

# **Estudio de ensayos clínicos asociados al cáncer en Estados Unidos.**

Nombre: Gervé, María Paula.  
Fecha:30/03/2023

## Índice:

- 1- Descripción de la temática.
- 2-Descripción de la hipótesis.
- 3- Objetivo del proyecto
- 4-Usuario final y nivel de aplicación del análisis.
- 5-Base de Datos
- 6-Listado de tablas y columnas
- 7-Diagrama de entidad-relación
- 8-Modelo relacional en Power BI
- 9-Visualización de Datos
  - 9.1-Carátula
  - 9.2-Datos censales
  - 9.3-Estudios Clínicos
  - 9.4-Estimación de eficiencia de estudios.
- 10-Medidas Calculadas
- 11-Conclusión

### 1-Descripción de la temática:

La base de datos seleccionada describe un análisis de datos censales relacionados con patologías de cáncer de Estados Unidos.

### 2-Descripción de la hipótesis:

El término cáncer es el nombre común que recibe un conjunto de enfermedades relacionadas en las que se observa un proceso descontrolado en la división de las células del cuerpo, lo cual si no se recibe tratamiento y según la malignidad del cáncer puede provocar a una desestabilización general del organismo como también a la muerte. Actualmente es una de las enfermedades que más preocupan a las sociedades.

Los cánceres más comunes (en orden descendente según el número estimado de casos nuevos en 2020) son: el cáncer de seno (mama), el cáncer de pulmón y bronquios, el cáncer de próstata, el cáncer de colon y recto, el melanoma (un cáncer de piel), el cáncer de vejiga, el linfoma no Hodgkin, el cáncer de riñón y pelvis renal, el cáncer de endometrio, la leucemia, el cáncer de páncreas, el cáncer de tiroides y el cáncer de hígado.

El cáncer es una de las causas principales de muerte en el mundo. En 2018, hubo 18,1 millones de casos nuevos y 9,5 millones de muertes por cáncer en el mundo. Se anticipa que en 2040 el número de casos nuevos de cáncer por año aumentará a 29,5 millones y el número de muertes por cáncer a 16,4 millones. Fuente: [Agencia Internacional para la Investigación del Cáncer](#).

Como hipótesis de trabajo se plantea analizar posibles relaciones censales como porcentaje de pobreza con porcentaje de muertes por cáncer como también, con cantidad de estudio clínicos que se llevan a cabo, y ver como influyen los factores sociales en el desempeño de ésta enfermedad.

### 3-Objetivo del proyecto.

Como hipótesis de trabajo se plantea analizar posibles relaciones censales como porcentaje de pobreza con porcentaje de muertes por cáncer como también, con cantidad de estudio clínicos que se llevan a cabo en regiones de los Estados Unidos realizados hasta el año 2015, y ver como influyen los factores sociales en el desempeño de ésta enfermedad. Además, se plantea analizar el avance de los estudios clínicos, en que etapa de estudio están, y características de los voluntarios que presentan los estudios.

### 4-Usuario final y nivel de aplicación del análisis:

Estratégico: El nivel de estratégico que se eligió realizar en este análisis permite orientar o destinar esfuerzos en el área de la clínica hacia las regiones que contienen mayor número de muertes por cáncer como también más desfavorecidas a nivel económico. Este análisis está destinado a personas con poder de decisión que tomará decisiones a largo plazo como pueden ser directores de centros de salud, como también dirigentes políticos del ámbito de la salud.

### 5-Base de Datos

La base de datos seleccionada contiene 3 tablas en formato csv.

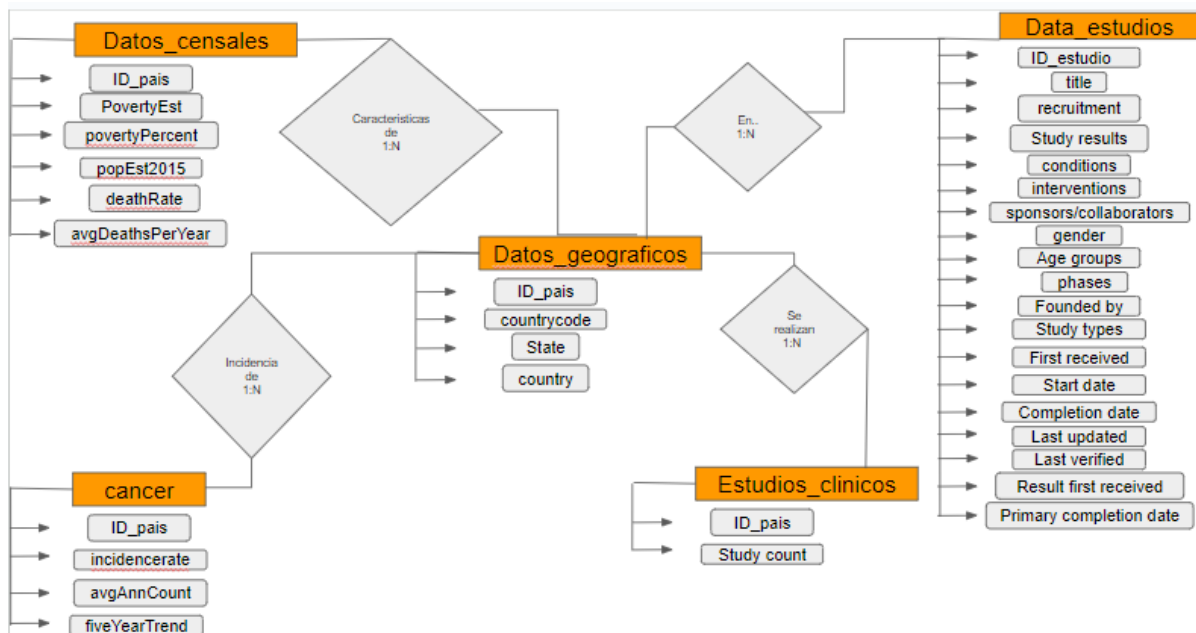
### 6-Listado de tablas y columnas:

Las tablas que se encontraban inicialmente en la tabla de datos se reorganizaron de la siguiente manera:

Data\_Estudios: Contiene información detallada de cada estudio clínico, grado de avance, fase en la que se encuentra, fechas de inicio y fin, como también características de los voluntarios participantes, genero , edad, entre otros.

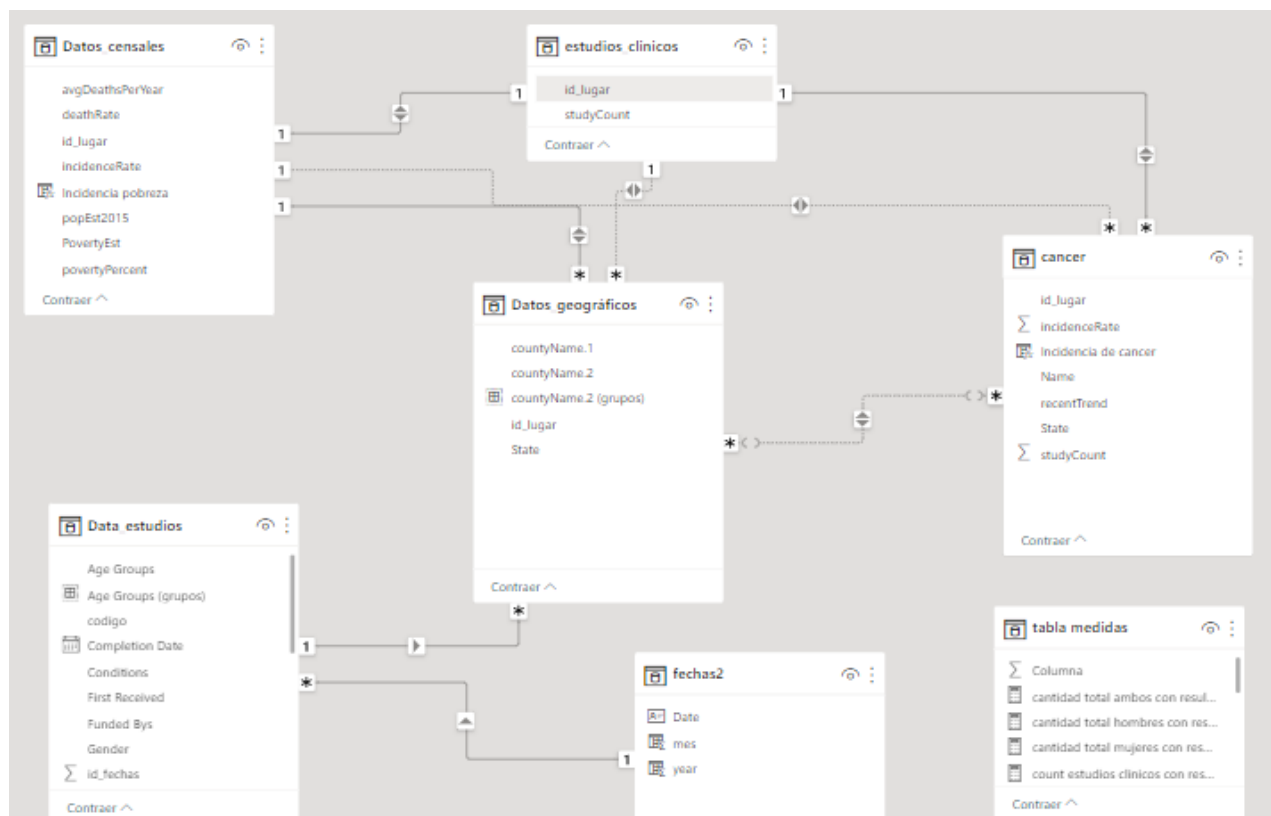
Data_Estudios	tipo de campo	tipo de clave	Datos_geograficos	tipo de campo	tipo de clave
ID_estudio	varchar	PK	Id_pais	int	
Title	varchar		State	varchar	PK,FK
Recruitment	varchar		County	varchar	
Study Results	varchar				
Conditions	varchar				
Interventions	varchar		Datos_censales	tipo de campo	tipo de clave
Sponsor/Collaborators	varchar		Id_pais	int	PK,FK
Gender	varchar		PovertyEst	int	
Age Groups	varchar		povertyPercent	int	
Phases	varchar		popEst2015	int	
Funded Bys	varchar		deathRate	int	
Study Types	varchar		avgDeathsPerYear	int	
First Received	date				
Start Date	date		cancer	tipo de campo	tipo de clave
Completion Date	date		Id_pais	int	PK,FK
Last Updated	date		incidenceRate	int	
Last Verified	date		avgAnnCount	int	
Results First Received	date		fiveYearTrend	int	
Primary Completion Date	varchar				
codigo	int	,FK	Estudios_clinicos	tipo de campo	tipo de clave
			Id_pais	int	PK,FK
			studyCount	int	

## 7-Diagrama entidad-relación.



## 8-Modelo relacional en Power BI

Una vez que los archivos planos fueron subidos a Power BI, se hicieron los siguientes cambios en los datos:



### Tabla Cancer:

- Se agrega la columna "incidencia de cancer" para calificar si la incidencia es alta o baja, según el valor de cada entrada sea mayor o igual, o menor que el promedio total.

Incidencia de cancer = `if(cancer[incidenceRate]>=AVERAGE(cancer[incidenceRate]),"Alta","Baja")`

#### Tabla Datos\_censales:

- Se agrega la columna "incidencia pobreza" para calificar si la incidencia es alta o baja, según el valor de cada entrada sea mayor o igual, o menor que el promedio total.

incidencia pobreza = IF(Datos\_censales[PovertyEst]>=AVERAGE(Datos\_censales[PovertyEst]),"Alta","Baja")

#### Tabla Datos\_geográficos:

- Se separa la columna de nombre de estado por medio del separador coma, de esta manera contengo una columna de estado, y otra de las localidades contenidas por cada estado.

Se creó además la tabla calendario que contiene fechas , y una tabla donde se depositaron las medidas mas adelantes descriptas.

Por otro lado, el diagrama entidad-relación quedó graficado en Power BI como se puede observar a continuación:

#### 9-Visualización de los datos

La visualización del tablero partió de un mockup que luego fue llevado a Power BI. A continuación, se detalla un breve resumen de cada una de las páginas creadas:

- Página de Caratula.

-Solapa 1:Datos censales:

Se creó esta página con el fin de obtener una rápida visualización de si los niveles de pobreza se relacionan con una alta incidencia de cáncer, y con la cantidad de estudios realizados por ubicación geográfica.

-Solapa 2:Estudios clínicos:

En esta parte se realizó con el objetivo de analizar la información relacionada a los estudios clínicos que se encuentran en ensayo.

-Solapa 3:Estimación de eficiencia de estudios:

En esta tercer solapa se tiene en cuenta como podría variar a futuro los datos si los tratamientos clínicos funcionaran con diferentes porcentajes promedios de eficiencia.

## 9.1. Carátula

### Estudio de ensayos clínicos asociados al cáncer en Estados Unidos.

1. Datos Censales



2. Estudios Clínicos



3. Estimación de eficiencia de estudios



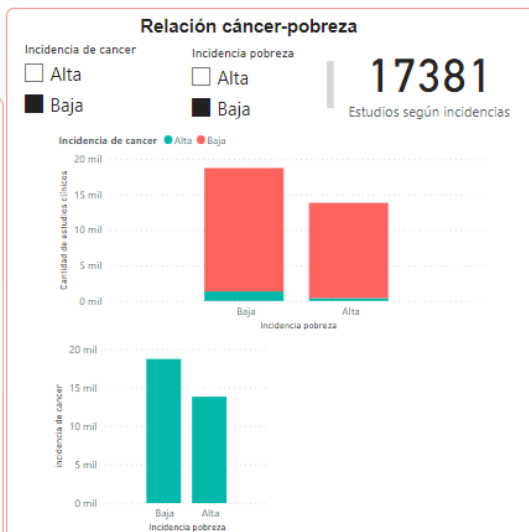
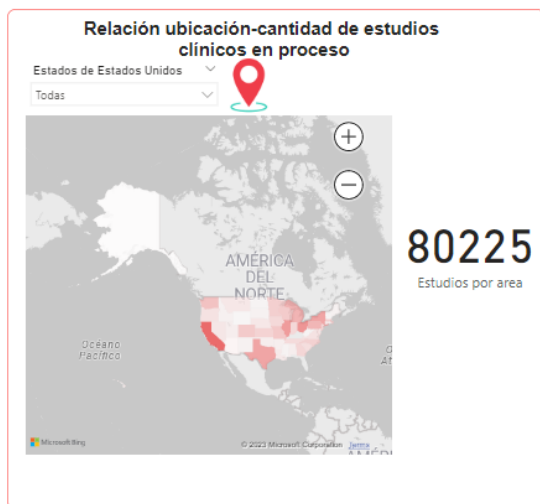
La caratula contiene un índice con íconos que contienen acción y permite dirigirnos a cada una de las secciones. Luego, cada una de las solapas, vuelve a tener los íconos con acciones para movernos en cada una de las solapas.

## 9.2 Datos censales.

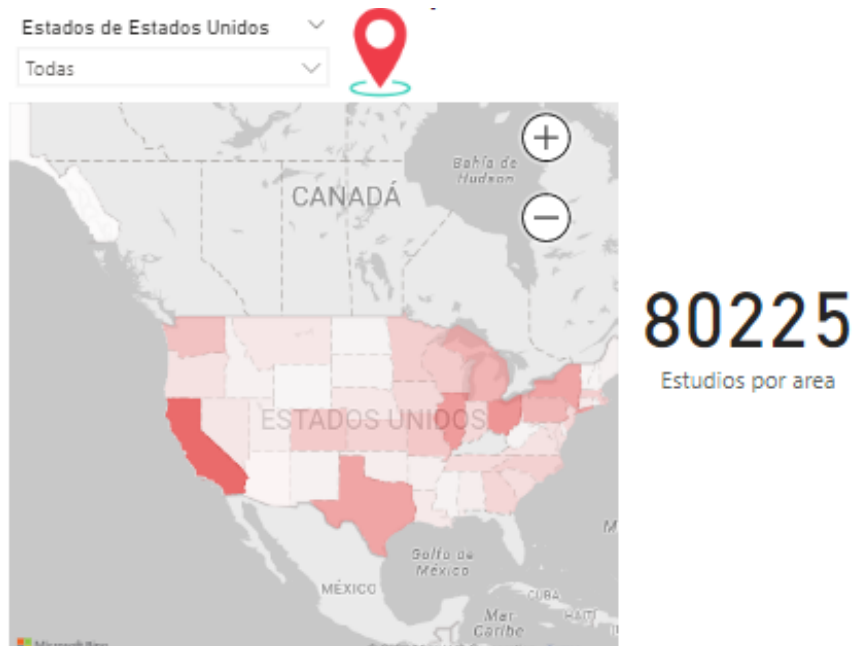
La solapa de Datos censales se divide en dos. Por un lado se muestra un análisis de la relación ubicación-cantidad de estudios clínicos que se encuentran en proceso, para esta parte, tenemos un segmentador de Estados de los Estados Unidos que permite dividir el análisis según la ubicación. Este segmentador es desplegable, y permite una selección múltiple.



### 1. Datos Censales



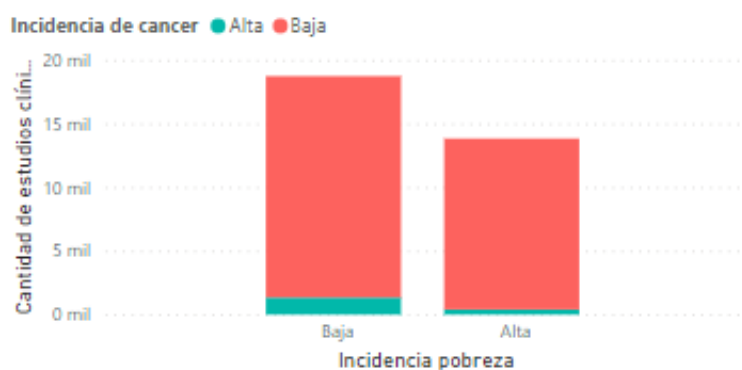
El mapa a su vez se encuentra contiene un degradado de color según la cantidad de estudios clínicos hechos o en proceso de cada Estado. A su vez hay una tarjeta a la derecha que indica la cantidad de estudios de la región seleccionada.



La segunda parte de la solapa contiene un análisis de la relación pobreza-cáncer. Para eso se colocó un segmentador que nos permite elegir las incidencias de pobreza y cáncer (Bajas o altas), y una tarjeta asociada que indica la cantidad de análisis clínicos en estudio según la característica seleccionada.

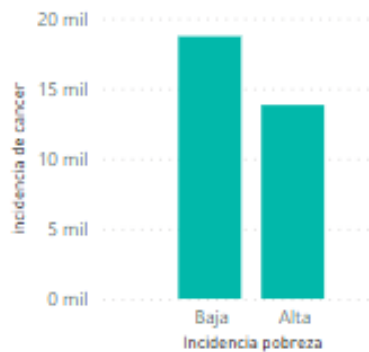


Además se realizó un gráfico donde se puede ver las relaciones visualmente de éstas incidencias en la cantidad de ensayos clínicos.

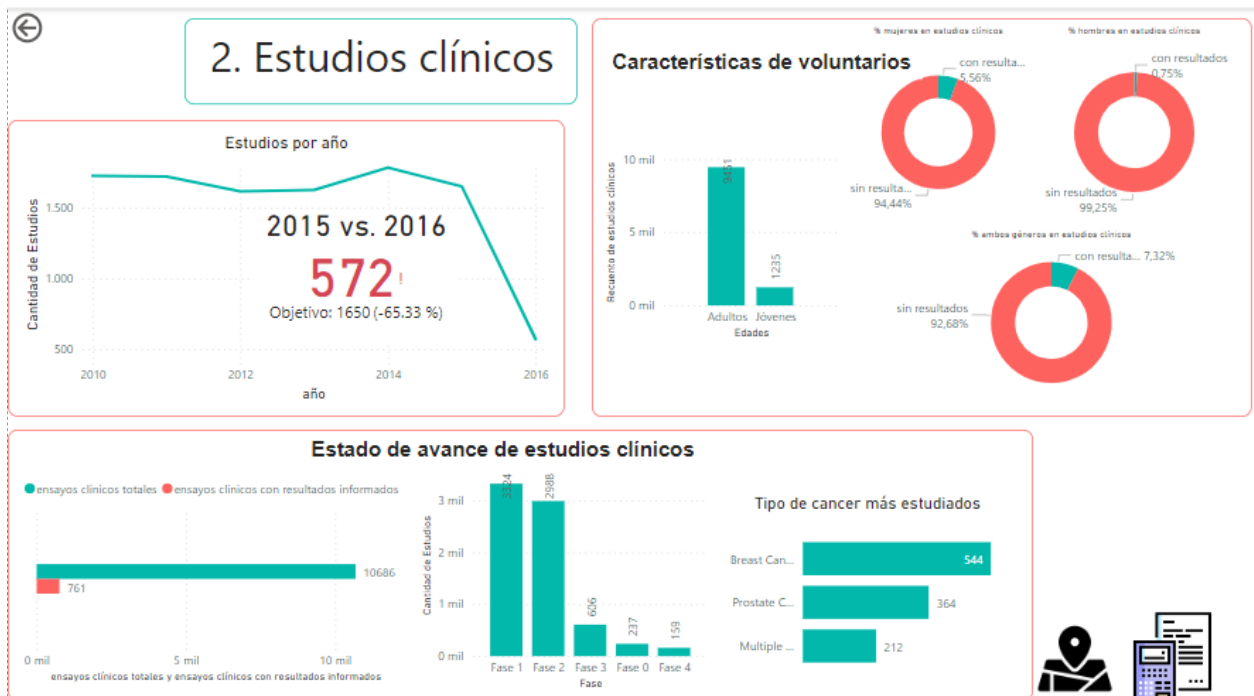




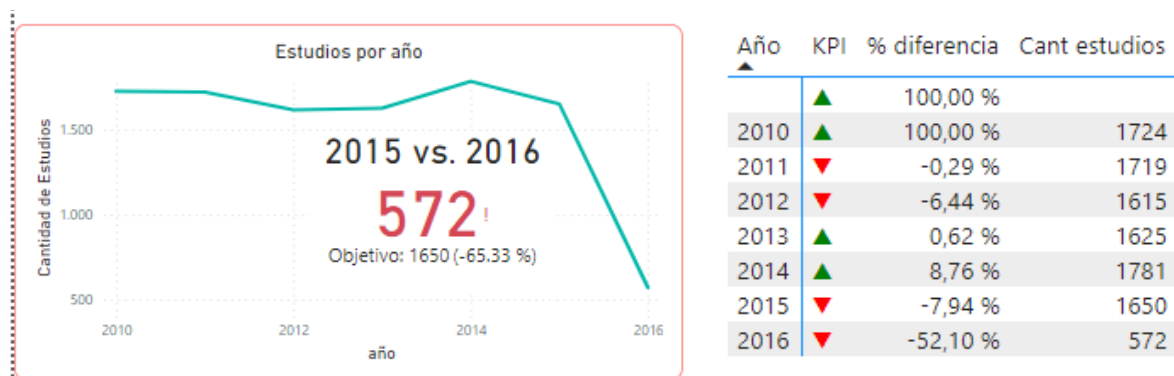
Por último, vemos en el gráfico de barras, la relación que existe entre incidencia de pobreza con incidencia de cáncer.



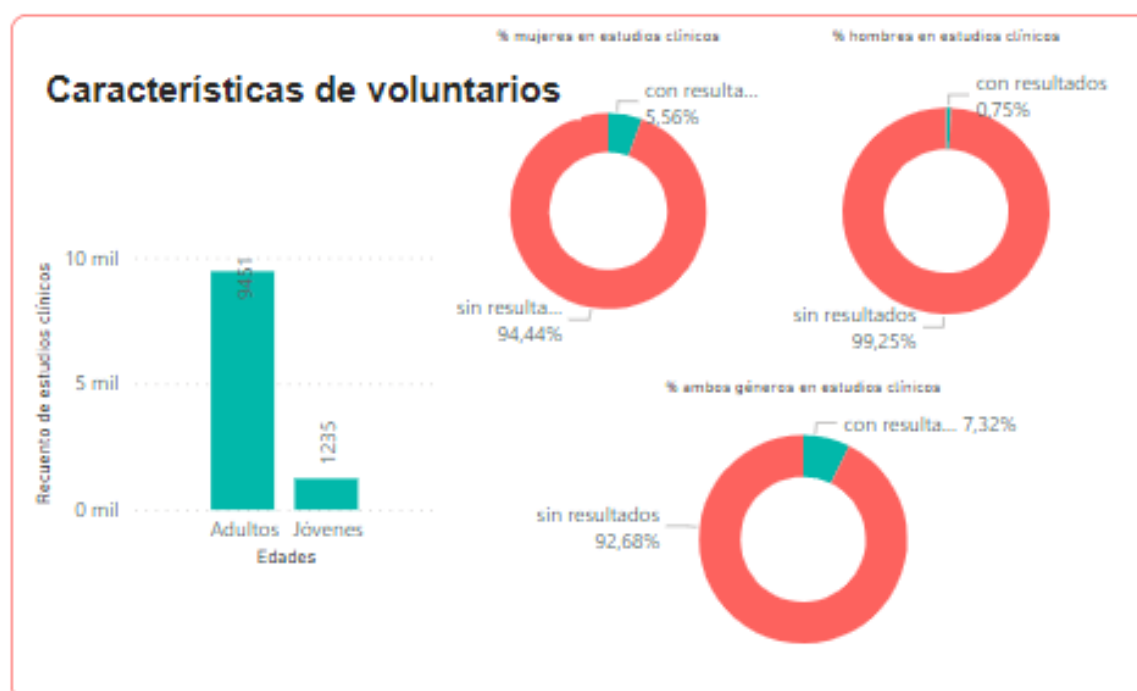
### 9.3 Estudios clínicos.



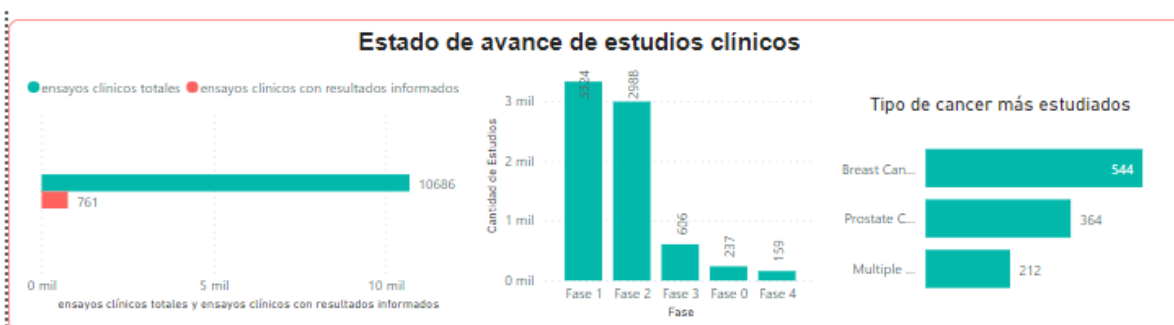
Esta solapa contiene por un lado, un gráfico de líneas de la evolución de la cantidad de los estudios clínicos por años, para ello se necesitó la creación de una tabla calendario. Debido a que es llamativa la baja de cantidad entre 2015 y 2016, se realizó un gráfico de KPI 2015 vs 2016 para destacar ese valor. A su vez, el gráfico de línea contiene un tooltip asociado en donde, mediante un cálculo de inteligencia de tiempo, se calculó el porcentaje de evolución año a año, y se determinó si aumenta o disminuye la cantidad de estudios, tomando para cada año su respectivo año anterior como target del KPI.



En la segunda parte de la solapa, se analizan las características de los voluntarios de los estudios clínicos, edades, género según estudios con y sin resultados actualmente publicados.



La tercer parte de la solapa, analiza el estado de avance de los estados clínicos, cantidad de estudios que actualmente han informado los resultados, los principales tipos de cánceres que se han abordado en los ensayos, y también cantidad de estudios en cada una de las fases, siendo la fase 0 la primer fase de estudio, y fase 4 la etapa final y más avanzada en una investigación clínica.



9.4 Estimación de eficiencia de estudios.

3. Estimación de eficiencia de estudios

Impacto de una futura eficiencia de los estudios clínicos en tasa de muerte e incidencia de cáncer



En esta solapa se analiza una estimación en donde, teniendo una mirada optimista a futuro podemos analizar en que medida la tasa de muerte y tasa de incidencia de cáncer. Para ello se realizó una medida de parámetro para estimar estos valores en función de 2 segmentadores.

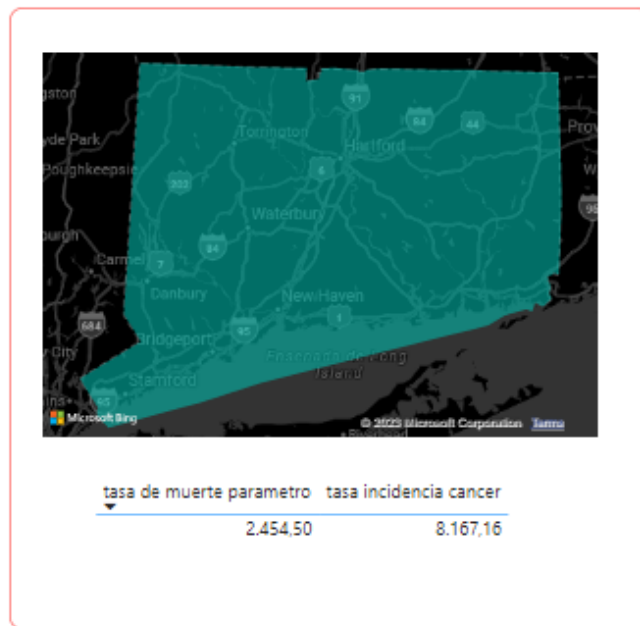
Mosaico: me permite seleccionar el valor de la futura predictiva eficiencia que se puede lograr con ensayos clínicos.



Lista vertical: de estados geográficos, me permite seleccionar ademas de la eficiencia la ubicación para ver los resultados generados.

- ☐ Indiana
- ☐ Iowa
- ☐ Kansas
- ☐ Kentucky
- ☐ Louisiana
- ☐ Maine
- ☐ Maryland
- ☐ Massachusetts
- ☐ Michigan

Los resultados estimados se observan en una doble tarjeta, y la ubicación se señala en el mapa.



#### 10-Medidas calculadas

Se calcularon las siguientes medidas y se agregaron a la 'tabla medidas':

1.Cantidad total ambos con resultados.

cantidad total ambos con resultado =  
`CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Both",Data_estudios[Study Results]="Has results")`

2.Cantidad total hombres con resultados.

cantidad total hombres con resultado =  
`CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Male",Data_estudios[Study Results]="Has results")`

3.Cantidad total mujeres con resultados.

cantidad total mujeres con resultado =  
`CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Female",Data_estudios[Study Results]="Has results")`

4. Count estudios clinicos con resultados:

count estudios clinicos con resultados = `(CALCULATE(COUNT(Data_estudios[Study Results]),Data_estudios[Study Results]="Has Results"))`

5.estudios2015 = `calculate(count(Data_estudios[codigo]),fechas2[year]=2015)`

6.estudios2016 = `calculate(count(Data_estudios[codigo]),fechas2[year]=2016)`

7.KPI = `if([Porcentaje de diferencia] >=0,UNICHAR(9650),UNICHAR(9660))`

8.KPI Color = `if([Porcentaje de diferencia]>=0,"Green","Red")`

9. Porcentaje de diferencia =  $\frac{A}{B}$

```

CALCULATE(count(Data_estudios[codigo]),DATEADD(fechas2[Date],-1,YEAR))
var B = calculate(count(Data_estudios[codigo]))
return 1-(A/B)

```

10. promedio tasa de muerte =  $\text{calculate}(\text{AVERAGE}(\text{Datos\_censales}[\text{deathRate}])))$

11. resta total menos con resultado ambos =  $\text{VAR}$  cantambos =  $\text{VAR}$

```

CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Both")
cantambosResultados= CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Both",Data_estudios[Study Results]="Has results") return cantambos-cantambosResultados

```

12. resta total menos con resultado hombres =  $\text{VAR}$  canthombres =  $\text{VAR}$

```

CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Male")
canthombresResultados= CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Male",Data_estudios[Study Results]="Has results") return canthombres-canthombresResultados

```

13. resta total menos con resultado mujeres =  $\text{VAR}$  cantmujeres =  $\text{VAR}$

```

CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Female")
cantmujeresResultados= CALCULATE(COUNT(Data_estudios[codigo]),Data_estudios[Gender]="Female",Data_estudios[Study Results]="Has results") return cantmujeres-cantmujeresResultados

```

14. tasa de muerte parametro =  $\frac{[\text{promedio tasa de muerte}]}{[\text{Valor eficiencia futuros tratamientos clínicos}]} * 100$

15. tasa incidencia cancer =  $\frac{\text{AVERAGE}(\text{cancer}[\text{incidenceRate}])}{[\text{Valor eficiencia futuros tratamientos clínicos}]} * 100$

16. Valor eficiencia futuros tratamientos clínicos =  $\text{SELECTEDVALUE}('eficiencia futuros tratamientos clínicos'[\text{eficiencia futuros tratamientos clínicos}])$

### 11-Conclusiones.

El cáncer es una de las enfermedades más influyentes en la actualidad alrededor del mundo. Abordar estrategias para tratar esta problemática sería muy importante particularmente para países tan relevantes y de gran población como Estados Unidos.

En el análisis de datos censales, se analiza que existen actualmente una gran cantidad de estudios clínicos en proceso distribuidos aleatoriamente; hay regiones con una gran cantidad mientras que en otras poseen muy baja cantidad de ensayos. En relación a los índices de pobreza y cáncer según el censo anual se ve que, como se suponía, que a mayor porcentaje de pobreza y de cáncer, menor cantidad de estudios clínicos asociados. De todas formas esta medida no es preocupante, ya que, los resultados de los estudios clínicos una vez que finalicen serán utilizados en cualquier parte del mundo sin importar la región donde se realizó el estudio. Además, era esperable que en regiones de mayor pobreza haya menos presupuesto destinado a la investigación. Por otro lado, un análisis que sí es interesante tener en cuenta es el gráfico que muestra que cuando la incidencia de pobreza es alta, también lo es la incidencia de cáncer. Éste gráfico nos puede ayudar a tomar acciones al respecto, si bien no es necesario que se invierta en investigaciones clínicas en regiones vulnerables, sí es indispensable invertir en infraestructura de los sistemas de salud, prevención, etc.

En el análisis de estudios clínicos, se analiza por un lado, los voluntarios que participan, en donde como era de esperar en la mayoría de los estudios participan adultos y jóvenes mayores de edad con consentimiento informado. También se analiza que hay estudios en donde participan hombres, en otros mujeres y en otros ambos sexos. En el gráfico de la evolución de la cantidad de estudios clínicos a lo largo de los años, se puede observar que, medianamente la cantidad de estudios varía muy poco, sin embargo es llamativo que el último año hubo una disminución abrupta de la cantidad de estudios clínicos. Si bien lo ideal sería que cada año se mantenga o aumente la inversión en investigación, también hay que tener en cuenta que son estudios que llevan muchos años, y recién muchos años posteriores al inicio de estudio se va a poder reflejar un posible resultado alentador, con lo cual a pesar de la menor cantidad de ensayos realizados en 2016, es posible tomar decisiones que permitan revertir esta situación de menor inversión en investigación científica asociada al cáncer de cara al futuro. Por último, se analiza el grado de avance de los ensayos clínicos, como era de esperar por la cantidad de años que llevan estas investigaciones como también las aprobaciones de los tratamientos, la mayoría de ellos no han informado resultados aún, y la mayoría de los estudios se encuentran en la fase 1 y 2.

Además, se realizó una estimación de la tasa de muerte y de cáncer según la eficiencia de los futuros ensayos clínicos, a partir de estos resultados como conclusión general vemos la importancia de que en la medida que el presupuesto destinado a salud lo permita, se invierta en salud e investigación de enfermedades que hasta el momento no tienen cura, eso evitaría las grandes tasas de incidencia y muerte, particular y preocupantemente de las regiones más vulnerables y en estado de pobreza, ya que, estas regiones son las más afectadas y presentan mayor incidencia de la enfermedad. Por tales motivos, se espera que luego de éste análisis se reestablezca el presupuesto que se venía invirtiendo en salud antes de 2015, ya que se produjo una baja notoria que podría afectar el avance que se venía teniendo anteriormente, y cuyos resultados se esperan para el futuro.