

Universidad de Buenos Aires

Facultad de Ciencias Exactas y Naturales

Departamento de Computación

Laboratorio de Datos

Trabajo Práctico 01 - Curso de Verano 2026

Análisis relacional entre las defunciones, las variables demográficas y el acceso a la salud de la República Argentina

Nombre del Grupo: labubuntu

Integrantes	LU
Lanabere, Delfina Daniela	246/24
Muhafra, Micaela Abril	1327/24
Gomez Arreaza, Catherine De Jesus	980/24

Febrero 2026

Resumen

En el presente informe se tratará el análisis de la relación entre causas de mortalidad, el acceso a la salud, la edad y las variables demográficas en la República Argentina.

Con este fin en mente, recopilamos los datos necesarios en páginas públicas del gobierno argentino que contenían los datos de los censos de 2010 y 2022, las defunciones a lo largo del tiempo y los establecimientos médicos nacionales de las que luego de evaluarse la calidad de los datos con metodología GQM, se armó un Diagrama de Entidad Relación (DER) y su Modelo Relacional correspondiente normalizado en Tercera Forma Normal (3FN). A partir de este modelo se limpiaron y filtraron los datos agrupando los atributos relevantes para el análisis.

Una vez terminado el diseño, se presentaron reportes en SQL y gráficos armados a partir del modelo proporcionado de los cuales pudimos concluir que existe una relación compleja y no lineal entre las enfermedades que causan la muerte, las variables demográficas y el acceso a la salud en Argentina. Donde la cantidad de establecimientos por persona por provincia influye en gran medida en la tasa de defunciones de la región, además los hombres tienen un tasa de defunción más alta que las mujeres en la juventud, hecho que se da vuelta para el grupo etario de 75 o más años y las causas de defunción son lideradas por los tumores y las enfermedades del sistema circulatorio.

Índice

1. Introducción	3
2. Procesamiento de Datos	4
2.1. Análisis de Calidad (GQM)	4
2.2. Análisis de Formas Normales de las tablas de Establecimientos de salud y Defunciones	5
2.3. Modelo Relacional, DER y Dependencias Funcionales	5
2.4. Importación de Datos	7
3. Decisiones Tomadas	9
4. Análisis de Datos	10
4.1. Reportes SQL	10
4.2. Visualizaciones	12
5. Conclusiones	16
6. Material suplementario	17

1 Introducción

Dada la importancia de las políticas públicas y las gestiones de salud poblacional, es de interés poder entender cómo se relacionan los distintos factores que a estas afectan. Donde el sistema de salud del país se encuentra en un constante desafío al adaptarse a una población desigual (con respecto a su distribución), se ve una realidad afectada por posibles problemáticas que relacionan causas de mortalidad con variables demográficas y el acceso a la salud pública o privada (o nulo acceso). Por lo tanto, el objetivo principal de este trabajo es evaluar el manejo de datos públicos (como centros de salud, defunciones por provincia, etc) y su visualización para determinar cómo han evolucionado las enfermedades que llevan a la muerte en relación con el acceso a la salud y las características de la población, principalmente, entre los años 2010 y 2022.

Esta ocasión se busca transformar datos estadísticos en información legible para comparar la evolución poblacional entre los períodos 2010 y 2022, identificando los principales polos de crecimiento. Evaluar la distribución de infraestructura, determinando si la cantidad de establecimientos médicos por departamento responde de manera proporcional a la densidad de habitantes. Y además identificar brechas de género y edad, observando cómo factores biológicos y sociales influyen en la frecuencia de las defunciones.

En resumen, este informe presenta técnicas de limpieza, modelado relacional y análisis de datos para llegar a una base sólida sobre la cual se podrá examinar resultados tanto visual como analíticamente.

2 Procesamiento de Datos

2.1 Análisis de Calidad (GQM)

Tabla: Defunciones

En el primer acercamiento con el archivo defunciones.csv, se detectó que una amplia cantidad de registros bajo la columna “jurisdiccion_residencia_nombre” contenían datos nulos (Null) o el valor “Sin información”, esto afecta la completitud de la calidad de datos siendo un problema de procesos.

Creemos que la causa principal es la fragmentación en la carga de datos. Al ser un sistema federal, la información pasa por médicos, registros civiles provinciales y administrativos nacionales. Por ende, si la residencia del fallecido no se sabe, o se pierde en algún pasaje de registros, el campo queda como “Sin información” o simplemente como nan o null.

Atributo de calidad afectado: Completitud

Meta (Goal)	Que el dato correspondiente a “jurisdiccion_residencia_nombre” este completo
Pregunta	¿Cuál es la proporción de jurisdicciones marcadas como “Sin información” o como null respecto a la cantidad de jurisdicciones bien definidas?
Métrica	(Cantidad de defunciones en Jurisdicciones sin definir / Total de defunciones en jurisdicciones bien definidas) * 100

Cuadro 1: Aplicación de métrica GQM sobre tabla Defunciones.csv.

Una vez aplicada la metodología GQM, se obtuvo que el resultado de la métrica fue 0,357603 % lo que significa que menos de 1 de cada 100 registros presentan ausencia o ambigüedad en el nombre de la jurisdicción. Si bien el porcentaje es bajo, estos casos representan una pérdida de precisión en la identificación territorial de las defunciones.

Dado que la magnitud del problema es mínima (< 0,5 %), para mantener la integridad del modelo relacional y cumplir con las restricciones de clave foránea en las tablas, se optó por filtrar los datos asociados a provincias sin definir.

Para solucionar este problema, proponemos que a la hora de ingresar jurisdicciones sea obligatorio completar el campo con una provincia válida.

Tabla: Establecimientos de Salud

Al analizar el archivo instituciones_de_salud.csv, se halló una falla en la consistencia en la relación entre identificadores y nombres de departamentos.

El caso más evidente es la Comuna 1 de la jurisdicción CABA, que en algunas filas aparece con ID 1 y en otras con ID 0. El origen más probable de esto, es de procesos y software, ya que al ser un sistema que integra diversas fuentes de datos detectamos la presencia de un software laxo que permite faltas de integridad en los datos que luego afectan a la consistencia de estos.

Atributo de calidad afectado: Consistencia

Meta (Goal)	Que no haya departamentos con igual nombre y distinto id en una misma provincia
Pregunta	¿Cuál es la proporción de departamentos con igual nombre y distinto id en una misma provincia?
Métrica	(Cantidad total de departamentos en una misma provincia con IDs distintos / Cantidad de departamentos en el país) * 100

Cuadro 2: Aplicación de métrica GQM sobre tabla *Instituciones_de_salud.csv*.

Una vez aplicada la metodología GQM, se obtuvo que el resultado de la métrica fue 0,56497 % y solo ocurrió con 3 departamentos de los 531. Aunque el porcentaje no es alto (< 1 %), esto no implica que no pueda modificar mucho los números de los análisis cuando se decide observar el comportamiento de las variables en función de los departamentos de cada provincia.

Para mantener la consistencia de los datos optamos por utilizar un único id en los departamentos, uniendo los resultados bajo una única clave.

Para solucionar este problema, proponemos que no sea posible cargar un departamento si el nombre de este no coincide con su id estándar.

2.2 Análisis de Formas Normales de las tablas de Establecimientos de salud y Defunciones

A continuación, se presenta el análisis de las formas normales para las tablas originales del dataset Defunciones y Establecimientos de Salud.

Tabla: Defunciones

- **Primera Forma Normal (1FN):** Aseguramos que la tabla se encuentra en Primera Forma Normal (1FN), dado que todos los atributos tienen valores atómicos.
- **Segunda Forma Normal (2FN):** La tabla **no cumple** con las condiciones de la 2FN. Al definir como clave primaria al conjunto compuesto por { *anio*, *jurisdiccion_de_residencia_id*, *cie10_causa_id*, *sexo_id*, *grupo_edad* }, se puede notar que existen dependencias parciales. Existen atributos no primos que dependen únicamente de una parte de la clave. Por ejemplo, el atributo *sexo* posee una dependencia funcional respecto a *sexo_id* que no es una clave primaria por sí sola.
- Notar que como la tabla no cumple con las condiciones de la Segunda Forma Normal, como cada forma normal incluye las condiciones de las formas anteriores a sí misma, la tabla tampoco cumple con las formas normales siguientes (3FN y BCFN).

Tabla: Establecimientos de Salud

- **Primera Forma Normal (1FN):** Aseguramos que la tabla **no cumple** los requerimientos para estar en Primera Forma Normal (1FN), dado que no todos los atributos tienen valores atómicos. Determinamos que el valor de *Domicilio* se podría separar en los valores *Calle* y *Altura* para que sean atómicos.
- Notar que como la tabla no cumple con las condiciones de la Primera Forma Normal, como cada forma normal incluye las condiciones de las formas anteriores a sí misma, la tabla tampoco cumple con las formas normales siguientes (2FN, 3FN y BCFN).

2.3 Modelo Relacional, DER y Dependencias Funcionales

DER

Luego de comprobar la calidad de los datos, se realizó un Diagrama Entidad-Relación en el que se pueden ver las entidades con sus atributos y claves y las relaciones entre estas.

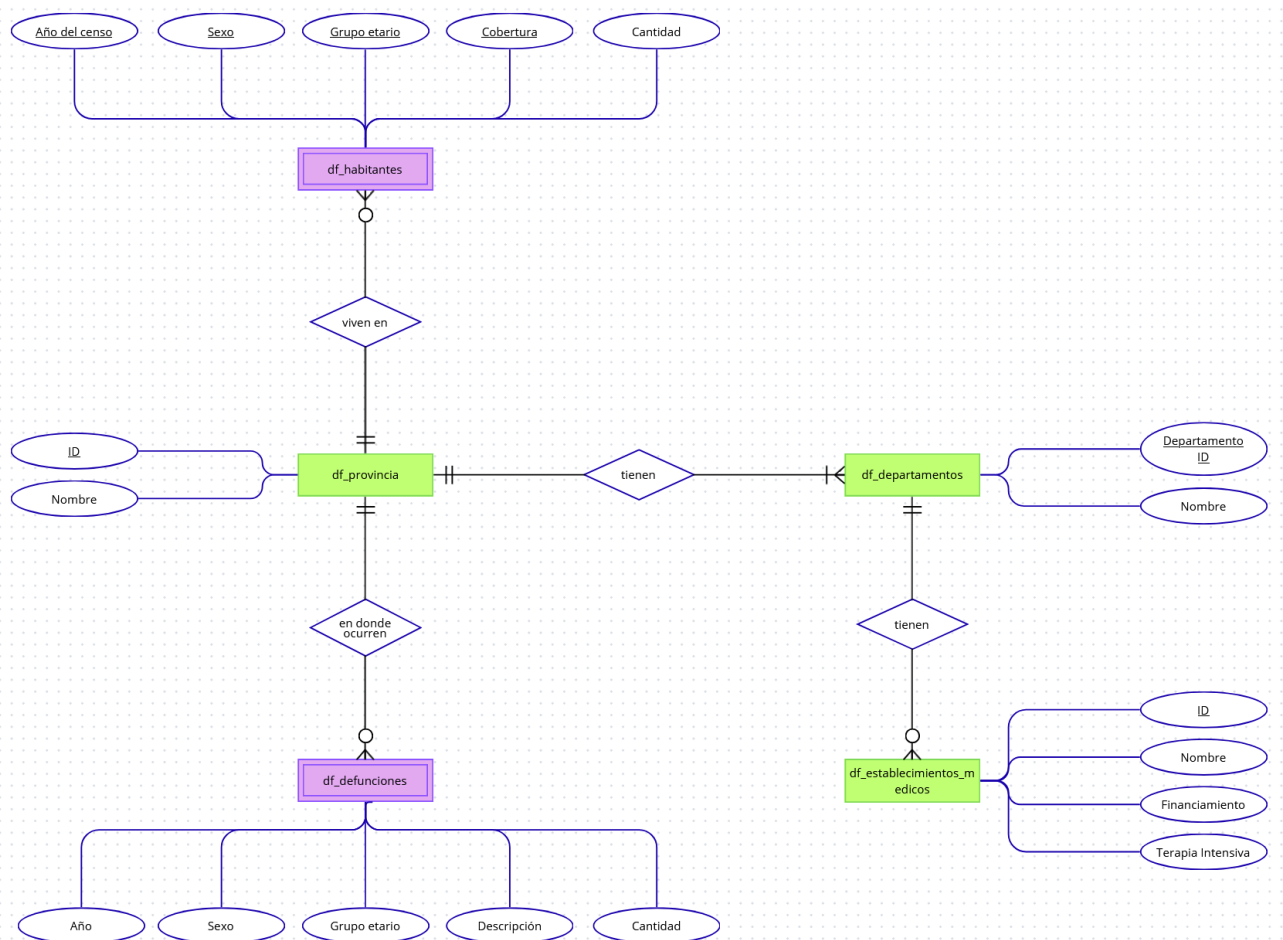


Figura 1: Diagrama Entidad-Relación.

Modelo Relacional

A partir del DER, se realizó el siguiente modelo relacional.

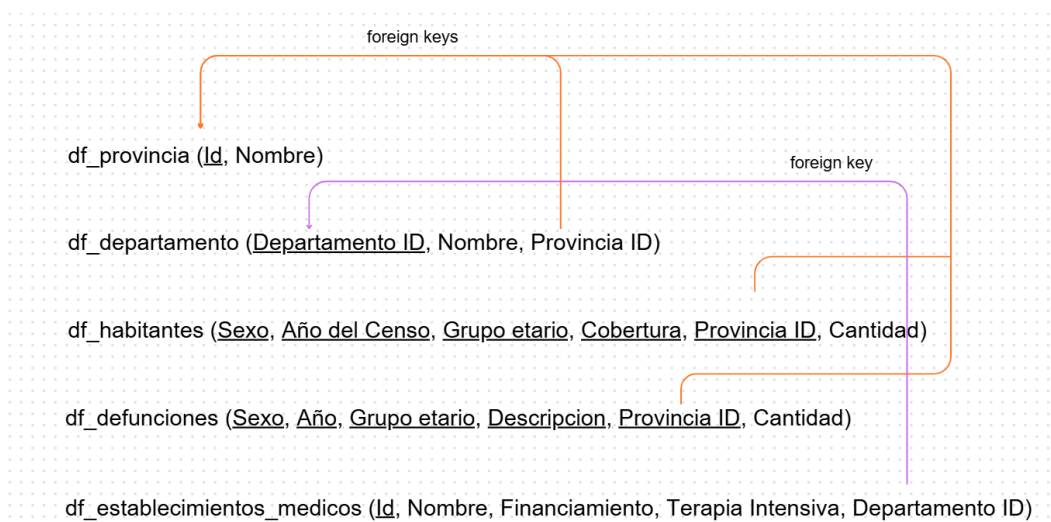


Figura 2: Modelo Relacional basado en el Diagrama Entidad-Relación.

Se puede ver en el DER que df.provincias, df.departamento y df.establecimientos_medicos

son Entidades Fuertes; `df_habitantes` y `df_defunciones` son Entidades Débiles de `df_provincias` y por lo tanto tienen `df_provincias.ID` una foreign key como parte de su primary key; como las relaciones entre `df_provincias` y `df_departamentos` es uno-a-uno, la primary key de `df_provincias` es un atributo de `df_departamentos`; analogamente como la relación entre `df_establecimientos_medicos` y `df_departamentos` es uno-a-uno, la clave primaria de `df_departamentos` es un atributo de `df_establecimientos_medicos`.

Dependencias Funcionales

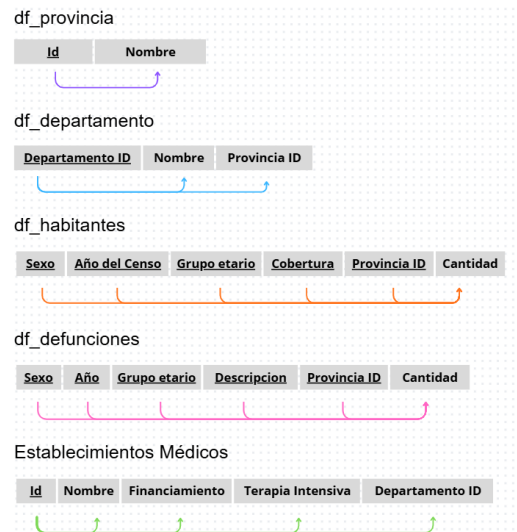


Figura 3: Dependencias funcionales.

Dado que todos los atributos son atómicos (1FN), que todo atributo no primo depende en su totalidad de la clave primaria (2NF) y que no existen dependencias funcionales transitivas (para toda dependencia funcional no trivial $X \rightarrow A$ del modelo, se cumple que X es una superclave (SK) o que A es un atributo primo) (3FN) podemos concluir que el modelo relacional está en Tercera Forma Normal.

2.4 Importación de Datos

Para la tabla **df_departamentos** tomamos las columnas “departamento_id” como “Departamento ID”, “departamento_nombre” como “Nombre” y “provincia_id” como “Provincia ID” de la tabla **instituciones_de_salud** que es la lectura del archivo **instituciones_de_salud.xlsx**

Para la tabla **df_provincias** tomamos las columnas “jurisdiccion_de_residencia_id” como “ID” y “jurisdiccion_de_residencia_nombre” como “Nombre” de la tabla **archivo_defunciones** que es la lectura del archivo **defunciones.csv**

Para la tabla **df_defunciones** tomamos las columnas “grupo_edad” que normalizamos el contenido (ej: si dice 01.De 0 a 14 años. pasa a ser 0-14) con el nombre “Grupo etario”, “Sexo” que normalizamos el contenido (si el valor es “femenino” pasa a ser “F” y si es “masculino” pasa a ser “M”) que conserva su nombre (“Sexo”), “anio” como “Año” y “jurisdiccion_de_residencia_id” como “Provincia ID” de la tabla **archivo_defunciones** que es la lectura del archivo **defunciones.csv**. También tomamos la columna “categorias” como “Descripción” de la tabla **categoriasDefunciones** que es la lectura del archivo **categoriasDefunciones.csv**. (Ambas tablas fueron unidas por la clave del ID de la defunción que respectivamente se llaman “cie10_causa_id” y “codigo_def”).

Para la tabla **df_establecimientos_medicos** tomamos las columnas “establecimiento_id” como “ID”, “establecimiento_nombre” como “Nombre”, “establecimiento_id” como “Departamento ID”, “terapia_intensiva” como “Terapia intensiva” y “origen_financiamiento” como “Financiamiento” de la tabla **instituciones_de_salud** que es la lectura del archivo **instituciones_de_salud.xlsx**

Para la tabla **df_habitantes** hace un proceso completamente análogo con las tablas **df_censo_2010** que es la lectura del archivo **censo2010.xlsx** y **df_censo_2022** que es la lectura del archivo **censo2022.xlsx** de las cuales tomamos las columnas “Cobertura” que conserva su nombre de la cual dividimos las coberturas existentes en “Privada”, “Pública” y “Sin cobertura” (se puede ver la categorización de la cobertura en la sección Desiciones3), la columna “Rango Etario” como “Grupo Etario” y luego pasamos el valor de “Mujer” y “Varón” a una nueva columna “Cantidad” y armamos otra columna “Sexo” cuyo valor es F para las cantidades de mujer y M para las cantidades de varón. Además armamos otra columna “Año del censo” que indica el año correspondiente a la tabla de censo que estamos usando. Luego con el nombre de la provincia dado por la columna “provincia” igualado la columna “Nombre” de la tabla **df_provincias** buscamos el Id de la provincia y armamos una nueva columna “Provincia ID” que reemplaza a “provincia” para que **df_habitantes** cumpla con el DER.

3 Decisiones Tomadas

Las fuentes de datos proporcionadas para la realización del trabajo incluían diversos atributos innecesarios para el análisis en función del objetivo planteado, así como registros inconsistentes e incompletos. En consecuencia, se tomaron las siguientes decisiones:

- De la fuente defunciones.csv no se consideraron atributos como *muerte_materna_clasificacion* y *muerte_materna_id*, debido a que sus valores eran mayoritariamente nulos y resultaban irrelevantes para el análisis desarrollado en este informe.

Tampoco se contemplaron atributos vinculados al tipo de residencia (con excepción del ID del departamento) provenientes de la fuente instituciones_de_salud.csv, dado que, si bien describían ciertas características demográficas, se determinó que no eran pertinentes a los objetivos específicos del análisis.

Además, se optó por utilizar las categorías presentes en categoriasDefunciones como parte de la clave de la tabla df_defunciones (en vez de usar el id de las defunciones) para poder respetar la Tercera Forma Normal con los contenidos vistos en la materia.

- En la construcción de df_defunciones, se eliminaron los registros cuyo *grupo_etario* o *sexo* no aportaban información relevante. La suma de dichos registros representaba menos del 0,4 % del total en cada caso, por lo que se consideró que su exclusión no afectaba significativamente los resultados y contribuía a una mayor claridad analítica.
- En la elaboración de df_provincia, se descartaron los casos correspondientes a *jurisdiccion_residencia_nombre* que no aportaban información significativa, ya que la suma total de dichos registros era inferior al 0,4 %. Se concluyó que su eliminación no alteraba sustancialmente el análisis.
- En el armado de df_departamentos, se decidió sobrescribir el ID de tres departamentos (ZAPALA (Neuquén), QUILMES (Buenos Aires) y COMUNA 1 (Ciudad Autónoma de Buenos Aires)) ya que en la tabla instituciones_de_salud.csv figuraban registrados con identificadores distintos. Esta inconsistencia afectaba el análisis, por lo que se procedió a su corrección para garantizar coherencia en los resultados.
- En la preparación de df_habitantes, se agruparon distintas categorías de la variable *Cobertura* con el fin de homogeneizar la información entre los censos disponibles. Las categorías *Obra Social (Incluye PAMI)*, *Prepaga a través de obra social* y *Prepaga sólo por contratación voluntaria* fueron reunidas bajo la categoría *Privada*. Por su parte, *Programas o planes estatales de salud* se clasificó como *Pública*, y *No tiene obra social, prepaga o plan estatal* se definió como *No tiene cobertura social*. Esta decisión respondió a que los censos no presentaban la misma cantidad de categorías, por lo que se consideró que esta clasificación permitía un análisis más consistente.
- Para el gráfico titulado *Defunciones por categoría a lo largo del tiempo*, se optó por presentar dos visualizaciones diferentes. La primera incluye las 10 categorías con mayor cantidad de defunciones. La segunda separa las 5 principales categorías (incluyendo la categoría *COVID-19*) en un gráfico, dejando el resto en otro. Esta decisión se fundamentó en la gran diferencia en la magnitud de las defunciones entre categorías, lo que dificultaba una correcta representación en un único gráfico. En particular, la categoría *COVID-19*, si bien no se encontraba estrictamente entre las cinco con mayor cantidad acumulada, presentó un pico excepcional en 2020 que comprimía visualmente el resto de las categorías, impidiendo su adecuada interpretación.

4 Análisis de Datos

4.1 Reportes SQL

Debido a la cantidad de filas de los reportes, solo serán presentados de forma ilustrativa. Se encuentran en completitud, como archivo csv, en la carpeta *Consultas* y su nombre es “re-
porte_” seguido por el número correspondiente a la consulta (i,ii,iii,iv,v).

REPORTE I

Para este reporte lo que se hizo fue comparar la cantidad de habitantes con y sin cobertura de salud entre los censos 2010 y 2022, separados además por grupo etario. Para ello, se construyó una única consulta que agrupa los datos por Provincia y Grupo etario, generando cuatro columnas calculadas mediante expresiones CASE dentro de funciones SUM: Habitantes con cobertura en 2010, habitantes sin cobertura en 2010, habitantes con cobertura en 2022 y habitantes sin cobertura en 2022.

Se consideró como población con cobertura a quienes registraban cobertura pública o privada, mientras que se clasificó como sin cobertura a quienes figuraban con la categoría “No tiene cobertura social”. Finalmente, los resultados se ordenaron por provincia y grupo etario para facilitar su lectura y posterior análisis comparativo entre ambos años de censo.

Índice	Provincia	Grupo etario	Habitantes con cobertura en 2010	Habitantes sin cobertura en 2010	Habitantes con cobertura en 2022	Habitantes sin cobertura en 2022
0	Buenos Aires	0-14	2.23051e+06	1.64307e+06	2.18354e+06	1.64513e+06
1	Buenos Aires	15-34	2.85005e+06	2.14599e+06	2.83238e+06	2.41821e+06
2	Buenos Aires	35-54	2.42343e+06	1.25518e+06	3.06877e+06	1.5331e+06
3	Buenos Aires	55-74	1.88994e+06	452068	2.34938e+06	490794

REPORTE II

En este reporte se buscó identificar, para cada provincia, la cantidad de establecimientos médicos que cuentan con terapia intensiva, diferenciando por tipo de financiamiento. Para lograrlo, se integraron tres tablas:

- *df_establecimientos_medicos*: que contiene información sobre financiamiento y disponibilidad de terapia intensiva.
- *df_departamentos*: que permite vincular cada establecimiento con su provincia.
- *df_provincias*: de donde se obtiene el nombre de la provincia.

Este reporte analiza cuántos establecimientos de salud con terapia intensiva existen en cada provincia, diferenciándolos según su tipo de financiamiento (privado o estatal).

La consulta permite evaluar cómo se distribuye la infraestructura de mayor complejidad entre el sector público y privado a nivel provincial, lo que ayuda a entender la capacidad sanitaria disponible ante situaciones críticas.

Índice	Nombre	Financiamiento	Cantidad de establecimientos
0	Buenos Aires	Estatal	307
1	Buenos Aires	Privado	591
2	Catamarca	Estatal	171
3	Catamarca	Privado	370

REPORTE III

Para el análisis de las causas de muerte, se buscó identificar los dos extremos de frecuencia de categoría de defunción por grupo etario y sexo (las 5 causas de defunción más y menos frecuentes). Para lograr esto, se utilizó una tabla temporal (*frecuencia_defunciones*) que salía de *df_defunciones*, utilizando a *ROW_NUMBER* se pudo particionar los datos por sexo y grupo etario, asignando simultáneamente dos rankings: uno descendente para las categorías más frecuentes y otro ascendente para las menos frecuentes. Por último, para la presentación de los datos, se terminó de ordenar con *CASE*, donde los rangos etarios estuvieran en un orden lógico (y no alfabético que es como lo haría SQL).

Índice	Sexo	Grupo etario	Descripción	Total de muertes	Frecuencia
0	F	0-14	Afecciones originadas en el periodo perinatal	28507	5 más frecuentes
1	F	0-14	Malformaciones congénitas, deformidades y anomalías cromosómicas	18550	5 más frecuentes
2	F	0-14	Accidentes y causas externas	8130	5 más frecuentes
3	F	0-14	Enfermedades del aparato respiratorio	6731	5 más frecuentes

REPORTE IV

Con respecto al cálculo de la tasa de mortalidad, se crearon dos estructuras temporales independientes: una para obtener el total de muertes y otra para obtener la población total a partir de los datos de los censos previamente procesados.

La integración de ambos conjuntos de datos se realizó mediante un *INNER JOIN* basado en la clave compuesta de Provincia ID y Grupo etario. Como se pidió en la consigna, se normalizo cada mil habitantes la tasa de mortalidad, lo que permite realizar comparaciones equitativas entre provincias con diferentes cantidades de población.

Índice	Provincia	Grupo etario	Tasa de mortalidad
0	Buenos Aires	0-14	0.580358
1	Buenos Aires	15-34	0.849047
2	Buenos Aires	35-54	2.80929
3	Buenos Aires	55-74	17.5961

REPORTE V

Se analiza la variación en la cantidad de defunciones por causa entre 2010 y 2022. A partir de la tabla *df_defunciones*, se agruparon los datos por descripción (causa de muerte) y se calcularon dos sumas condicionales: Total de defunciones en 2022 y total de defunciones en 2010.

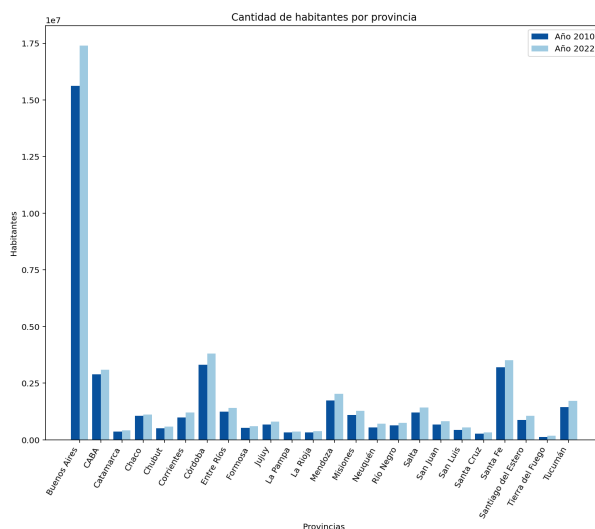
Luego, se restó el valor de 2010 al de 2022 para obtener una nueva métrica: “Diferencia de defunciones”, que indica cuánto aumentó o disminuyó cada causa en el período analizado. El reporte permite medir el cambio en la mortalidad por causa entre 2010 y 2022, identificando qué enfermedades o eventos tuvieron mayor crecimiento o reducción en términos de defunciones totales.

Índice	Descripción	Diferencia de defunciones
0	Enfermedades del aparato respiratorio	25846
1	COVID-19	23812
2	Enfermedades del aparato circulatorio	13002
3	Enfermedades del sistema genitourinario	5787

4.2 Visualizaciones

GRÁFICO I:

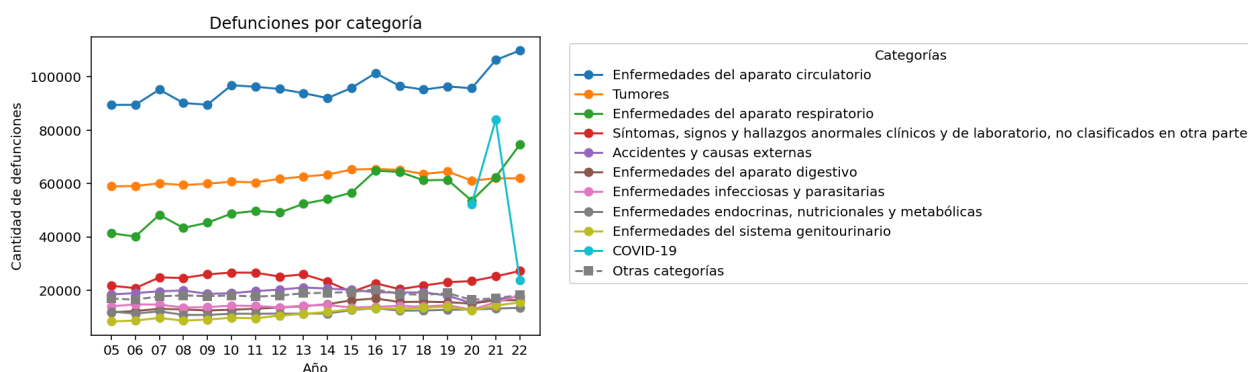
Para este análisis se optó por un gráfico de barras agrupadas, que permite comparar de forma directa la asimétrica densidad de población en Argentina separada por jurisdicciones y el crecimiento ocurrido en el 2022 a comparación del año 2010. Es sumamente visible la concentración de población que presenta Buenos Aires a comparación del resto del país. Todas las jurisdicciones muestran un incremento en su población con respecto al 2010, mostrando una tendencia de crecimiento demográfico. Y los extremos se mantienen en posición, donde jurisdicciones como Tierra del Fuego y La Pampa se mantienen con los volúmenes más bajos, mientras que Buenos Aires y Córdoba siguen dominando.



El gráfico demuestra que el crecimiento poblacional no es uniforme.

GRÁFICO II:

Para el análisis de la mortalidad a lo largo del tiempo, se desarrollaron dos propuestas similares 6, pero visualmente distintas que buscan equilibrar la cantidad de información con la legibilidad del gráfico. Acá se muestra la primera versión, la cual consistió en la elaboración de un gráfico de líneas único donde se priorizaron las 10 categorías con mayor volumen de muertes totales. Se utilizó un agrupamiento por descripción y año, identificando el "top 10" mediante un orden descendente. El resto de las categorías de defunción se consolidaron en una única línea llamada "Otras Causas".



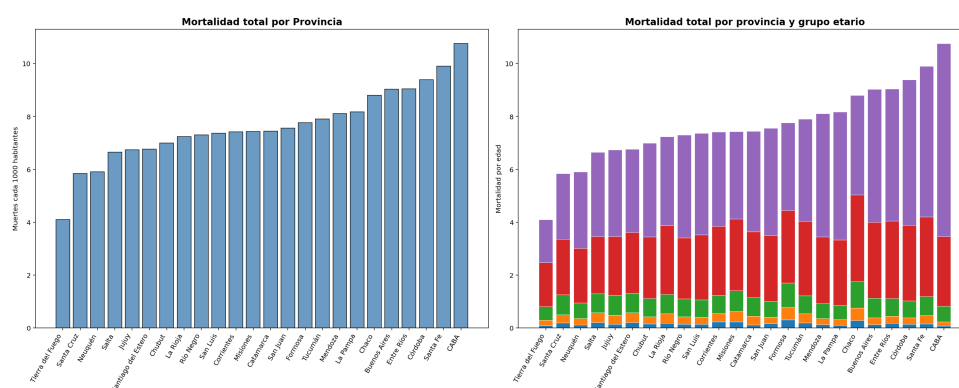
Ambas versiones del gráfico dos nos muestran cómo las enfermedades del aparato circulatorio y tumores lideran históricamente la estadística, mostrando un crecimiento estable y

sostenido. Se observa un comportamiento particular con un pico abrupto en 2021 donde el Covid 19 igualó en magnitud a las causas principales, alterando la estructura habitual de defunciones.

GRÁFICO III:

Primero se crearon dos tablas temporales, una de muertes por provincia y otra de población total por provincia. Luego, se tomó una tasa de mortalidad total cada mil habitantes por provincia (mil porque tiene buena divisibilidad y facilita el cálculo y la visualización).

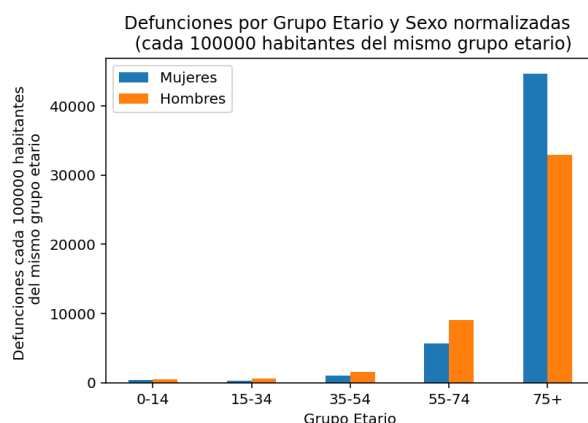
Luego se crean otras dos tablas temporales pero esta vez para el segundo gráfico, donde se tuvo en cuenta el grupo etario. Este fue elegido como variable ya que permite realizar un análisis sobre la distribución de la mortalidad donde, al separar por edades, podemos identificar si la tasa responde a un proceso natural de envejecimiento o si existen factores de riesgo que afectan a la población más joven.



Este enfoque nos permite observar como provincias con poblaciones más envejecidas mostrarán una mayor contribución en el grupo '75+' (como sucede con CABA), mientras que provincias con mayores índices de mortalidad infantil o accidentes en jóvenes mostrarán lo contrario (grupos '0-14' o '15-34').

GRÁFICO IV:

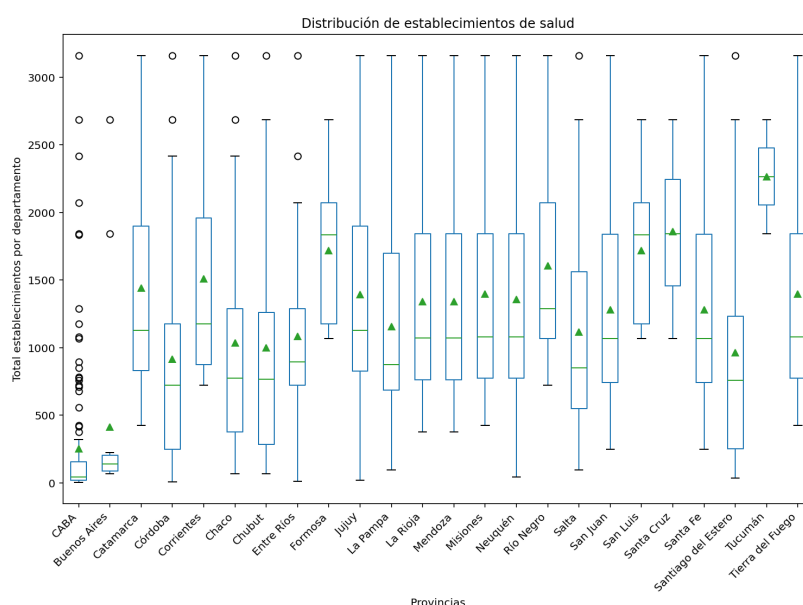
Se agruparon previamente los datos por grupo etario y sexo, sumando la cantidad total de defunciones en cada combinación. Luego se reorganizó la tabla para tener, para cada grupo etario, dos columnas: una correspondiente al sexo femenino y otra al masculino. Por último se calculó el total de habitantes por grupo etario y se normalizaron las columnas con estos datos y cada 100000 habitantes. El gráfico resultante permite visualizar no solo la evolución de la mortalidad según la edad, sino también las diferencias relativas entre hombres y mujeres en cada grupo etario.



La mortalidad está fuertemente concentrada en los grupos de mayor edad, presenta diferencias de género que varían según el respectivo rango etario. Este gráfico complementa el análisis del Gráfico iii de tasas por provincia y grupo etario, ya que acá se observa la estructura general por edad y sexo.

GRÁFICO V:

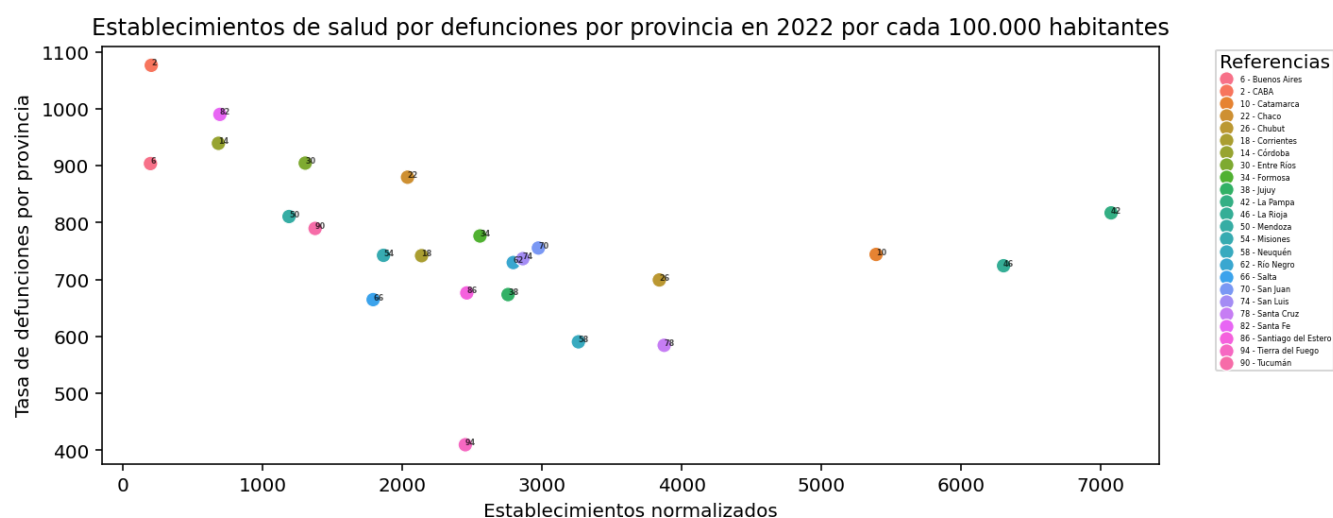
Para este análisis se utilizó un boxplot por provincia, donde cada caja representa la distribución de la cantidad de establecimientos de salud por departamento dentro de cada jurisdicción. Este tipo de gráfico resulta adecuado porque no solo muestra un valor promedio, sino que permite visualizar la mediana, la dispersión de los datos, los valores extremos y la media, marcada explícitamente en el gráfico. En primer lugar, se observa una gran desigualdad en la distribución de establecimientos dentro de muchas provincias. Algunas jurisdicciones presentan cajas muy amplias y muchos valores atípicos, lo que indica que ciertos departamentos concentran una cantidad significativamente mayor de centros de salud respecto al resto, como se puede ver con la Ciudad de Buenos Aires. Provincias como Buenos Aires, Córdoba o Santa Fe muestran una dispersión considerable, lo cual es esperable dado que cuentan con departamentos muy poblados y urbanizados que concentran infraestructura sanitaria, mientras que otros departamentos más rurales presentan valores más bajos. Por otro lado, provincias con menor densidad poblacional o menor cantidad de departamentos tienden a mostrar distribuciones más compactas, lo que sugiere una menor variabilidad interna.



El gráfico evidencia que la infraestructura sanitaria en Argentina presenta desigualdad territorial entre provincias definida por la concentración de habitantes y las diferencias específicas (que no se analizan en este informe) entre departamentos.

GRÁFICO VI:

Para este análisis se eligió utilizar un gráfico de dispersión, donde cada punto representa una provincia. El eje X representa la cantidad normalizada de establecimientos de salud, mientras que el eje Y la tasa de defunciones, ambos valores están expresados por cada 100.000 habitantes en el año 2022. El objetivo es observar si existe algún tipo de relación entre la disponibilidad de establecimientos de salud y la mortalidad por provincia. Las provincias con mayor población, si bien tienden a tener muchas más instituciones de salud, al normalizar se observa que la disponibilidad relativa es considerablemente menor que en otras provincias menos pobladas y como eso afecta en la tasa de defunciones.



El gráfico ilustra cómo la infraestructura de salud impacta directamente en la mortalidad. Existe una relación inversa donde las provincias con menos establecimientos por habitante sufren las mayores tasas de defunciones mientras que al incremento de centros de salud se hace visible una reducción sostenida en la tasa de fallecimientos.

5 Conclusiones

A partir del análisis realizado, se puede concluir que existe una relación compleja y no lineal entre las enfermedades que causan la muerte, las variables demográficas y el acceso a la salud en Argentina:

Relación Enfermedad-Demografía: Históricamente, las enfermedades del aparato circulatorio y los tumores lideran las estadísticas de mortalidad, mostrando un crecimiento estable y sostenido en el tiempo. La mortalidad está fuertemente concentrada en los grupos de mayor edad (75+ años), lo que refleja un proceso natural de envejecimiento poblacional. Por otro lado, los hombres presentan sistemáticamente mayor mortalidad en su juventud sin embargo, a partir de los 75 años o más, la tendencia se invierte y las mujeres concentran una mayor proporción de defunciones.

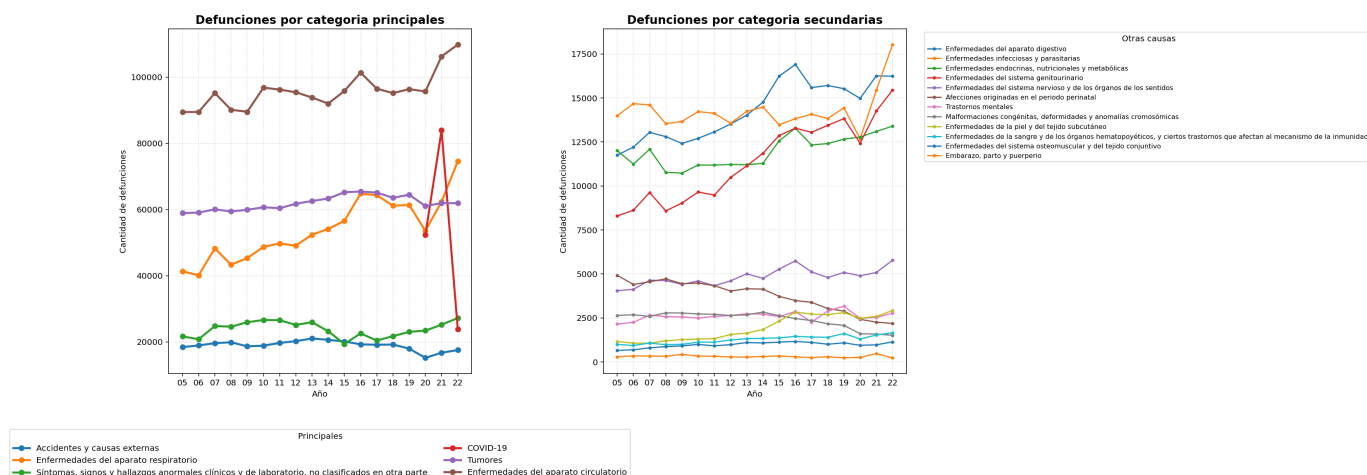
Variaciones Temporales y COVID-19: Se observó una alteración significativa en la estructura habitual de defunciones en 2021 debido a la pandemia de COVID-19, la cual alcanzó magnitudes similares a las causas principales de muerte en el país. Además, todas las jurisdicciones mostraron una tendencia de crecimiento demográfico entre los censos de 2010 y 2022, aunque este crecimiento no fue totalmente uniforme.

Infraestructura y Salud: La disponibilidad de recursos sanitarios es un factor determinante en la tasa de mortalidad. Si bien existe una correlación entre el volumen poblacional y la cantidad absoluta de centros de salud, los datos demuestran que en las jurisdicciones con mayor densidad demográfica, no aumenta de manera proporcional, resultando en una menor cantidad de establecimientos por habitante. Esta falta de infraestructura relativa coincide directamente con los picos más elevados en las tasas de mortalidad. Por el contrario, aquellas provincias que han logrado mantener una mayor densidad de establecimientos por habitante presentan tasas de defunciones más bajas y controladas.

6 Material suplementario

GRAFICO II - versión 2:

La segunda versión pasa a ser una estructura de dos subplots. Acá, en lugar de agrupar mas de 10 categorías en “Otras Causas”, se decidió mostrarlas por separado. El gráfico de la izquierda contiene las causas de mayor impacto, mientras que el segundo muestra todas las causas “secundarias” independientemente. Esta separación permite que cada gráfico tenga su propia escala en el eje Y, ajustándose automáticamente al rango de valores de su grupo. Un punto importante a resaltar en el diseño de esta versión fue cómo se trató a la categoría Covid-19. A pesar de no ser una causa histórica, se decidió forzar su inclusión en el primer gráfico por estas dos razones: Su pico en el 2021, alcanzando cantidades altas de cantidad de defunciones. Y para preservar la escala en el gráfico de la derecha, ya que si Covid-19 se hubiese quedado allí, su gran pico iba a complicar su comprensión.



Ubicación de archivos

- Los *reportes* se encuentran en la carpeta *Consultas* en el zip.
- Los *gráficos* se encuentran en la carpeta *Graficos* en el zip.
- Las *tablas originales* se encuentran en la carpeta *Archivos originales*