

COVID-19

Proyecto Final Data Analytics

CoderHouse

Moya Facundo

Matías José Nougues

Yanina Natalia Mojavik

1. Introducción

1.1 Covid-19

1.2 Objetivos

1.3 Herramientas tecnológicas utilizadas

1.4 Futuras líneas

1.5 Base de datos

2. Diagramas y tablas

2.1 Diagrama de Entidad-Relación

2.2 Tablas

2.3 Tabla Ubicación de provincias

2.4 Tabla de versionado

4. Medidas y visualizaciones

4.1 Medidas

4.2 Medidas avanzadas y parámetros

4.3 Tabla Calendario

5. Visualizaciones

5.1 Solapa Inicio

4.2 Solapa país

4.3 Solapa provincia

1. Introducción

1.1 Covid-19

A lo largo de los dos últimos años, el mundo se vio inmerso en una nueva pandemia la cual genera un enorme desafío para la administración de la misma. En este caso la argentina no fue la excepción.

Se trata del COVID-19, una enfermedad que todavía sigue generando incertidumbre y por lo visto también requiere de medidas rápidas. El virus puede propagarse desde la boca o la nariz de una persona infectada en forma de pequeñas partículas líquidas que expulsa cuando tose, estornuda, habla, canta o respira. Estas partículas pueden ser desde pequeños aerosoles hasta gotitas respiratorias más grandes.

Puedes contagiarte de COVID-19 si respiras cerca de una persona infectada o si tocas una superficie contaminada y, seguidamente, te tocas los ojos, la nariz o la boca. El virus se propaga más fácilmente en espacios interiores o en aglomeraciones de personas.

Es por esto que a continuación se presentarán una serie de datos para el análisis comprensión y, posteriormente a través de los resultados obtenidos tomar las decisiones necesarias hacia donde dirigir los recursos que permitan reducir la propagación del virus y aminorar sus efectos

1.2 Objetivos

El objetivo de este análisis es representar la cantidad de contagio producidos en el país, distinguiéndose por provincias, y cantidades de contagios mensuales.

Se busca que el análisis exponga información necesaria para tomar decisiones sobre donde es necesario reforzar medidas de prevención y asignar recursos.

1.3 Herramientas tecnológicas utilizadas

Se utilizaron las herramientas vistas en clases o recomendadas por los tutores, destacándose las siguientes:

- Power BI
- Excel
- Power Point
- SQL

1.4 Futuras líneas

En este caso el proyecto presentado por supuesto que tiene un desarrollo a futuro. Esto se debe a que aún al día de hoy seguimos en pandemia, así que se puede continuar con un estudio de este tema, llevando a cabo un análisis correcto, teniendo en cuenta que nuestro análisis se realizó entre el año 2020 y noviembre del 2021.

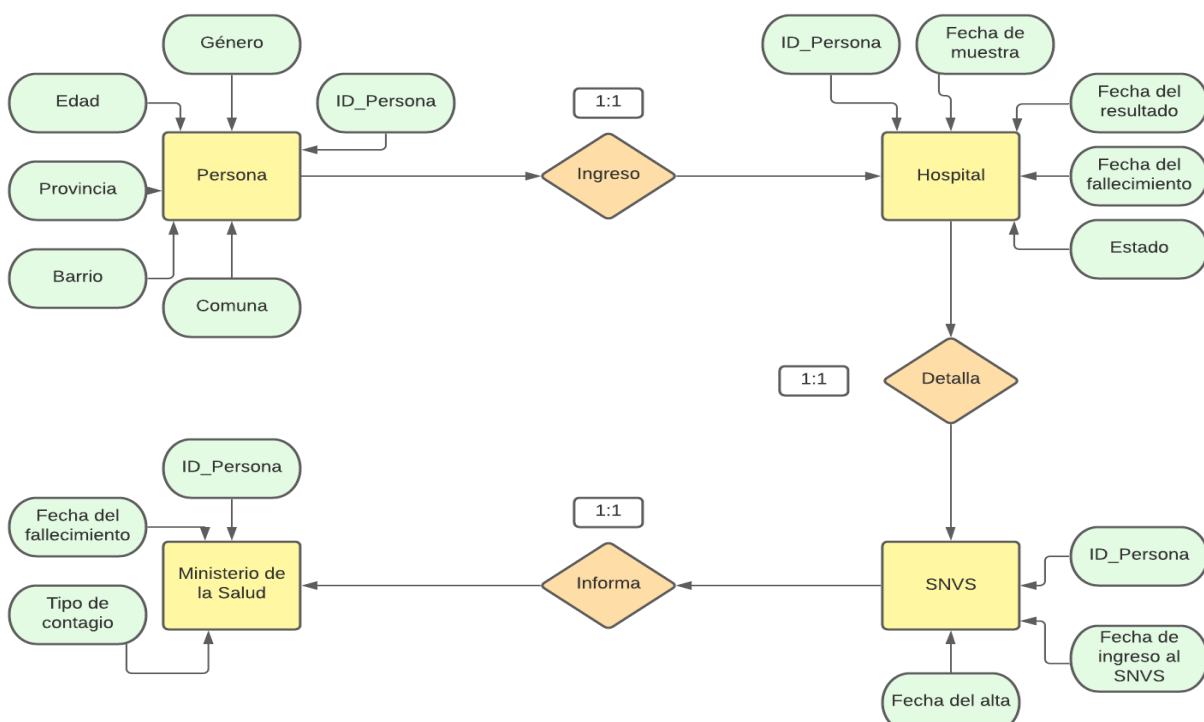
1.5 Base de datos

La base de datos se obtuvo del Ministerio de Salud (www.data.gov.ar). Es una base de datos en la que se debieron hacer muchos cambios y transformaciones (más adelante se mostrarán las mismas). Si bien la base de datos es una base completa, se debieron corregir ciertas filas, ya que esta cuenta con un poco más de un millón de datos. La base de datos está adjunta con el proyecto final, en donde se muestran correctamente las tablas y campos de la misma, para una definición mucho mejor se utiliza el DER (Diagrama de Entidad-Relación). Nuestra base de datos se encuentra dividida en las siguientes tablas con sus respectivos campos.

- **Persona:** ID_Persona, Género, Barrio, Comuna, Provincia, Edad
- **Hospital:** ID_Persona, Fecha de muestra, Fecha del resultado, Fecha del fallecimiento, Estado
- **SNVS:** ID_Persona, Fecha de ingreso al SNVS, Fecha del alta
- **Ministerio de la Salud:** ID_Persona, Fecha del fallecimiento, Tipo de contagio

2. Diagramas y tablas

2.1 Diagrama de Entidad-Relación



A nivel conceptual, tomamos el **COVID-19** puntualmente en **Argentina**, como afecta en el país, distinguiendo por provincia, barrio, comuna, y distintas fechas que permiten llevar un análisis mayor. Además, distinguimos tablas muy específicas y una relación simple pero directa, que nos da la posibilidad de profundizar en el tema de manera sencilla.

2.2 Tablas

Tabla N°1: Persona

Columna	Descripción	Tipo de dato	Es PK	Es Fk
ID Persona	ID único por persona contagiada	Int	Si	No
Provincia	Provincia en donde ocurrió el contagio	Varchar (100)	No	No
Barrio	Barrio en donde ocurrió el contagio	Varchar (100)	No	No
Comuna	Comuna en donde ocurrió el contagio	Varchar (100)	No	No
Género	Especifica el genero de la persona contagiada	Varchar (50)	No	No
Edad	Edad de la persona contagiada	Int	No	No

En nuestro proyecto, una **Persona** es aquel individuo que contrajo COVID-19. En esta tabla se encuentran las principales características de aquellas personas que fueron contagiadas.

Tabla N°2: Hospital

Columna	Descripción	Tipo de dato	Es PK	Es Fk
ID Persona	ID único por persona contagiada	Int	Si	No
Fecha de muestra	Fecha de la toma de muestra nasofaringea del paciente con COVID-19	datetime	No	No
Fecha del resultado	Fecha de entrega del resultado de la muestra nasofaringea en el paciente con COVID-19	datetime	No	No
Fecha del fallecimiento	Fecha del fallecimiento del paciente con COVID-19	datetime	No	No
Estado del paciente	Muestra en que estado de salud se encuentra el paciente contagiado	Varchar(100)	No	No

En la tabla **Hospital** se especifica la información relevante cargada en la base de datos de los hospitales en donde se realizaron los **hisopados/test rápidos** y la **fecha de fallecimiento** de aquellos pacientes con COVID-19.

Tabla N°3: Base de Datos del Sistema Nacional de Vigilancia de la Salud (SNVS)

Columna	Descripción	Tipo de dato	Es PK	Es Fk
ID Persona	ID único por persona contagiada	Int	Si	No
Tipo de Contagio	Vía identificada de contagio de COVID-19	Varchar(100)	No	No
Provincia	Provincia en donde ocurrió el contagio	Varchar(100)	No	No
Fecha del ingreso al SNVS	Fecha de ingreso al SNVS del paciente con COVID-19	datetime	No	No
Fecha del resultado	Fecha de entrega del resultado de la muestra nasofaringea en el paciente con COVID-19	datetime	No	No
Fecha del alta	Fecha del alta del paciente contagiado	datetime	No	No
Edad	Edad de la persona contagiada	Int	No	No
Género	Género de la persona contagiada	Varchar(100)	No	No

La **Tabla N°3** muestra los datos que son ingresados a la **Base de Datos del SNVS (Sistema Nacional de Vigilancia de la Salud)**.

Tabla N°4: Ministerio de Salud

Columna	Descripción	Tipo de dato	Es PK	Es Fk
ID Persona	ID único por persona contagiada	Int	Si	No
Tipo de Contagio	Vía identificada de contagio de COVID-19	Varchar(100)	No	No
Provincia	Provincia en donde ocurrió el contagio	Varchar(100)	No	No
Fecha del fallecimiento	Fecha de fallecimiento de la persona contagiada	datetime	No	No
Fecha del alta	Fecha del alta del paciente contagiado	datetime	No	No
Edad	Edad de la persona contagiada	Int	No	No
Género	Género de la persona contagiada	Varchar(100)	No	No

La **Tabla N°4** muestra la información relevante en la base de datos del **Ministerio de Salud**.

Tabla N°5: INDEC

Columna	Descripción	Tipo de dato	Es PK	Es Fk
Provincia	Provincia en donde ocurrió el contagio	Varchar(100)	No	No
Edad	Edad de la persona contagiada	Int	No	No
Género	Género de la persona contagiada	Varchar(100)	No	No
Fecha del fallecimiento	Fecha de fallecimiento de la persona contagiada	datetime	No	No
Tipo de Contagio	Vía identificada de contagio de COVID-19	Varchar(100)	No	No

La **Tabla N°5** muestra la información relevante para estudios estadísticos por parte del **INDEC** en cuestiones relacionadas con la salud.

2.3 Tabla Ubicación de provincias

Esta tabla es una tabla específica, ya que solo fue creada para proyectar de forma correcta las provincias en el mapa de la segunda solapa. Dado que en dicha solapa se lleva a cabo un análisis por provincias, se ubicó las provincias en el mapa del país, para ello fue necesario la creación de una tabla con el nombre *Ubicación de provincias*, ya que esta contiene las latitudes y longitudes de las provincias.

2.4 Tabla de versionado

03/01/2022 Desafío 9	Se realiza la conexión entre la BD y el Power BI, encontrando obvios errores en el modelo relacional.
03/01/2022 Desafío 10	Se realiza la primera conexión correcta en el modelo relacional, encontrando posteriormente errores en la cardinalidad.
10/01/2022 Desafío 11	Se corrige dicha cardinalidad para poder relacionar bien las tablas
17/01/2022 Desafío 13	Se elimina totalmente del proyecto la tabla COE, ya que no sirve para aportar nada al análisis
26/01/2022	Se realiza la correcta conexión de la tabla calendario con la tabla Hospital, y se corrige la forma de encontrar dichos datos. También se hace la conexión de la tabla Ubicación provincias para el mapa de la segunda solapa.

3. Transformación de datos

Este apartado es vital para el proyecto ya que muestra todos los cambios que se realizarán en el archivo **.pbix** como en cualquier otro cambio relacionado con la entrega del proyecto final.

- Se realizó la conexión de la base de datos con Power BI.
- El modelo relacional al ser ejecutado de manera automática por Power BI, presenta errores, los cuales se modifican realizando las líneas de conexión de manera correcta.
- Se realizaron cambios en el modelo relacional uniendo de manera correcta las líneas de conexión.
- Los tipos de relaciones de nuestro proyecto son todas 1:1 y solo una tuvo que ser creada
- A su vez creamos la Tabla Calendario a partir de la fecha del resultado que está relacionada con la Tabla Hospital y jerarquizamos en año y mes.
- Utilizamos la tabla Ubicación de provincias que tienen las longitudes y latitudes de las provincias para el mapa de la segunda solapa.
- Se modificaron fechas mal escritas que causaban errores en el análisis.
- Se corrigieron la ubicación de contagio (provincias), ya que también había errores de escritura que complicaban el análisis del proyecto.

4. Medidas y visualizaciones

4.1 Medidas

Este apartado es para mostrar las mediciones y visualizaciones de nuestro proyecto, se muestran imágenes como datos concretos.

Medida	DAX
Edad promedio de muerte	Edad promedio de muerte = <code>CALCULATE(AVERAGE(Persona[Edad]),Hospital[Falleció]="Si")</code>
Total de	Tasa de recuperados = <code>DIVIDE([Total de recuperados],[Total de contagiados])</code>

Recuperados	
Total de contagiados	Total de contagiados = <code>COUNT(Persona[ID Persona])</code>
Muertes por COVID-19	Muertes por covid = <code>CALCULATE(COUNT(Hospital[Falleció]),Hospital[Falleció]="si")</code>
Tasa de Mortalidad	Tasa de mortalidad = <code>DIVIDE([Muertes por covid],[Total de contagiados])</code>

4.2 Medidas avanzadas y parámetros

En nuestro proyecto se encuentra una tabla denominada **BMedidasAvanzadas**, en donde se encuentran todas las medidas que cumplen con este requisito.

Medida	Dax
Segmentación por edades	0 a 18 años = <code>CALCULATE(COUNT(Persona[Edad]),Persona[Edad]<=18)</code> 19 a 60 años = <code>CALCULATE(COUNT(Persona[Edad]),Persona[Edad]>18,Persona[Edad]<=60)</code> 61 o mas = <code>CALCULATE(COUNT(Persona[Edad]),Persona[Edad]>=61)</code>
Contagiados por región	Contagiados Cuyo = <code>var A = CALCULATE([Total de contagiados],Persona[Provincia]="Mendoza")</code> <code>var B = CALCULATE([Total de contagiados],Persona[Provincia]="San Luis")</code> <code>var C = CALCULATE([Total de contagiados],Persona[Provincia]="San Juan")</code> <code>var D = CALCULATE([Total de contagiados],Persona[Provincia]="La Rioja")</code> <code>RETURN A+B+C+D</code> Contagiados NEA = <code>var A = CALCULATE([Total de contagiados],Persona[Provincia]="Chaco")</code> <code>var B = CALCULATE([Total de contagiados],Persona[Provincia]="Corrientes")</code> <code>var C = CALCULATE([Total de contagiados],Persona[Provincia]="Misiones")</code> <code>var D = CALCULATE([Total de contagiados],Persona[Provincia]="Formosa")</code> <code>RETURN A+B+C+D</code> Contagiados NOA = <code>var A = CALCULATE([Total de contagiados],Persona[Provincia]="Tucumán")</code> <code>var B = CALCULATE([Total de contagiados],Persona[Provincia]="Salta")</code> <code>var C = CALCULATE([Total de contagiados],Persona[Provincia]="Jujuy")</code>

	<pre> var D = CALCULATE([Total de contagiados],Persona[Provincia]="Catamarca") var E = CALCULATE([Total de contagiados],Persona[Provincia]="La Rioja") RETURN A+B+C+D+E Contagiados Pampeana = var A = CALCULATE([Total de contagiados],Persona[Provincia]="Buenos Aires") var B = CALCULATE([Total de contagiados],Persona[Provincia]="CABA") var C = CALCULATE([Total de contagiados],Persona[Provincia]="Entre Ríos") var D = CALCULATE([Total de contagiados],Persona[Provincia]="Santa Fe") var E = CALCULATE([Total de contagiados],Persona[Provincia]="Córdoba") var F = CALCULATE([Total de contagiados],Persona[Provincia]="La Pampa") RETURN A+B+C+D+E+F Contagiados Patagonia = var A = CALCULATE([Total de contagiados],Persona[Provincia]="Neuquén") var B = CALCULATE([Total de contagiados],Persona[Provincia]="Santa Cruz") var C = CALCULATE([Total de contagiados],Persona[Provincia]="Río Negro") var D = CALCULATE([Total de contagiados],Persona[Provincia]="Chubut") var E = CALCULATE([Total de contagiados],Persona[Provincia]="Tierra del Fuego") RETURN A+B+C+D+E </pre>
Contagiados según tipo de contagio	<pre> Comunitarios = CALCULATE(COUNT('Ministerio de la Salud'[Tipo de contagio]),'Ministerio de la Salud'[Tipo de contagio]="Comunitario") Contacto = CALCULATE(COUNT('Ministerio de la Salud'[Tipo de contagio]),'Ministerio de la Salud'[Tipo de contagio]="Contacto") Importados = CALCULATE(COUNT('Ministerio de la Salud'[Tipo de contagio]),'Ministerio de la Salud'[Tipo de contagio]="Importado") Investigación = CALCULATE(COUNT('Ministerio de la Salud'[Tipo de contagio]),'Ministerio de la Salud'[Tipo de contagio]="En Investigación") NA = CALCULATE(COUNT('Ministerio de la Salud'[Tipo de contagio]),'Ministerio de la Salud'[Tipo de contagio]="NA") Trabajadores de la Salud = CALCULATE(COUNT('Ministerio de la Salud'[Tipo de contagio]),'Ministerio de la Salud'[Tipo de contagio]="Trabajador de la Salud") </pre>
Cantidad de contagiados según sexo	<pre> Femenino = CALCULATE([Total de contagiados],Persona[Género]="femenino") Masculino = CALCULATE([Total de contagiados],Persona[Género]="masculino") </pre>
Edad promedio de contagios y si pertenece a la	<pre> Edad promedio de la población de riesgo = VAR A = SUM(Persona[Edad])/[Total de contagiados] RETURN IF(A<=65,"La edad promedio de contagiados NO pertenece a la población de riesgo","La edad promedio de contagiados SI pertenece a la </pre>

población de riesgo	población de riesgo")
---------------------	-----------------------

En el área de **parámetros** utilizamos los siguientes:

- Un parámetro para ir directamente a la base de datos, llamado *Conexión con la base de datos*
- Y el último parámetro para crear una baja porcentual que se encuentra dentro de la tabla *Baja de contagios* y que se utiliza en un solo gráfico que es el medidor de la misma hoja

La medida que se utilizó en el medidor es:

Baja porcentual de contagios = $[Total\ de\ contagiados] * (1 - ([Valor\ Baja\ de\ contagios]) / 100)$
--

4.3 Tabla Calendario

La tabla calendario fue creada con el propósito de filtrar y segmentar correctamente los datos. En este caso, en nuestro proyecto es muy importante dicha filtración y segmentación ya que el análisis lo hacemos por año y meses. Nuestra tabla se encuentra dividida en:

- Año
- Mes
- Bimestre
- Semestre
- Fecha ISO
- Día
- Cuatrimestre
- Semana

Pero únicamente se prioriza la división por año y meses para nuestro proyecto

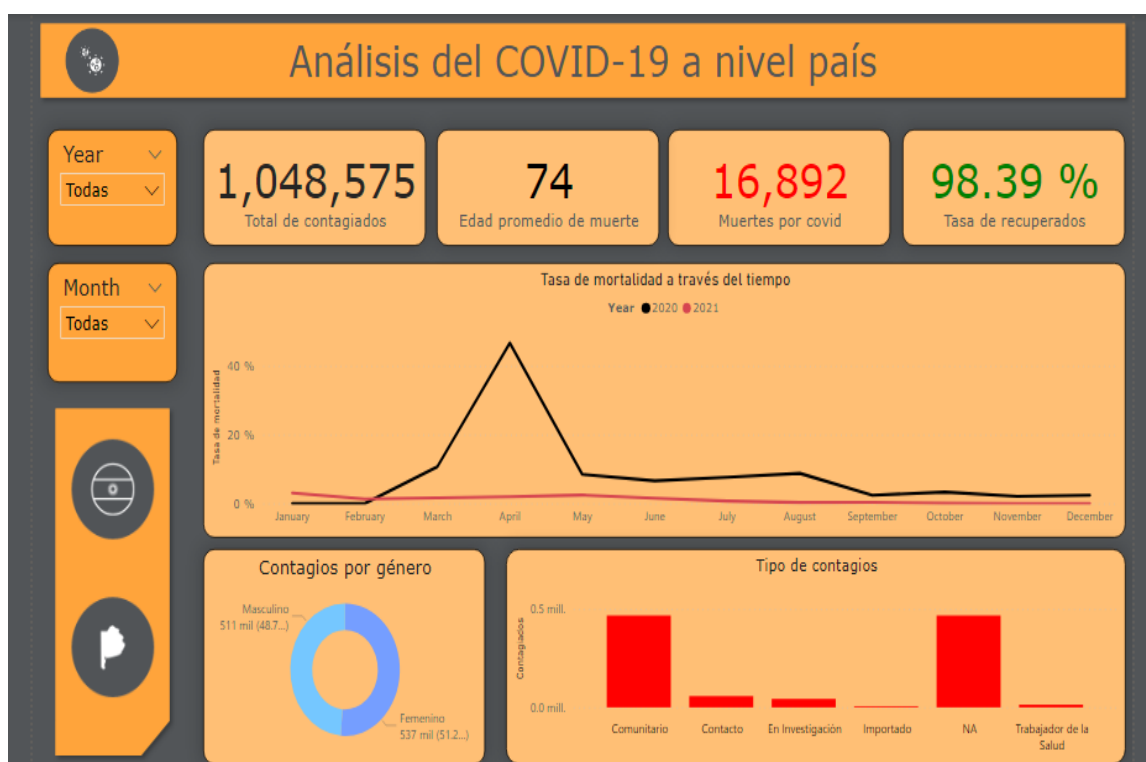
5. Visualizaciones

5.1 Solapa Inicio



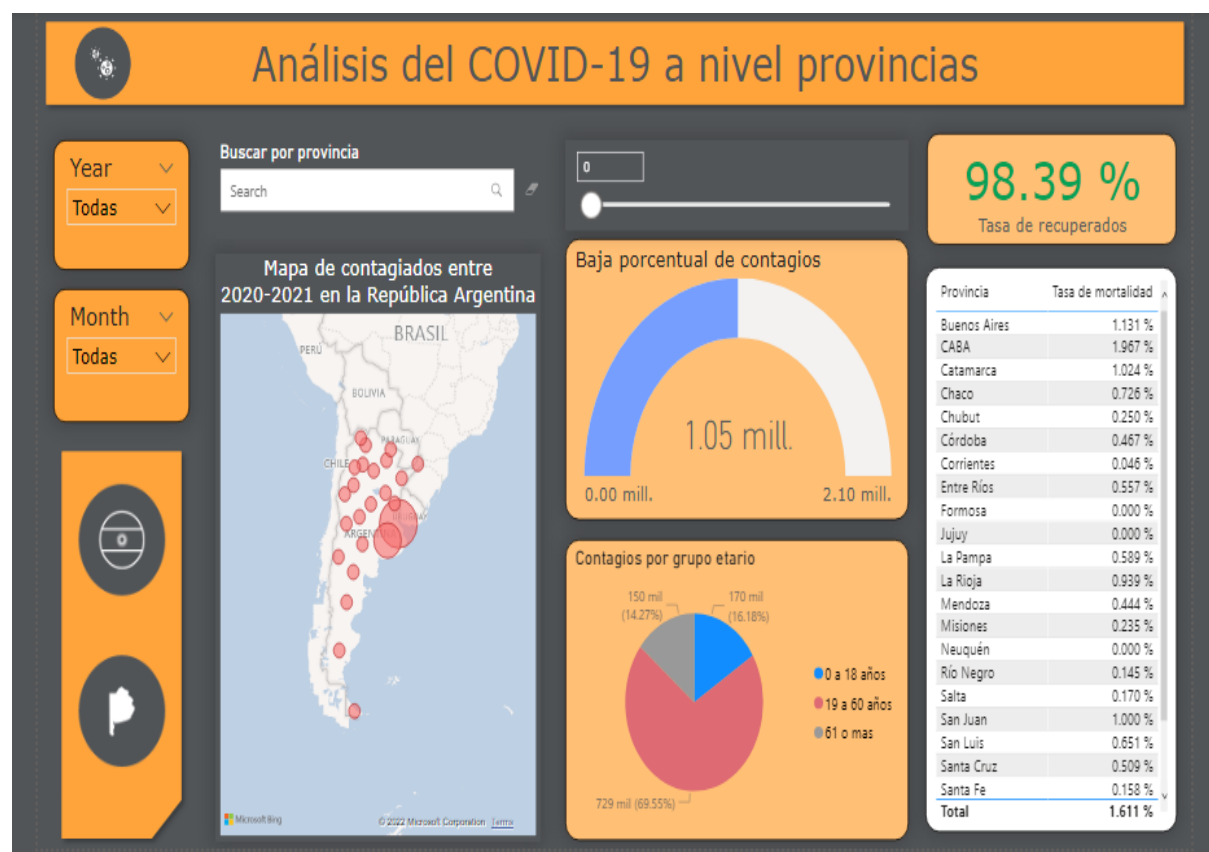
En la solapa **Inicio** se encuentra el título de nuestro proyecto, un logo y dos marcadores que dirigen al usuario a los dos principales ejes de análisis del tablero.

4.2 Solapa país



En esta solapa se ve un análisis más detallado, todo a **nivel país**. Cuenta con un título y un respectivo logo agregado hacia la izquierda del mismo. A su vez tenemos una pequeña región con los respectivos marcadores que nos llevan a la solapa a la que queramos ir. En la parte superior contamos con cuatro tarjetas, que muestran los contagios totales, la edad promedio de muerte, la cantidad total de fallecidos y la tasa de recuperados. Debajo de este contamos con tres gráficos que toman prácticamente toda la parte superior de la solapa. Hacia la izquierda un filtro que nos da la posibilidad de filtrar las cosas ya sea por mes o por año, o las dos al mismo tiempo.

4.3 Solapa provincia



En esta solapa se ve un análisis más detallado, todo a nivel provincial. Cuenta con un mapa muy detallado de las ubicaciones de cada provincia, con sus respectivas burbujas, dependiendo del tamaño de los contagios. A su vez tiene el mismo formato de cambio de solapa a través de íconos y filtración de datos que en la solapa PAÍS. Cuenta con una tarjeta que muestra el porcentaje de recuperados, además de un buscador en donde se

debe ingresar la provincia y todos los gráficos cambian dependiendo la provincia ingresada. En el centro podemos encontrar una proyección de baja porcentual de contagios representado en un medidor, arriba del mismo se puede interactuar de tal manera que se muestre la proyección deseada. Hacia la derecha encontramos una tabla con todas las provincias y su tasa de mortalidad, otra vez, estos datos pueden ser filtrados, por los mismos filtros de la izquierda, usando el mapa o el buscador en la parte superior.