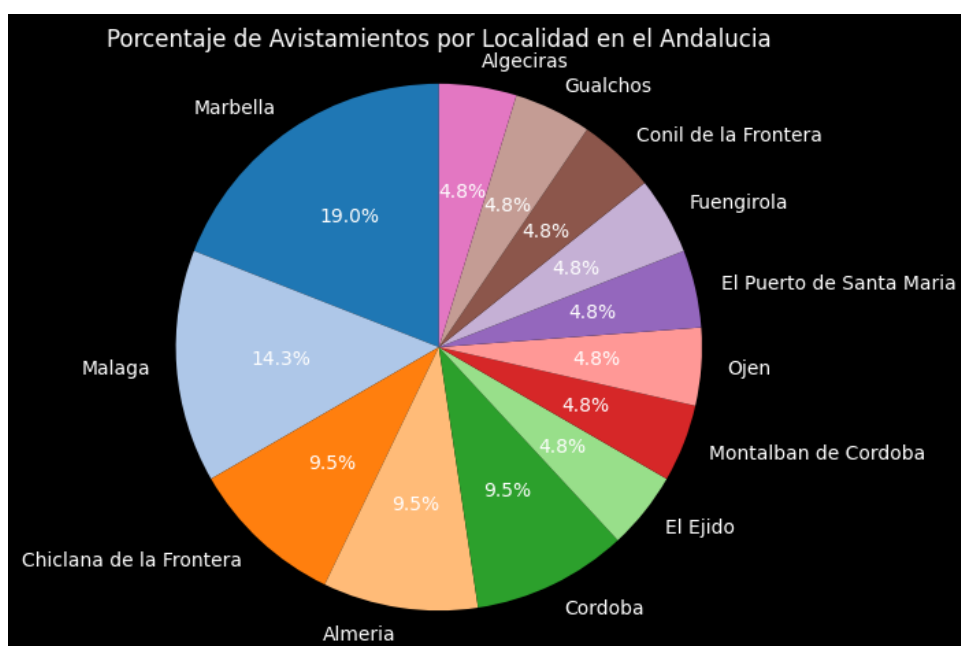
En nuestro mapa relacionado con toda la geografía española, se puede ver como la mayoría se registran por la zona de Andalucía, concretamente Málaga, seguida de la zona de Murcia y las islas Canarias.





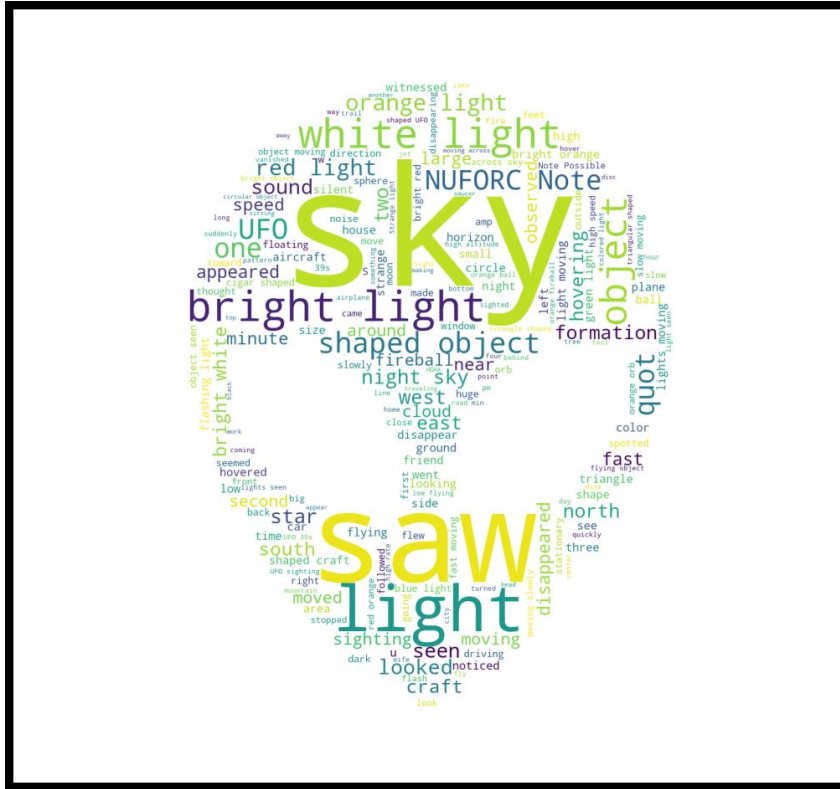
Avistamientos en Andalucía

Hemos querido profundizar y mostrar los datos referentes a Andalucía ya que es la zona con mayor número de avistamientos de toda la geografía española, aunque en nuestro análisis también hemos entrado a desglosar otras comunidades como Cataluña y la Comunidad de Madrid.



Análisis de las palabras más repetidas en las descripciones

Como se puede ver, las palabras más repetidas son sky, saw y light, que vendrían a corresponderse con lo analizado en el apartado del tipo de ovni más visto, ya que eran luces.



Información extra sobre el Área 51

La Área 51 se encuentra en la base de la cordillera Groom Lake, en el desierto de Mojave, al noroeste de Las Vegas, Nevada.

Fue establecida durante la Guerra Fría en la década de 1950 como una base de pruebas y entrenamiento para aeronaves experimentales y de reconocimiento. Su existencia se mantuvo en secreto durante décadas, y el gobierno de Estados Unidos no reconoció oficialmente la base hasta la década de 1990.

La falta de información oficial sobre las actividades de la Área 51 generó numerosas teorías de conspiración.

Algunas teorías sugieren que la base ha sido utilizada para estudiar tecnología alienígena y realizar experimentos con naves espaciales no identificadas (OVNIs). Aunque la Área 51 sigue siendo una base militar activa, su enfoque ha evolucionado con el tiempo.

Se cree que actualmente se utiliza para probar aeronaves y tecnologías clasificadas. El acceso a la Área 51 está altamente restringido. La base está rodeada por vallas y señales de advertencia, y la entrada está prohibida para el público en general.

Upgrade-hub > ovni > Jobs > UFO2

UFO2

☆

+ Create job (preview)

↺ Refresh

✕ Cancel

🗑 Delete

⚙ View options

Default

📊 Dashboard view

☰ Flat list of Jobs

🔍 Search

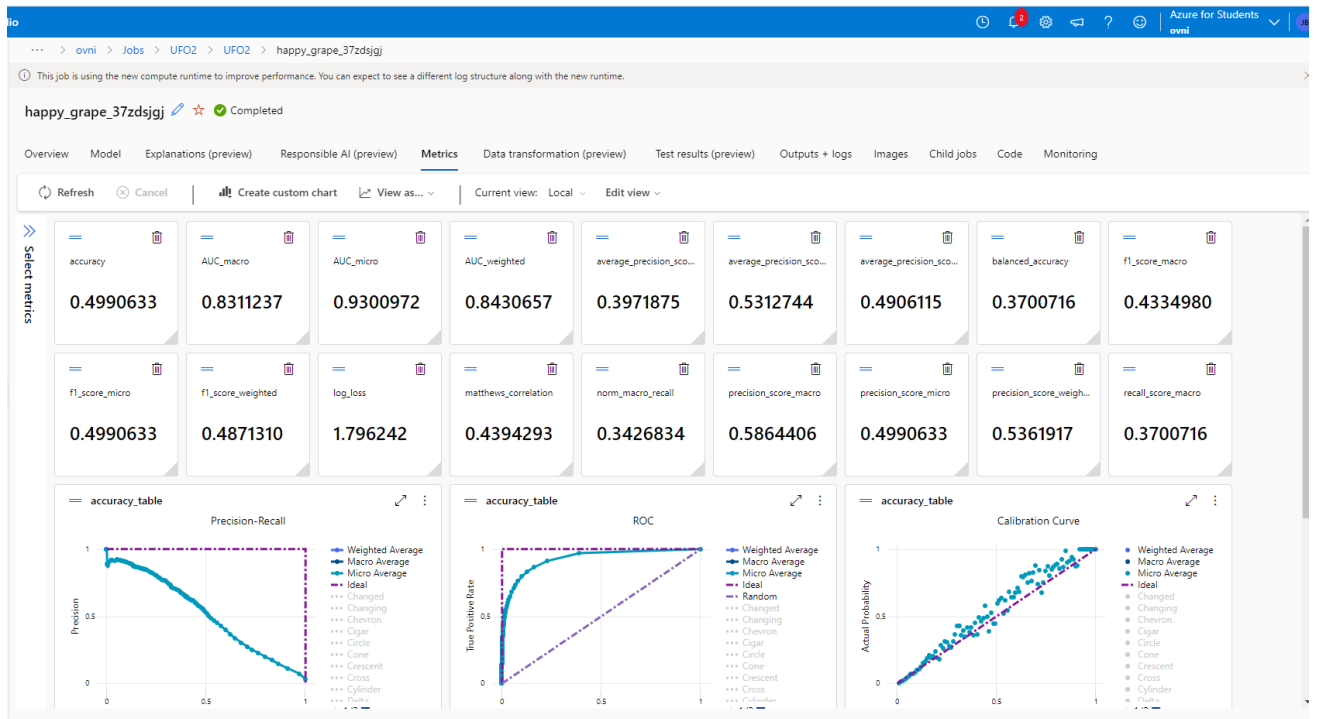
🔍 Only my jobs

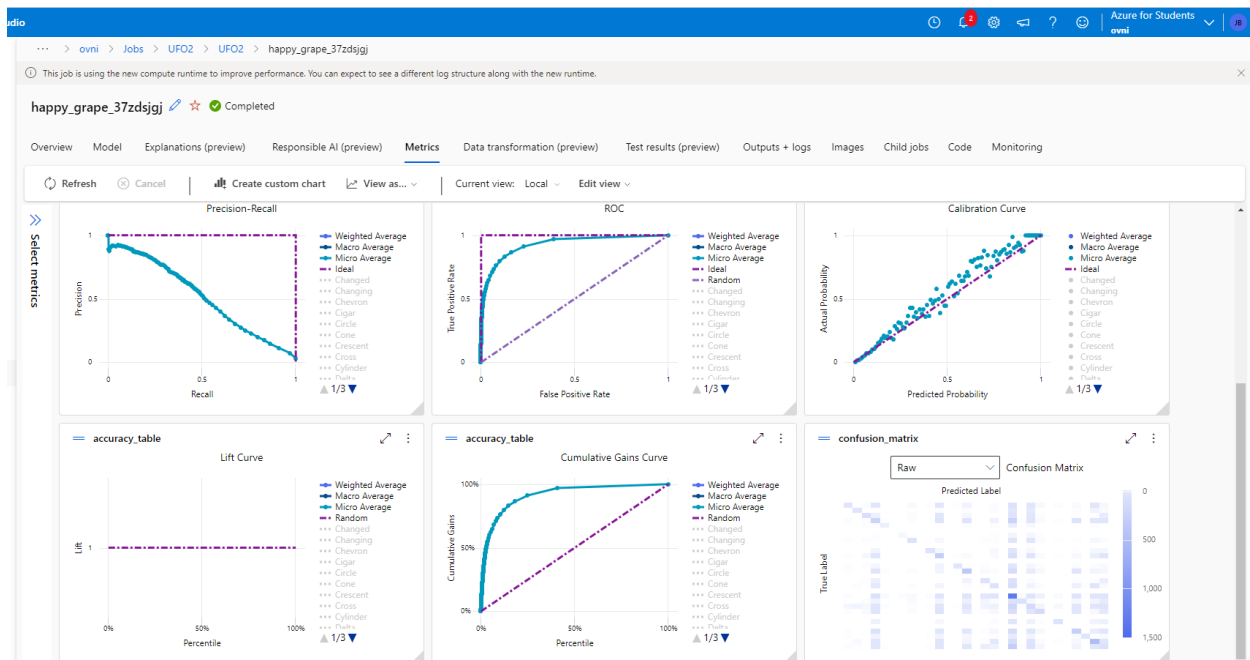
⌵ Filter

📄 Columns

🔍 Display name (1 visualized)	Status	Created on ↓	☆ Duration	Created by	Compute target	Job type	Tags
🔍 > UFO2 (38)	✅ Completed	Nov 22, 2023 6:05 PM	6h 17m 46s	Javier Muñoz Balle...	Serverless	Automat...	dynamic_allowlisting_iterations : ...

Upgrade-hub > omni > Jobs > UFO2 > UFO2							
UFO2 Completed							
Overview Data guardrails Models Outputs + logs Child jobs							
Refresh Deploy Download Explain model # View generated code View options							
<div> <input type="text" value="Search"/> <div>Filter Columns</div> </div>							
Algorithm name	Explained	Responsible AI	AUC weighted ↓	Sampling	Created on	Duration	Hyperparameter
VotingEnsemble		View responsible AI dashboard	0.84307	100.00 %	Nov 23, 2023 12:16 AM	6m 20s	algorithm : [XGBoostClassifier', ...
MaxAbsScaler, XGBoostClassifier			0.83499	100.00 %	Nov 22, 2023 6:12 PM	28m 23s	tree_method : auto ...
SparseNormalizer, XGBoostClassifier			0.83426	100.00 %	Nov 22, 2023 10:29 PM	12m 14s	booster : gbtree colsample_byt ...
StandardScalerWrapper, XGBoostClassifier			0.83175	100.00 %	Nov 22, 2023 6:12 PM	7m 57s	booster : gbtree colsample_byt ...
SparseNormalizer, XGBoostClassifier			0.82888	100.00 %	Nov 22, 2023 6:12 PM	8m 12s	booster : gbtree colsample_byt ...
SparseNormalizer, XGBoostClassifier			0.82808	100.00 %	Nov 22, 2023 6:12 PM	4m 47s	booster : gbtree colsample_byt ...
StandardScalerWrapper, XGBoostClassifier			0.82806	100.00 %	Nov 22, 2023 6:12 PM	4m 55s	booster : gbtree colsample_byt ...
StandardScalerWrapper, XGBoostClassifier			0.82787	100.00 %	Nov 22, 2023 9:19 PM	5m 11s	booster : gbtree colsample_byt ...
SparseNormalizer, XGBoostClassifier			0.82738	100.00 %	Nov 22, 2023 11:50 PM	3m 48s	booster : gbtree colsample_byt ...
StandardScalerWrapper, XGBoostClassifier			0.81684	100.00 %	Nov 22, 2023 6:12 PM	3m 21s	booster : gbtree colsample_byt ...
SparseNormalizer, LightGBM			0.81126	100.00 %	Nov 22, 2023 11:55 PM	10m 16s	boosting_type : gbdtr colsample ...
StandardScalerWrapper, RandomForest			0.79773	100.00 %	Nov 22, 2023 6:12 PM	12m 57s	



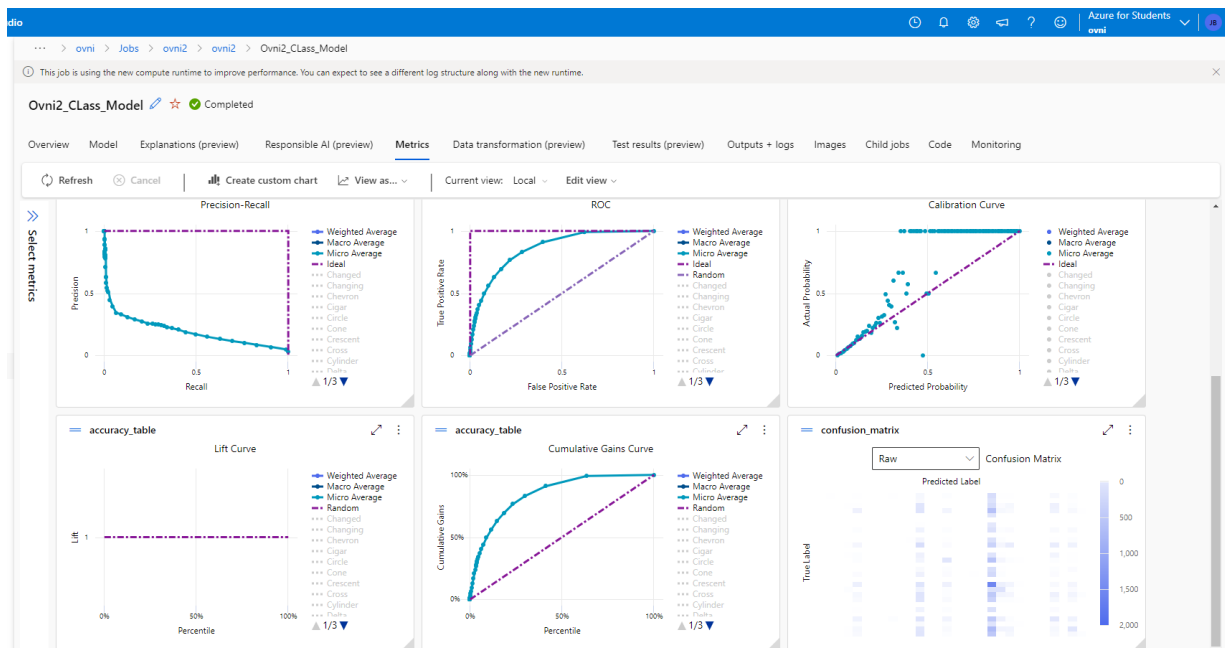


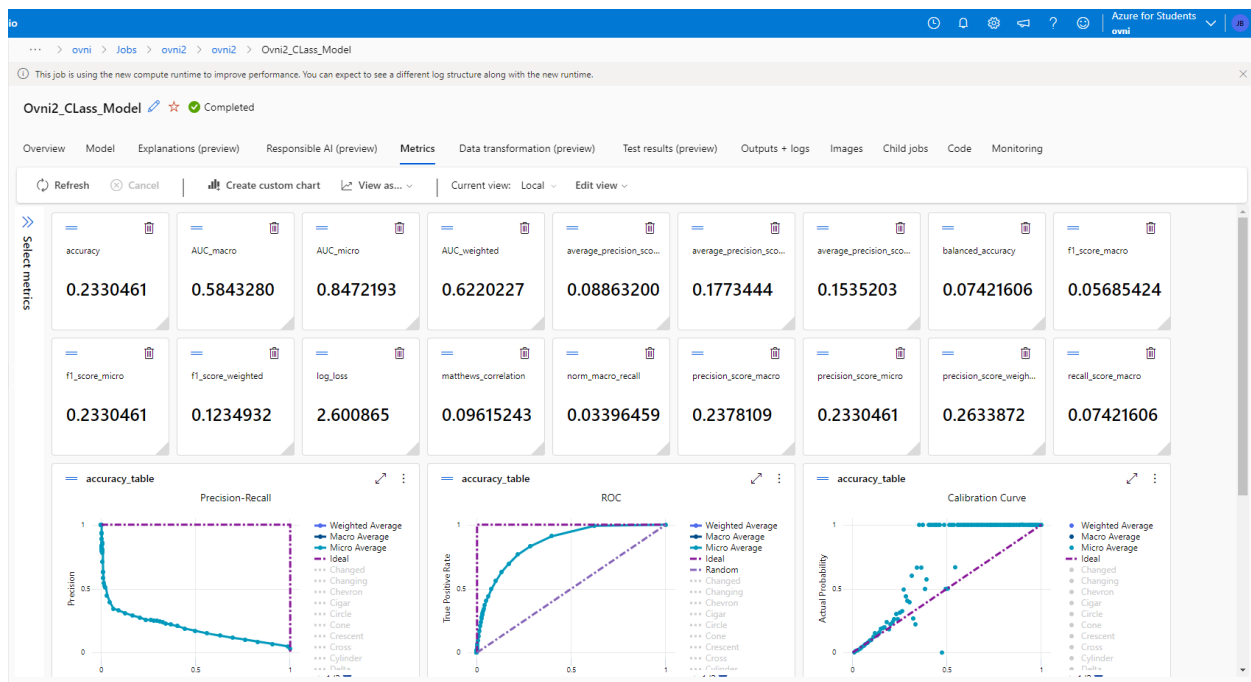
Segundo análisis con sólo algunas de nuestras columnas

Para este segundo intento, debido a los problemas que nos ocasionaba la descarga, se realizó con sólo 10 columnas en lugar de las 16 que se utilizaron en el anterior. Como podemos ver, también nos sacó que el mejor modelo se corresponde con un XGBoost Classifier y la duración del mismo fue mucho menor, aproximadamente 1/3.

Upgrade-hub > ovni > Jobs > ovni2									
ovni2 ☆									
<div> <div>Create job (preview)</div> <div>Refresh</div> <div>Cancel</div> <div>Delete</div> <div>View options</div> <div>Default</div> <div>Dashboard view</div> <div>Flat list of Jobs</div> </div>									
<div> <div>Search</div> <div>Only my jobs</div> <div>Filter</div> <div>Columns</div> </div>									
Display name (1 visualized)	Status	Created on ↓	☆	Duration	Created by	Compute target	Job type	Tag	
ovni2 (47)	Completed	Nov 23, 2023 6:44 PM		2h 16m 32s	Javier Muñoz Balle...	Serverless	Automat...		

Algorithm name	Explained	Responsible AI	AUC weighted ↓	Sampling	Created on	Duration	Hyperparameter
VotingEnsemble	View explanation		0.62202	100.00 %	Nov 23, 2023 8:56 PM	4m 47s	algorithm : [XGBoostClassifier, 1, ...
StandardScalerWrapper, XGBoostClassifier			0.60872	100.00 %	Nov 23, 2023 7:24 PM	1m 29s	booster : gbtree colsample_by ...
StandardScalerWrapper, LightGBM			0.60523	100.00 %	Nov 23, 2023 7:51 PM	1m 34s	boosting_type : gbdtr colsample ...
MaxAbsScaler, LightGBM			0.60454	100.00 %	Nov 23, 2023 6:53 PM	1m 11s	boosting_type : goss colsample ...
SparseNormalizer, XGBoostClassifier			0.60248	100.00 %	Nov 23, 2023 8:10 PM	1m 23s	booster : gbtree colsample_by ...
StandardScalerWrapper, XGBoostClassifier			0.60206	100.00 %	Nov 23, 2023 7:23 PM	50s	booster : gbtree colsample_by ...
MaxAbsScaler, LightGBM			0.60162	100.00 %	Nov 23, 2023 7:26 PM	1m 3s	boosting_type : gbdtr colsample ...
MaxAbsScaler, XGBoostClassifier			0.60148	100.00 %	Nov 23, 2023 6:52 PM	1m 47s	tree_method : auto
SparseNormalizer, XGBoostClassifier			0.60144	100.00 %	Nov 23, 2023 6:53 PM	1m 4s	booster : gbtree colsample_by ...
StandardScalerWrapper, XGBoostClassifier			0.60141	100.00 %	Nov 23, 2023 6:53 PM	1m 5s	booster : gbtree colsample_by ...
MaxAbsScaler, LightGBM			0.60112	100.00 %	Nov 23, 2023 8:49 PM	48s	boosting_type : goss colsample ...
StandardScalerWrapper, LightGBM			0.60038	100.00 %	Nov 23, 2023 8:55 PM	52s	boosting_type : qbdtr colsample ...





Streamlit

En el proyecto hemos utilizado la biblioteca Streamlit creando un archivo “.py” para la visualización interactiva. Se presentan imágenes y gráficos para hacer el análisis más accesible, y se invita a los usuarios a participar mediante la encuesta. Dentro del archivo, el desglose realizado sería este:

Secciones Principales:

- **Introducción:**
 - Presentación del equipo y elección del tema (OVNIs).
 - Uso de datos de NUFORC, con más de 80k registros desde 1906 hasta 2014.
- **Contexto:**
 - Exploración de datos interesantes y contexto visual mediante imágenes.
- **Secciones Específicas:**
 - Datos interesantes, eventos, mapas, análisis por regiones, épocas, días y horas, tipos de OVNIs, y una sección dedicada al Área 51.
 - Minería de datos con conclusiones.
 - Encuesta interactiva para los usuarios con almacenamiento en un CSV.
- **Barra Lateral Interactiva:**
 - Exploración de tipos de OVNIs con imágenes asociadas.
 - Estadísticas de avistamientos con filtro de años.
- **Agradecimiento y Compartir:**
 - Agradecimiento y enlaces a perfiles de LinkedIn del equipo.
 - Fuentes de datos mencionadas.

APP

La aplicación Streamlit desarrollada tiene como objetivo predecir la duración de avistamientos de Ovnis utilizando un modelo de regresión previamente entrenado con la biblioteca PyCaret. Esta aplicación proporciona una interfaz de usuario que nos permite ingresar parámetros como el mes, la hora, el país y la forma de OVNI para obtener una estimación de la duración del avistamiento.

Conclusiones

En este proyecto, se observa un aumento constante en los avistamientos de ovnis a lo largo de los años, con algunas variaciones dependiendo de algunos eventos específicos. Este patrón podría indicar un mayor interés, conciencia o reporte de avistamientos a lo largo del tiempo.

El verano es la estación con más avistamientos, seguido de cerca por el otoño. La explicación sugerida es la claridad del cielo y la menor cobertura de nubes durante estas estaciones, lo que facilita la observación de fenómenos en el cielo.

Especificando un poco más, los meses de junio, julio y agosto son los más propensos a los avistamientos, mientras que los de la primavera son los menos habituales. Los fines de semana y especialmente los sábados tienen el mayor número de avistamientos. Esto podría deberse a que la gente tiene más tiempo libre para observar el cielo. Y si queremos concretar la hora, varía a lo largo del día, con un aumento desde las 10a.m hasta las 9pm. Puede estar relacionado con la actividad humana y la visibilidad.

Otra de las conclusiones que hemos sacado ha sido, que hay una diferencia bastante impactante en la distribución de avistamiento entre las luces intermitentes y los demás fenómenos, porque son los más comunes. Luego van las formas triangulares y circulares con poca diferencia entre las dos. Esto sugiere que tipos de ovnis pueden tener una mayor propensión a ser observados en ciertos momentos.

Es muy claro que Estados Unidos lidera con un porcentaje abrumador del 88.52% de todos los avistamientos que están registrados en estos datos, mientras que España representa un pequeño porcentaje del 0.09%.

Si excluimos Estados Unidos, observaremos una amplia variación en la cantidad de avistamientos por países. Hay diferencia entre los dos primeros (Canadá y Reino Unido) y el resto. Esta información nos dice que hay otros países que también han registrado un número significativo.

Considerando la concentración de los avistamientos, especialmente en América del Norte, esta región es la más propensa a los avistamientos de ovnis hasta 2014. Estos resultados estarán influenciados por los informes presentados y la disposición de la población.

Para ir acabando con las conclusiones, lo que nos queda para comentar sería sobre los avistamientos en Europa, España y concretamente en Andalucía.

Los avistamientos en Europa revelan patrones interesantes. Reino Unido destaca como una región con un alto número de avistamientos, que nos podría afectar en la percepción general

de la distribución. Si entramos en España, vemos que pasa lo mismo con Andalucía, específicamente en Málaga. Otras áreas destacadas pueden ser Murcia o las Islas Canarias. Y para acabar profundizaremos un poco Andalucía. Se confirma que esta región es la que tiene el mayor número de avistamientos en toda España. Este análisis más detallado puede ser útil para comprender las áreas específicas dentro de Andalucía. Sin embargo, para obtener una comprensión completa, también sería beneficioso explorar otros lugares destacados como Cataluña y la Comunidad de Madrid.

[Panel Trello](#)