

# Proyecto Final: “Casos de COVID-19 confirmados 2020-2021”

## Integrantes:

- Sandoval Pablo

Curso: Data Analytics

Año: 2022



## Índice:

<b>Índice:</b>	<b>2</b>
<b>Descripción de la temática de los datos</b>	<b>2</b>
Hipótesis	2
Datasets	3
Diagrama Entidad - Relación (DER)	3
<b>Listado de tablas</b>	<b>3</b>
<b>Listado de columnas por tablas</b>	<b>4</b>
<b>Transformación los datos</b>	<b>7</b>

Integrantes:

Sandoval Pablo

## 1. Descripción de la temática de los datos

Para el presente proyecto, se definió como temática a tratar el análisis de los diversos casos confirmados de contagio causado por el COVID-19 detectados y registrados en diversos centros de testeo en Argentina.

Estos datos fueron extraídos durante el periodo 2020 - 2021 y los mismos se centraron en relevar los diversos casos confirmados, construyendo un perfil de cada paciente basado en su género, edad, y el tipo de contagio que sufrió, como también aspectos geográficos de los centros de testeo limitados en la provincia, barrio y comuna a la que está dirigido.

## 2. Objetivo del Proyecto

El presente proyecto tiene como objetivo resaltar de forma comparativa la dimensión de los casos confirmados en el periodo definido 2020 - 2021 clasificando los mismos a fin de que dicho análisis sirva o sea aplicable en casos de medición de impactos de las medidas sanitarias implementadas en esos plazos para futuras correcciones u observaciones.

## 3. Alcance

El corriente proyecto describe el desarrollo de un informe desarrollado con la herramienta de Microsoft Power BI, el cual estará conectado con una base de datos trabajada en SQL Server que cuenta con un total aproximado de 16 mil registros de casos confirmados y tablas relacionados a los mismos. Estas serán tratadas y desarrolladas con el fin de obtener gráficas que transmitan la información necesaria para alcanzar el objetivo formal del proyecto.

## 4. Usuario Final y nivel de aplicación del análisis:

Se tiene como usuario final a los miembros del ministerio de salud de la nación argentina a fin de poder dar seguimiento a los resultados de las medidas implementadas en el periodo y provincias muestreadas. Se aplicará un análisis tanto descriptivo como diagnóstico exponiendo información por medio de gráficos relacionada a los pacientes, centros de testeos y los casos como lo es la cantidad de pacientes registrados siendo clasificados de diversas formas (edad, género, provincia, tipo de contagio, etc).

## 5. Hipótesis

Por medio de los diversos centros de testeo públicos, ubicados en los barrios de CABA, BSAS entre otras provincias de Argentina (Chubut, Formosa, Misiones, Salta, Santa Fe, Catamarca, Chaco, Tucuman, San Luis, La Pampa, Neuquen, Cordoba, Tierra del Fuego, Santa Fe, Catamarca, Chaco, Tucuman, San Luis, La Pampa, Neuquen, Cordoba, Tierra del

Integrantes:

Sandoval Pablo

Fuego, Santa Cruz, Rio negro, Mendoza, Entre Rios, La Rioja San Juan, Santiago del Estero y NA en caso de no estar documentado) , se busca comprender el nivel de magnitud que trajo consigo el COVID 19 clasificándolos en diversos tipos de contagios a lo largo del tiempo, pudiendo medir a su vez en función de la edad y en la cantidad de fallecidos según cada categoría por caso registrado. Esto servirá en última instancia para poder tener una visión estadística acerca de la problemática y el alcance que tuvo en ese periodo; por medio de la clasificación se puede determinar si las medidas que en ese tiempo eran aplicadas sirvieron para responder a esta o es necesario realizar ajustes en dichas estrategias de sanidad a fin de tener información fehacientes ante potenciales olas de la pandemia.

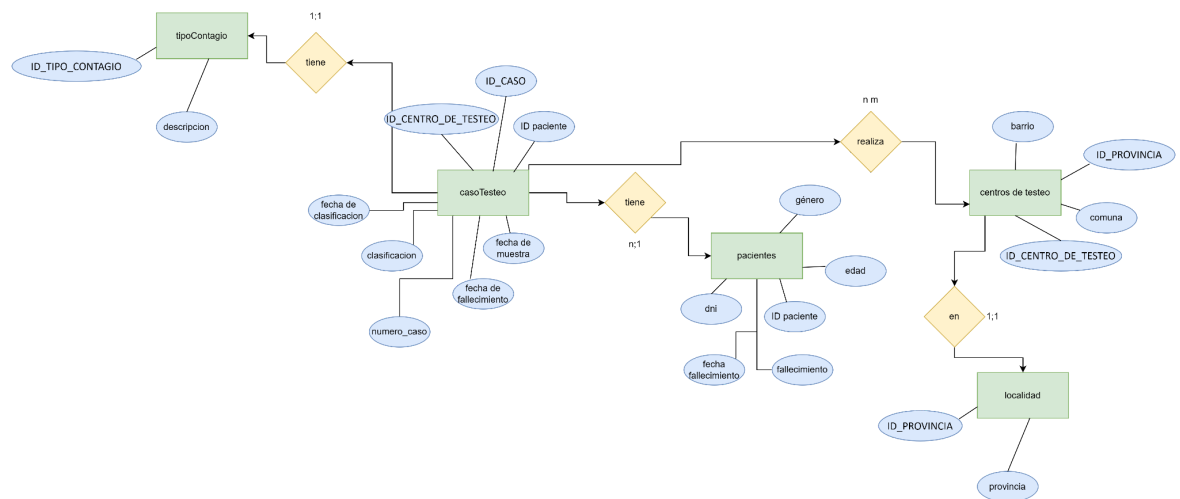
## 6. Datasets

En el presente apartado se adjuntará el enlace del conjunto de datos tabulados o Dataset que se empleara para realizar el análisis del corriente trabajo:

[Dataset covidFinal.xlsx](#)

## 7. Diagrama Entidad - Relación (DER)

En el siguiente apartado, se presenta a detalle el diagrama entidad relación del base de datos a emplear:



Integrantes:

Sandoval Pablo

## 8. Listado de tablas

En esta sección, se brinda una descripción de cada tabla disponible en la base de datos a trabajar, incluyendo cuales son sus claves primarias y secundarias:

- Casos de Testeo: contiene las fechas en el que se tomó las muestras como también que también la clasificación del caso como la causa del tipo de contagio.
  - PK: ID\_CASO
  - FK: ID\_CENTRO\_DE\_TESTEO
  - FK: ID\_TIPO\_CONTAGIO
- Pacientes: contiene el género, la edad, y si el paciente ha fallecido o no. En caso de que así sea, está computada a su vez la fecha del deceso.
  - PK: ID\_PACIENTE
  - FK: ID\_CASO
- Centro de testeo: contiene información tanto de la provincia a la que pertenece como, a un nivel más detallado, en que barrio y comuna se sitúa.
  - PK : ID\_CENTRO\_DE\_TESTEO
  - FK : ID\_PROVINCIA
- Tipo de contagio: en este se presenta a detalle los tipos de contagio en los que se puede clasificar los casos de COVID-19.
  - PK: ID\_TIPO\_CONTAGIO
- Localidad: contiene las diversas localidades en las que se sitúan los centros de muestreo.
  - PK: ID\_LOCALIDAD

## 9. Listado de columnas por tablas

A continuación se presentará, por tabla, los diversos campos con su tipo de datos correspondiente que las componen:

**TABLA CASOS DE TESTEO**

Nombre del campo	Primar y key	Candidate key	Foreing key	Tipo de dato
ID_CASO	x			text

Integrantes:

Sandoval Pablo

Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"

ID_CENTRO_DE_TESTEO			x	text
ID_PACIENTE				int
fecha de muestra				date
fecha de fallecimiento				date
fecha_clasificacion				date
clasificacion				varchar

**TABLA PACIENTES**

Nombre del campo	Primar y key	Candidate key	Foreing key	Tipo de dato
género				text
edad				int
dni				int
Id_paciente	x			int

**TABLA CENTRO DE TESTEO**

---

Integrantes:

Sandoval Pablo

Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"

Nombre del campo	Primar y key	Candidate key	Foreing key	Tipo de dato
CODIGO_CENTRO_DE_TESTEO	X			varchar
comuna				int
barrio				text
ID_PROVINCIA			X	text

**TABLA TIPO DE CONTAGIO**

Nombre del campo	Primar y key	Candidate key	Foreing key	Tipo de dato
ID_TIPO_CONTAGIO	x			text
descripción				text

**TABLA LOCALIDAD**

Nombre del campo	Primar y key	Candidate key	Foreing key	Tipo de dato
------------------	--------------	---------------	-------------	--------------

Integrantes:

Sandoval Pablo

ID_PROVINCIA	X			text
provincia				text

## 10. Transformación los datos

Para el diseño de la base de datos que se vinculó con el proyecto de power BI (.pbix), se tuvo que realizar cambios en las tablas de "CentroTesteo" y "CasosTesteo" a fin de poder vincular las tablas y construir el modelo de forma apropiada; esto se hizo cambiando las columnas de "id\_provincia" y "id\_tipo\_contagio" por los valores de las columnas del mismo nombre de las tablas "Localidad" y "TipoContagio". A su vez, posteriormente se cambió el tipo de dato que por defecto se seteo como varchar(50) a int.

Para dichas modificaciones se empleó SQL Server por medio de las siguiente transacciones:

```
BEGIN TRAN a
UPDATE CasosTesteo
SET id_tipo_contagio =CASE
WHEN ct.id_tipo_contagio='Contacto' THEN '1'
WHEN ct.id_tipo_contagio='Comunitario' THEN '2'
WHEN ct.id_tipo_contagio='En Investigación' THEN '4'
WHEN ct.id_tipo_contagio='Trabajador de la Salud' THEN '5'
ELSE '3'
END FROM CasosTesteo AS ct

ROLLBACK TRAN a
--commit TRAN a
SELECT * FROM CasosTesteo AS ct;
PRINT @@TRANCOUNT
BEGIN TRAN a
ALTER TABLE CasosTesteo
ALTER COLUMN id_tipo_contagio INT
sp_help CasosTesteo
```

Integrantes:

Sandoval Pablo



```
BEGIN TRAN a
UPDATE CentroTesteo
SET
    id_provincia = CASE
        WHEN ct.id_provincia='CABA' THEN '1'
        WHEN ct.id_provincia='Buenos Aires' THEN '2'
        WHEN ct.id_provincia='Chubut' THEN '3'
        WHEN ct.id_provincia='Formosa' THEN '4'

        WHEN ct.id_provincia='Misiones' THEN '6'
        WHEN ct.id_provincia='Salta' THEN '7'
        WHEN ct.id_provincia='Santa Fe' THEN '8'
        WHEN ct.id_provincia='Catamarca' THEN '9'
        WHEN ct.id_provincia='Chaco' THEN '10'
        WHEN ct.id_provincia='Tucuman' THEN '11'
        WHEN ct.id_provincia='San Luis' THEN '12'
        WHEN ct.id_provincia='La Pampa' THEN '13'
        WHEN ct.id_provincia='Neuquen' THEN '14'
        WHEN ct.id_provincia='Cordoba' THEN '15'
        WHEN ct.id_provincia='Tierra del Fuego' THEN '16'
        WHEN ct.id_provincia='Santa Cruz' THEN '17'
        WHEN ct.id_provincia='Rio Negro' THEN '18'
        WHEN ct.id_provincia='Mendoza' THEN '19'
        WHEN ct.id_provincia='Entre Rios' THEN '20'
        WHEN ct.id_provincia='La Rioja' THEN '21'
        WHEN ct.id_provincia='San Juan' THEN '22'
        WHEN ct.id_provincia='Santiago Del Estero' THEN '23'
        ELSE '5'
    END FROM CentroTesteo AS ct
ROLLBACK TRAN a
--commit TRAN a
SELECT * FROM CentroTesteo AS ct
PRINT @@TRANCOUNT
BEGIN TRAN a
ALTER TABLE CentroTesteo
ALTER COLUMN id_provincia INT
sp_help CentroTesteo
```

Integrantes:

Sandoval Pablo

## 11. Medidas calculadas:

A continuación se remarcaran las diversas medidas desarrolladas y los campos empleados para desarrollarlas:

Medida	Detalle	Descripción
Cantidad de centro de testeo	DISTINCTCOUNT(CentroTesteo[id_centro_de_testeo])	campo creado para expresar la cantidad total de centros de testeo captados en la muestra
Cantidad de pacientes	DISTINCTCOUNT(CasosTesteo[id_paciente])	campo creado para expresar la cantidad total de pacientes captados en la muestra
cantidad de pacientes masculinos	var a= CALCULATE(count(CasosTesteo[id_paciente]),Pacientes[genero]="masculino") return a	campo creado para expresar la cantidad total de pacientes de género masculino, captados en la muestra
Cantidad de tipos de contagio	DISTINCTCOUNT(TipoContagio[descripcion])	campo creado para expresar la cantidad total de tipos de contagio clasificados en la muestra
Pacientes Genero Femenino	COUNTX(filter(Pacientes,Pacientes[genero]="femenino"),Pacientes[id_caso])	campo creado para expresar la cantidad total de pacientes de género femenino, captados en la muestra
Pacientes genero Masculino	COUNTX(filter(Pacientes,Pacientes[genero]="masculino"),Pacientes[id_paciente])	campo creado para expresar la cantidad total de pacientes de género masculino, captados en la muestra
total de casos por fecha	total de casos por fecha = var b = count(CasosTesteo[id_caso]) var a =TOTALMTD(b,CasosTesteo[Fecha de fallecimiento]) return a	total de casos registrados por cada fecha registrada.

Integrantes:

Sandoval Pablo

Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"

Total de pacientes confirmados	COUNT(CasosTesteo[id_caso])	campo creado para expresar la cantidad total de pacientes captados en la muestra
Total de pacientes confirmados / total población BSAS	DIVIDE([Total de pacientes confirmados],15625084)	porcentaje de la muestra obtenida frente a la población total de buenos aires
total de pacientes sobre total de pacientes masculinos	'Parámetro'[Valor Parámetro]/[Cantidad de pacientes]	porcentaje del total de pacientes masculinos registrados frente al total de los pacientes registrados.
cantidad de pacientes de 2020	calculate(DISTINCTCOUNT(CasosTesteo[id_caso]),FILTER(CasosTesteo, year(CasosTesteo[Fecha de fallecimiento])=2020))	cantidad total de pacientes registrados en 2020
cantidad de pacientes de 2021	calculate(DISTINCTCOUNT(CasosTesteo[id_caso]),FILTER(CasosTesteo, year(CasosTesteo[Fecha de fallecimiento])=2021))	cantidad total de pacientes registrados en 2021
cantidad de pacientes de Buenos Aires 2020	calculate(DISTINCTCOUNTNOBLANK(CasosTesteo[id_caso]),FILTER(CasosTesteo,year(CasosTesteo[Fecha de fallecimiento])=2020), FILTER(CasosTesteo,CasosTesteo[provincia del centro de testeo]="buenos aires"),FILTER(CasosTesteo,CasosTesteo[provincia del centro de testeo]<>BLANK()))	cantidad de pacientes registrados en Buenos Aires en 2020
cantidad de pacientes de Buenos Aires 2021	calculate(DISTINCTCOUNTNOBLANK(CasosTesteo[id_caso]),FILTER(CasosTesteo,year(CasosTesteo[Fecha de fallecimiento])=2021), FILTER(CasosTesteo,CasosTesteo[provincia del centro de testeo]="buenos aires"),FILTER(CasosTesteo,CasosTesteo[provincia del centro de testeo]<>BLANK()))	cantidad de pacientes registrados en Buenos Aires en 2021

Integrantes:

Sandoval Pablo

Total de pacientes confirmados CABA	COUNTAX(FILTER(CasosTesteo,CasosTesteo[provincia del centro de testeo]="CABA"),CasosTesteo[id_caso])	total de pacientes registrados en CABA
Cantidad de centro de testeo (caba)	CALCULATE(DISTINCTCOUNT(CentroTesteo[id_centro_de_testeo]),FILTER(CentroTesteo,CentroTesteo[Provincia]="CABA"),FILTER(CentroTesteo,CentroTesteo[Provincia]<>BLANK()))	cantidad de centros de testeo en CABA
cantidad de centro de testeo (BSAS)	CALCULATE(DISTINCTCOUNT(CentroTesteo[id_centro_de_testeo]),FILTER(CentroTesteo,CentroTesteo[Provincia]="buenos aires"))	cantidad de centros de testeo en BSAS
cantidad de centro de testeo en otras provincias	CALCULATE(DISTINCTCOUNT(CentroTesteo[id_centro_de_testeo]),FILTER(CentroTesteo,CentroTesteo[Provincia]<>"buenos aires"),FILTER(CentroTesteo,CentroTesteo[Provincia]<>"CABA"))	cantidad de centros de testeo en otras provincias

## 12. Análisis funcional del tablero:

En este apartado se desarrollará una explicación para cada solapa presente en el proyecto:

- **Pacientes:** Se detalla varios aspectos relacionados a los pacientes registrados de los casos de covid como son:
  - cantidad total de pacientes registrados



- el porcentaje de pacientes registrados en CABA y BSAS frente a la población total

Integrantes:

Sandoval Pablo

## Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"



- Se empleó un gráfico de anillo para comparar el total de pacientes por género



- Se empleó un gráfico de línea para evaluar, de los pacientes registrados, la distribución media por rangos etarios, es decir, se pueda visualizar a mayor detalle el rango de edades que mayor cantidad de pacientes (se añadió un filtro para considerar los pacientes con más de un 1 año de edad)

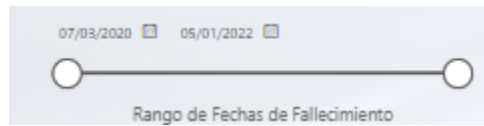


Integrantes:

Sandoval Pablo

## Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"

- Cabe destacar por último que se usó un segmentador de datos en forma de deslizador a fin de poder evaluar los cambios de forma dinámica en función de las fechas de fallecimientos registradas.



- **Casos confirmados:** Se analizará en detalle varias características de los casos registrados como son:
  - indicador clave de desempeño entre la cantidad de casos registrados en el año 2020 comparados con la cantidad de casos registrados en el año 2021



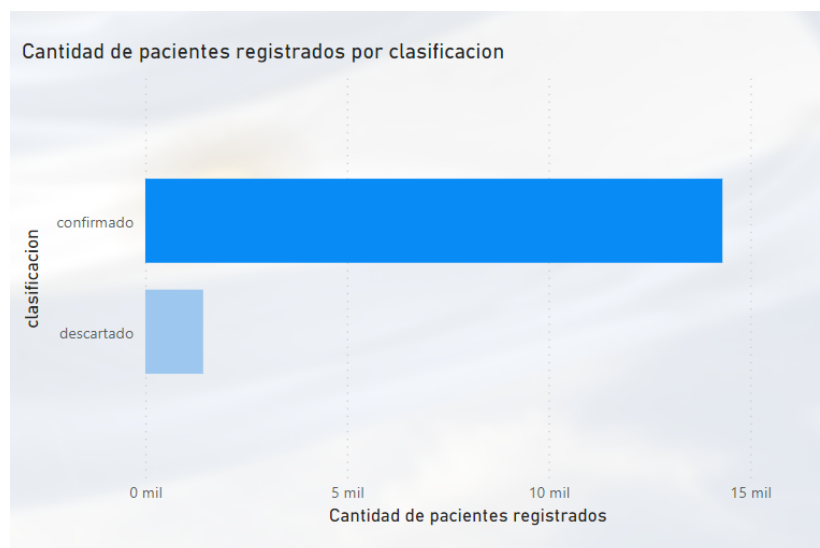
- se empleó el gráfico de barras para representar tanto la cantidad de casos por tipos de contagio como por la clasificación de confirmado y descartado



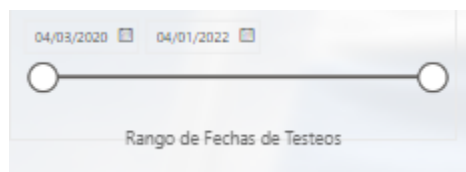
Integrantes:

Sandoval Pablo

## Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"



- Cabe destacar por último que se usó un segmentador de datos en forma de deslizador a fin de poder evaluar los cambios de forma dinámica en función de las fechas de muestreo registradas



- **centros de testeos:** Se mostrará varios aspectos registrados y relacionados con los centros de testeos y las provincias como son:
  - Por medio de gráfico de torta se expone de forma porcentual la cantidad de pacientes registrados en las diversas provincias

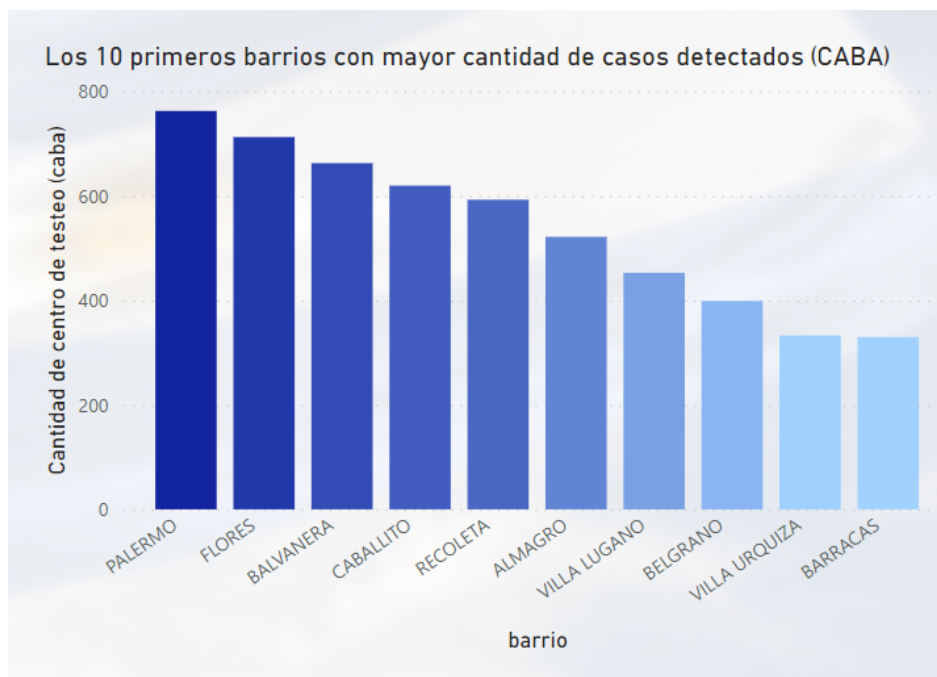


Integrantes:

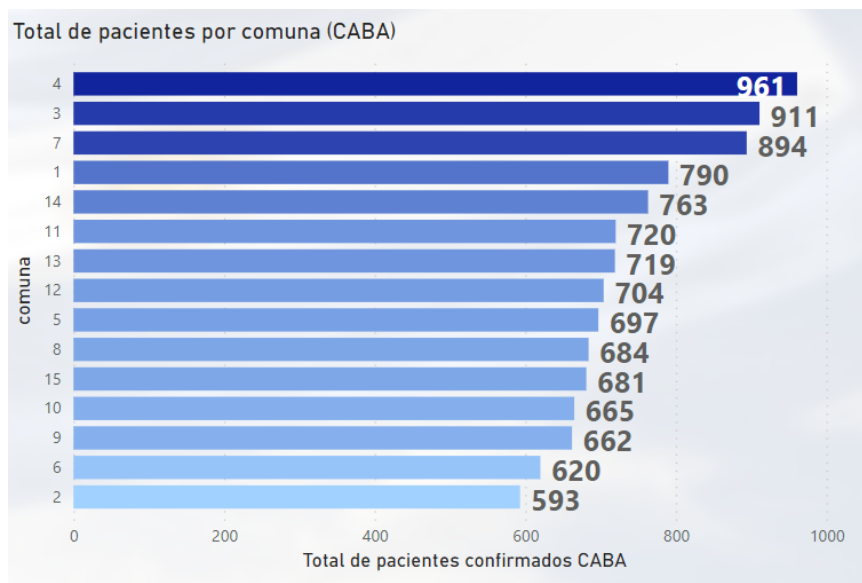
Sandoval Pablo

## Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"

- Se expone por medio de gráficos de barras 10 barrios de CABA con la mayor cantidad de casos registrados (se añadió un filtro para visualizar a los barrios con mas de 300 casos)



- Se expone por medio de un "gráfico de barras agrupado" la cantidad de casos detectados por comuna de CABA (se añadió un filtro para que no aparezcan los campos en blanco)



Integrantes:

Sandoval Pablo



## Proyecto Final: "Casos de COVID-19 confirmados 2020-2021"

- indicador clave de desempeño entre la cantidad de casos registrados en Buenos Aires en el año 2020 comparados con la cantidad de casos registrados en el año 2021



- Se emplea un segmentador de datos en forma de lista para segmentar los gráficos de provincias y "comunas"

Tipos de contagio

- ☐ Comunitario
- ☐ Contacto
- ☐ En Investigación
- ☐ NA
- ☐ Trabajador de la Salud

A list-based data filter titled "Tipos de contagio" with five unchecked checkboxes: Comunitario, Contacto, En Investigación, NA, and Trabajador de la Salud.

- Cabe destacar por último que se usó un segmentador de datos en forma de deslizador a fin de poder evaluar los cambios de forma dinámica en función de las fechas de muestreo registradas

04/03/2020 04/01/2022

Rango de Fechas de Testeos

A slider-based date range filter showing a timeline from 04/03/2020 to 04/01/2022. The slider is represented by a horizontal line with circular endpoints. Below the slider, the text "Rango de Fechas de Testeos" is displayed.

Integrantes:

Sandoval Pablo

## 13. Conclusión:

A partir de lo analizado por medio del tablero, se arribó a la conclusión de que la cantidad de casos de la población tomada como referencia para el relevamiento de datos en 2021 fue mucho menor al captado en 2020, demostrando la efectividad de las medidas vigentes en ese tiempo en una considerable disminución de casos considerables. Además de esto, se puede detectar el rango etario por medio de los picos denotados en el grafico de lineas de la solapa de paciente, que puede servir como referente a la hora de proponer alguna contramedida en respuesta a esto.

Es notable también mencionar las clasificaciones por tipo de contagio disponible en el tablero, que permite distinguir que la mayoría de casos fue "comunitario", es decir, personas que no estuvieron en contacto directo con alguien que tiene el virus, pero trabajan en el mismo lugar; por lo que fueron casos inevitables debido a la transición post-cuarentena en el periodo trabajado en este proyecto. Esto bien también puede ser reforzando las medidas de distanciamiento en dichos espacios públicos.

En conclusión, el tablero logró transmitir de manera satisfactoria el objetivo del presente proyecto el cual estaba centrado en analizar qué impacto tuvieron las medidas de salubridad en el periodo 2020-2021 en los casos de COVID-19.

Integrantes:

Sandoval Pablo