

Nombre: José Ramón Moratalla Muñoz

StackBlitz: <https://js-zt1hnm.stackblitz.io>

GitHub: <https://github.com/jramon-mm/08MBID-Visualizacion-de-Datos>

Google Colab: <https://colab.research.google.com/drive/1o2IeG2SShYn2u8AbvCToNvivLTQ8FcdT?usp=sharing>

Google Looker: <https://lookerstudio.google.com/reporting/53350cf7-4755-479f-8f6b-b88d1d03c151>

08MBID - Trabajo Práctico

1. Introducción

El término "OVNI" (Objeto Volador No Identificado) o, en inglés, UFO (Unidentified Flying Object) se utiliza para referirse a cualquier objeto en el cielo que no pueda ser identificado por el o los observadores. Los informes de avistamientos de ovnis se remontan a siglos atrás, pero, fue en la década de 1940 cuando el interés por este tipo de fenómenos creció debido a uno de los incidentes OVNI más famosos, el incidente OVNI de [Roswell](#).

Estados Unidos es uno de los países que más fenómenos de este tipo registra. De hecho, para una definición más amplia, el término ha sido bautizado a UAP en inglés (Unidentified Flying Objects) por y FANI en español (Fenómenos Aéreos No Identificados), tal y cómo se indica en [este artículo](#) del periódico "El País".

En España también se tienen registros de incidentes de este tipo desde hace bastante tiempo. El objetivo de este trabajo es conocer más acerca de la historia de este tipo de fenómenos en España. Se hablará, por tanto de, evolución de este tipo de fenómenos a lo largo del tiempo, sitios y ubicaciones en los que se han producido y posibles características comunes que estos incidentes han reportado a lo largo del tiempo.

Para ello, se utilizarán distintos tipos de visualizaciones y de herramientas de visualización. Además, como es obvio, necesitaremos un conjunto de datos. El conjunto de datos seleccionado para trabajar en el tema que nos ocupa es [este dataset](#) de [Kaggle](#).

2. Metodología

Como se ha comentado, la idea para este trabajo es utilizar varias herramientas de visualización de datos. Para ello, veamos, en primer lugar, las herramientas de visualización de datos que hemos visto a largo de esta asignatura.

- **[Google Sheets](#):** Herramienta fácil y sencilla proporcionada por Google. Está alojada en la nube y es gratuita. Permite crear gráficos sencillos pero no son interactivos. Sin embargo, la cantidad de datos que permite manejar no es muy grande. También permite compartir los gráficos para poder integrarlos en otras plataformas, como por ejemplo, StrackBlitz.
- **[Datawrapper](#):** De nuevo, herramienta en la nube. Muy sencilla de utilizar e intuitiva. Permite generar código HTML para integrar los gráficos en otras plataformas. Muy utilizada en medios de comunicación.
- **[Power BI](#):** Herramienta más avanzada. Permite tener un programa local (Power BI Desktop) pero también tiene modalidad en nube. Permite manejar grandes conjuntos de datos e incluso cruzar varias fuentes de información. Al igual que las anteriores también posee diferentes formas de pulicar o compartir los gráficos.
- **[D3](#):** Biblioteca open source de JavaScript. Permite crear visualizaciones interactivas pero, sin embargo, la curva de aprendizaje es bastante elevada. Permite diseños responsives y una de sus grandes ventajas es el vínculo entre los datos y el gráfico ya que si se cambia el dato, se cambia el gráfico.
- **[Python](#):** Python es un conocido lenguaje de programación. Posee múltiples librerías para el análisis y trabajo con conjuntos de datos. Una de las librerías para trabajar con dataset más conocidas es [Pandas](#). De igual manera, posee numerosas librerías para la visualización de datos. En este sentido, [Seaborn](#) es una de las más famosas.

Ahora, hagamos algo de investigación acerca de las herramientas y/o plataformas que hay disponibles en el mercado para visualización de datos. En [este artículo de Forbes](#) se recogen las mejores herramientas de visualización de 2023 (según Forbes). También se indican pros y contras de cada una y una anotación que indica para qué ámbito de visualización de datos en concreto es muy recomendable cada herramienta. Hagámos una revisión rápida del top 5 de estas herramientas según Forbes:

- **[Power BI](#):** Según Forbes esta es la mejor herramienta de 2023. Destacan su precio competitivo e integración

con productos Microsoft y sus servicios cloud. Como desventajas, reportan un alto uso de CPU y la incompatibilidad de la versión de escritorio con el sistema operativo MacOS.

- **Tableau:** La definen como la mejor herramienta para visualizaciones interactivas. Destacan su capacidad para crear Dashboard narrativos y su integración con [Slack](#). Como desventajas reportan una curva de aprendizaje dura y su alto precio.
- **Qlik Sense:** Lo señalan como mejor opción para dashboard personalizados. Como ventajas destacan su integración con numerosas plataformas, su plan gratuito y las métricas de uso que ofrece. Como desventajas reportan que el plan gratuito no permite descargar reportes en PDF.
- **Qlik Sense:** Lo señalan como mejor opción para inteligencia artificial. Como ventajas destacan la posibilidad de automatizar acciones y facilidades de cara a AI insights. Como desventajas reportan que se necesita mejorar la interfaz de usuario.
- **Klipfolio:** Lo señalan como mejor opción para dashboard personalizados. Como ventajas destacan su integración con numerosas plataformas, su plan gratuito y las métricas de uso que ofrece. Como desventajas reportan que el plan gratuito no permite descargar reportes en PDF.
- **Looker:** Herramienta de Google. Como ventajas destacan la presencia de plugins que mejorarán nuestras visualizaciones, su interfaz drag-and-drop y sus bonitos diseños visuales. Como desventajas reportan limitaciones de personalización y un mal funcionamiento con conjunto de datos grandes.

Para concluir esta parte, daré mi opinión subjetiva acerca de una de las herramientas mencionadas anteriormente ya que he podido trabajar con ella en la empresa para la que trabajaba. Esta herramienta es Tableau. Para mí Tableau es la mejor herramienta de visualización y tratamiento de datos. Aunque es cierto que es bastante cara, ofrece muchísimas alternativas para tratar con datos, visualizarlos y, además, lo hace de manera rápida incluso con conjuntos de datos grandes. Esto, en parte, es por el motor de datos que tiene por debajo. Dicho motor, se llama [Hyper](#) y ayuda a agilizar las consultas e integraciones entre distintas fuentes de datos. Otra de las características que más me gusta es su facilidad a la hora de preparar los datos para la visualización. Gracias a [Tableau Prep](#), podemos crear flujos que transformen y limpien nuestros datos. Posteriormente, podremos crear las visualizaciones necesarias con [Tableau Desktop](#).

Para terminar este apartado indicaré las herramientas que voy a usar y el motivo. Las herramientas que voy a usar son: Python y Power BI. La razón para usar Python es que, al venir de Ingeniería Informática, me siento cómodo programando y dada la versatilidad de Python podré probablemente acceder a todo lo necesario para optar a toda la nota de este trabajo. De hecho, la idea es utilizarlo para visualizar datos de manera interactiva e, incluso, ver si se puede añadir algún filtro.

En segundo lugar, usaré Power BI para ver si, en lo personal, me gusta más que Tableau. También tengo curiosidad de comprobar si es tan buena como la revista Forbes indica ya que, las veces que lo he usado para hacer alguna prueba, no me ha terminado de convencer.

3. Resultados

3.1 Número de avistamientos a lo largo de los años (Python con Seaborn y Plotty)

En primer lugar, puesto que los datos que tenemos constituyen una serie temporal en la cuál se indican avistamientos de OVNI en España reportados a lo largo de los años, las preguntas que vamos a tratar de contestar, todas acotadas en términos de territorio al territorio español son:

- ¿Cómo se distribuyen los avistamientos a lo largo de los años?
- ¿Se producen muchos avistamientos?
- ¿Ha podido afectar la aparición de tecnología como cámaras de fotos, cámaras de vídeo o teléfonos inteligentes al reporte de este tipo de fenómenos?

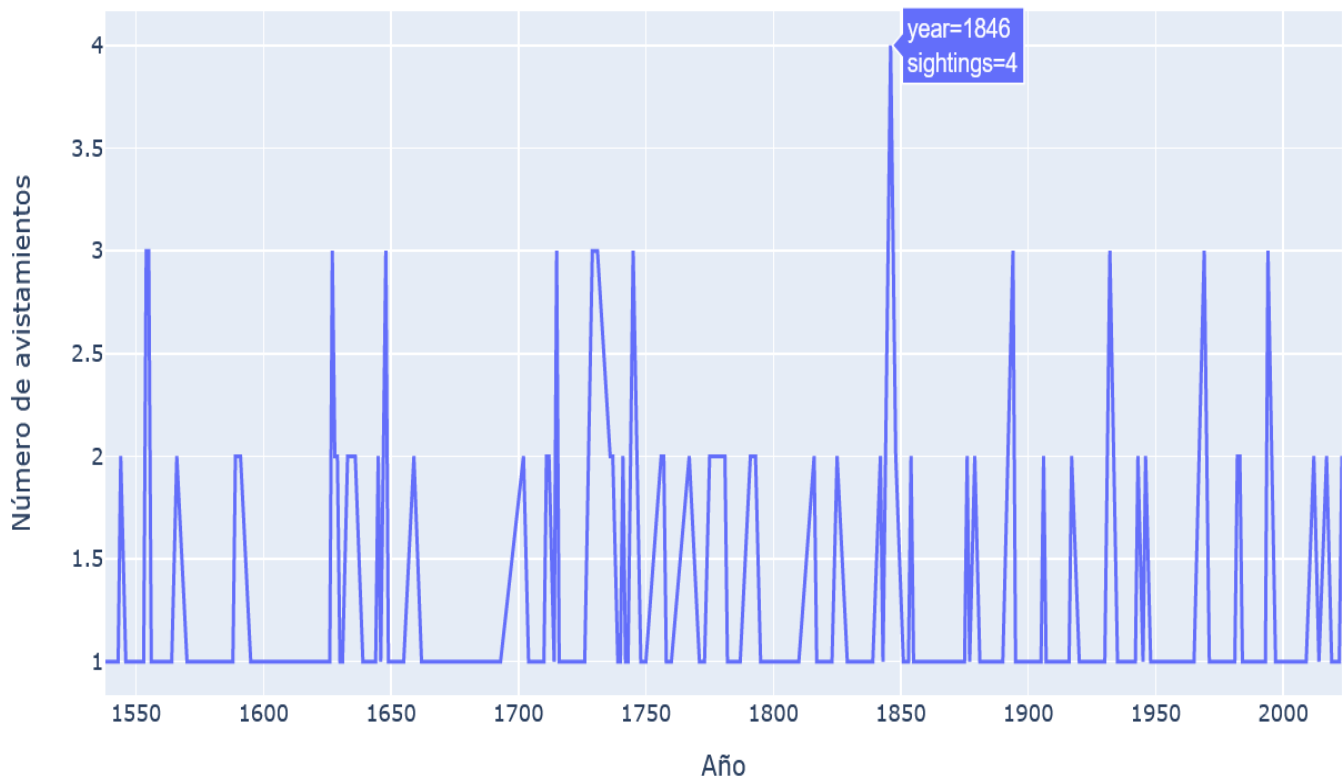
Para responder a estas preguntas es evidente que necesitamos ser capaces de graficar el número de avistamientos que se producen a lo largo de los años. Para ello, podríamos usar un gráfico de columnas o un gráfico de barras pero, lo que, en mi opinión, mejor se ajusta a lo que queremos representar es un gráfico de líneas dónde el eje x sea el tiempo, en este caso los años, y el eje y sea el número de avistamientos de esos años. Todo esto está hecho y explicado en el cuaderno de Google Colab por lo que, en este punto, recomiendo visitarlo a través del link que está al inicio de este PDF. En cualquier caso, procedo a explicar el proceso que he hecho en el Colab:

1. Carga de los datos desde un CSV.

2. Agrupamiento de los datos en función del año. El resultado es un conjunto de datos con dos columnas "year" y "sightings".
3. Creación de gráfico de líneas no interactivo con Seaborn.
4. Creación de gráfico de líneas interactivo con Plotly.

Por último, adjunto el gráfico final interactivo con Plotly mediante captura de pantalla. De nuevo, recomiendo visitar el notebook de Google Colab para así poder ver como se interactúa con el gráfico de Plotly.

Número de avistamientos de ovnis en España por año



Con la imagen anterior ya seríamos capaces de contestar las preguntas que hemos formulado. En primer lugar vemos que el año que más avistamientos se producen es 1846 y se producen 4 avistamientos. El resto de los años oscila entre 2, 3 y 1. De hecho la mayoría de los años se produce tan sólo 1 avistamiento. Por lo tanto, podemos decir que no se producen muchos avistamientos. De hecho, personalmente, cuando estaba visualizando los datos me he sorprendido ya que esperaba que hubiese más avistamientos.

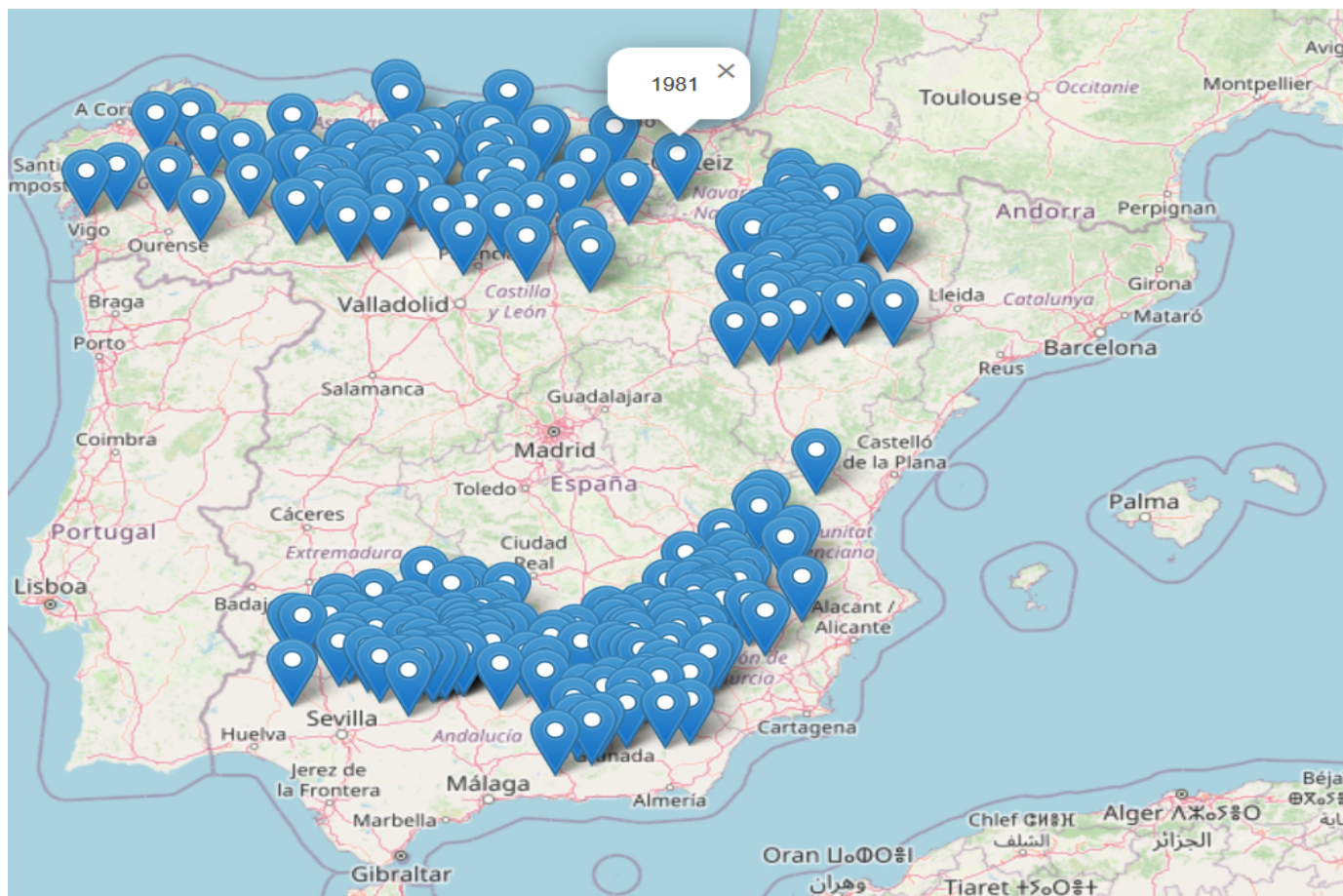
Para acabar con esta ronda de preguntas, podemos afirmar que la aparición de tecnologías como cámara de fotos, cámara de vídeo o smartphones no incrementó el número de este tipo de fenómenos al menos en España. Esta última afirmación me resulta bastante extraña pero es lo que los datos nos dicen. Para contrastarlo mejor habría que buscar otras bases de datos que contenga información acerca de este tipo de fenómenos en España y comparar resultados.

3.2 Lugar de los avistamientos (Python con Seaborn - Plotly)

El siguiente paso, ahora que hemos visto la distribución de los avistamientos y, puesto que tenemos información de la ubicación de los mismos, sería visualizar geográficamente dónde se han producido cada uno de ellos. Para ello, de nuevo utilizaremos Python. Las preguntas a responder son:

- ¿Dónde se producen los avistamientos?
- ¿Hay alguna zona de España dónde sea más frecuente este tipo de fenómenos?

El gráfico a utilizar en este caso está bastante claro, utilizaremos un gráfico de Mapa. Para ello utilizaremos la librería Folium. Esta librería la hemos utilizado en clase. El proceso está, de nuevo, explicado en Google Colab por lo que recomiendo visitarlo para verlo. No obstante, el gráfico obtenido es:



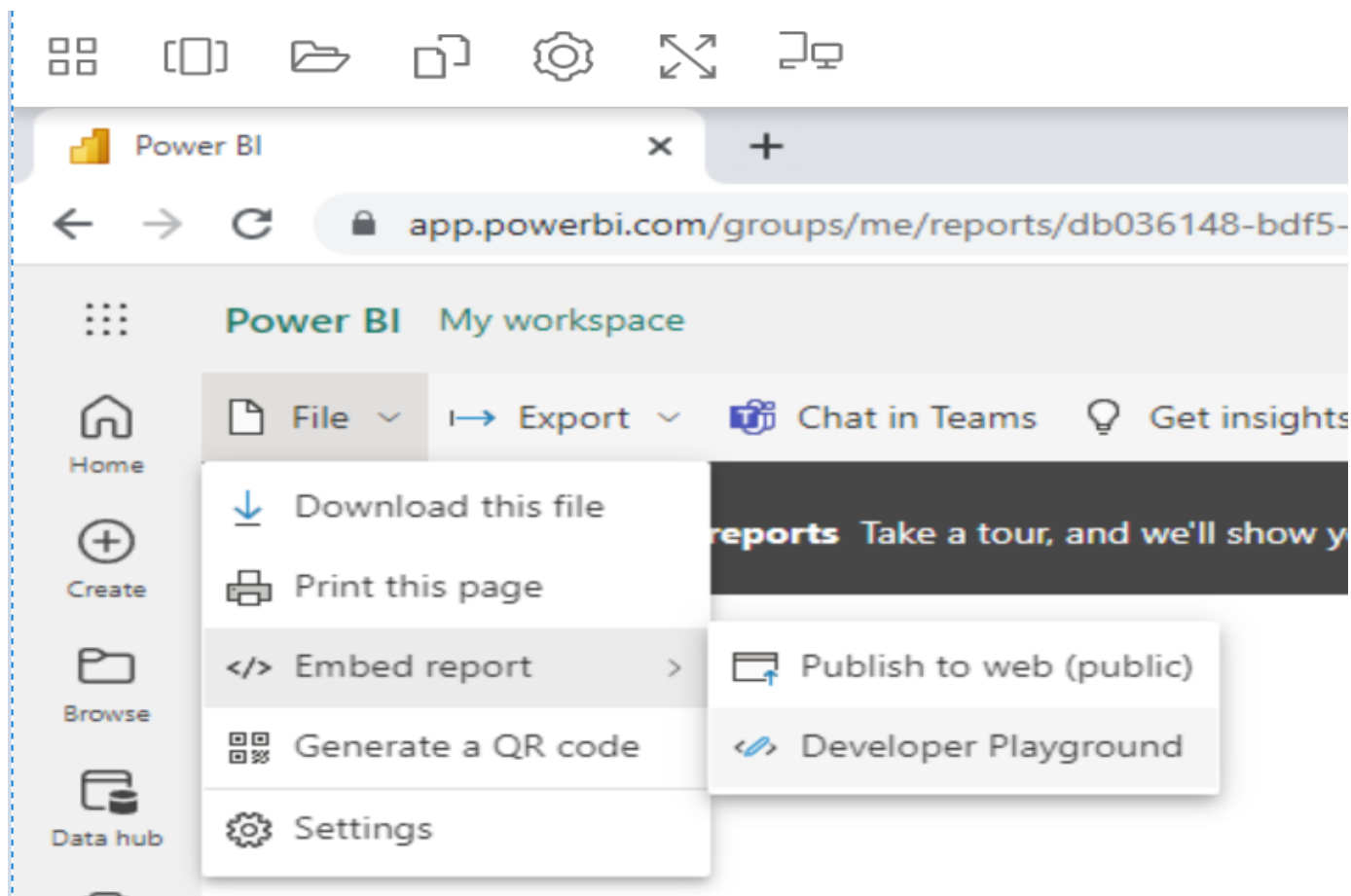
Como se puede ver, se ha añadido como "popup" el año en el que se produjo el evento. En cuanto a la respuesta a las preguntas, la respuesta a la primera es el propio gráfico en sí, mientras que, respecto a la segunda pregunta, podemos ver en el gráfico que hay dos agrupamientos claros. Uno al sur y sureste de España, otro al norte y noreste de España. Vemos también que en la zona centro no hay ningún caso. Extremadura, Castilla-La Mancha y Madrid serían las comunidades que carecen de este tipo de fenómenos.

3.3 Comparación de los incidentes más comunes (Power BI)

Ahora que hemos visto la distribución de los eventos a lo largo de los años y la ubicación de los mismos vamos a focalizar en qué ha ocurrido en los eventos. Para ello, nuestro conjunto de datos tiene una serie de eventos comunes que suelen ocurrir en este tipo de fenómenos y, para cada registro, se indica si este evento ha ocurrido o no. Las preguntas para este apartado podrían ser:

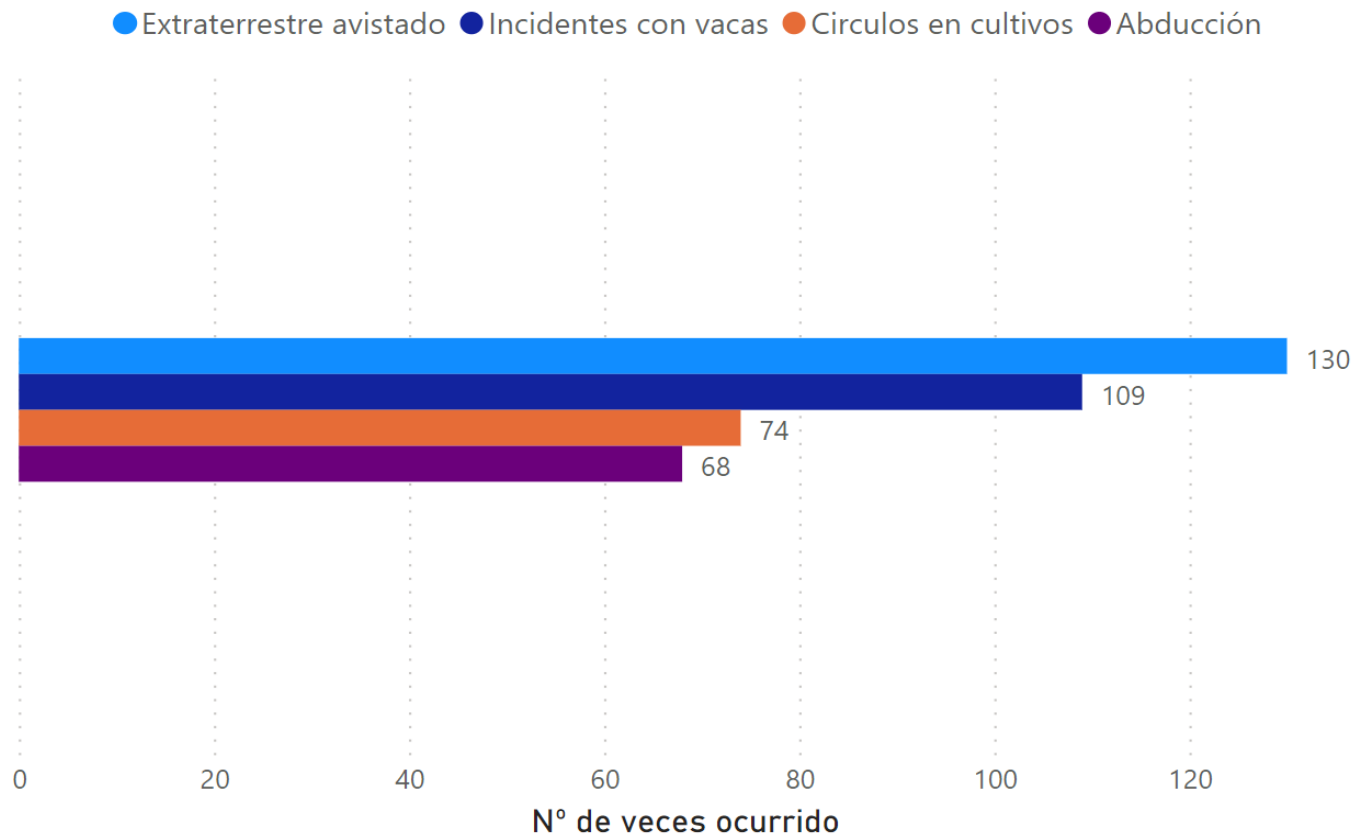
- ¿Cuál es el evento más común?
- ¿Podría tener algún tipo de relación el evento más común con la zona en la que se producen los avistamientos?

La idea para adjuntar este gráfico era adjuntarlo mediante un iframe. Sin embargo, VirtualLabs carece de la opción que, según las diapositivas, se necesita para poder generar el código del iframe. Adjunto captura para que quede claro que la opción no está presente.



Volviendo al trabajo, lo ideal para representar la ocurrencia de cada una de las características en los avistamientos sería un gráfico de barras o uno de columnas. Simplemente se quiere ver la cantidad de veces que ha ocurrido cada uno de los eventos. Por dicho motivo, cualquiera de los dos tipos (barras o columnas) valdría. Personalmente, me gustan más los diagramas de barras que los de columnas. Por dicho motivo haré un diagrama de barras. Además, ordenaré los datos de mayor frecuencia a menor frecuencia. El gráfico resultante es:

Eventos ocurridos durante los avistamientos



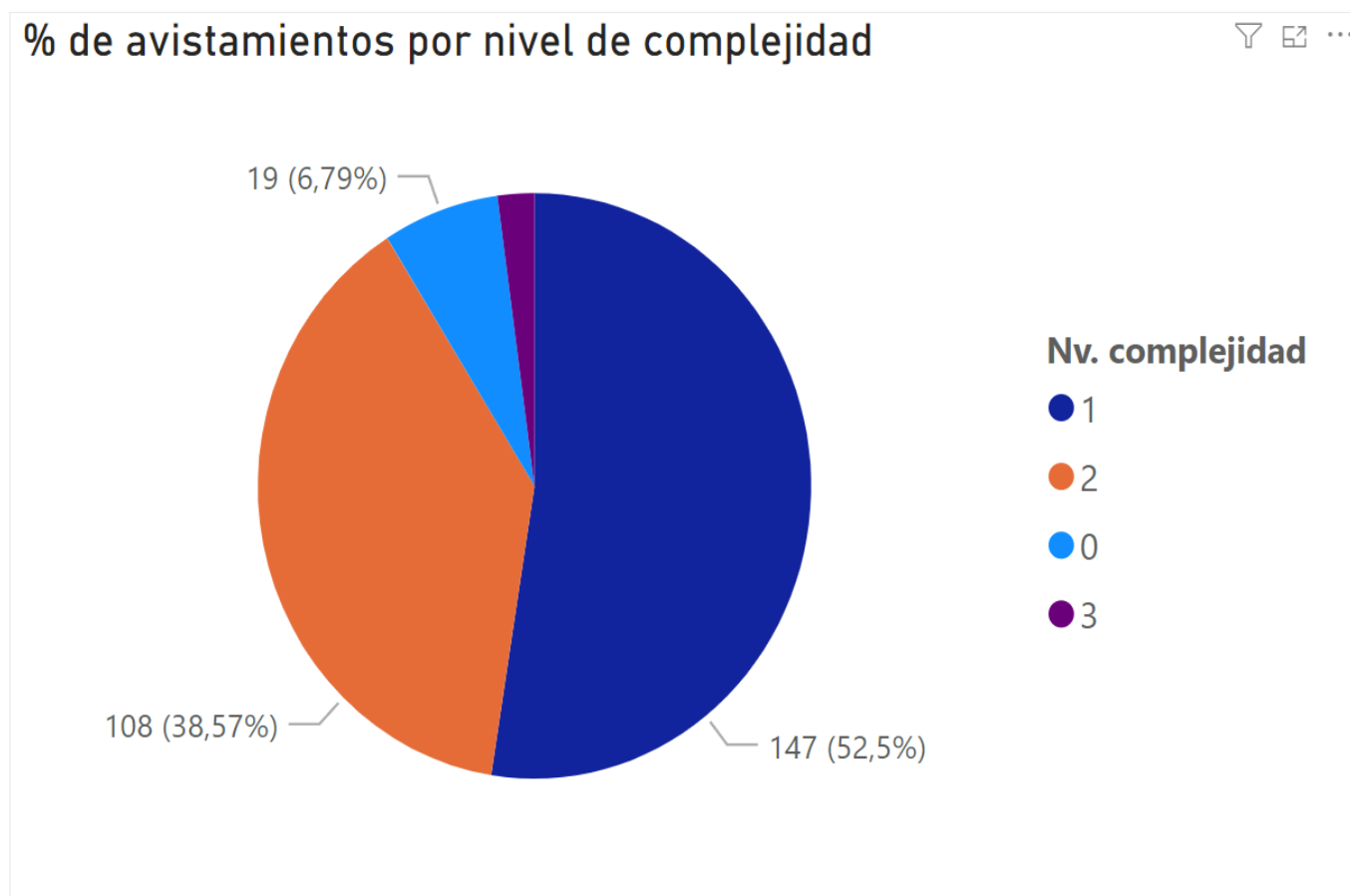
Con este gráfico, podemos responder a las preguntas. El evento más común es el avistamiento un extraterrestre y lo menos común son las abducciones. El segundo lugar es para los incidentes con vacas. A falta de un análisis más exhaustivo, esto puede llegar a tener sentido teniendo en cuenta que mucha concentración de avistamientos estaba en la parte norte de España dónde, comunmente, están la mayoría de las ganaderías bovinas.

3.4 Complejidad de los avistamientos (Power BI)

En este último apartado, lo que se va a tratar de dotar de algún tipo de medida de calidad a los avistamientos. Hemos visto que por cada avistamiento hay varios eventos que pueden suceder en dicho avistamiento. Con Google Colab, se ha generado una nueva columna para categorizar el nivel de complejidad de los avistamientos, siendo esta variable un entero del 1 al 4 indicando qué cantidad de estos eventos se da en un mismo registro. Las preguntas que se quieren responder durante este apartado son:

- ¿Se producen varios eventos en un mismo avistamiento?
- ¿Se podría categorizar los avistamientos en niveles de complejidad, siendo la complejidad mayor cuanto más eventos ocurren?
- ¿De qué nivel serían más comunes los avistamientos?
- ¿Hay algún avistamiento en el que se hayan producido todos los eventos?

Para este gráfico hay dos opciones. La primera de las dos opciones es un gráfico de sectores. Hay que tener cuidado con este tipo de gráficos ya que si hay sectores muy pequeños no se distinguen bien los sectores. En nuestro caso, sólo hay 4 sectores. Veamos como sale y decidamos si lo mantenamos o no. El gráfico sería:

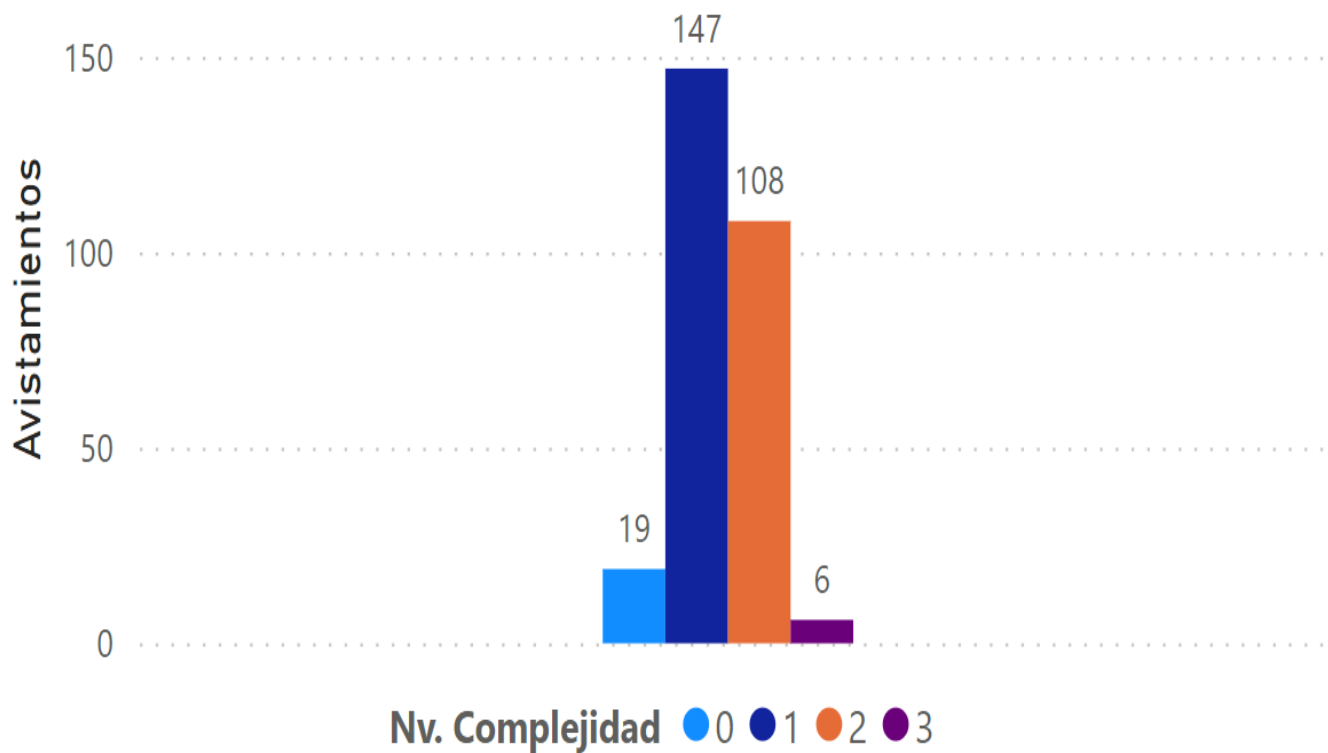


Como se puede comprobar el gráfico de sectores no es del todo malo. Ciertamente es que hay un sector que sin etiqueta debido a que es muy pequeño pero el resto de sectores están bastante bien diferenciados. De hecho, podemos responder a las preguntas. Vemos que sí se producen varios eventos en un mismo avistamiento, que, tal y como hemos hecho si se puede categorizar los avistamientos en niveles de complejidad y que el nivel de complejidad más común entre todos los avistamientos es el nivel uno.

Lo que también hemos visto, es que hay avistamientos en los que no se producen ningún tipo de evento de los tipos que tenemos registrados. Estos avistamientos serían los de complejidad 0. Por último, podemos afirmar que no hay ningún avistamiento en el que se hayan producido todos los eventos a la vez. Este gráfico aporta bastante valor ya que, si fuera un investigador de este tipo de eventos, querría analizar un avistamiento en el que se hayan producido muchos eventos a la vez antes que uno menos complejo.

Finalmente, adjunto un gráfico de columnas con esta misma información para ver cómo se quedaría y elegir uno de los dos como el definitivo.

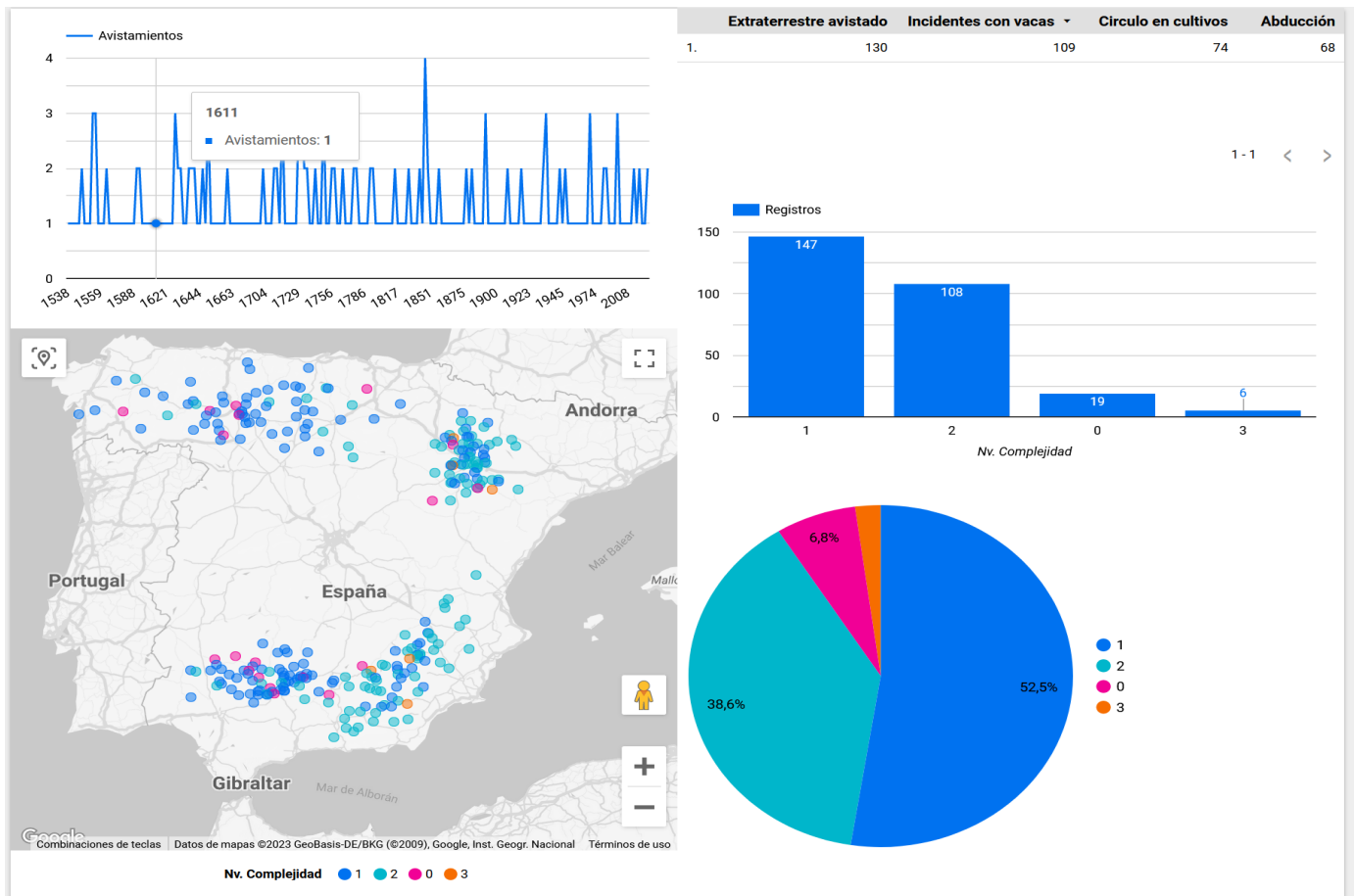
Avistamientos por Nv. Complejidad



Personalmente y, aunque cualquiera de los dos podría valer, me quedaría antes con el gráfico de columnas. El motivo es que no me termina de gustar que la etiqueta no se vea. Por ello, elijo el gráfico de columnas.

3.5 Dashboard (Google Looker Studio)

Para crear el dashboard con los gráficos se usará Google Looker Studio. Aunque se adjuntará a la entrega un PDF y se ha adjuntado el link del proyecto de Google Looker al inicio del documento, adjunto captura del dashboard:



No he podido crear el diagrama de eventos ocurridos de otra forma que no fuese una tabla. Probablemente esto es debido al desconocimiento del manejo de Google Looker. He de decir que me ha gustado bastante aunque he de profundizar más en la herramienta.

4. Conclusiones

Como conclusiones, podemos sacar:

- Al menos según la fuente de datos con la que hemos trabajado, parece que el fenómeno OVNI en España no es muy común ya que hay pocos casos (el año que más casos ha habido han sido 4 casos).
- Los fenómenos se dan sobre todo en la parte norte y sur de la península. La parte central de la península (Extremadura, Castilla-La Mancha y Madrid) no presentan ningún caso. De nuevo, esto, según la fuente de datos utilizada.
- Los avances tecnológicos como cámaras de vídeo, cámaras de fotos y smartphones no han ayudado a incrementar este tipo de fenómenos en España, al menos, según la fuente de datos utilizada.
- El evento más común es el avistamiento de un extraterrestre seguido de los incidentes con las vacas.
- La mayor parte de los casos sólo corresponden a un evento, probablemente, a un simple avistamiento de un extraterrestre. No hay ningún caso con complejidad extrema.