

Proyecto Final

Área: Data Analytics.

Integrante: Germán López Búcar.

-Alcance del trabajo:

El tema elegido, es la problemática del COVID-19, tanto en Argentina como en los diferentes países del mundo. El objetivo es hacer un análisis informativo teniendo en cuenta la cantidad de casos, recuperados y muertes en los mismos a lo largo de estos últimos tres años. Así como ver la incidencia que tuvo las vacunas en los parámetros antes mencionados, e identificar si existe alguna relación. Hacer series temporales, para poder ver con más claridad la evolución de dichos parámetros en el tiempo por país. Además de tener una segunda vista a nivel de Argentina, analizando otras variables como el sexo, edad, ubicación local y tipo de contagio.

Los datos fueron sacados:

- Una tabla de la página “data.buenosaires” (la tabla de Argentina).
- Dos tablas de “Kaggle” (los datos en los diferentes países del mundo y la tabla de las vacunas).

-Análisis y transformaciones de las tablas:

COVID19_ARGENTINA

Data set extraído en la página “data.buenosaires”, donde detalla las muestras por covid-19 que se realizaron a través del tiempo en las distintas provincias. Así como también el género, tipo de contagio y edad de la persona.

Transformación:

- Le agregue un Id a la tabla “IdOpA”: Para que cada línea tenga una identificación única.
- Le agregue un campo “IdP” (FK), para poder vincularla con la tabla “PAISES”, y de esta forma tener la posibilidad de realizar futuras medidas o relaciones.
- En la parte de las fechas, las mismas estaban en formato texto con la parte del mes en inglés, y también con su correspondiente horario. Utilice varias fórmulas de Excel para su normalización y además de la eliminación de los “null”.
- Segregue la tabla en diferentes tablas dimensionales, para una mejor organización de los datos. Dichas tablas son: “PROVINCIAS”, “BARRIOS”, “GENERO”, “CLASIFICACION” Y “TIPOS_CONTAGIOS”. Para de esta forma poder agregar o modificar datos a estas tablas que produzca un impacto en la tabla principal de manera más sencilla y segura.

| COVID19_ARGENTINA | | |
|---------------------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdOpA | Int | PK-Index |
| IdP | Int | FK |
| IdProv | Int | FK |
| IdB | Int | FK |
| Fecha_Apertura | Datetime | - |
| Fecha_Muestra | Datetime | - |
| Fecha_Clasificacion | Datetime | - |
| Fecha_Alta | Datetime | - |
| Fecha_Fallecimiento | Datetime | - |
| IdCla | Int | FK |
| IdG | Int | FK |
| Edad | Int | - |
| IdCon | Int | FK |

| BARRIOS | | |
|---------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdB | Int | PK-Index |
| Barrios | Text(25) | - |

| CLASIFICACION | | |
|---------------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdCla | Int | PK-Inex |
| Clasificacion | Text(25) | - |

| PROVINCIAS | | |
|------------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdProv | Int | PK-Index |
| Provincias | Text(25) | - |
| Cod Map | Text(25) | - |

| GENERO | | |
|--------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdG | Int | PK-Index |
| Genero | Text(25) | - |

| TIPOS_CONTAGIOS | | |
|-----------------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdTCon | Int | PK-Index |
| Tipos_Contagios | Text(25) | - |

COVID19_MUNDO:

Data set extraído en la página “Kaggle”, donde detalla la cantidad de contagios, recuperados y muertes que se registraron en los diferentes países del mundo a través del tiempo (últimos tres años).

Transformación:

- Le agregue un Id a la tabla “IdOpM”: Para la identificación de las filas de la tabla.
- Normalización del campo fecha.
- Agregue el campo “IdP” (FK), para poder vincularla con la tabla “PAISES”.

| COVID19_MUNDO | | |
|---------------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdOpM | Int | PK-Index |
| IdP | Int | FK |
| Fecha | Datetime | - |
| Casos | Int | - |
| Recuperados | Int | - |
| Muertes | Int | - |

VACUNAS_MUNDO:

También extraído en la página “Kaggle”, donde muestra la cantidad de vacunas aplicadas, así como también el promedio de vacunados con al menos una dosis y con la vacunación completa que tiene registrado cada país del mundo.

Transformación:

- Le agregue un Id a la tabla “IdVac”, identificando las filas de esta.
- Normalización de los campos que contienen números, ya que anteriormente tenían formato texto.
- Modifique el campo “IdP” (FK), para poder vincularla con la tabla “PAISES”.

| VACUNAS_MUNDO | | |
|-------------------|---------------|------------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdVac | Int | PK-Index |
| IdP | Int | Clave Candidatas |
| Dosis_Aplicadas | Int | - |
| Vac_1-Dosis(%) | Decimal(16,2) | - |
| Vac_Full-Dosis(%) | Decimal(16,2) | - |

PAISES:

Diseñe esta tabla con todos los países que se mencionaban en las distintas tablas antes mencionadas, para poder vincularlas. Además, busque en internet con el fin de agregar a la lista de países su continente y su respectiva población.

Transformación:

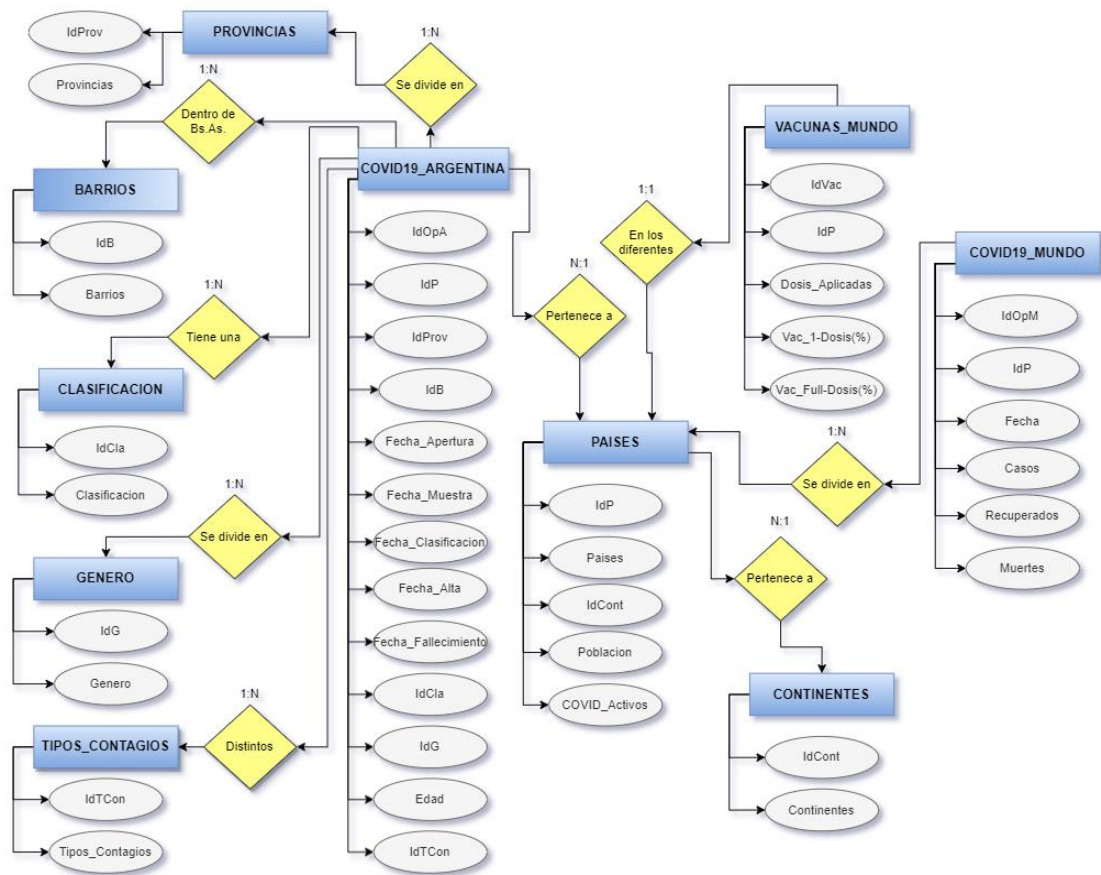
- Le agregue un Id a la table “IdP”, la cual es la PK de las tablas “COVID19_ARGENTINA”, “COVID19_MUNDO” y “VACUNAS_MUNDO”. Y de esa forma lograr la relación entre todas.
- En el campo “países”, realice la traducción de estos y junte todos los países de las tablas en esta columna.
- Segregue la tabla, creando una tabla de dimensión llamada “CONTINENTES”.

| PAISES | | |
|---------------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdP | Int | PK-Index |
| Países | Text(25) | - |
| IdCont | Int | FK |
| Poblacion | Int | - |
| COVID_Activos | Int | - |
| Cod Mapa | Text(25) | - |

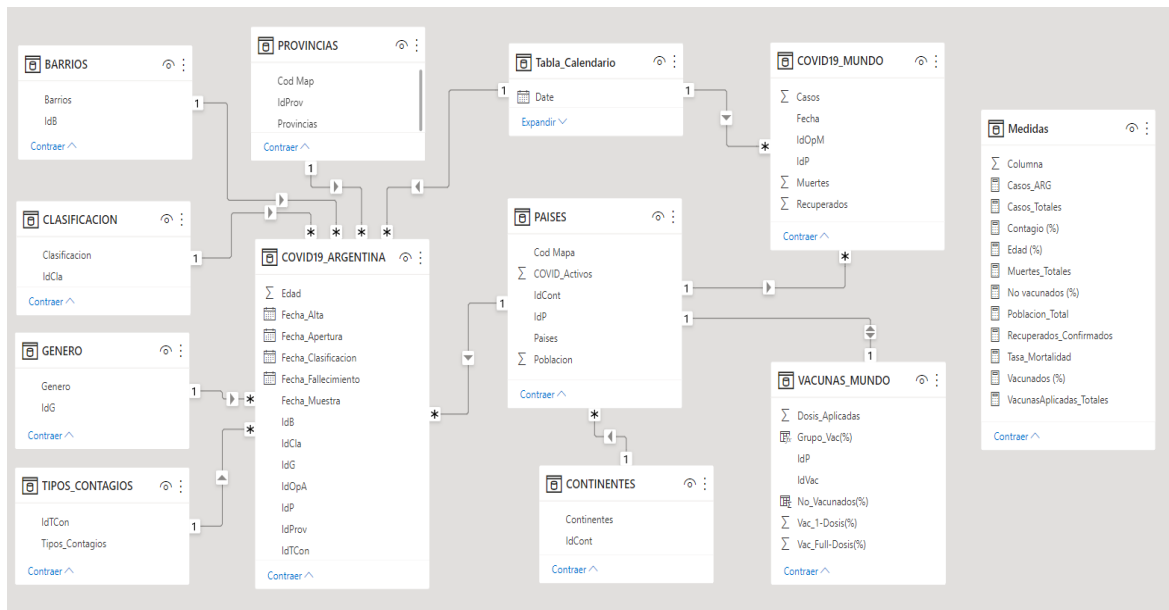
| CONTINENTES | | |
|-------------|---------------|---------------|
| Campo | Tipo de Campo | Tipo de Clave |
| IdCont | Int | PK-Index |
| Continentes | Text(25) | - |

Una vez que termine la normalización de las tablas, importe las mismas del Excel al Power BI, luego verifique y controle las relaciones designadas entre ellas.

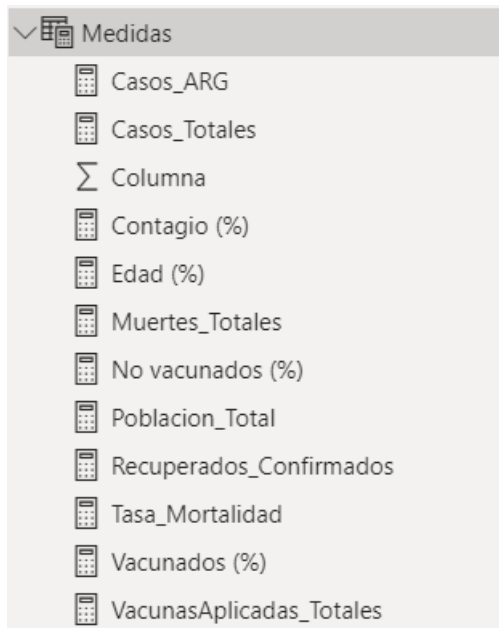
-Diagrama de Entidad- Relación realizado anteriormente:



-Modelo creado en Power BI:



-Medidas realizadas (Power BI):



- 1) Calcula la cantidad de casos que dieron positivo según su clasificación en argentina.

`Casos_ARG = CALCULATE(COUNT(COVID19_ARGENTINA[IdCla]), COVID19_ARGENTINA[IdCla] = 1)`

- 2) Calcula los casos totales por país.

`Casos_Totales = CALCULATE(SUM(COVID19_MUNDO[Casos]), COVID19_MUNDO[IdP])`

- 3) Divide dos funciones dax para calcular el contagio promedio sobre una población.

`Contagio (%) = DIVIDE([Casos_Totales], [Poblacion_Total])`

- 4) Edad promedio de las personas contagiadas.

`Edad (%) = CALCULATE(AVERAGE(COVID19_ARGENTINA[Edad]), COVID19_ARGENTINA[IdCla] = 1)`

- 5) Muertes totales por país.

`Muertes_Totales = CALCULATE(SUM(COVID19_MUNDO[Muertes]), COVID19_MUNDO[IdP])`

- 6) Calcula el promedio de los “No vacunados”.

`No vacunados (%) = AVERAGE(VACUNAS_MUNDO[No_Vacunados(%)])`

- 7) Población total por país.

`Poblacion_Total = CALCULATE(SUM(PAISES[Poblacion]), PAISES[IdP])`

- 8) Recuperados confirmados por país.

`Recuperados_Confirmados = CALCULATE(SUM(COVID19_MUNDO[Recuperados]), COVID19_MUNDO[IdP])`

- 9) Calcula la tasa de mortalidad de cada país, teniendo en cuenta las personas que se contagiaron.

Tasa_Mortalidad = `CALCULATE(DIVIDE([Muertes_Totales],[Casos_Totales]),COVID19_MUNDO[IdP])`

- 10) Promedio de vacunados con al menos una dosis sobre la población total.

Vacunados (%) = `AVERAGE(VACUNAS_MUNDO[Vac_1-Dosis(%)])`

- 11) Cantidad total de vacunas aplicadas.

VacunasAplicadas_Totales = `SUM(VACUNAS_MUNDO[Dosis_Aplicadas])`

-Campos calculados en la tabla "VACUNAS MUNDO":

- 1) Promedio de no vacunados por país.

No_Vacunados(%) = `1 - VACUNAS_MUNDO[Vac_1-Dosis(%)]`

- 2) Para crear 10 grupos entre todos los países según el promedio de vacunados que tienen sobre su población.

```



1 Grupo_Vac(%) = IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.90,"91% - 100%",
2 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.80,"81% - 90%",
3 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.70,"71% - 80%",
4 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.60,"61% - 70%",
5 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.50,"51% - 60%",
6 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.40,"41% - 50%",
7 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.30,"31% - 40%",
8 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.20,"21% - 30%",
9 IF(VACUNAS_MUNDO[Vac_1-Dosis(%)>0.10,"11% - 20%",
10 "0% - 10%"
11 )))))))

```

-Tabla calendario:


Tabla_Calendarario = `CALENDAR("01/01/2020","31/05/2022")`

-Marcadores:

| Marcadores |
|---|
|  Agregar  Vista |
| Pagina Portada |
| Pagina Mundo |
| ^ Visualización Solo |
| Solo casos Mundo |
| Solo recuperados Mundo |
| Solo muertes Mundo |
| Mapa Mundo |
| ^ Mapa |
| Recuperados Mundo |
| Casos Mundo |
| Muertes Mundo |
| Pagina Vacunas |
| ^ Visualización Vacunas |
| Porcentaje Vacunas |
| Cantidad Vacunas |
| Pagina ARG |
| Mapa ARG |
| ^ Consultas |
| Consulta Mundo |
| Consulta Portada |
| Consulta Vacunas |
| Consulta ARG |
| ^ Sin Filtro |
| Vacunas (SF) |
| ARG (SF) |
| Mundo (SF) |

| |
|--------------------------------|
| Actualizar |
| Cambiar nombre |
| Eliminar |
| Agrupar |
| <u>Datos</u> |
| ✓ Mostrar |
| ✓ Página actual |
| ✓ Todos los objetos visuales |
| Objetos visuales seleccionados |

Solapas:

-Portada: Contiene 2 marcadores (“Pagina Portada” y “Consulta Portada”), con el objetivo que si el usuario necesita una guia de pagina puede apretar el boton  (el mismo se situa en el extremo superior derecho), al precionarlo le aparecera una breve explicacion de la solapa. Los marcadores son para llevarlo a la consulta y traerlo de vuelta a la pagina.

Dicho botón aparece en todas las solapas disponibles, con la intención que el usuario pueda entender con mas facilidad el contenido de los gráficos. Agrupé los marcadores en el grupo “Consultas”.

Ademas cree un grupo de marcadores “Sin Filtro”, que se aplica en las ultimas 3 solapas, con la finalidad que el usuario pueda borrar rapidamente todos los filtros seccionados en la misma, cuando haya elegido mas de uno y quiera reiniciar los mismos.

-En el Mundo: En esta solapa existen varios marcadores como el grupo de “Visualización Solo”, donde la intención es mostrar gráficos e información de manera individual de los casos, recuperados ó las muertes.

Tambien esta el grupo “Mapa”, donde le permite al usuario interactuar con un mapa de color y poder ver de manera visual los casos, recuperados y muertes a traves del tiempo.

Por ultimo tenemos el marcador que nos lleva a la pagina principal (“Pagina Mundo”), y el que nos permite acceder al mapa antes mencionado (“Mapa Mundo”).

-Vacunas: En esta solapa ademas de los marcadores en común, se pueden ver dos mas en el grupo “Visualización Vacunas”, donde permiten el acceso a dos gráficos de barras atraves de sus respectivos botones.

-En Argentina: Podemos ver dos marcadores más (“Pagina ARG” y “Mapa ARG”), los cuales permite el acceso a un mapa de color, como así la vuelta a la pagina principal.

En varios marcadores, le destilde la opcion “Datos”, para que el usuario pueda interactuar con los botones sin perder los filtros seleccionados.

En otros marcadores se lo deje, ya que en esos casos es necesario el reinicio de los filtros.

-Visualizaciones descargadas:



Advance Card iconos asociados: 10



Tornado 2.1.0 iconos asociados: 1



Synoptic Panel by OKViz iconos asociados: 4

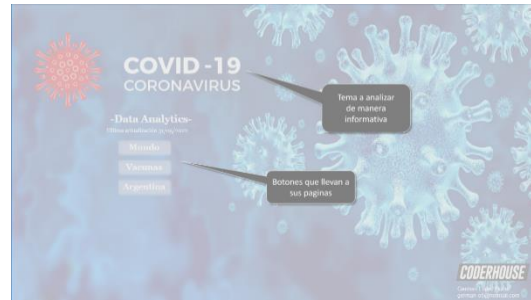
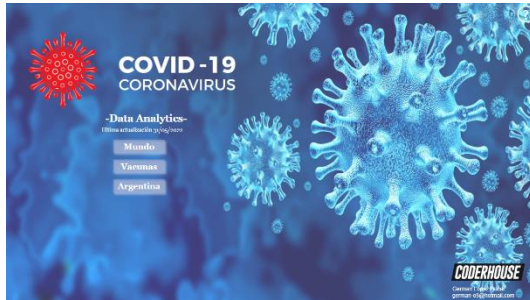
- 1) **Advance Card:** Son tarjetas de información donde se puede modificarlas visualmente, darle formato.
- 2) **Tornado 2.1.0:** Utilizado en la solapa “En Argentina”, en la parte del mapa. El cual muestra las provincias de Argentina mas afectadas con su promedio de edad.
- 3) **Synoptic Panel by Okviz:** Lo utilicé para lograr un mapa de color, mostrando una visualización más atractiva sobre la cantidad de casos, recuperados y muertes en los diferentes lugares del mundo a través del tiempo, ya que incorporé a la visualización una segmentación de tiempo, con el fin de ver la evolución de los parametros mencionados.
Para que el mapa funcione correctamente tuve que agregar un campo a la tabla “PAISES” llamado “Cod Mapa”, con el fin que la visualización descargada pueda identificar cada lugar a traves de un “Id”. Y en la tabla “PROVINCIAS” un campo “Cod Map” con el mismo objetivo.



El proyecto esta compuesto por 4 solapas, donde la primera indica el tema a tratar de manera descriptiva- informativa. Luego las dos solapas que le siguen analizan datos a nivel mundial, donde en “En el Mundo” compara los diferentes paises teniendo en cuenta la cantidad de casos, recuperados y muertes a traves del tiempo; asi como en “Vacunas”, relaciona el porcentaje de vacunados y no vacunados con los parametros de la sopala anterior, como lo son el porcentaje de contagios de una población y la tasa de mortalidad de los paises. Por ultimo tenemos la sopala “En Argentina”, que analiza los casos en las distintas provincias a través del tiempo.

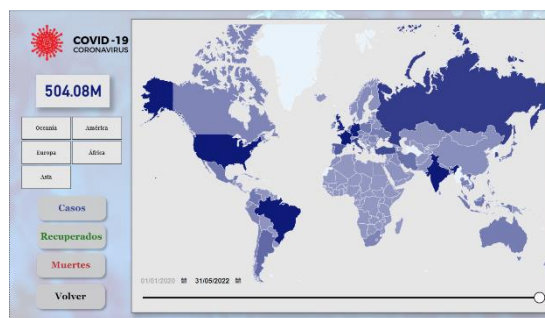
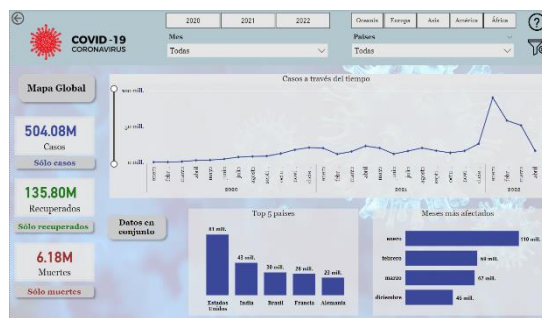
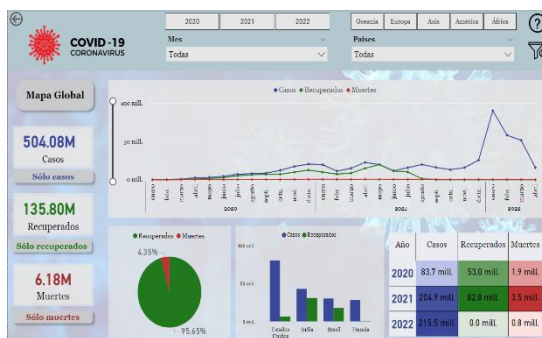
-Algunas imágenes del trabajo

Portada:



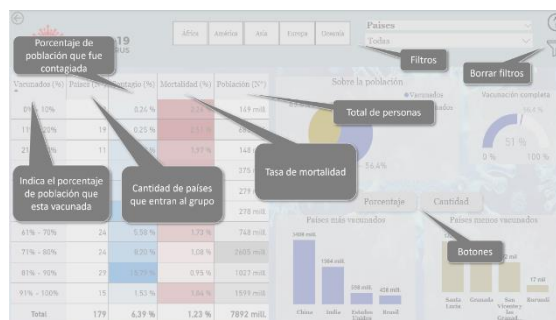
En el Mundo:

Segmentación: Por tiempo (Año y mes) y por lugar (Continente y País).



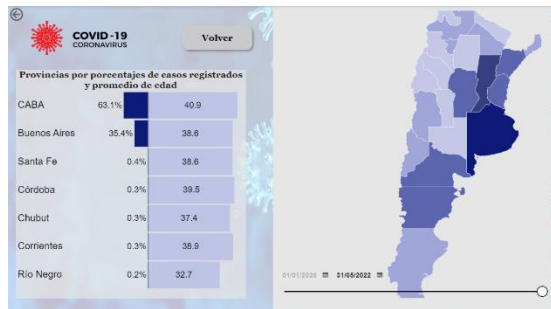
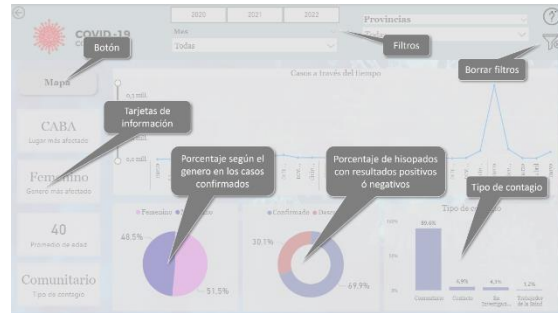
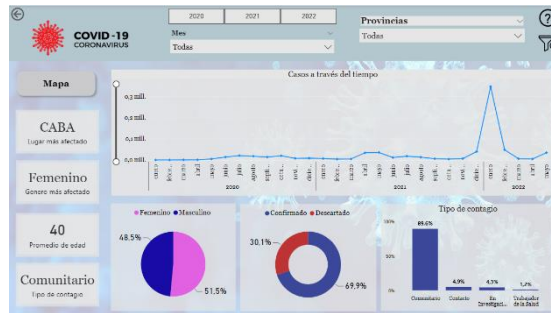
Vacunas:

Segmentación: Por lugar (Continente y País).



En Argentina:

Segmentación: Por tiempo (Año y mes) y por lugar (Provincia).



-Link del trabajo:

<https://app.powerbi.com/view?r=eyJrIjojYWU1YzlkODAtNDgzOCooNmIoLWIwY2EtNWMyYjM4MzAzZmQ4IiwidCI6IjFmN2Q0Njlk5LWNjZDctNDVhNi1iYzc4LWI4Yjk1MGJjYWVhYyIsImMiOjR9>