Tercera entrega para el proyecto final de Data analytics

**Proyecto Ferias del Libro en la Argentina**

Alumno Participante: Molinari Matías.

Comisión: 41175.

Academia: Coder House.

Tutor a Cargo: Rodríguez German.

Fecha de presentación: 5/9

**INDICE**

[1-Introducción: 2](#_Toc113229191)

[2-Alcance: 2](#_Toc113229192)

[3-Hipótesis: 2](#_Toc113229193)

[4-Herramientas Utilizadas: 2](#_Toc113229194)

[5- DataSets: 3](#_Toc113229195)

[6- Diagrama entidad-relación: 4](#_Toc113229196)

[7- Listado de tablas: 4](#_Toc113229197)

[8- Listado de columnas por tablas: 5](#_Toc113229198)

[9-Modelo Relacional En POWER BI: 6](#_Toc113229199)

[10-Transformaciones Realizadas: 7](#_Toc113229200)

[11-Segmentaciones: 8](#_Toc113229201)

[12) Tooltips 14](#_Toc113229202)

[13) Conclusión: 16](#_Toc113229203)

[14)Futuras Líneas: 16](#_Toc113229204)

[15) Bibliografía: 17](#_Toc113229204)

# Introducción

Es un evento social, económico y cultural, que se realiza con cierta periodicidad.

El responsable de la misma es el Sistema de Información Cultural de la Argentina -Ministerio de Cultura de la Nación.

La base de datos fue creada el 27 de diciembre de 2021 y actualizada el 12 de enero de este 2022.

Dentro contiene información de diferentes ferias de libros de años anteriores también donde podemos encontrar el nombre de ellas donde se realizó cada una como otros tipos de datos

Provincia en donde las podemos encontrar, el departamento que se encargó, las localidades que dieron el lugar para hacerlas, la información necesaria (Dirección, mail, teléfono, página web (si es que usaban), mes, año, cuando fue realizada la primera feria y la última hasta el momento, nombre que llevaba esa feria, modalidad y tipo de gestión).

El objetivo de las ferias es difundir y comercializar libros que promueven la lectura y difusión comercial de editoriales.

# 2) Alcance

El DashBoard que tiene este proyecto es hacia las personas que sepan el manejo de la herramienta para poder navegar y entender el mismo

La utilización es para poder ver en qué provincias se han realizado en el tiempo diferentes ferias y así poder observarlas si cuentan con los datos necesarios para futuras actualizaciones. Podremos contar con mapas para que se puedan ubicar en donde se hicieron las mismas.

# 3) Hipótesis

Como objetivo es mostrar al usuario con los datos encontrados y plasmados en nuestro tablero las ferias realizadas a través del tiempo en la Argentina así se puede saber en qué provincia se realizó, su departamento y localidad como también así la dirección de las mismas.

Deberemos ver si podemos utilizar los datos para demostrar por ejemplo qué cantidad de ferias se realizaron en cierta cantidad de años, si podemos ubicarlas de alguna manera gráfica, para que al final del trabajo podamos concluir si nos falto información para hacer un análisis más profundo sobre otras cuestiones que se nos presenten.

# 4) Herramientas Utilizadas

* Excel para la lectura y limpieza del dataset.
* ERDPlus para la creación del diagrama entidad-relación (<https://erdplus.com/>).
* Power BI Desktop

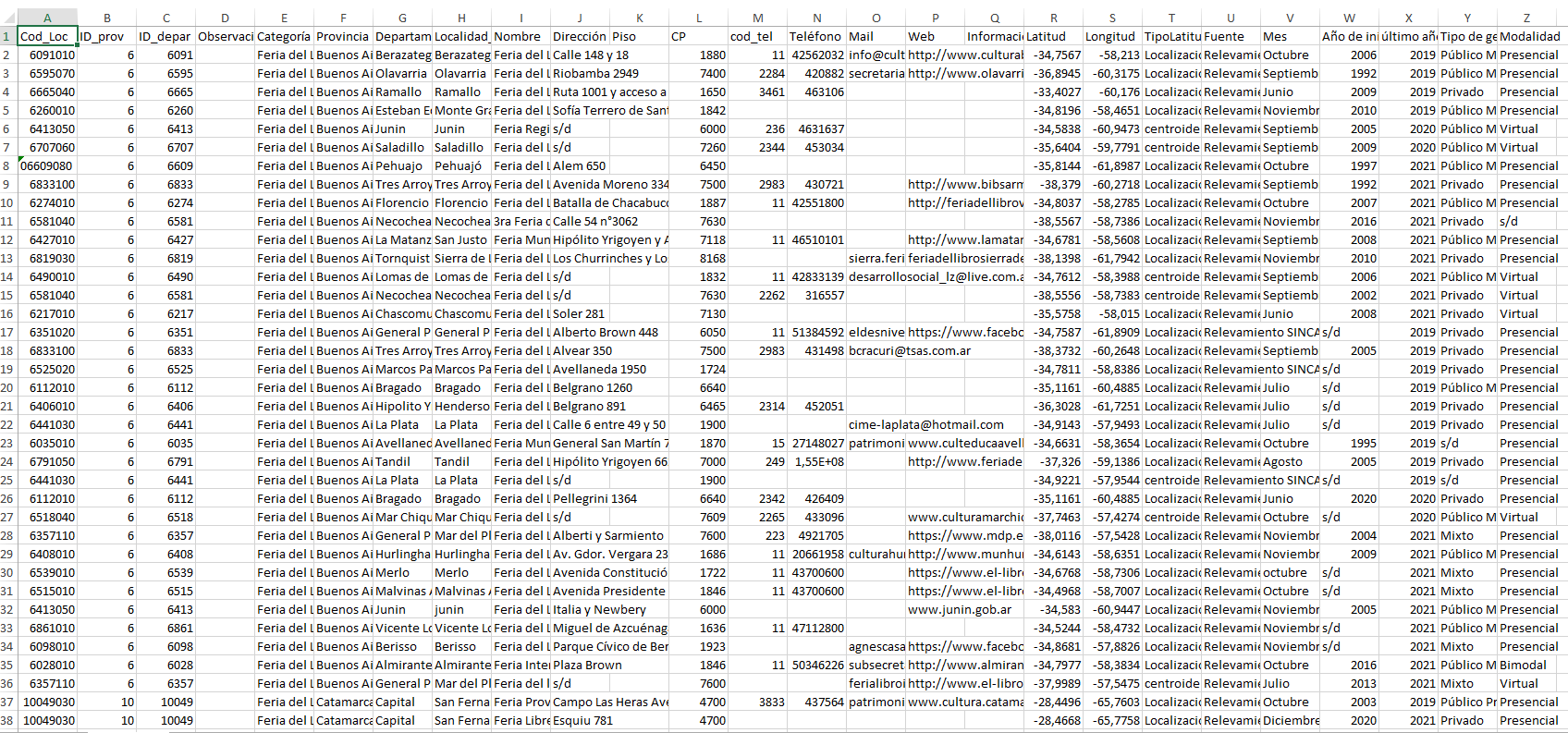
# 5) DataSets

Primero digamos que la base de datos original está ordenada de esta manera:

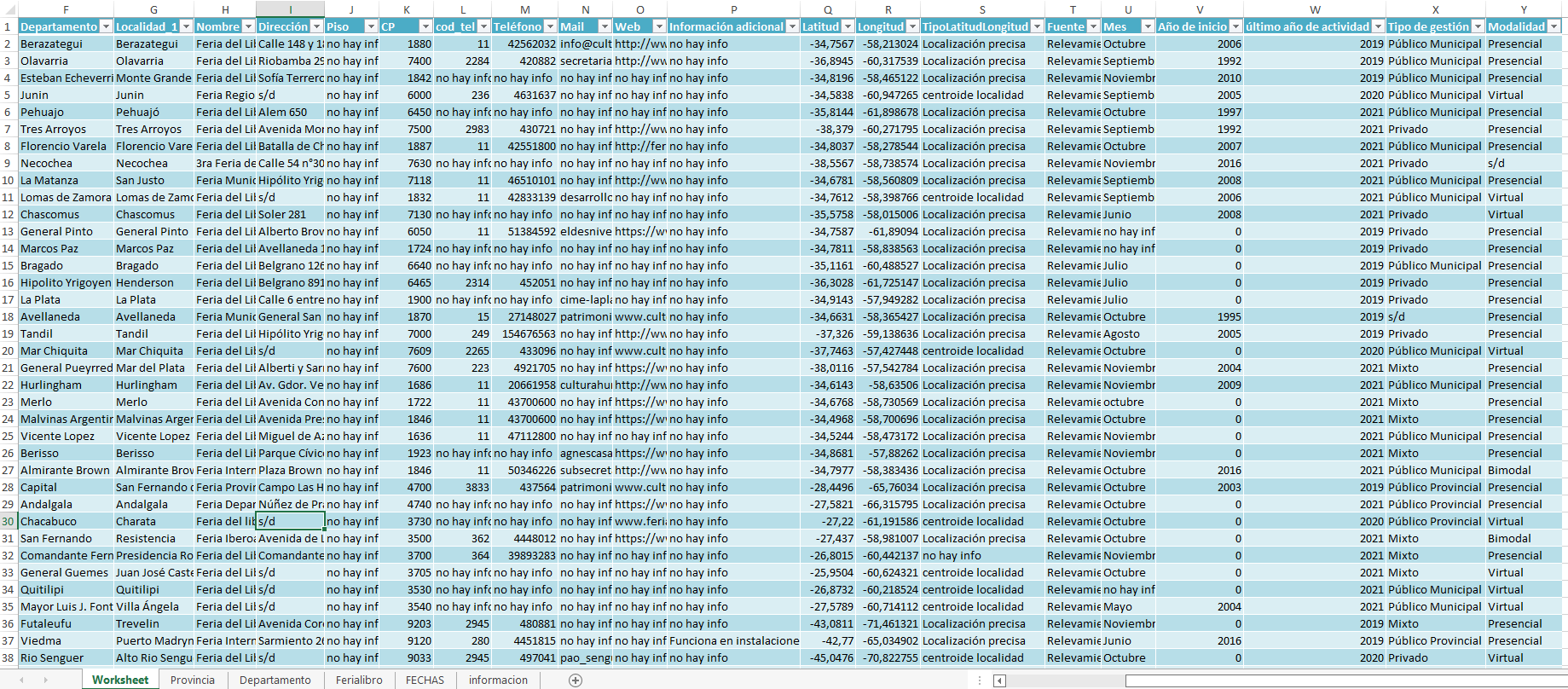
* Título de la columna
* Tipo de dato
* Descripción



El cual cuando la abrimos la tenemos así:



Antes de llevarla al modelo final para relacionarla con Power BI, limpiamos, ordenamos para poder generar el modelo entidad de la manera más cómoda y no generar problemas en la herramienta de trabajo, lo cual nos quedó así:



De esta manera pudimos acomodarlo para luego cuando lo usemos no tengamos inconvenientes, el modelo relacional del mismo se podrá ver en el punto 9 de este informe.

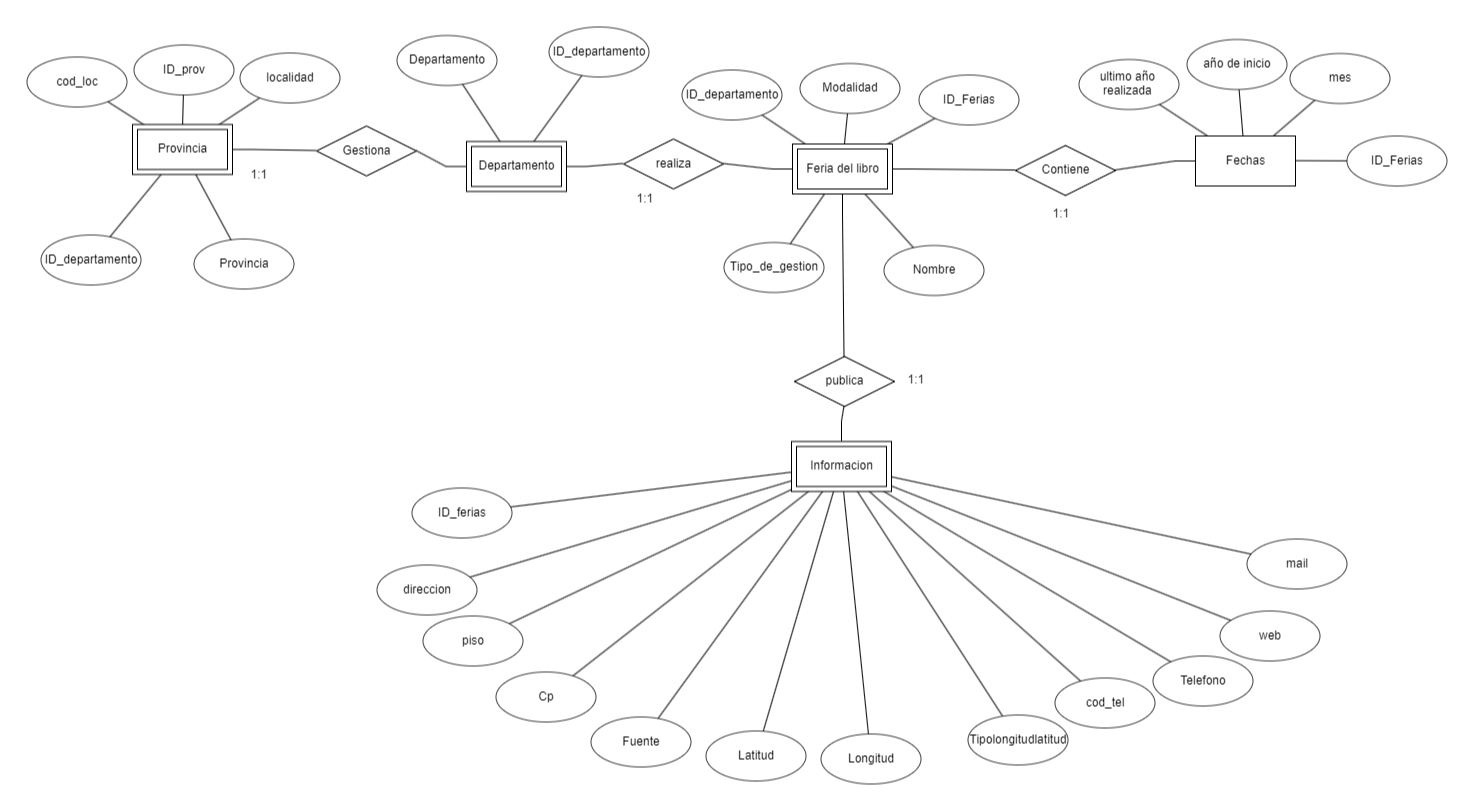
Como por ejemplo en la tabla Año de inicio donde teníamos valores null los reemplazamos por un 0.

Y en las demás tablas donde no tuviéramos información lo reemplazamos por "No hay info" Ahora con todo esto guardado el nuevo archivo con el que trabajaremos será el siguiente:



# 6) Diagrama entidad-relación

Acá dejamos el primer diagrama creado al principio del proyecto:



Donde podemos observar las relaciones como, por ejemplo:

Una Provincia gestiona muchos departamentos. Un Departamento realiza diversas ferias del libro.

La Feria del libro contiene diferentes fechas en cuanto se realizan a lo largo de los años como, así también la misma proporciona y publica información para que se pueda asistir.

# 7) Listado de tablas

Provincia: Contiene Las provincias, localidades y el código de las mismas.

● PK: ID\_Prov

● PK: Cod\_loc

● FK: ID\_departamento

Departamento: Departamentos que se encargaron de la realización de las ferias en sus respectivas localidades.

● PK: ID\_departamento

Feria del libro: Nombre con las cuales las ferias fueron realizadas como también su modalidad y el tipo de gestión.

● PK: ID\_Ferias

● FK: ID\_departamento

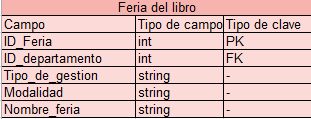
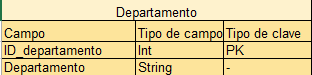
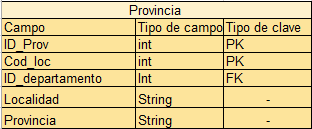
Fechas: Año desde que se realizó la primera feria, mes en que se, hizo, también su último año vigente hasta el momento.

● FK: ID\_Ferias

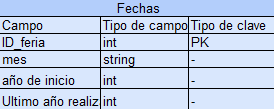
Información: En esta última tabla encontraremos toda la información necesaria de la base como el teléfono (su código de área también), dirección, página web o red social que la promueve, la dirección con su respectivo código postal, piso y por último la latitud, longitud, fuente y tipo de longitudlatitud donde se encuentra realizada la feria.

● FK: ID\_Ferias

# 8) Listado de columnas por tablas

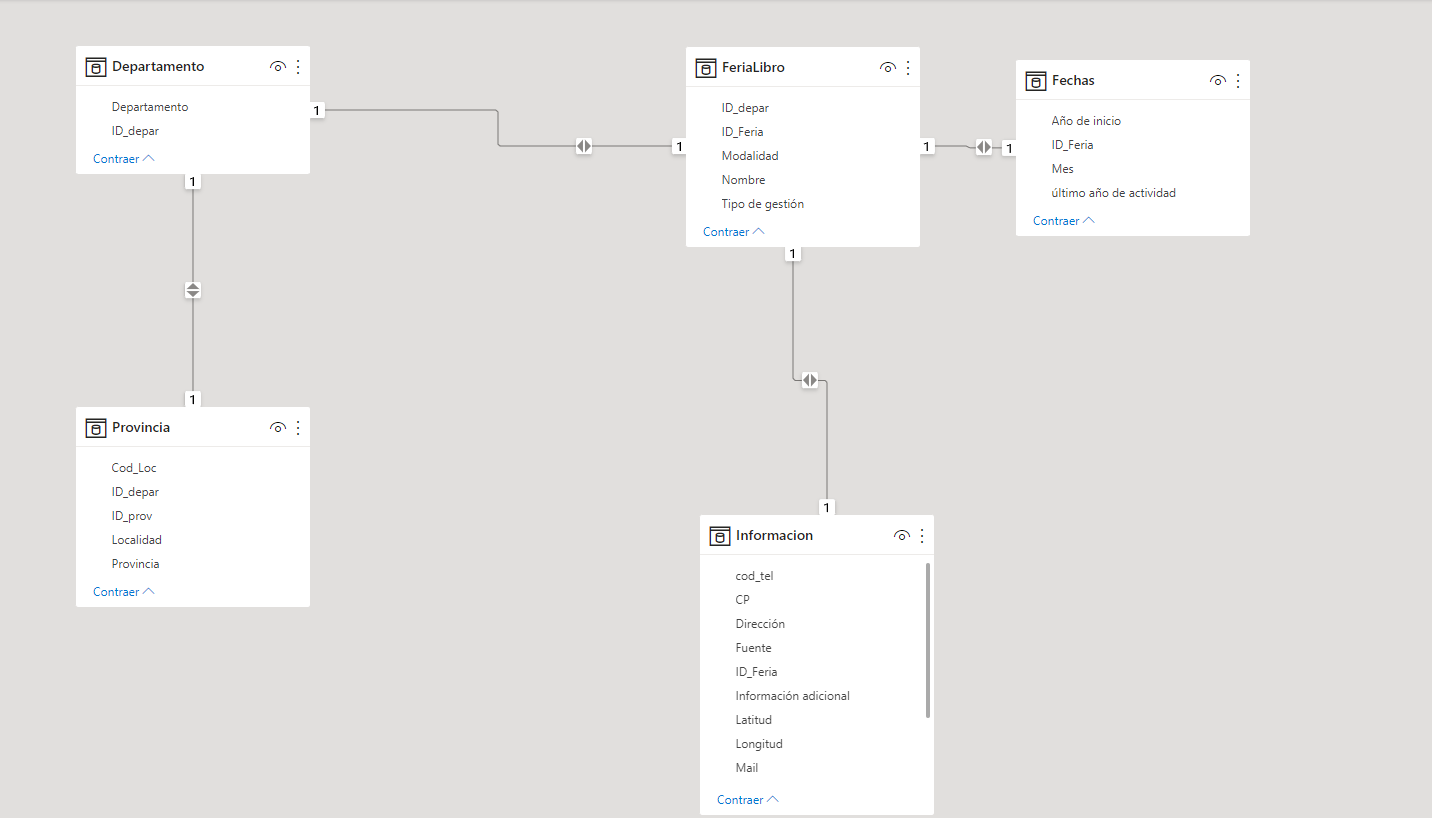






# 9) Modelo Relacional En POWER BI

Después de limpiar la base de datos en Excel, separar por carillas las tablas, crearlas y llevarlas al programa nos quedó de la siguiente manera:



# 10) Transformaciones Realizadas

Cuando importamos tablas de Excel a la herramienta POWER BI

Nos encontramos que algunos datos de las tablas estaban mal clasificados como, por ejemplo:

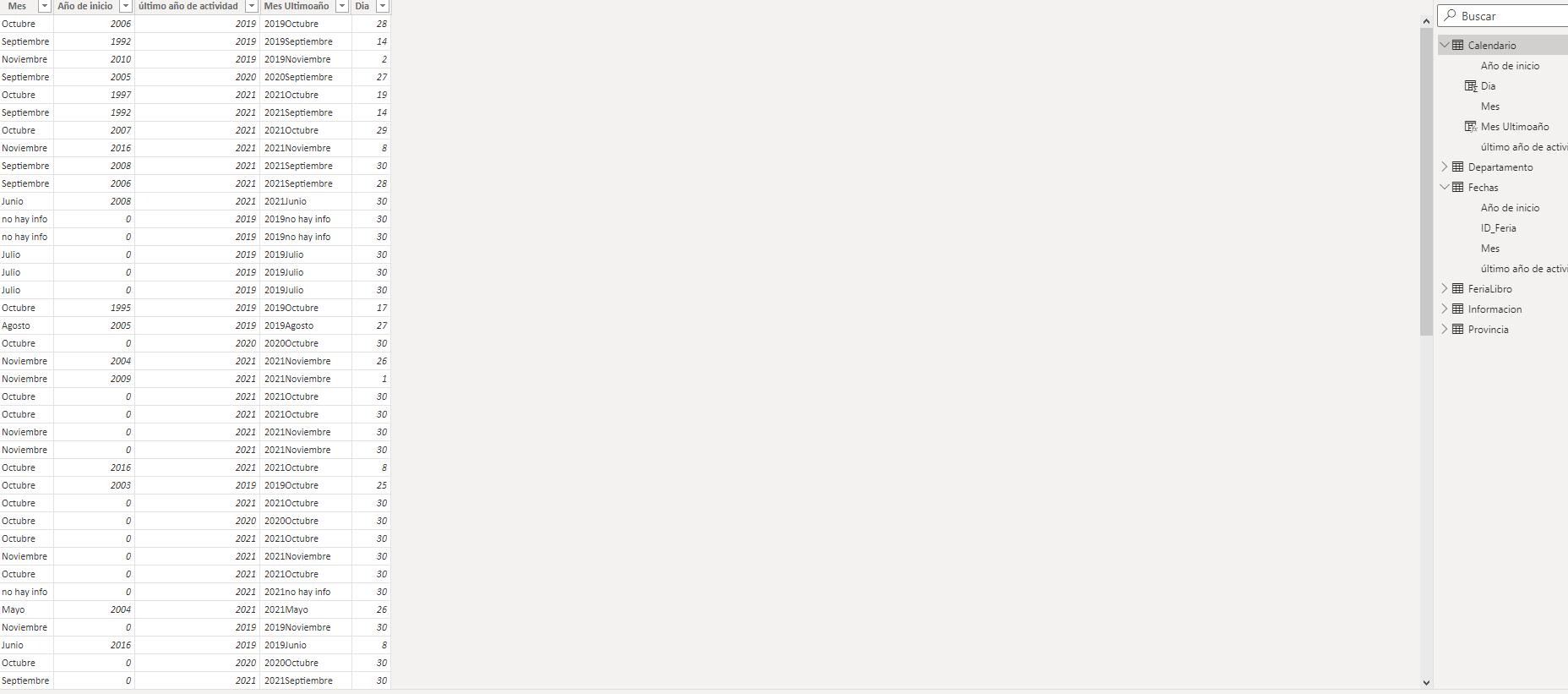
De la tabla Provincia:

1. La columna Cod\_loc y ID\_prov figuraban en el resumen con el signo Σ paso seguido lo dejamos sin resumir (como figura en la imagen del punto 9).

De la tabla Información:

1. La columna Latitud, Longitud y CP también contenía el signo Σ por lo cual lo cambiamos por “no resumir”.

Luego generamos una Tabla Calendario

.

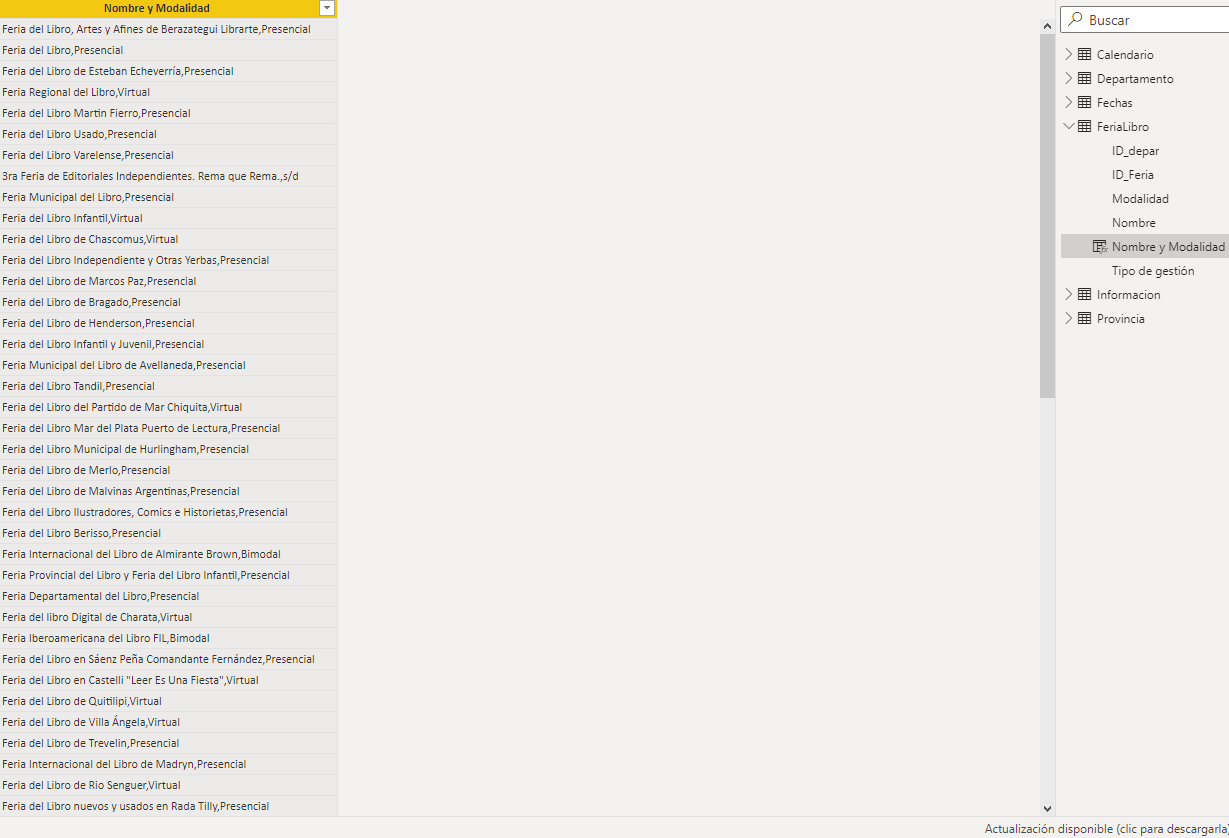
La cual la sacamos duplicando la tabla “Fechas”, agregamos las columnas “mes ultimoaño”, la creamos usando el siguiente comando:

Mes Ultimoaño = Calendario[último año de actividad] & FORMAT(Calendario[Mes],"mm")

La columna “Día” cual fue creada:

Dia = DAY(Calendario[Año de inicio])

En la tabla FeriaLibro Creamos la columna “Nombre y Modalidad”, así al momento de plasmar la información era más sencillo adjuntar estos datos que tenerlos por separado



# 11) Segmentaciones

Si bien no tomamos como primera segmentación la presentación dado que solo cumple la función de presentar el tablero si es la primera página de nuestro trabajo.



1. **En esta primera segmentación denominada "Localizaciones".**

En el gráfico de mapa insertamos los datos de las columnas pertenecientes a las siguientes Tablas:

De la tabla Información mostramos Latitud y Longitud

De la tabla Provincia mostramos Localidad

Aparte pusimos otro gráfico de mapa, pero esta vez el coroplético

Donde nos muestra la provincia y su nombre para ello utilizamos La tabla provincia y la columna del mismo nombre.

Que la relacionamos con el gráfico de Tabla que contiene:

De la tabla Ferialibro , Nombre .

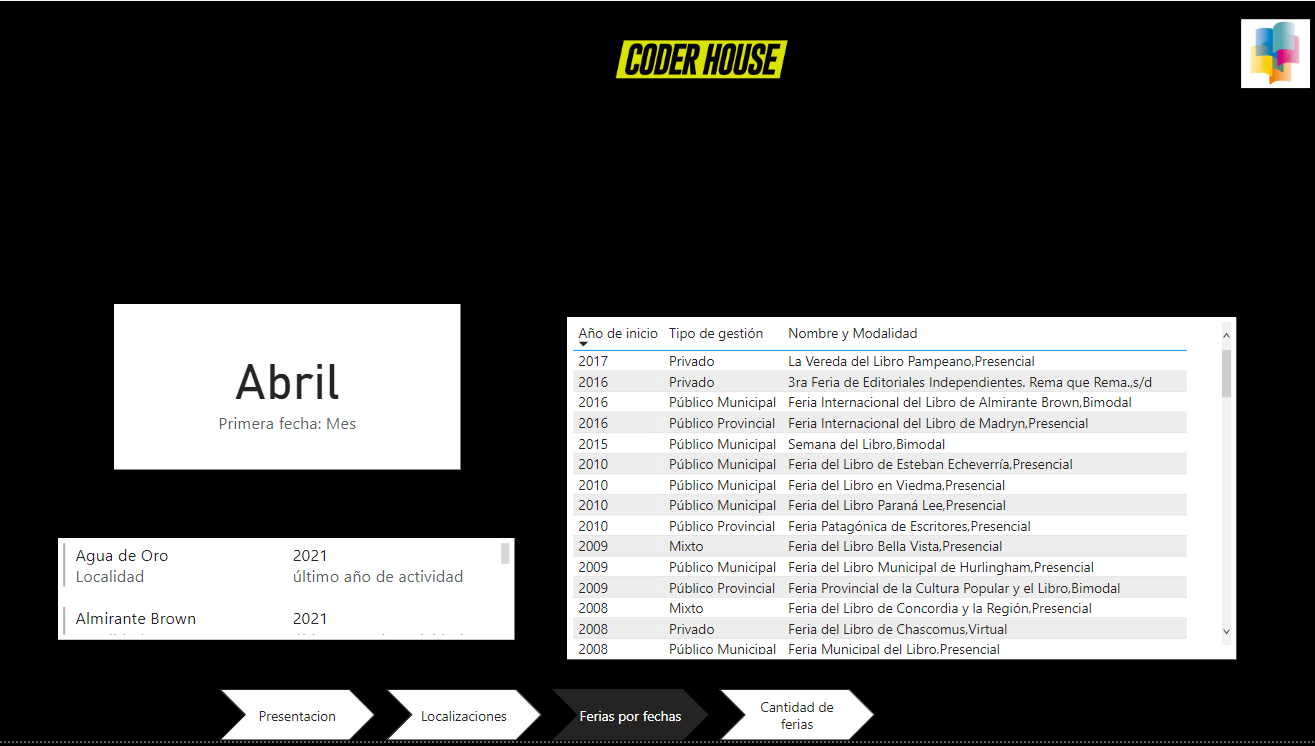
De la tabla Provincia, Cod\_Loc .

, De la tabla Información, dirección.

De esta manera con estas 2 relaciones podemos interactuar tanto con el mapa como con la tabla para poder seleccionar y mirar los datos que queramos para hacerlo de una manera más prolija deberíamos de hacerlo desde el gráfico “Tabla”.

Y por último para qué para que podamos terminar de entender bien en qué Departamento se realizó cada feria hicimos un gráfico más de “Tarjeta” donde introdujimos de la Tabla “Departamento “La columna del mismo nombre dando así a entender que cada vez que mostremos una feria tengamos su dirección, localidad, su localización más exacta posible, provincia donde se realizó, el nombre de la festividad y el departamento que la llevó a cabo.

1. **En la segunda solapa “Feria por fechas”**



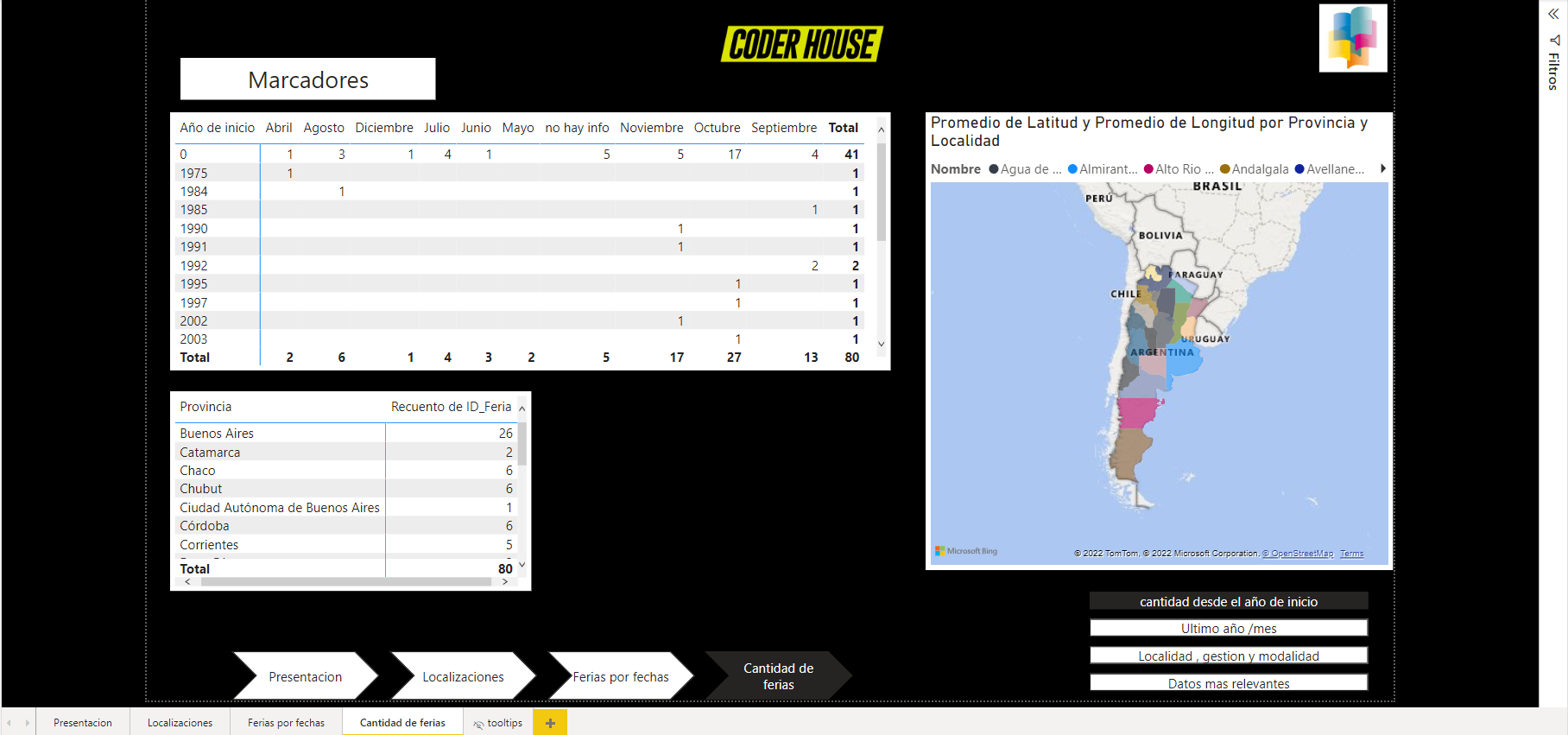
Podemos observar la siguiente segmentación:

Usamos los datos para poder ordenar respectivamente por el nombre de las ferias también su modalidad, pero lo importante es mostrar el año de inicio, el mes y su último año realizada hasta la fecha de este informe (29/08/2022).

Podemos encontrar el gráfico de “Tarjeta” el cual tiene los datos de la tabla fechas y columna Mes. La siguiente gráfica “Tarjeta de varias filas” que posee los datos de las tablas Provincia con su columna localidad y de la tabla Fechas último año de actividad.

Por último, volvimos a utilizar el grafico de “Tabla” para poder mostrar de la tabla fechas la columna Año de inicio y de la tabla FeriaLibro utilizamos las columnas Tipo de gestión y Nombre y modalidad, así desde este gráfico podemos interactuar para ver por el nombre o fecha la feria que queramos dándonos así los datos de su último año, mes, localidad, gestión y modalidad.

1. **En la tercer y última solapa “Cantidad de ferias”**



En esta última solapa intentamos poder explayar en su totalidad la información que no pusimos en la anterior como así también implementar la herramienta de marcadores podemos ver la cantidad de ferias realizadas a lo largo de su año de inicio como de su último registro, como también su mes, su modalidad y tipo de gestión. Para hacerlo de manera más prolija utilizamos los marcadores los cuales los dividimos en 4 partes.

Las primeras dos las incluimos con un gráfico de mapa coroplético donde podemos ver a partir del año la provincia con su localidad también sumamos el dato que poseemos de latitud y longitud donde se realizó dado que en la primera solapa no mostramos ese dato.

En nuestro primer marcador el cual se llama “Cantidad desde el año de inicio “contiene 3 gráficos distintos 2 de matriz y uno de mapa coroplético.

En los gráficos de matrices están compuestos de la siguiente manera:

En el gráfico superior encontramos de tabla “Fechas” las columnas, año de inicio,mes ,como también la de ID\_Feria.

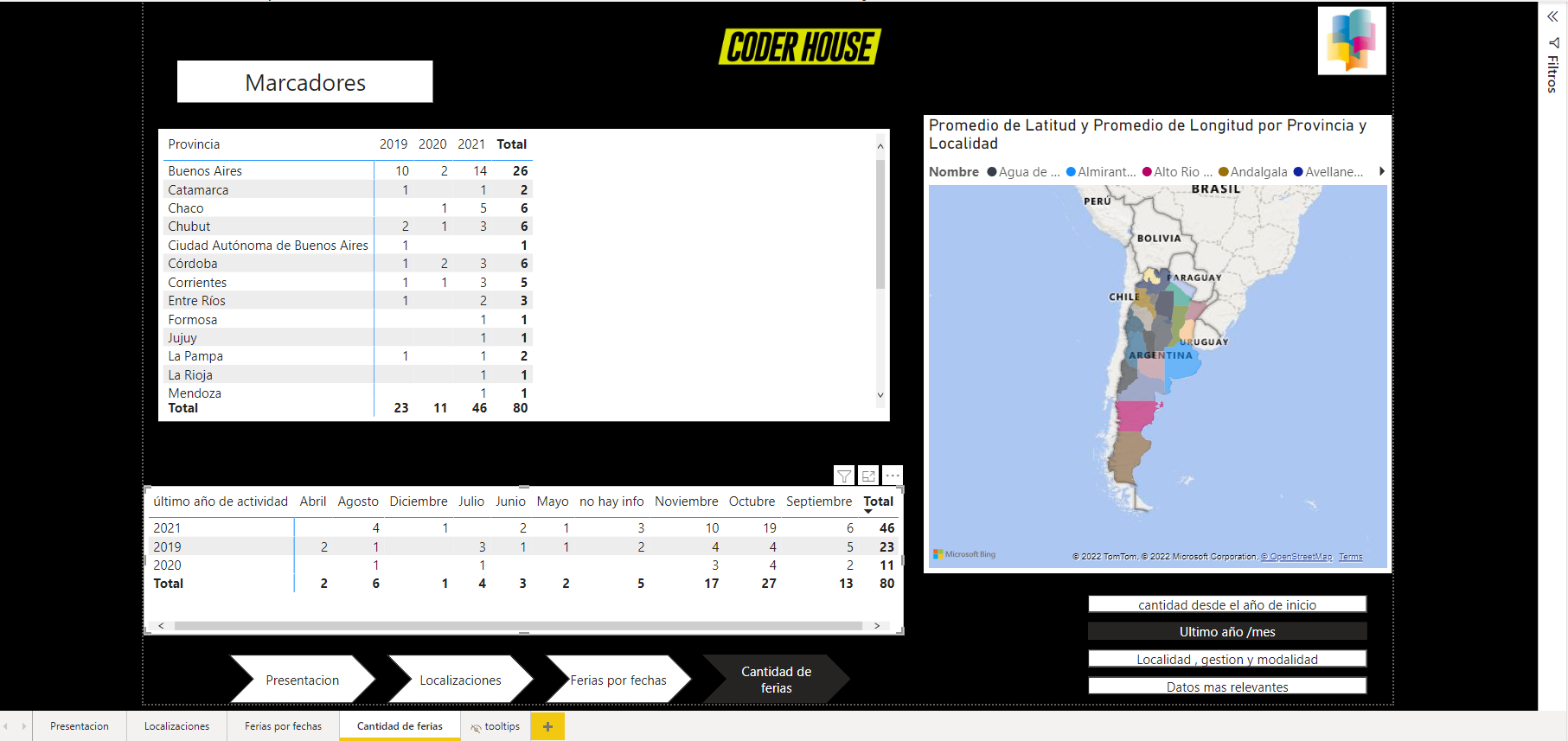
Luego en el de abajo tenemos de la tabla “Provincia” la columna Provincia y también de la tabla “Fechas”, ID\_Feria.

En el siguiente gráfico de mapa podemos encontrar de la tabla “Provincia” los datos de “Provincia” y “Localidad”, de la tabla “Información” tenemos el dato de “Latitud “y “Longitud”.

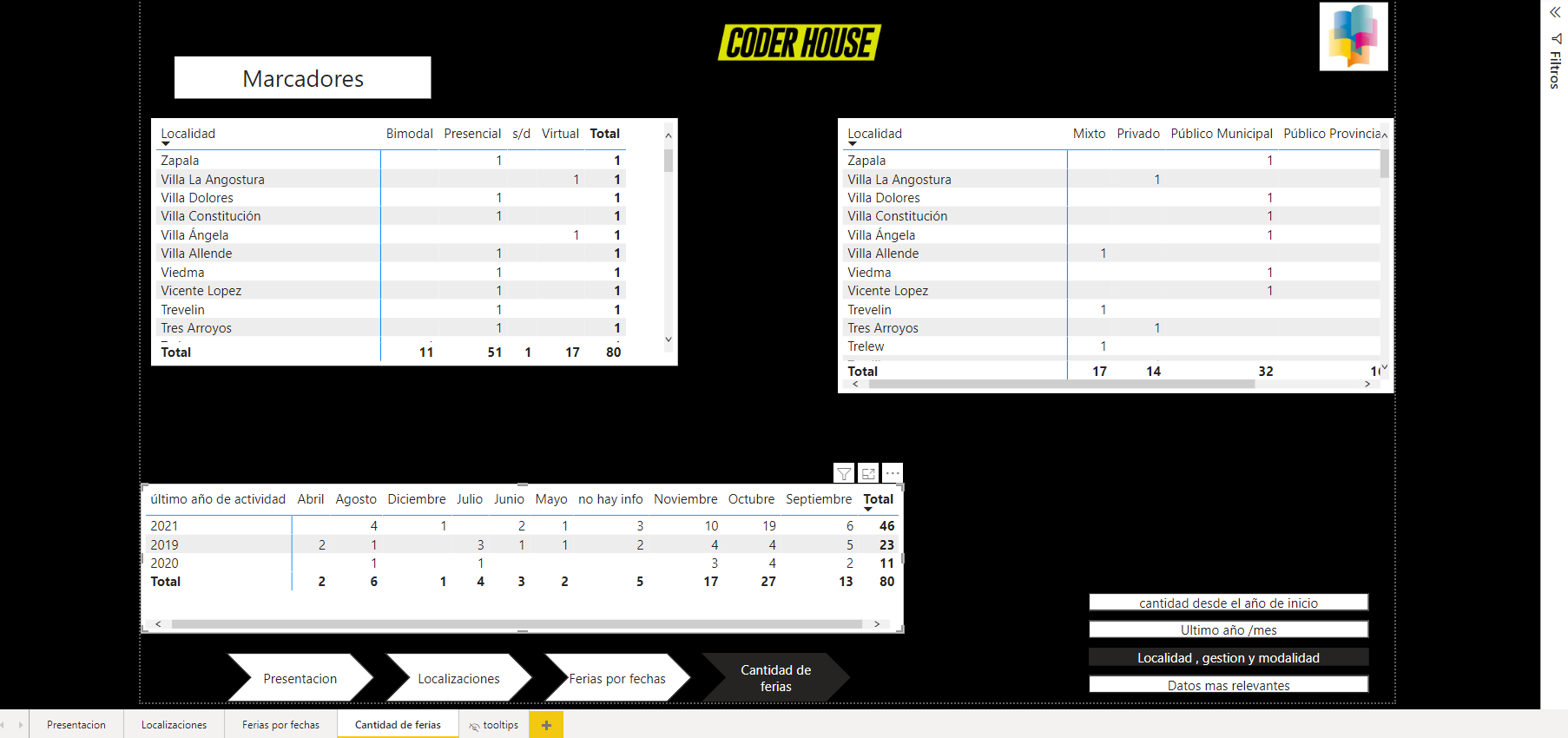
El primer objetivo de este marcador es mostrar cuantas ferias se hicieron durante sus años de inicio a lo largo de los meses en las provincias, lamentablemente vemos que tenemos el año “0” como también en la parte del mes “No hay info” esto lo tuvimos desde la base de datos original como también cuando la limpiamos antes de pasarla al tablero, tanto el año como el mes no tenemos los datos de esos momentos, pero si en la provincia que se hicieron por ejemplo por eso los mantuvimos.

En este segundo marcador “Último año /mes” poseemos otra vez como en el primero 2 gráfico de matriz y 1 de mapa coroplético. El objetivo es el mismo que el primero donde buscamos mostrar e identificar la cantidad de ferias realizadas hasta su último año con su mes y en la provincia con su respectiva localidad.

Son las mismas tablas y columnas que el anterior marcador solo que varía el del segundo gráfico de matriz que se muestra por debajo de él de provincia.Para ello usamos de la tabla “Fechas” las columnas “Último año de actividad “, “mes” y “ID\_Feria” de la tabla “FeriaLibro”.



Para el tercer marcador “Localidad, gestión y modalidad” quisimos apuntar a mostrar más en profundidad la cantidad de ferias realizadas pero esta vez con su tipo de gestión y modalidad en las localidades que se llevaron a cabo durante su último año de actividad.



En este utilizamos 3 gráficos de matrices

El primero que sería el de la izquierda superior está conformado:

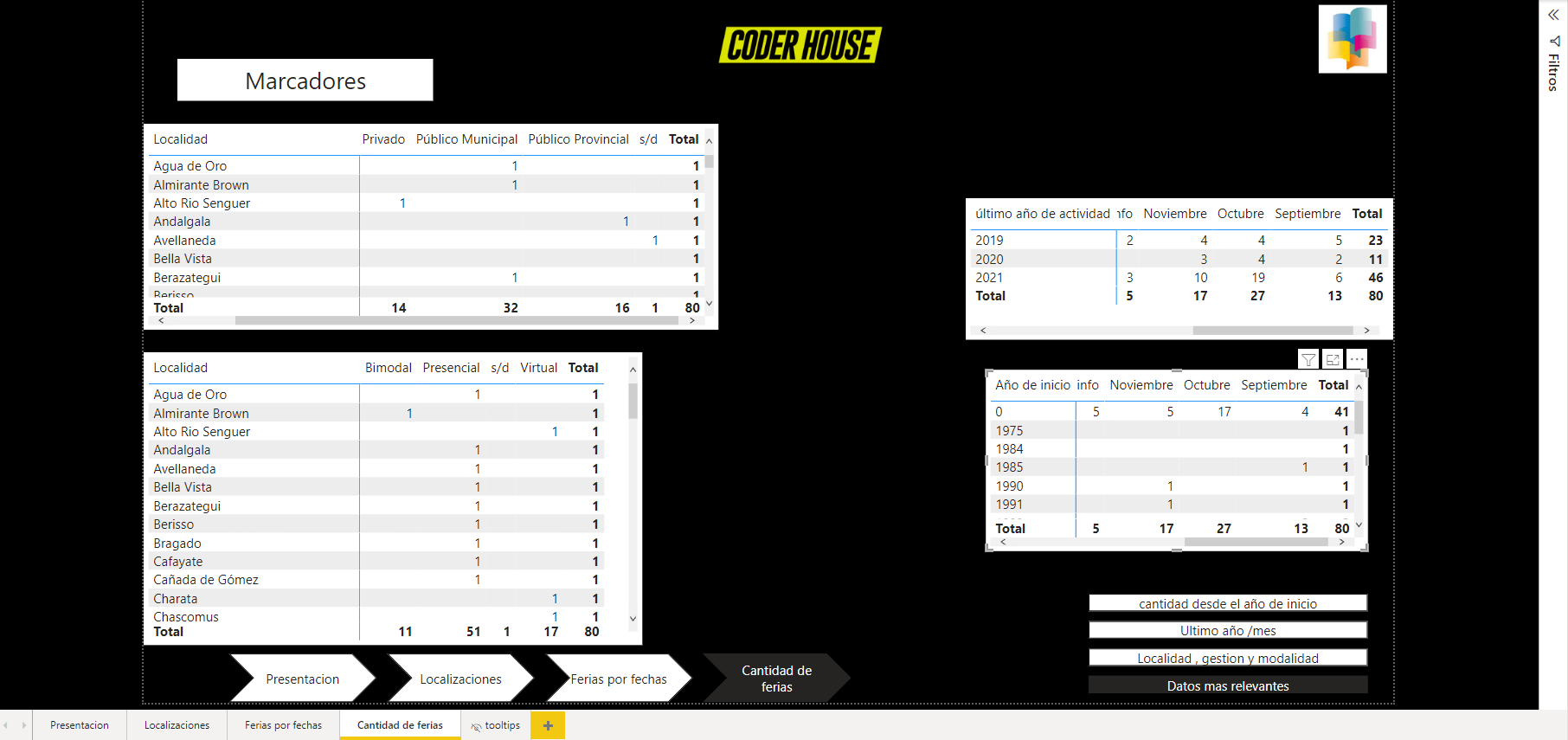
Por las tablas “FeriaLibro” y “Provincia”.

Donde de la primera usamos las columnas “ID\_Feria”, “Modalidad” y de la segunda solamente la de “Localidad”.El gráfico que se encuentra por debajo mantiene los mismos datos de los marcadores 1 y 2 .

Por último, el gráfico que se posiciona en el lado derecho contiene de las tablas “Ferialibro” y “Provincia” de las cuales las columnas utilizadas fueron “ID\_Feria”, “Tipo de gestion” y nuevamente “Localidad”.

Ya en el último marcador “Datos más relevantes “quisimos mostrar de los marcadores anteriores y sumando un gráfico más de matriz es poder observar en un solo espacio por el tipo de modalidad, gestión, año de inicio, mes y último año de actividad toda la cantidad de ferias realizadas hasta el momento sin dejar ningún dato suelto de los que tenemos en disposición para el tablero.

Poseen las mismas tablas y columnas utilizadas en los anteriores marcadores.



# 12) Tooltips

Utilizamos 2 páginas de tooltips diferentes para nuestro tablero los cuales se pueden ver en los gráficos de “mapa” donde utilizamos el primero para tipo de Latitud - Longitud ya que nos quedaba afuera este dato y se nos vio conveniente instalarlo de esta manera, para el mapa coroplético para el de nombre de ferias así queda a la vista con más detalle las mismas.

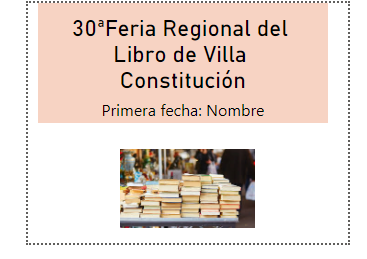
En esta primera imagen vemos el tooltip creado para el gráfico de mapa:



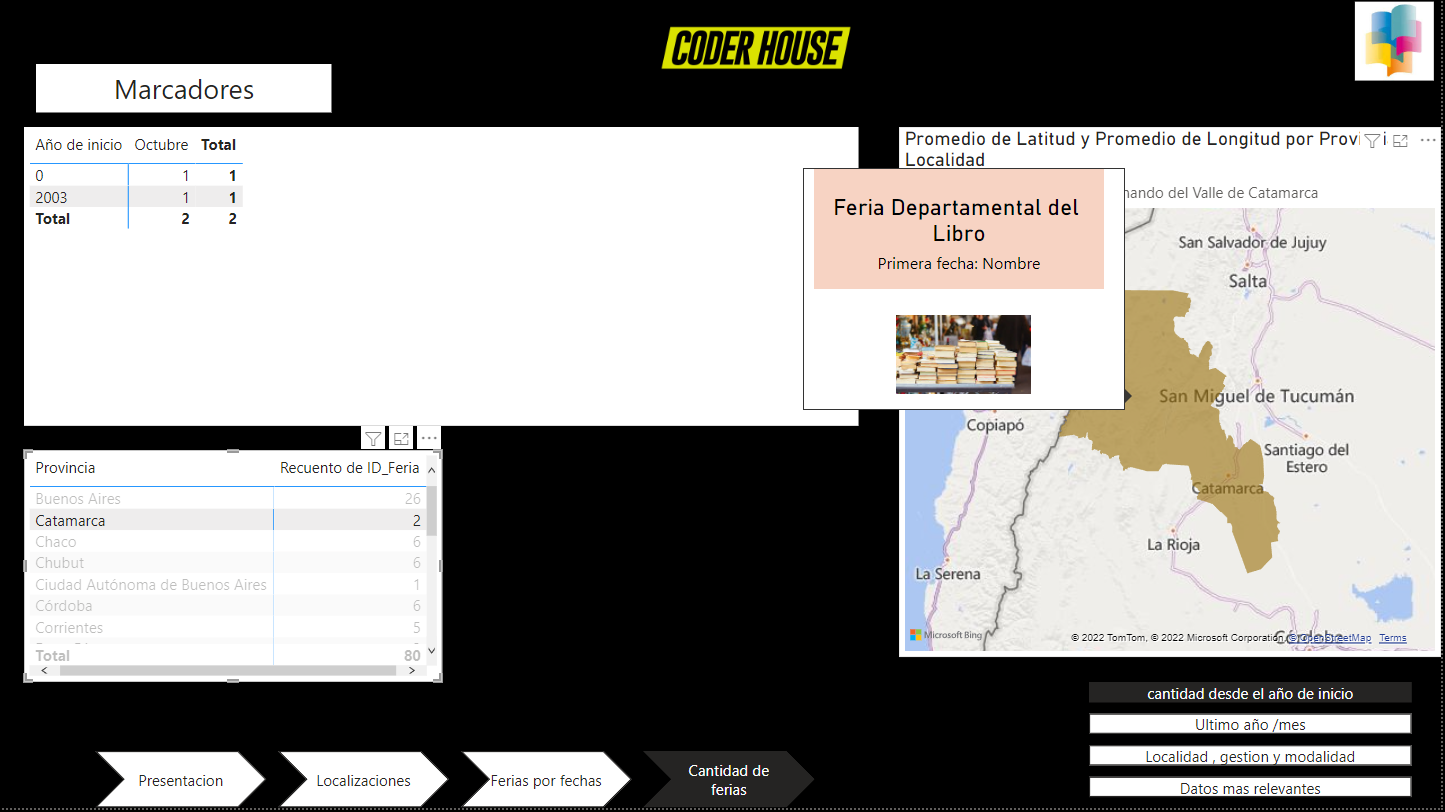
Y aquí en la otra en otra imagen su uso:



Por último, el segundo tooltip:



Y su uso:



# 13) Conclusión

Se pudo observar la cantidad de ferias realizadas en los diferentes años tanto de inicio como su último año de actividad, lo que nos llamó la atención fue en el año 2021 se llevaron a cabo un total de 46 ferias un número que supera tanto a las del 2019 (23) luego del 2020 (que tuvo 11) que fue el comienzo de la pandemia lo cual demuestra como el impulso hacia la reactivación social, cultural y económica de estas festividades.

Como también en la Provincia de Buenos Aires es donde más ferias se realizan (26 si contamos los últimos 3 años ) podemos pensar como es la capital de Argentina llega a movilizar muchas más que en otras en cuanto al turismo y la divulgación mediante los diferentes tipos de publicidad (Redes sociales , propagandas por radio y televisión , página web del gobierno de la ciudad y carteles publicitarios ) ,también se llevó a cabo la Primer Feria Internacional del Libro en el año 1975 en la ciudad de Buenos Aires .

Lamentablemente no contamos con información de ventas/compras, ingreso de personas durante los días de las festividades, editoriales que hayan asistido dejándonos así sin poder hacer un análisis más profundo aparte de no poder generar medidas calculadas. También nos encontramos con cierta falta de datos sobre algunos años de inicio y mes por lo cual no sabemos en qué momento se iniciaron, pero si tenemos su última fecha de actividad por lo cual pudimos llevar adelante el TP.

# 14) Futuras Líneas

Podemos tomar en cuenta a futuro que en la base de datos original que intenten buscar el registro de las ferias que figuran sin año de inicio como también así el dato del mes en el que se hizo, aparte también se podría agregar datos sobre el costo de la feria, cuanto se recaudó, cuantos libros en total se vendieron en las ediciones, cuantas personas concurrieron en las mismas entre otras sugerencias para poder utilizar otro tipo de herramientas dentro del tablero como el generar medidas calculadas , mejor dicho poder trabajar con datos numéricos para así darle un análisis mas profundo y completo.

# 15) Bibliografía

<https://www.datos.gob.ar/dataset/cultura-mapa-cultural-agentes-actividades-culturales/archivo/cultura_eb517a71-ccc5-400d-b489-455bccbbb0b4>

Fecha de consulta 20/06/2022.