

¿Practicamos?

BOOKLET

DATA ANALYTICS

CODER HOUSE

¡Bienvenido/a!



Que bueno encontrarte por acá.

Este espacio ha sido creado para los entusiastas de la práctica y, es que todo lo que conseguirás en este documento ha sido pensado para que puedas **ejercitarse sobre los contenidos trabajados a lo largo de la cursada**.

Este Booklet (o cuadernillo de ejercicios) está compuesto de **diferentes actividades** por clase, con sus **respectivas resoluciones** de manera que puedas conseguir aquí una herramienta complementaria que te ayude profundizar y afianzar conocimientos. 💪 Toda la información y ejercicios que a continuación se presentan están vinculados a datos tomados de páginas públicas y libres, lo cual te permitirá tener una mirada del mercado, de lo que otros comparten y sobre lo que otros trabajan, para que puedas ir ajustando los productos que entregas y tu Proyecto Final.

En la siguiente slide te contamos un poco más sobre la estructura de este cuadernillo que **promete ayudarte** en tu proceso de aprendizaje. 🤝

SOBRE EL BOOKLET

Elementos que componen el Booklet:

- **Actividades:** son propuestas de ejercitación práctica, basadas en problemáticas comunes de la analítica de datos. Algunas consignas son específicas, pero otras son un poco más abiertas, con la intención de familiarizarnos el día a día de un analista de datos.
- **Resoluciones propuestas:** el booklet viene con resoluciones para que las compares con tus respuestas. Es propuesta por que **no es perfecta**, en casi ninguna rama de la programación encontrarás una solución única e infalible, ya que mucho depende del contexto. Por ende, si tu resolución presenta un funcionalidad similar o mejor a la propuesta en el booklet, ¡felicitaciones, resolviste la actividad!
- **Scripts SQL:** por cada actividad de SQL, contamos con un script de referencia, con el cual puedes comparar si lo que codificaste es similar a la funcionalidad esperada.
- **Power BI:** por cada actividad de Power BI se adjuntará una captura de pantalla de una resolución para comparar con tu resolución.

TABLERO DE CLASES

1
Data Analytics

2
Bases de datos

3
Bases de datos
relacionales

4
Sublenguaje
SQL DDL

5
Sublenguaje
SQL DML

6
Lenguaje SQL

7
Funciones
escalares y
subconsultas

8
Join & Union

9
Workshop

10
Charla de un
especialista en
base de datos

11
Historias con
Datos

12
Gráficos
eficientes

13
Introducción a
Power BI

14
Transformación
de datos

15
Dimensiones y
medidas

16
Gráficos y
segmentaciones

17
Medidas
avanzadas

18
Workshop
Power Bi

19
Charla de
especialista
visualización de
datos

20
Medidas de
tendencia
central

21
Tips and tricks

22
Proyecto Final

23
Presentación del
Proyecto Final

Toma un atajo para saltar directamente hasta la clase que deseas ejercitarte, haciendo clic en los cuadros presentados.



CODER HOUSE

¡EMPECemos!.. 🚀

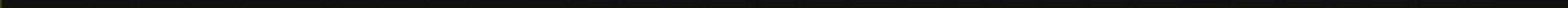


CODER HOUSE

CODER HOUSE

DATA ANALYTICS

Clase 1 - Data Analytics



ACTIVIDADES





FUENTES DE DATOS

Actividad N° 1

datos.gob.ar

Datos Argentina

Ponemos a tu alcance datos públicos en formatos abiertos para que puedas usarlos, modificarlos y compartirlos. Estos datos son tuyos. Podés crear visualizaciones, aplicaciones y grandes herramientas con ellos.

Ingresá al portal de datos públicos de la República Argentina, navega por las distintas secciones y encuentra datasets en tres formatos distintos.



¿Cuáles encontraste?



Notas actividad 1:

Datos Públicos de Argentina: <https://datos.gob.ar/>



EN BUSCA DE DATASETS

Actividad N° 2

Dataset Search

Search for Datasets 

Try [boston education data](#) or [weather site:noaa.gov](#)

[Learn more](#) about including your datasets in Dataset Search.

Encontrar un conjunto de datos parece fácil, 😐 pero no lo es pues hay muchos lugares donde buscar.

Navega por la web y encuentra 3 fuentes de datos fiables.



Notas actividad 2:

Datos sobran, la calidad escasea. Prioriza la calidad y los sitios de renombre. Buenos datos generan resultados fiables 💪

INFORMACIÓN Y DATOS



Actividad N° 3

ID	BILL DATE	VENDOR	BAN	MATCH TO	MATCH STATUS	AUDIT
1	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"REF ONLY"	"EXCEPT"	
2	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"EX"	
3	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"FC"	
4	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"EX"	
5	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"FC"	
6	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"FC"	
7	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"FC"	
8	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"FC"	
9	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"VA"	
10	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
11	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
12	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
13	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
14	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
15	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
16	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
17	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
18	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"V"	
19	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
20	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
21	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
22	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
23	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	
24	11/8/2008 0:00:00	"ATT-NRC"	"3107020003552"	"INV TO MIROR"	"F"	

Tener los datos no es lo mismo que tener información. Con datos solos no hacemos nada; necesitamos generar información. 

Busca un ejemplo de datos y cómo han sido transformados en información interesante.



Notas actividad 3:

La información siempre implica procesar los datos de forma tal que se obtenga conocimiento relevante.

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Fuentes de datos

datos.gob.ar/dataset/enacom-total-nacional-accesos-telefonia-fija

datos.gob.ar

Datasets

Datos Argentina / Dataset

Total nacional de accesos a telefonía fija

Ente Nacional de Comunicaciones (ENACOM) 

Número total de accesos al servicio de telefonía fija (hogares, comerciales, gobierno y otros) (trimestral)

Recursos del dataset

 Total nacional de accesos a telefonía fija
Número total de accesos al servicio de telefonía fija (hogares, comerciales, gobierno y otros) (trimestral) CONSULTAR  DESCARGAR 

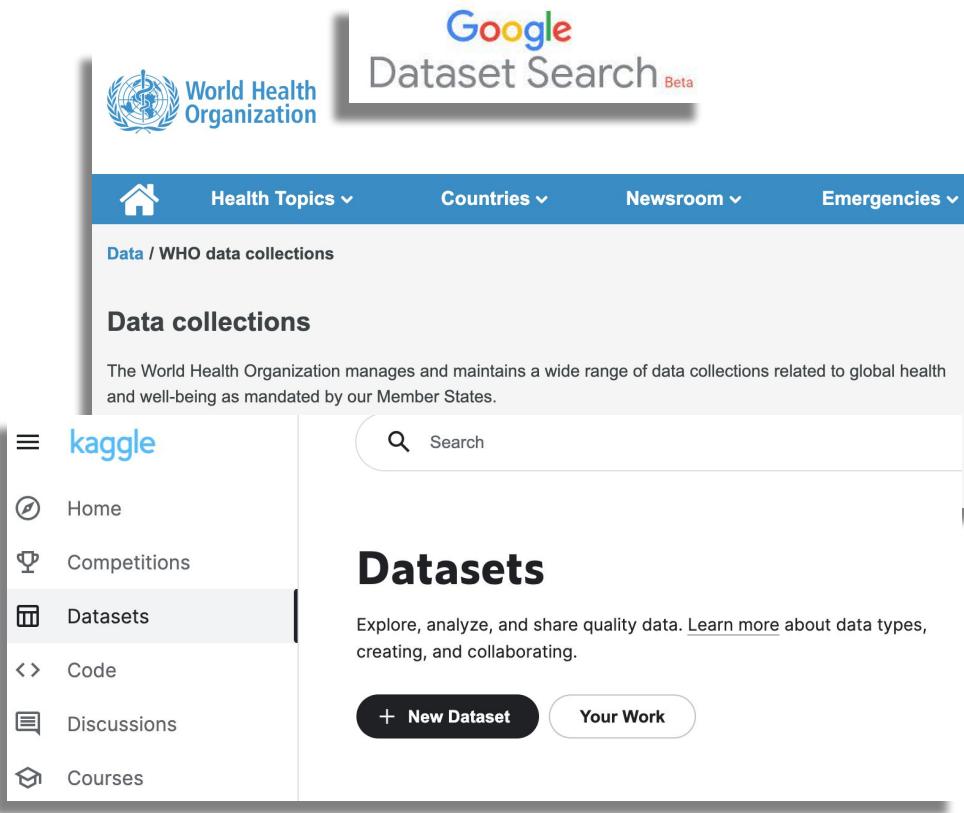
 Total nacional de accesos a telefonía fija
Número total de accesos al servicio de telefonía fija (hogares, comerciales, gobierno y otros) (trimestral) CONSULTAR  DESCARGAR 

 Total nacional de accesos a telefonía fija
Número total de accesos al servicio de telefonía fija (hogares, comerciales, gobierno y otros) (trimestral) CONSULTAR  DESCARGAR 

Conocer este tipo de información nos permitirá identificar los distintos formatos y reconocer cuál nos conviene más en cada caso. En el ejemplo de la imagen podemos ver el mismo dataset en tres formatos diferentes: CSV, XLS y JSON. Recuerda que los datos dentro de los archivos son los mismos, lo único que cambia es el formato en el cual se presentan. 😊

SOLUCIÓN PROPUESTA

Actividad N° 2 - En busca de Datasets

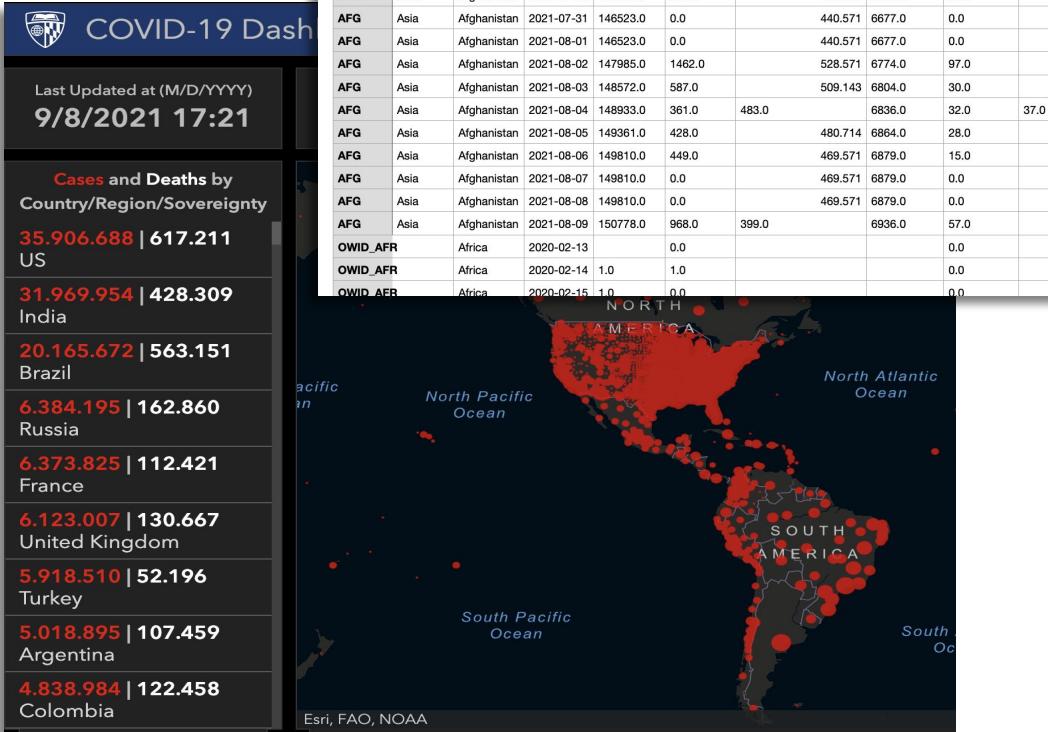


The screenshot shows the Google Dataset Search interface. At the top, there's a banner for the World Health Organization (WHO) with the text "Data / WHO data collections". Below the banner, there's a search bar with the placeholder "Search". On the left side, there's a sidebar with navigation links: Home, Competitions, Datasets (which is currently selected), Code, Discussions, and Courses. The main content area has a title "Datasets" and a sub-instruction "Explore, analyze, and share quality data. [Learn more](#) about data types, creating, and collaborating." There are two buttons at the bottom: "+ New Dataset" and "Your Work".

A la hora de buscar datos para tu proyecto final, no escatimen ni en tiempo ni en calidad. Sitios de renombre como [Google Dataset Search](#), [Kaggle](#) o la [OMS](#) brindan datos fiables y de gran reusabilidad. Además, muchos países tienen sitios propios con fuentes de datos públicas (ver  [aquí](#) y [aquí](#)).

SOLUCIÓN PROPUESTA

Actividad N° 3 - Información y datos

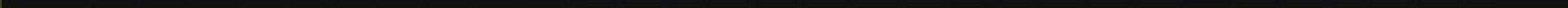


Por cada persona infectada con COVID-19 se ingresa un registro en una base de datos. Miles de registros son procesados y generan información valiosa sobre, por ejemplo, qué variante prevalece en determinada zona geográfica y permite a los científicos mejorar las vacunas.

CODER HOUSE

DATA ANALYTICS

Clase 2 - Bases de datos



ACTIVIDADES





HERRAMIENTAS DE BASE DE DATOS

Actividad N° 1



Si bien Microsoft SQL es el software más conocido para manejo de bases de datos, hay muchos otros. Busca en la web qué otras herramientas permiten administrar bases de datos.



Notas actividad 1:

Hay infinidad de herramientas disponibles, solamente busca algunas a modo de ejemplo ¡No hay respuestas incorrectas!



POR LA NUBES

Actividad N° 2



Cloud Processing

Un cambio radical de los últimos años es la migración hacia el almacenamiento y procesamiento de datos en la nube, pero esto tiene ventajas y desventajas. Investiga cómo funciona BigQuery y cuál es la principal y costosa diferencia respecto a las herramientas de la actividad anterior.



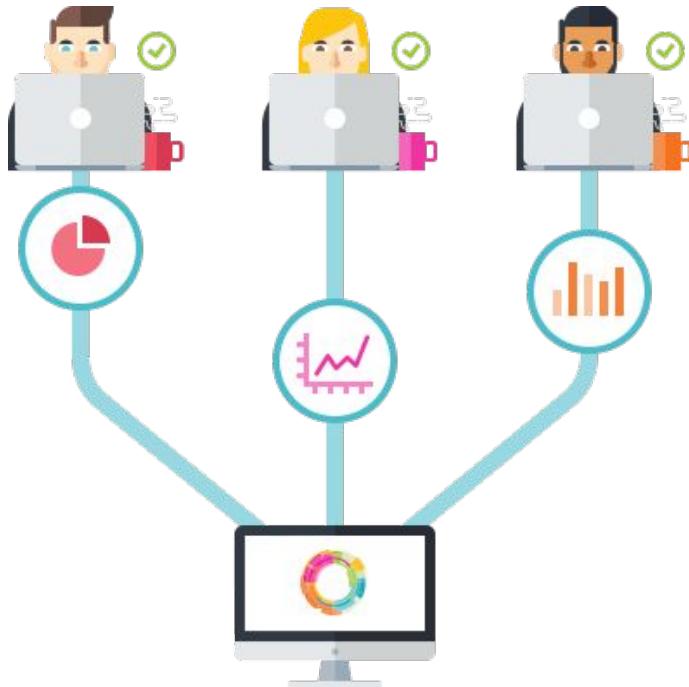
Notas actividad 2:

Diferencias hay muchas, pero hay una que es relevante y notoria entre las herramientas “terrenales” y las de la “nube”.



JUNTANDO DATOS

Actividad N° 3



A veces es necesario conseguir datos externos que complementen los datos que ya poseemos. Por ejemplo, necesitaríamos la cotización diaria del dólar USD para transformar los precios de las propiedades en Argentina a pesos. 💰

Encuentra un conjunto de datos que pueda ser complementado con otro.



Notas actividad 3:

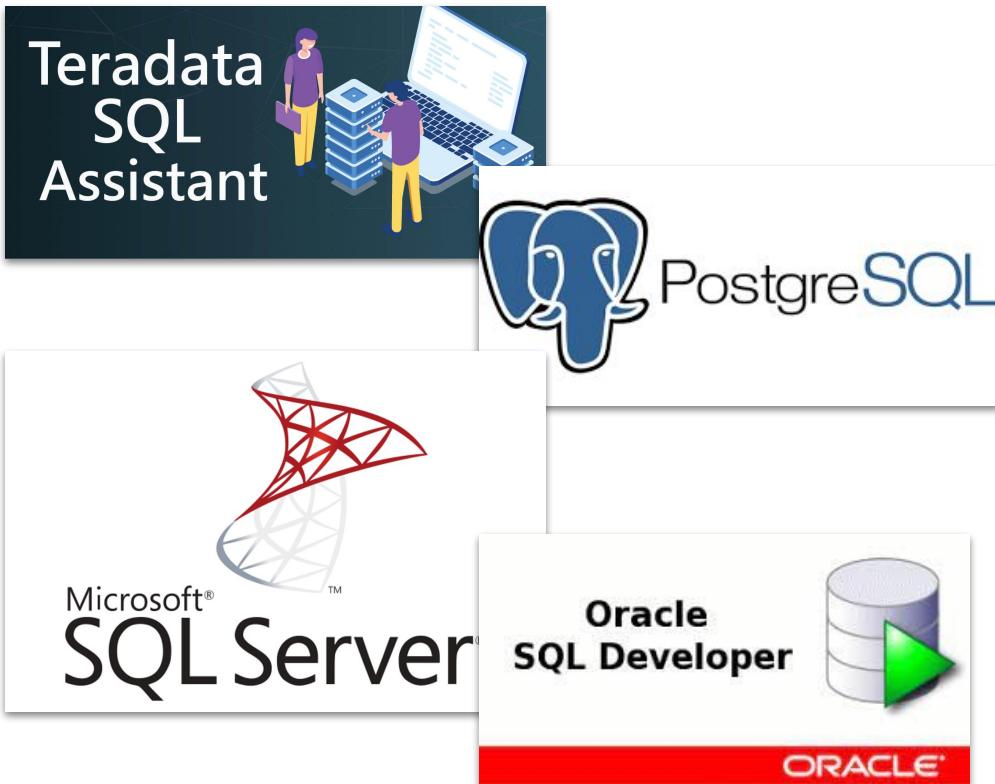
En nuestras labores cotidianas es importante buscar datos de diversas fuentes. Estas nos darán información extra que nos ayudará a tener una nueva perspectiva. 🧐

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Herramientas de bases de datos



SQL es casi un lenguaje universal. Si bien hay montones de herramientas y cada una tiene sus particularidades, todas manejan el mismo idioma: SQL.



SOLUCIÓN PROPUESTA

Actividad N° 2 - Por la nubes

BigQuery tiene entre sus características cobrar por cada operación que se realice. La mayoría de sus servicios cloud se abonan por transacción, cant. de datos procesados, etc. (ver calculadora de precios [acá](#)).

Cada vez que se ejecute un comando en BigQuery se detallará la cantidad de datos a procesar y con ello se podrá tener una idea de cuanto sale realizar esa operación.



Unsaved query Edited HIDE EDITOR

```
1 SELECT * FROM irs_990.irs_990_ein
```

Run Save query Save view Schedule query More

This query will process 383.1 MB when run. ✓

SOLUCIÓN PROPUESTA

Actividad N° 3 - Juntando datos

La ciudad de Buenos Aires registra los viajes realizados en bicicletas públicas. 

👉 Si cruzamos esos datos con los del servicio meteorológico podríamos saber si los días de lluvia o de altas temperaturas la gente viaja más (o menos).

👉 Podríamos obtener INFORMACIÓN sobre cuándo conviene reparar las bicicletas (porque habrá pocas en uso) o cuando tener más disponibles (porque habrá más demanda).

Todo esto simplemente mirando el pronóstico del clima.

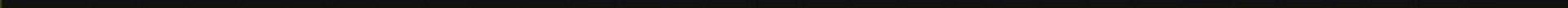
- Bicicletas: [dataset](#)
- Clima: [dataset](#)

Unir ambos datasets por fecha.



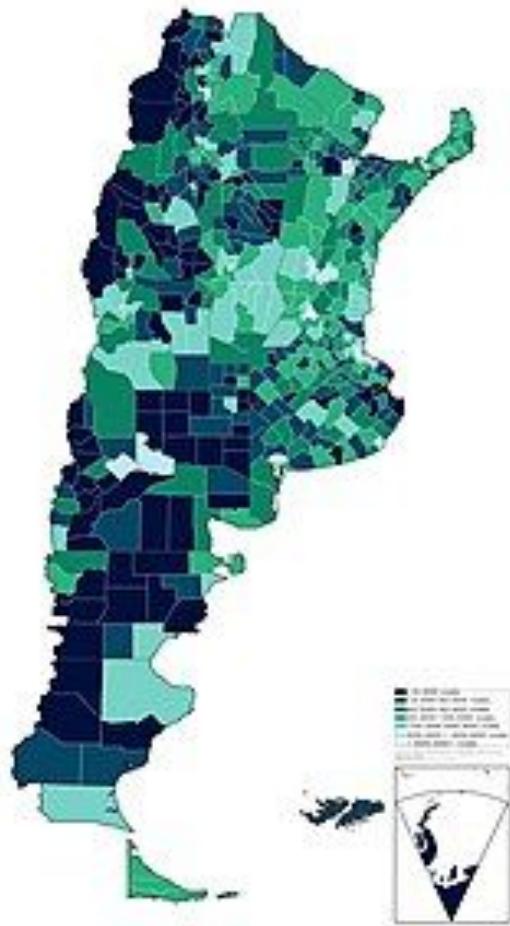
DATA ANALYTICS

Clase 3 - Bases de datos relacionales



ACTIVIDADES





CLAVES PRIMARIAS

Actividad N° 1



Descarga el archivo de [municipios de Argentina](#). Una vez descargado mira su contenido. El mismo detalla todos los municipios/departamentos que componen la República Argentina.

¿Qué campo tomarías como clave primaria (PK) para esa tabla si tuvieras que diseñarla desde cero? 😊



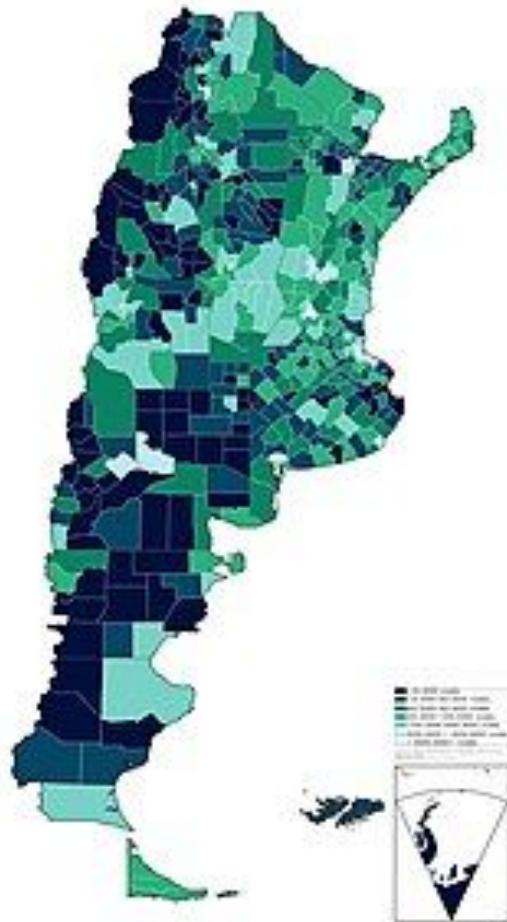
Notas actividad 1:

Analizar el diseño de una base de datos es algo que debe ser pensado al detalle y no tomado a la ligera. Lo que nos parece obvio a veces no lo es. 🧠



CLAVES FORÁNEAS

Actividad N° 2



Sigamos con el archivo de [municipios de Argentina](#). Hay dos columnas que en su nombre contienen las siglas **ID**. Parece que hacen referencia a un identificador, a una clave pero ¿A qué harán referencia realmente? ¿Qué tablas externas crearías para que estos ID sean FK (foreign key)? ¿Porqué serán necesarias estas tablas?



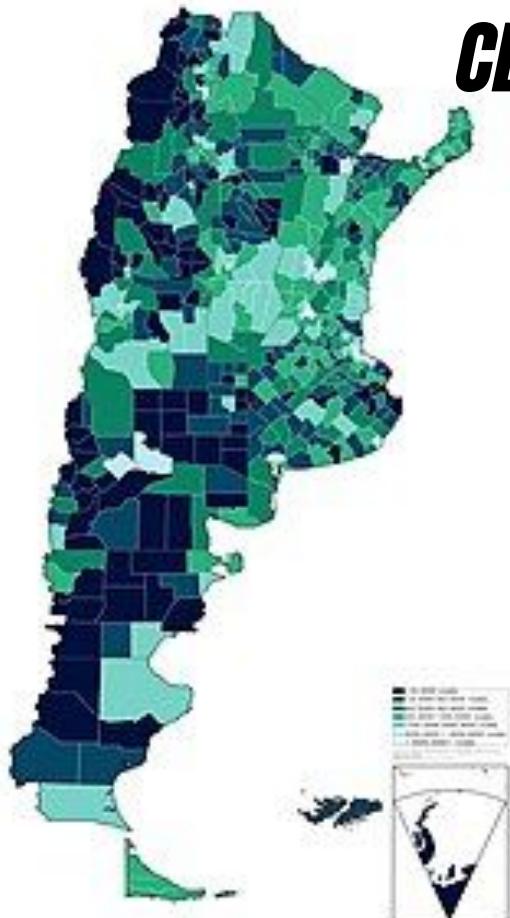
Notas actividad 2

Analizar el diseño de una base de datos es algo que afecta el presente pero se debe hacer mirando al futuro, por eso es importante pensar en qué cambios podrían afectar mis datos.



CLAVES Y DATOS EXTERNOS

Actividad N° 3



Sigamos, todavía más, con el archivo de [municipios de Argentina](#) 😊 ¿Qué pasaría si unieras los municipios de Argentina con los de otro país? ¿Qué precauciones deberíamos tener en cuenta?



Notas actividad 3

Analizar el diseño de una base de datos es también pensar cómo van a crecer nuestros datos. Un buen diseño resiste el paso del tiempo. ⏳

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Claves primarias

La primera impresión al analizar el archivo con los datos es considerar al nombre del departamento como clave primaria dado que de eso se trata la base. Sin embargo, podrían existir departamentos o municipios con igual nombre y distinta provincia.  En ese caso debiéramos considerar una clave primaria compuesta, es decir, conformada por dos campos. Por ejemplo, hay dos municipios MERLO, uno en Buenos Aires, otro en San Luis. La forma de diferenciarlos como PK será: Merlo-San Luis, Merlo-Buenos Aires (por municipio - provincia).

👉 Para pensar: ¿Qué función cumple la columna ID?

SOLUCIÓN PROPUESTA

Actividad N° 2 - Claves foráneas

Hay dos columnas ID: una para los municipios/departamentos, otra para las provincias. A la hora de definir claves primarias es conveniente usar números dado que estos son más “baratos” para almacenar y procesar. Nuestra tabla de municipios tiene datos redundantes, el ID hace referencia al nombre, con lo cual una de las columnas sobra.  El espacio cuesta dinero, el procesamiento cuesta dinero, por lo que al diseñar nuestra base debemos ahorrar todo el espacio posible. Así que podríamos generar dos tablas externas, una con los nombres de las provincias y otra con los nombres de los municipios. Entonces ganaremos reusabilidad y ahorraremos espacio. 

SOLUCIÓN PROPUESTA

Actividad N° 3 - Claves y datos externos

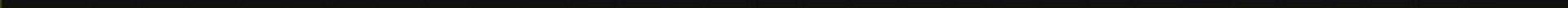
No muchos lo saben, pero hay un Londres en Argentina, 😱 precisamente en la provincia de Catamarca. Si uniéramos nuestros datos con los condados de Inglaterra, tendríamos dos Londres. Datos duplicados en una base de datos, uno de los peores errores.

Entonces debiéramos generar una nueva tabla que contenga los datos de los países y una nueva columna en nuestro conjunto de datos original que haga las veces de Foreign Key a la tabla de países. De esa manera podremos saber si se trata de Londres Catamarca, Argentina, o Londres, Gran Londres, Inglaterra. ¿Se imaginan cuántos Londres hay en el mundo? Miren y piensen los problemas que se pueden evitar al saber esto:
<https://es.wikipedia.org/wiki/London>



DATA ANALYTICS

Clase 4 - Sublenguaje SQL DDL I



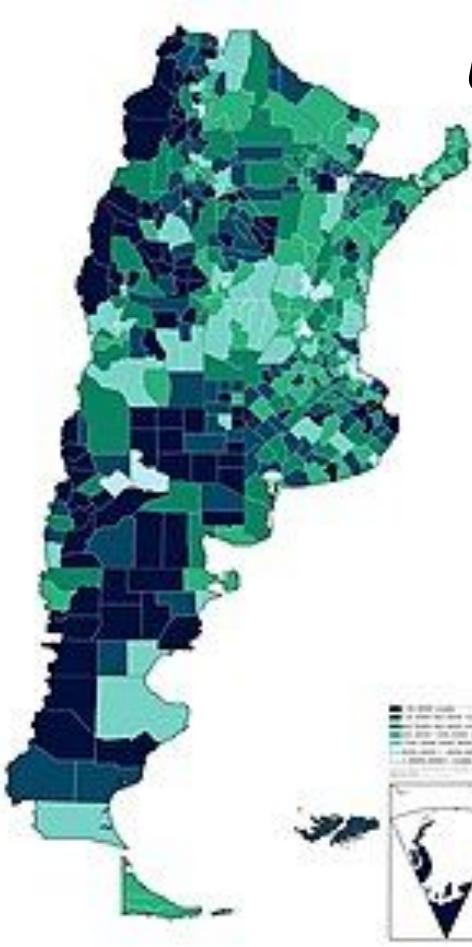
ACTIVIDADES





CREACIÓN DE BASES DE DATOS

Actividad N° 1



Ya que estamos conociendo en profundidad el archivo de municipios, vamos a crear unas tablas ¿Te animas? 😊

Recuerden descargar los datos de [municipios de Argentina](#).

Vamos a crear una base de datos llamada *MUNICIPIOS*.



Notas actividad 1:

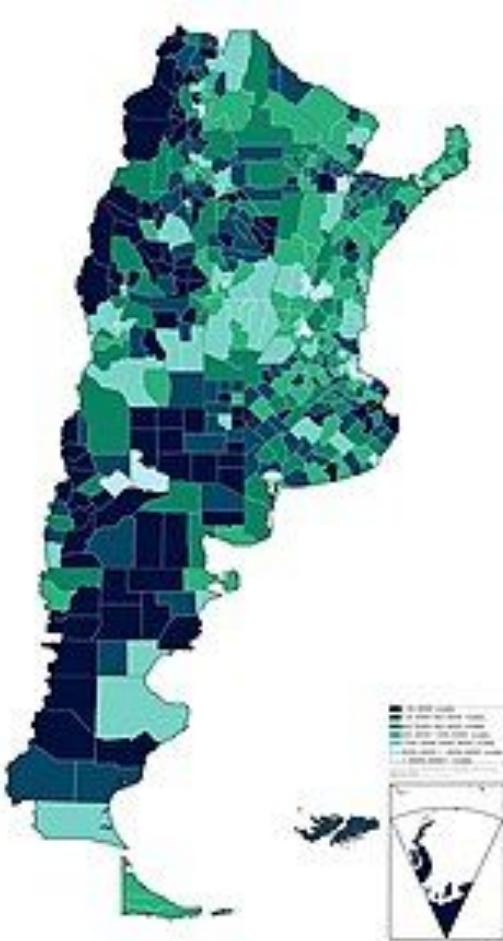
La creación de la base de datos es la piedra fundamental sobre la cual trabajaremos. Es el lugar donde crearemos nuestras tablas.





CREACIÓN DE TABLAS I

Actividad N° 2



Siguiendo con el archivo de municipios, vamos a crear más tablas. 😊 Recuerden descargar los datos de [municipios de Argentina](#).

Vamos a crear una tabla PAIS que contenga dos campos, un **ID de país** y el **nombre del país**. Ninguno de ellos puede ser nulo.



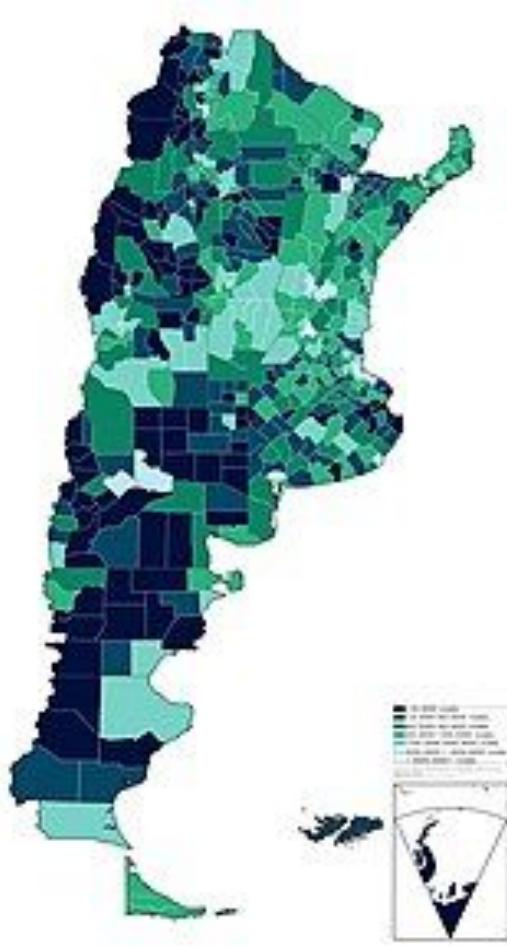
Notas actividad 2:

👉 La creación de tablas es el paso inicial para un buen análisis. Datos mal almacenados darán malos resultados.



CREACIÓN DE TABLAS II

Actividad N° 3



Vamos también a crear una tabla llamada PROVINCIA 😊 que contenga cuatro campos: un **ID de provincia**, el **nombre de la provincia**, el **ID del país** al que pertenece y si **limita con otro país**. Ninguno de ellos puede ser nulo.



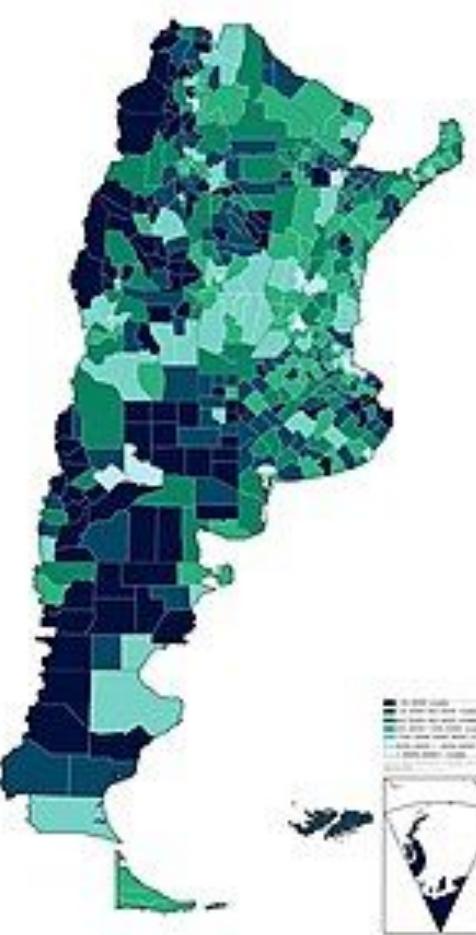
Notas actividad 3:

Los campos booleanos son muy útiles y muy usados. Solamente pueden tomar dos valores. A pesar de contener letras o números su definición es diferente. 👍



IMPORTAR ARCHIVOS

Actividad N° 4



Muchas veces necesitamos importar datos que ya están creados en otro formato ¿Cierto? 😊 Vamos a subir nuestro archivo de Municipios a SQL de forma tal que se cree una tabla. Investiga cómo cargar el archivo de municipios en SQL.



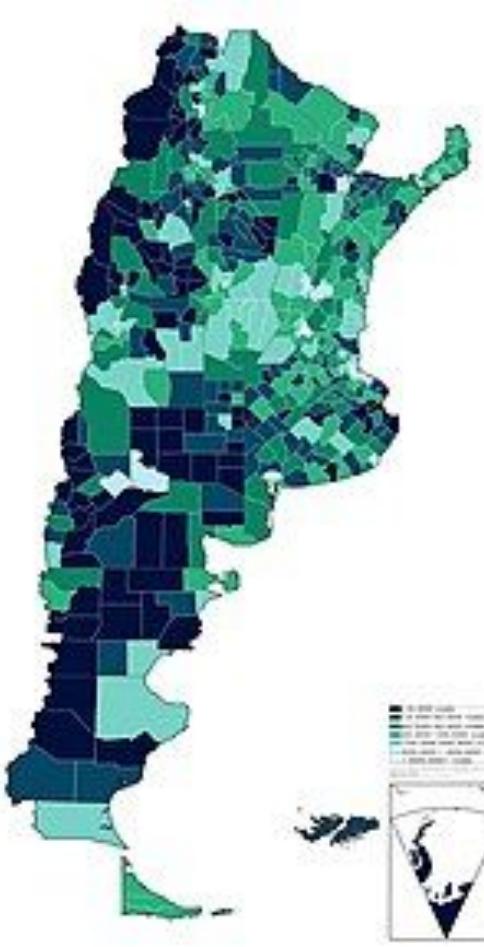
Notas actividad 4:

Prácticamente todas las herramientas de bases de datos permiten importar y exportar datos de forma tal que puedan comunicarse entre sí. ☺



MODIFICACIÓN DE TABLAS

Actividad N° 5



Ahora nos encargaremos de modificar las **tres tablas** que hemos creado. Vamos a cargar las claves primarias a cada una de ellas pero ¡Ojo! Hay que recordar que puede haber claves compuestas o concatenadas.



Notas actividad 5:

Todos los comandos en SQL permiten hacer modificaciones a excepción del DROP o DELETE, con los cuales debemos ser muy cuidadosos.

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Creación de bases de datos

```
CREATE DATABASE MUNICIPIOS;
```

SOLUCIÓN PROPUESTA

Actividad N° 2 - Creación de tablas I

```
CREATE TABLE PAIS (
ID INT NOT NULL,
PAIS varchar(255) NOT NULL
);
```

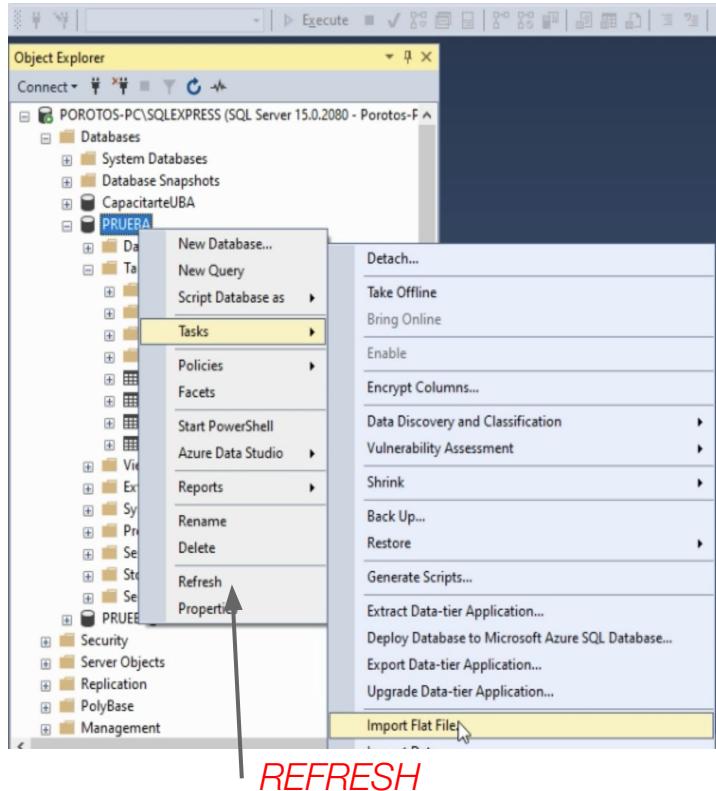
SOLUCIÓN PROPUESTA

Actividad N° 3 - Creación de tablas II

```
CREATE TABLE PROVINCIA (
ID INT NOT NULL,
PROVINCIA varchar(255) NOT NULL,
ID_PAIS INT NOT NULL,
LIMITROFE BIT
);
```

SOLUCIÓN PROPUESTA

Actividad N° 4 - Importar archivos



Debemos ir a nuestra base de datos, hacer clic con el botón derecho y seleccionar *TASKS* (Tareas) y luego *IMPORT FLAT FILE* (importar archivo plano). Se abrirá una ventana donde buscaremos nuestro archivo y le pondremos el nombre a la tabla que se creará. Le daremos *NEXT* hasta finalizar y, ¡listo! Nuestra tabla ya estará creada e importada. Recuerda hacer *REFRESH* de la base para ver los cambios. 🚀

Puedes ver un ejemplo en: 
<https://tinyurl.com/DAimportfile>

SOLUCIÓN PROPUESTA

Actividad N° 5 - Modificación de tablas

```
ALTER TABLE PROVINCIA (  
ADD PRIMARY KEY (ID));
```

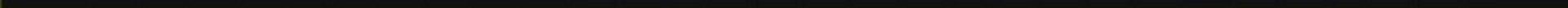
```
ALTER TABLE PAIS (  
ADD PRIMARY KEY (ID));
```

```
ALTER TABLE MUNICIPIOS (  
ADD PRIMARY KEY (id,provincia_id));
```



DATA ANALYTICS

Clase 5 - Sublenguaje SQL DDL II



ACTIVIDADES





INSERCIÓN DE DATOS I

Actividad N° 1

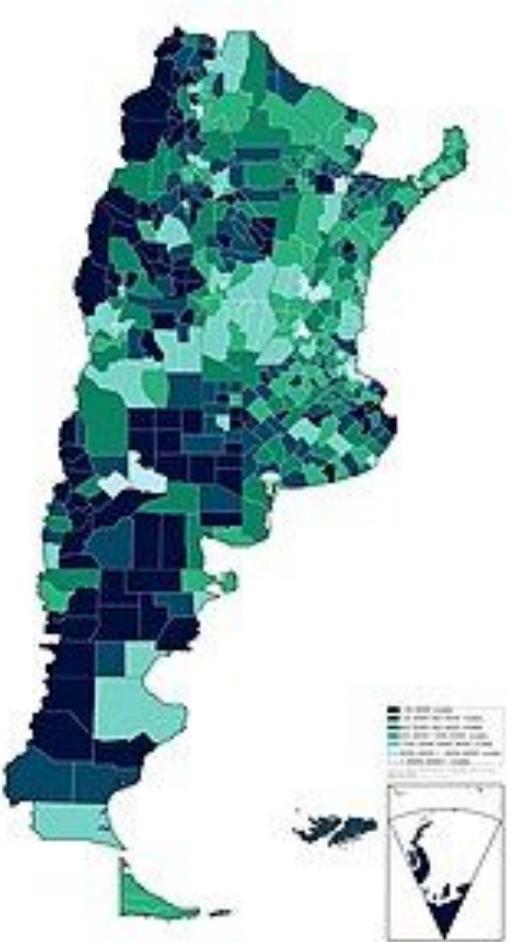
Seguiremos aprovechando nuestro conjunto de datos sobre Municipios. 😎 La tarea ahora es cargar el siguiente conjunto de datos en la tabla de países.

- 1, Argentina
- 2, Uruguay
- 3, Inglaterra
- 4, Londres
- 5, Cardiff
- 6, Reino Unido



Notas actividad 1:

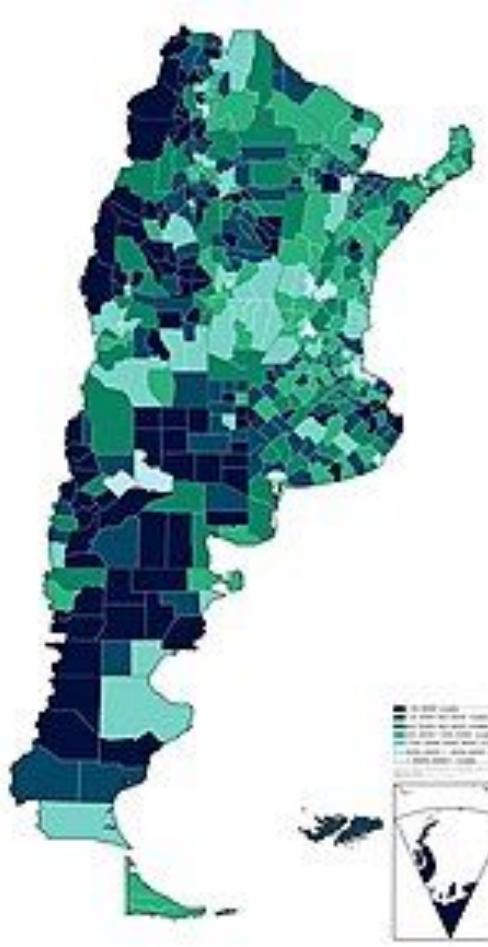
Te parecerá que hay un error en los datos, no te preocupes, cárgalos así. Los datos siempre tienen errores y el analista debe detectarlos a tiempo. 🧐





INSERCIÓN DE DATOS II

Actividad N° 2



Continuando con el ejercicio anterior, ahora vamos a cargar datos en la tabla de provincias. + En esta ocasión deberán insertar 5 provincias de Argentina y 5 Departamentos de Uruguay. A cada uno deberán asignarle un ID, determinar el ID de país que corresponda y cargar si es o no limítrofe.



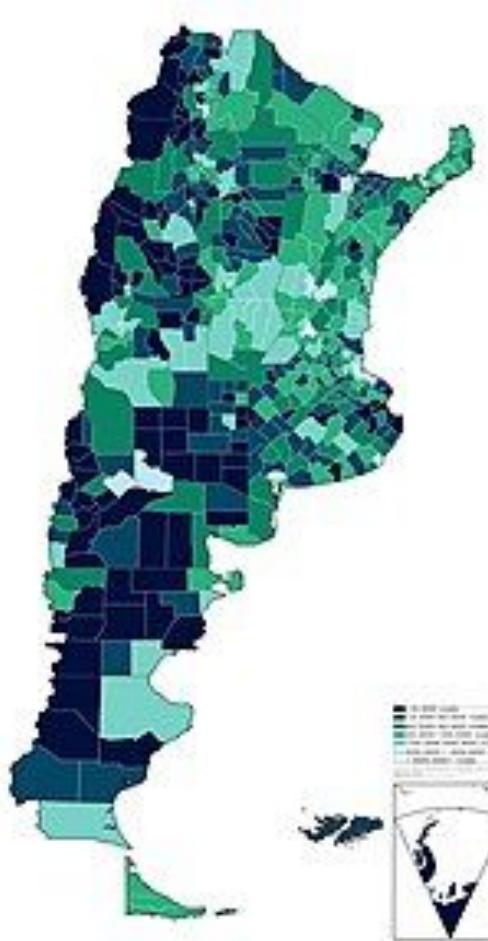
Notas actividad 2:

Recuerda que el campo LIMITROFE es solamente una marca que indica si limita con otro país. Puede ser un 1 o un 0, como encendido y apagado. Carga al menos 10 provincias. 😊



BORRADO DE DATOS I

Actividad N° 3



En la [actividad 1](#) de este capítulo hemos cargado datos erróneos en la tabla PAIS. Por ejemplo, Reino Unido e Inglaterra podrían considerarse duplicados (aunque no lo sean del todo). Tu tarea es eliminar el registro correspondiente a Reino Unido de la tabla PAIS.



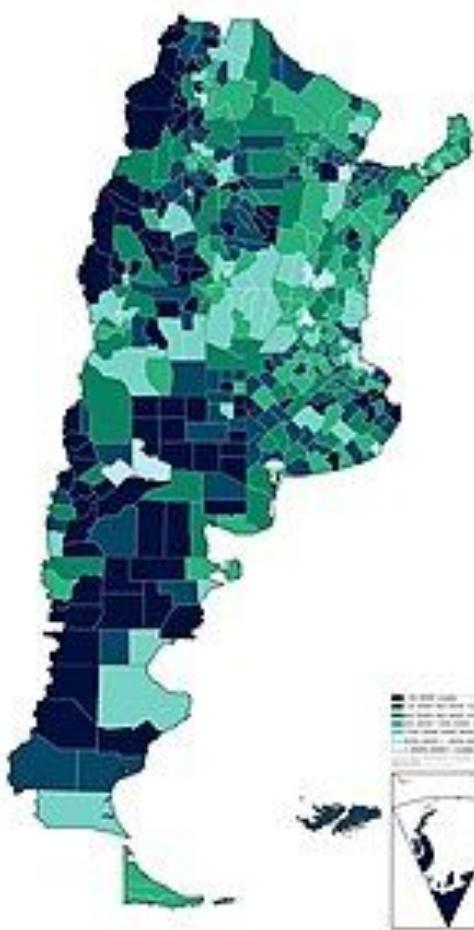
Notas actividad 3:

El comando DELETE es muy útil pero muy peligroso. Debe ser usado con mucha cautela para evitar borrar tablas enteras o datos en uso.



BORRADO DE DATOS II

Actividad N° 4



Hemos agregado algunos valores incorrectos en la tabla PAIS. En esta ocasión vamos a corregirlos. 
Reemplaza el valor CARDIFF por GALES dentro de la tabla PAIS.



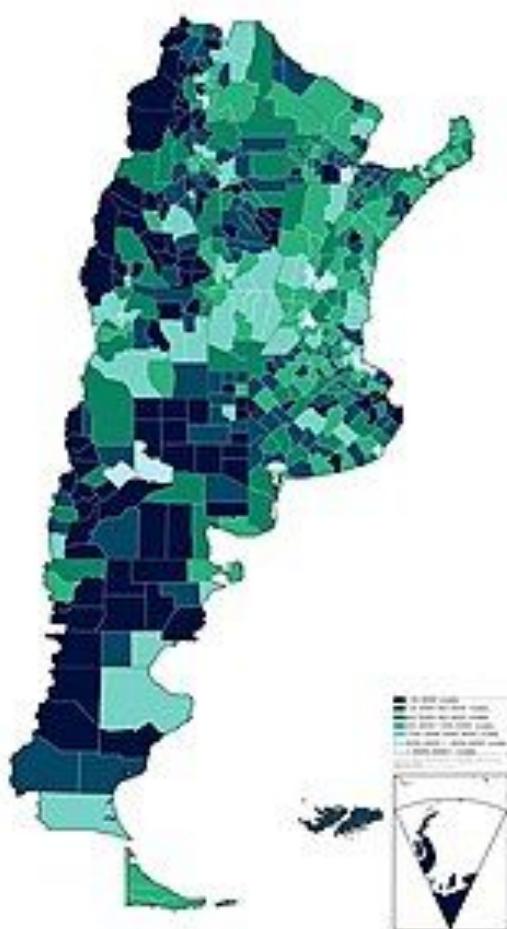
Notas actividad 4:

Es recomendable comentar todo el código que se escriba. Esto mejora la legibilidad y permite que sea entendido por todo el equipo de trabajo. 



CONSULTA GENERAL

Actividad N° 5



Hagamos un repaso general de las tablas creadas: PAIS y PROVINCIA. Consultemos el contenido de ambas tablas y verifiquemos si ha quedado todo correcto. Si hay errores, es el momento de corregirlos. 



Notas actividad 5:

Los comentarios se pueden insertar con --doble guión medio o con /*barra y asterisco*/ 

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Inserción de datos I

```
INSERT INTO PAIS  
VALUES  
(1, 'ARGENTINA'),  
(2, 'URUGUAY'),  
(3, 'INGLATERRA'),  
(4, 'REINO UNIDO'),  
(5, 'CARDIFF')
```

SOLUCIÓN PROPUESTA

Actividad N° 2 - Inserción de datos II

--1 significa límitrofe, 0 significa no límitrofe

INSERT INTO PROVINCIA

VALUES

```
(1, 'SANTA FE', 1, '0'),  
(2, 'MISIONES',1,'1'),  
(3, 'CATAMARCA',1,'1'),  
(4, 'LA RIOJA',1,'1'),  
(5, 'MALDONADO',2,'0'),  
(6, 'RIVERA',2,'1'),  
(7, 'CHUBUT',1,'1'),  
(8, 'SANTA CRUZ',1,'1'),  
(9, 'NEUQUEN',1,'1'),  
(10, 'FLORES',2,'0')
```

SOLUCIÓN PROPUESTA

Actividad N° 3 - Borrado de datos !

--según el editor de SQL que usen puede que sea necesario ingresar las
--cadenas de caracteres entre comillas dobles ("Reino Unido") o simples --('REINO
UNIDO')

```
DELETE  
FROM PAIS  
WHERE PAIS LIKE 'REINO UNIDO'
```

SOLUCIÓN PROPUESTA

Actividad N° 4 - Borrado de datos II

--para insertar comentarios puedes usar el doble guión, como al principio
--de estas líneas

```
UPDATE PAIS
SET PAIS = 'GALES'
WHERE PAIS LIKE 'CARDIFF'
```

SOLUCIÓN PROPUESTA

Actividad N° 5 - Consulta general

/* Comando select */

SELECT *

FROM PAIS

SELECT *

FROM PROVINCIA

CODER HOUSE

DATA ANALYTICS

Clase 6 - Sublenguaje SQL

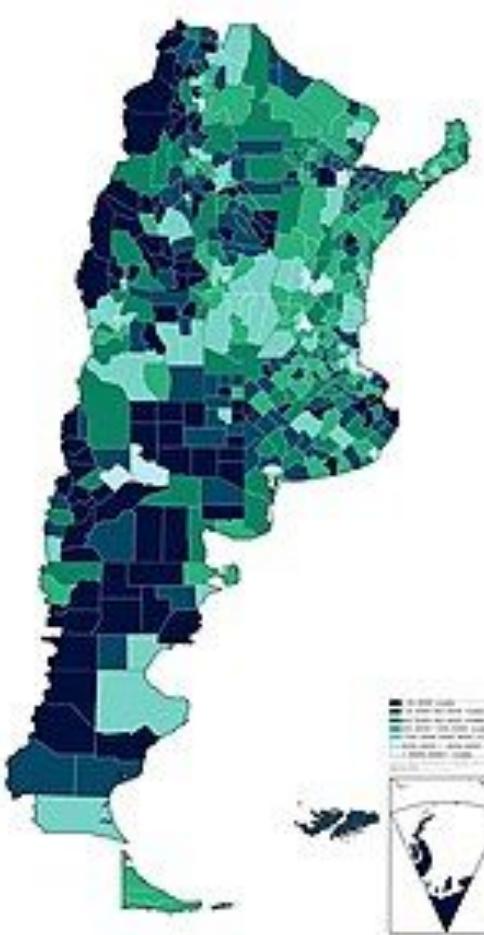
ACTIVIDADES





CONSULTA DE DATOS I

Actividad N° 1



Vamos a refinar y empezar a trabajar a fondo con las consultas. De la tabla de Municipios vamos a visualizar todos los valores distintos de la columna CATEGORÍA. ☀️☁️



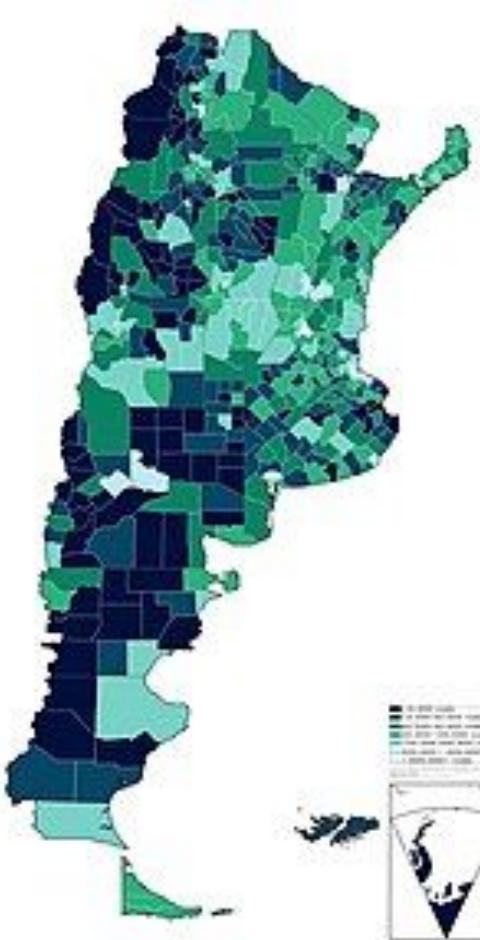
Notas actividad 1:

La cláusula DISTINCT permite obtener los valores únicos existentes en una columna. ↗



CONSULTA DE DATOS II

Actividad N° 2



En esta ocasión tomaremos de la tabla de Municipios todas las columnas con categoría MUNICIPIO de la provincia con ID 10. 



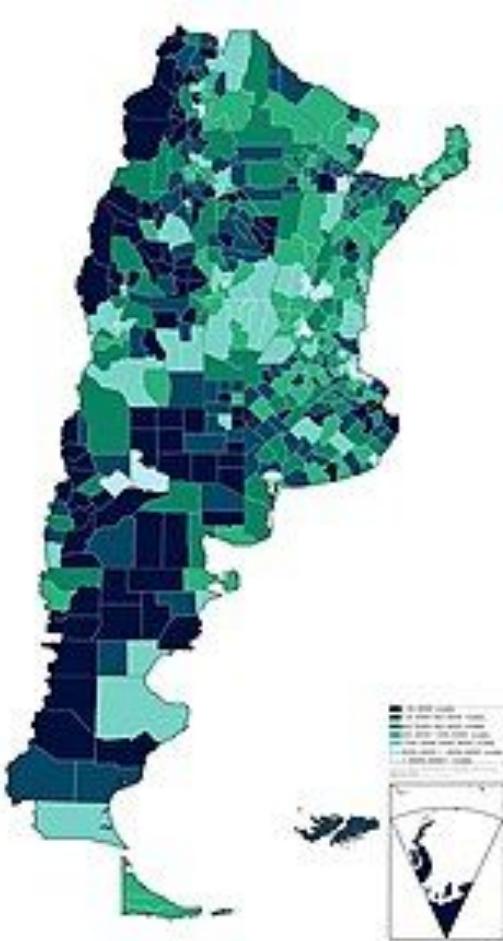
Notas actividad 2:

La cláusula WHERE es la que nos permite aplicar filtros a nuestras consultas. 



CONSULTA DE DATOS III

Actividad N° 3



Continuando con la consulta de Municipios ahora deberás generar una consulta que traiga las provincias que empiecen con la letra S o que el ID de provincia sea mayor a 50. 😊



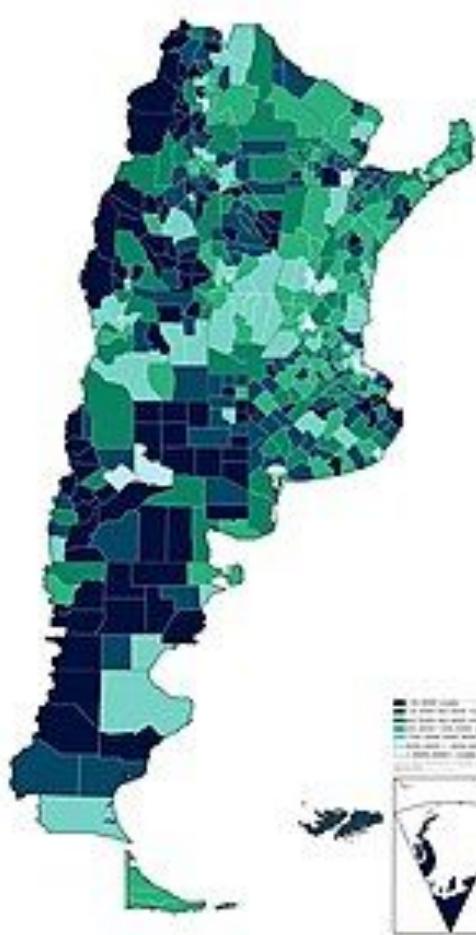
Notas actividad 3:

Dentro de una cláusula WHERE puedo agregar infinidad de filtros. Recuerda usar una línea nueva para cada condición y así mejorar la legibilidad. 😊



FUNCIONES Y AGRUPACIONES

Actividad N° 4



Sigamos con la consulta de Municipios. En esta ocasión vamos a contar cuantos municipios tiene cada provincia.

Ordena los resultados por cantidad de forma descendente. La columna que cuenta los municipios debe llamarse “CANTIDAD”.



Notas actividad 4:

Cada vez que apliquemos una función (contar, sumar, máximo, mínimo) debemos aplicar la cláusula GROUP BY.



CONSULTA DE DATOS AVANZADA

Actividad N° 5

Ahora, trabajando bajo el mismo archivo (municipios), vamos a buscar aquellas provincias con menos cantidad de categorías, para ello deberás buscar cuántas categorías distintas hay por provincia. Estas deben estar ordenadas en forma descendente, por cantidad y solo mostrar aquellas cantidades menores a 10.



Notas actividad 5:



Los requerimientos son siempre realizados por personas, y a veces no son claros. Hay muchas formas de responder una pregunta. Ante la duda, es importante brindar una respuesta pero también dejar en claro las ambigüedades.

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Consulta de datos !

```
/* Comando select distinct*/
```

```
SELECT DISTINCT CATEGORIA  
FROM MUNIC
```

SOLUCIÓN PROPUESTA

Actividad N° 2 - Consulta de datos II

```
SELECT *
FROM MUNICIPIOS
WHERE CATEGORIA = 'MUNICIPIO' AND PROVINCIA_ID = 10
```

SOLUCIÓN PROPUESTA

Actividad N° 3 - Consulta de datos III

```
SELECT *
FROM MUNICIPIOS
WHERE PROVINCIA_NOMBRE LIKE 'S%'
OR PROVINCIA_ID > 50
```

SOLUCIÓN PROPUESTA

Actividad N° 4 - Funciones y Agrupaciones

--en group by y en order by puedo usar el nombre de la columna o su posición
--en esta ocasión se ordena por la segunda columna, que es CANTIDAD

```
SELECT PROVINCIA_NOMBRE, COUNT(*) AS CANTIDAD
FROM MUNICIPIOS
WHERE CATEGORIA = 'MUNICIPIO'
GROUP BY PROVINCIA_NOMBRE
ORDER BY 2 DESC
```

SOLUCIÓN PROPUESTA

Actividad N° 5 - Consulta de datos avanzada

- en group by y en order by puedo usar el nombre de la columna o su posición
- en esta ocasión se ordena por la tercera columna, que es CANTIDAD
- y se agrupa por la primera y segunda columna (NOMBRE y CATEGORIA)

```
SELECT PROVINCIA_NOMBRE, CATEGORIA, COUNT(*) AS CANTIDAD  
FROM MUNICIPIOS  
GROUP BY 1, 2  
HAVING COUNT(*) < 10  
ORDER BY 3 DESC
```



DATA ANALYTICS

Clase 7 - Funciones escalares y Subconsultas

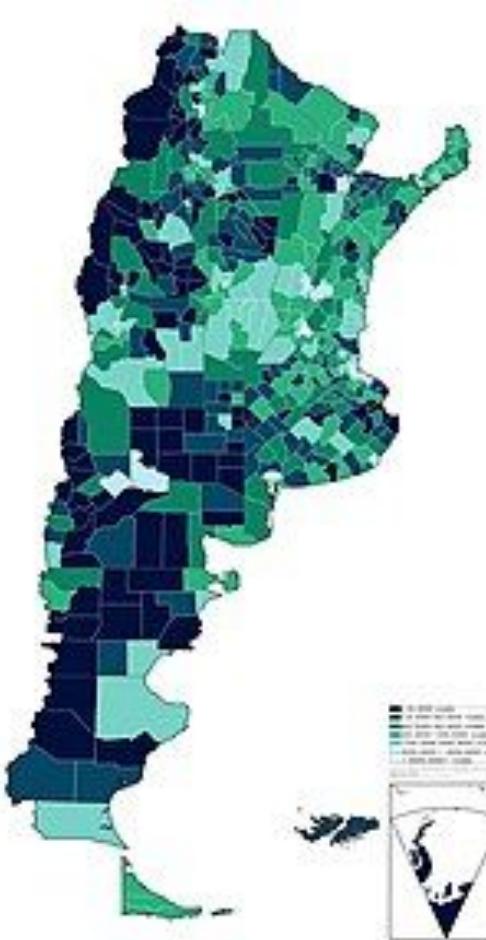
ACTIVIDADES





CONVERSIONES DE DATOS

Actividad N° 1



En ocasiones los datos no están en el formato deseado y se deben hacer conversiones. ↪

Genera una consulta donde las columnas CENTROIDE_LAT y CENTROIDE_LON se muestren como tipo de dato FLOAT.



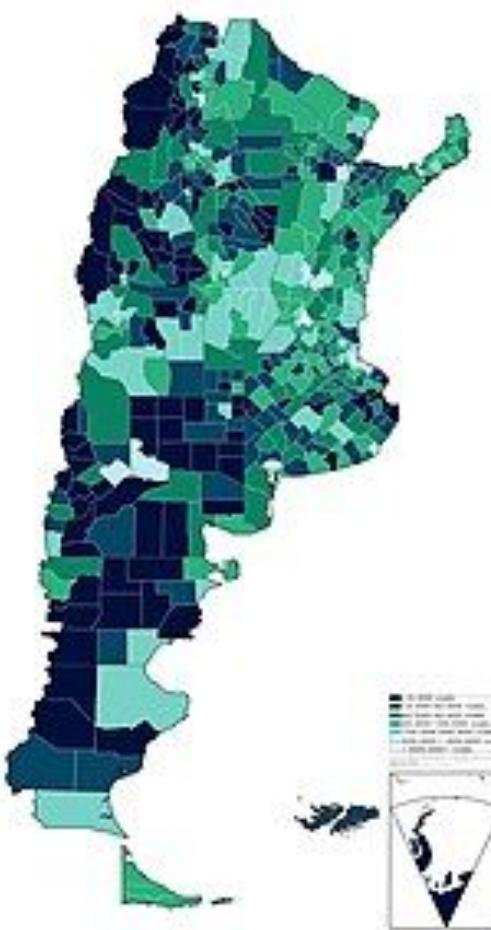
Notas actividad 1:

- 🚫 Muchas veces los datos son importados en formatos incorrectos con lo cual debemos realizar una conversión al operar con ellos.



FUNCIONES

Actividad N° 2



Vamos a utilizar algunas funciones; para ello necesitamos obtener el valor absoluto $|z|$ de todas las latitudes.

En el siguiente link encontrarás una definición de “valor absoluto”: https://es.wikipedia.org/wiki/Valor_absoluto



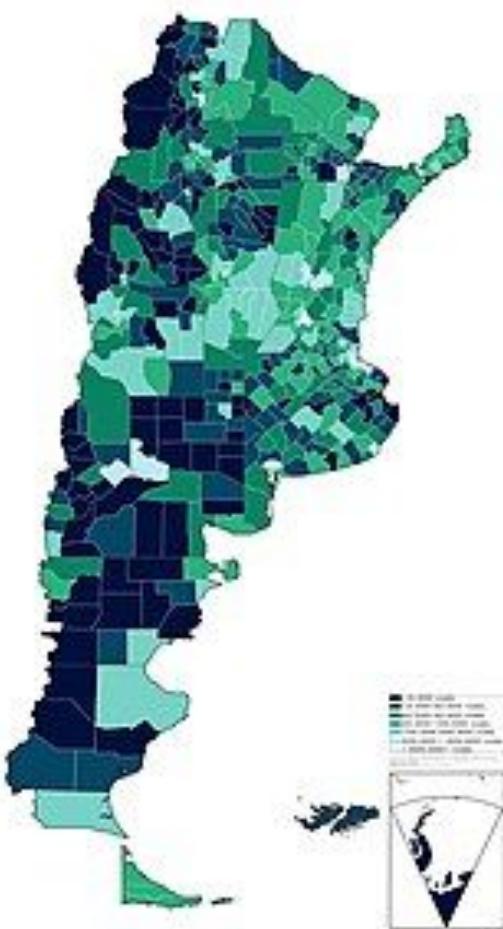
Notas actividad 2:

Las funciones son muy útiles para ahorrar tiempo de programación. Es importante consultar la documentación para saber sus usos. [100](#)



SUBQUERIES

Actividad N° 3



En esta ocasión estamos buscando las categorías, la provincia y los municipios/departamentos de aquella provincia cuyo ID sea el máximo de todos los existentes. Asegúrate de usar una subconsulta (o subquery).

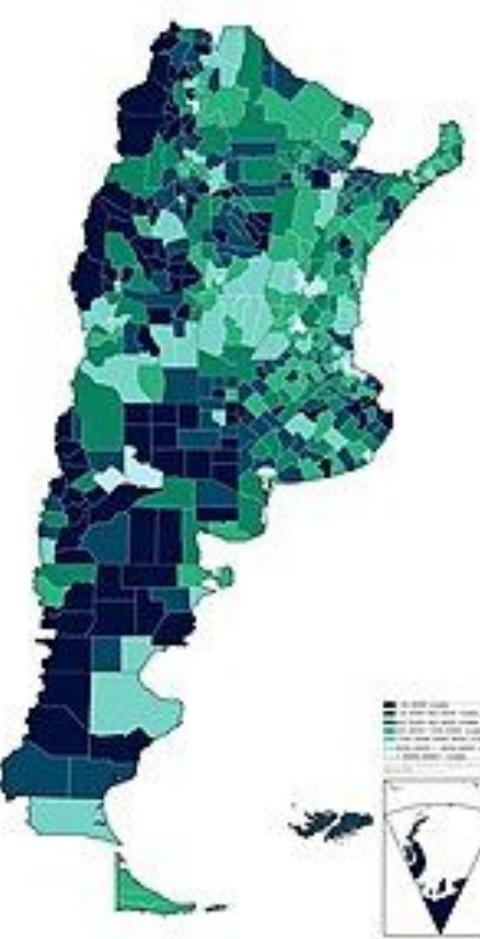
Notas actividad 3:

Las subconsultas aportan muchas soluciones pero también complejidad. Siempre mantén buenas prácticas al escribir el código.



SUBQUERIES Y FUNCIONES I

Actividad N° 4

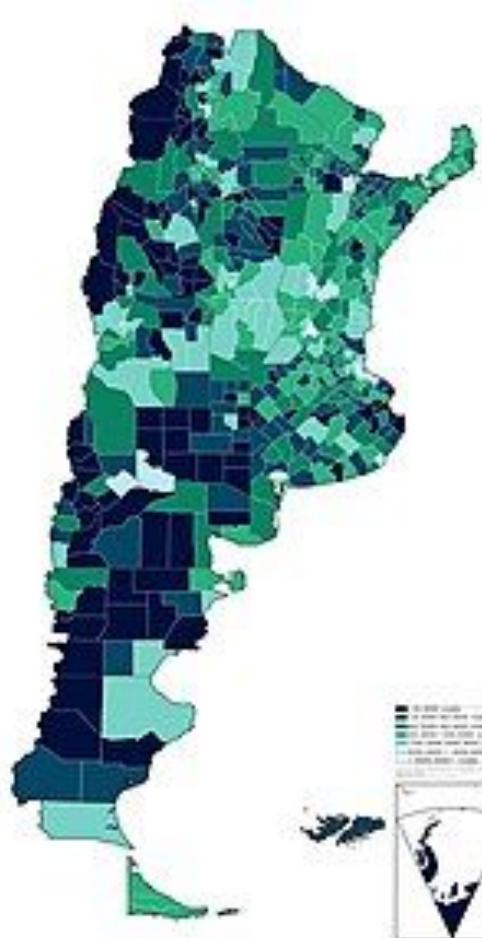


Las consultas se ponen cada vez más complejas, pero nosotros cada vez aprendemos más. Ahora necesitamos obtener el valor entero (sin decimales) y absoluto de la longitud de aquellas provincias que contengan la sigla SAN en su nombre. Que no haya valores repetidos. 😞



Notas actividad 4:

Trabajar con texto es complicado. 😞 Siempre conviene usar las funciones upper o lower para evitar diferencias entre mayúsculas y minúsculas.



SUBQUERIES Y FUNCIONES II

Actividad N° 5

Generemos una consulta donde se muestre para cada provincia su máxima **latitud** y máxima **longitud**. Las columnas deberán llamarse LATITUD y LONGITUD. Además se mostrará el nombre de cada provincia antecediendo la palabra ARGENTINA.



Notas actividad 5:

La clave del éxito es una buena legibilidad del código. 😎 La prolijidad y buenas prácticas son más de la mitad de la solución.

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Conversiones de datos

```
SELECT CAST(CENTROIDE_LAT AS FLOAT), CAST(CENTROIDE_LON AS FLOAT)  
FROM MUNICIPIOS
```

SOLUCIÓN PROPUESTA

Actividad N° 2 - Funciones

- si el campo existe en la tabla como FLOAT no será necesario hacer el cast
- si el campo está como VARCHAR se debe hacer el CAST antes del ABS -- -- para convertir el texto a número

```
SELECT ABS(CAST(CENTROIDE_LAT AS FLOAT))  
FROM MUNICIPIOS
```

SOLUCIÓN PROPUESTA

Actividad N° 3 - Subqueries

```
SELECT CATEGORIA, PROVINCIA_NOMBRE, NOMBRE  
FROM MUNICIPIOS  
WHERE PROVINCIA_ID = (SELECT MAX(PROVINCIA_ID)  
                      FROM MUNICIPIOS)
```

SOLUCIÓN PROPUESTA

Actividad N° 4 - Subqueries y funciones /

/* TRUNC= quita decimales; ABS= obtengo el valor absoluto; CAST= convierte a float

Prueba sacar la cláusula UPPER. Qué sucede? Porqué? Pista: Googlear CASE SENSITIVE */

```
SELECT DISTINCT TRUNC(ABS(CAST(CENTROIDE_LON AS FLOAT))),  
PROVINCIA_NOMBRE  
FROM MUNICIPIOS  
WHERE UPPER(PROVINCIA_NOMBRE) LIKE 'SAN%'
```

SOLUCIÓN PROPUESTA

Actividad N° 5 - Subqueries y funciones II

```
SELECT DISTINCT CONCAT('Argentina - ', PROVINCIA_NOMBRE),  
MAX(CAST(CENTROIDE_LAT AS FLOAT)) AS LATITUD,  
MAX(CAST(CENTROIDE_LON AS FLOAT)) AS LONGITUD  
FROM MUNICIPIOS  
GROUP BY 1
```

CODER HOUSE

DATA ANALYTICS

Clase 8 - Join & Union

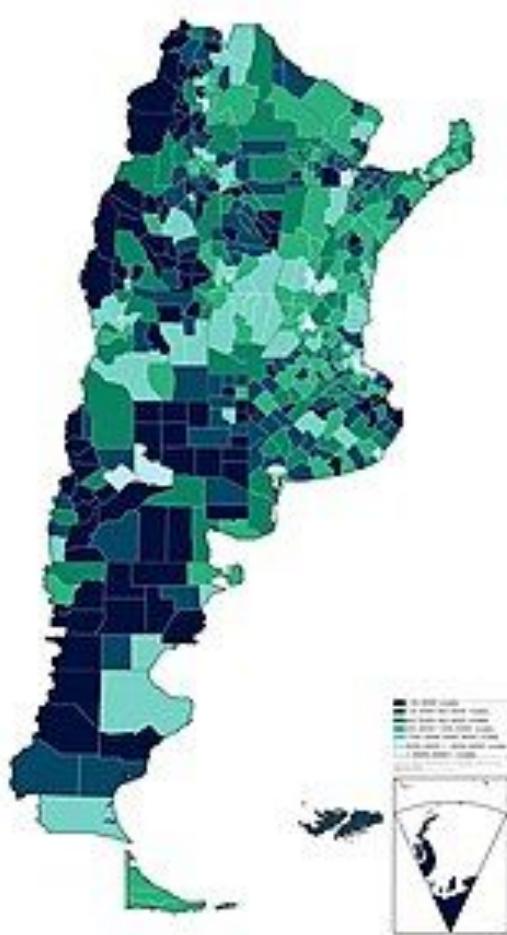
ACTIVIDADES





UNION DE TABLAS I

Actividad N° 1



La tabla de municipios contiene datos sobre provincias. También nosotros, en un ejercicio anterior, creamos una tabla de provincias. 🙌 Unamos los resultados únicos de cada una de las tablas (provincia y ID) y ordena los resultados por nombre de Provincia.



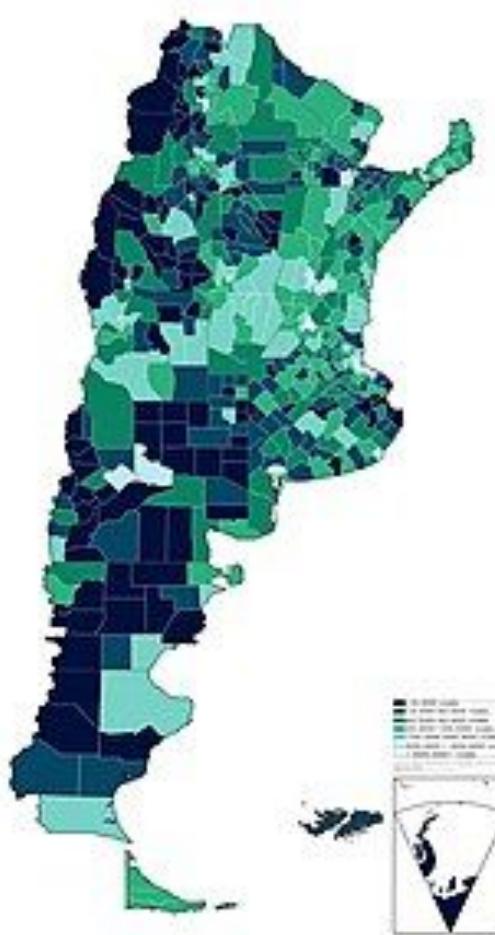
Notas actividad 1:

La cláusula UNION nos permite concatenar el resultado de dos consultas. ☀



UNION DE TABLAS II

Actividad N° 2



El resultado del ejercicio anterior seguramente arrojó provincias con igual nombre y distinto ID. Actualiza la tabla PROVINCIA para que los ID coincidan con los ID de la tabla MUNICIPIOS. Este ejercicio puede hacerse de muchas maneras diferentes.



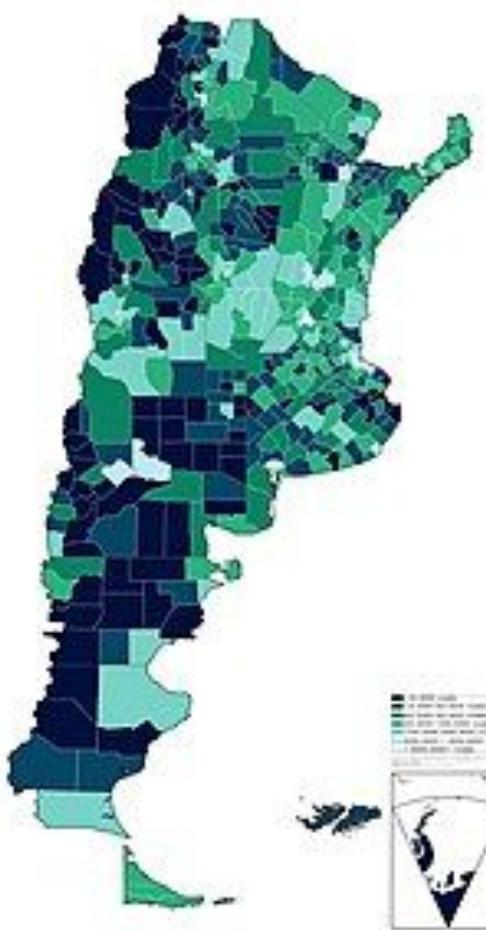
Notas actividad 2:

Mientras el resultado sea correcto, cualquier camino sirve.
 Lo importante es analizar el resultado y verificar que no haya falencias.



JOIN I

Actividad N° 3



Hemos equiparado los códigos de las tablas PROVINCIA y MUNICIPIOS. Ahora junta ambas tablas con un LEFT JOIN. Trae solamente los nombres de las provincias de ambas tablas. La cláusula FROM que sea con MUNICIPIOS, **LEFT JOIN** con PROVINCIA ¿Qué sucede? 😐



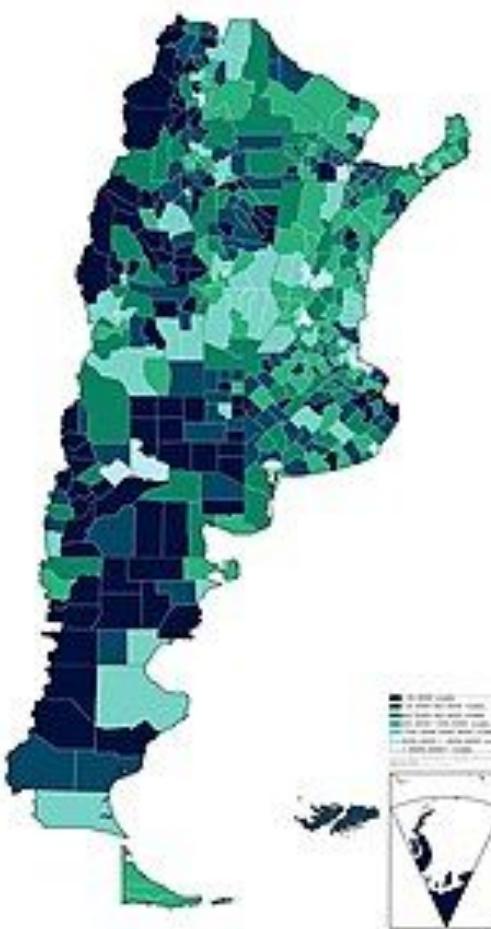
Notas actividad 3:

La cláusula JOIN es muy versátil pero muy peligrosa. 😷 A veces una mala conexión entre tablas puede traer malos resultados.



JOIN II

Actividad N° 4



Repita el ejercicio anterior pero con un RIGHT JOIN. Trae solamente los nombres de las provincias de ambas tablas. La cláusula FROM que sea con MUNICIPIOS, **RIGHT JOIN** con PROVINCIA ¿Qué sucede? 🤔



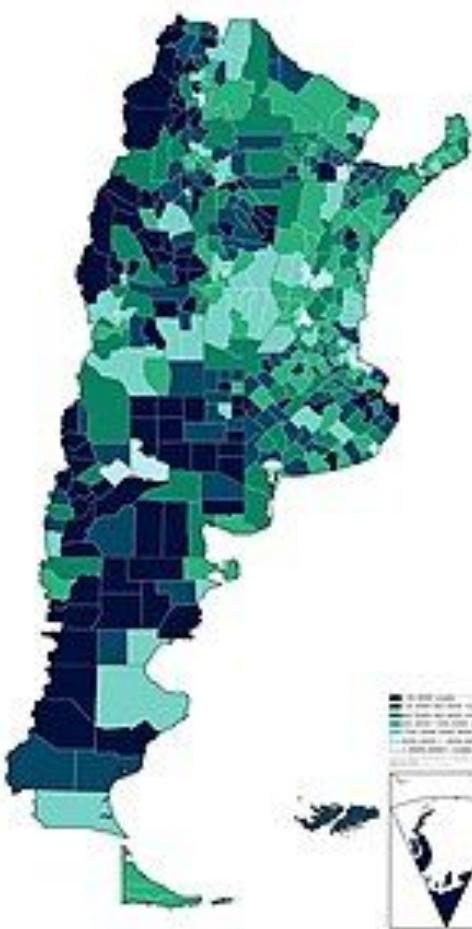
Notas actividad 4:

Left, Right, Inner. Son muchas palabras y marean. Todas funcionan de manera similar, cambia solamente el lugar que se les da a los valores NULOS. 🤓



JOIN III

Actividad N° 5



En un ejercicio previo creamos una tabla PAÍS. Consultala y reflexiona sobre ☺ ¿Cuáles serán los campos que traerán NULL en un LEFT JOIN con PROVINCIA? ¿Qué pasará en un INNER JOIN? ¡Ahora sí! Haz un LEFT, un RIGHT y un INNER JOIN con PAIS y PROVINCIA.



Notas actividad 5:

🏆 La mejor forma de aprender es jugar, practicar, ver qué sucede si aplicamos una cláusula o la otra.

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Union de tablas

--obtengo las provincias y ID desde la tabla de municipios

```
SELECT DISTINCT PROVINCIA_NOMBRE, PROVINCIA_ID  
FROM MUNICIPIOS
```

UNION ALL

--obtengo las provincias y los ID desde la tabla creada en otro ejercicio previo

```
SELECT DISTINCT PROVINCIA, ID  
FROM PROVINCIA
```

SOLUCIÓN PROPUESTA

Actividad N° 2 - Union de tablas II

--los cambios se pueden hacer con un UPDATE o también manualmente
--Aquí usaremos un UPDATE

UPDATE PROVINCIA

SET ID = PROVINCIA_ID

FROM MUNICIPIOS

WHERE MUNICIPIOS.PROVINCIA_NOMBRE = PROVINCIA.PROVINCIA

SOLUCIÓN PROPUESTA

Actividad N° 3 - Join I

--Nota que los resultados de la consulta tendrá datos completos a la --IZQUIERDA (LEFT). Los campos vacíos (DERECHA) serán completados con --NULL

```
SELECT DISTINCT PROVINCIA_NOMBRE, P.PROVINCIA  
FROM MUNICIPIOS AS M  
LEFT JOIN PROVINCIA AS P  
ON P.ID = M.PROVINCIA_ID
```

	provincia_nombre character varying (255)	provincia character varying (255)
1	Buenos Aires	[null]
2	San Juan	[null]
3	La Pampa	[null]
4	Mendoza	[null]
5	Formosa	[null]
6	Santa Fe	Santa Fe
7	Chubut	[null]
8	Córdoba	[null]
9	Río Negro	[null]

SOLUCIÓN PROPUESTA

Actividad N° 4 - Join II

--Noten que los resultados de la consulta tendrá datos completos a la --DERECHA (RIGHT). La campos vacíos (IZQUIERDA) serán completados con --NULL, pero en esta ocasión no hay.

```
SELECT DISTINCT PROVINCIA_NOMBRE, P.PROVINCIA  
FROM MUNICIPIOS AS M  
RIGHT JOIN PROVINCIA AS P  
ON P.ID = M.PROVINCIA_ID
```

	provincia_nombre character varying (255)	provincia character varying (255)
1	Santa Fe	Santa Fe
2	La Rioja	La Rioja
3	Misiones	Misiones
4	Catamarca	Catamarca
5	Tucumán	Tucumán

SOLUCIÓN PROPUESTA

Actividad N° 5 - Join III

```
SELECT *
FROM PAIS AS A
LEFT JOIN PROVINCIA AS B
ON A.ID = B.ID_PAIS
```

	id integer	pais character varying (255)		id integer	provincia character varying (255)		id_pais integer	limítrofe bit
1	1	ARGENTINA		82	Santa Fe		1	0
2	1	ARGENTINA		54	Misiones		1	1
3	1	ARGENTINA		90	Tucumán		1	0
4	1	ARGENTINA		46	La Rioja		1	1
5	1	ARGENTINA		10	Catamarca		1	1
6	2	URUGUAY		[null]	[null]		[null]	[null]
7	3	INGLATERRA		[null]	[null]		[null]	[null]
8	5	GALES		[null]	[null]		[null]	[null]

```
SELECT *
FROM PAIS AS A
INNER JOIN PROVINCIA AS B
ON A.ID = B.ID_PAIS
```

	id integer	pais character varying (255)		id integer	provincia character varying (255)		id_pais integer	limítrofe bit
1	1	ARGENTINA		82	Santa Fe		1	0
2	1	ARGENTINA		54	Misiones		1	1
3	1	ARGENTINA		90	Tucumán		1	0
4	1	ARGENTINA		46	La Rioja		1	1
5	1	ARGENTINA		10	Catamarca		1	1



DATA ANALYTICS

Clase 9 - Workshop (Clase especial)

Sin actividades 😊



DATA ANALYTICS

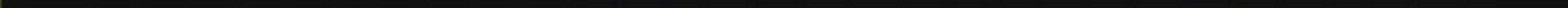
Clase 10 - Charla de un especialista en Base de Datos

Sin actividades 😊



DATA ANALYTICS

Clase 11 - Historias con datos



ACTIVIDADES





TABLEROS I

Actividad N° 1



En épocas de COVID-19 sobran los ejemplos de tableros brindando información. En esta ocasión trabajaremos sobre el [Monitor de Vacunación](#) en Argentina, generado por el Ministerio de Salud de este país. Identifica las partes que lo componen (al menos 3 distintas).



Notas actividad 1:

Identificar las partes nos permitirá construir un mejor todo.

No solo es tirar datos sobre la mesa, se trata de contar una buena historia.





TABLEROS II

Actividad N° 2



Ahora comparemos el [Monitor de Vacunación](#) del Ministerio de Salud con el generado por la [provincia de Buenos Aires](#) y el del [Ministerio de Sanidad del Gobierno de España](#) ¿Cuál te resulta más lindo? ¿Cuál más informativo? ¿Cuál cuenta mejor una historia? 🤔



Notas actividad 2:

La estética es subjetiva, y no debe llevarse los laureles. Lo importante es la información, la estética debe acompañar y ser simple pero efectiva. 📊



TABLEROS III

Actividad N° 3



Notas actividad 3:

Nunca abandonen el papel y el lápiz. Es preciso sentarse a pensar y dibujar antes de volcar ideas sobre la pantalla.

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Tableros I

Fecha Actualización

Última actualización de datos
5:59:22 am

Seguimiento plan de vacunación Covid

Fuente de datos: NOMIVAC



Ministerio de Salud
Argentina

Título

Indicador

Dosis aplicadas	Dosis distribuidas
43,221,929 Aplicaciones	49,300,024 Dosis
Vacunados con una (1) dosis	Vacunados con dos (2) dosis
28,100,621 Personas	15,121,308 Personas

Tabla



Provincia	Distribuidas total	Aplicadas 1*	Aplicadas 2*	Aplicadas total
Buenos Aires	19,010,595	10,832,928	5,538,350	16,371,278
CABA *	3,664,741	2,210,017	1,420,700	3,630,717
Catamarca	450,752	283,780	108,070	391,850
Chaco	1,310,062	658,235	316,659	974,894
Chubut	665,588	353,086	180,676	533,762
Corrientes	1,213,430	659,332	294,970	954,502
Córdoba	3,985,259	2,359,401	1,401,742	3,761,143
Entre Ríos	1,501,433	826,421	381,878	1,208,299
Formosa	656,533	410,733	190,504	601,237
Jujuy	812,787	486,201	214,418	700,619
La Pampa	393,827	236,052	168,780	404,832
La Rioja	416,372	259,828	142,950	402,778
Mendoza	2,105,642	1,176,844	663,682	1,840,526
Misiones	1,367,400	607,913	296,753	904,666
Neuquén	711,529	430,276	232,547	662,823
Río Negro	789,748	482,649	274,564	757,213
Salta	1,499,016	769,825	360,755	1,130,580
San Juan	853,125	481,015	294,166	775,181
San Luis	555,454	326,078	114,913	544,8
Santa Cruz	407,556	197,564	312,477	

¿Cómo podemos ayudarte?



CODER HOUSE

SOLUCIÓN PROPUESTA

Actividad N° 2 - Tableros II

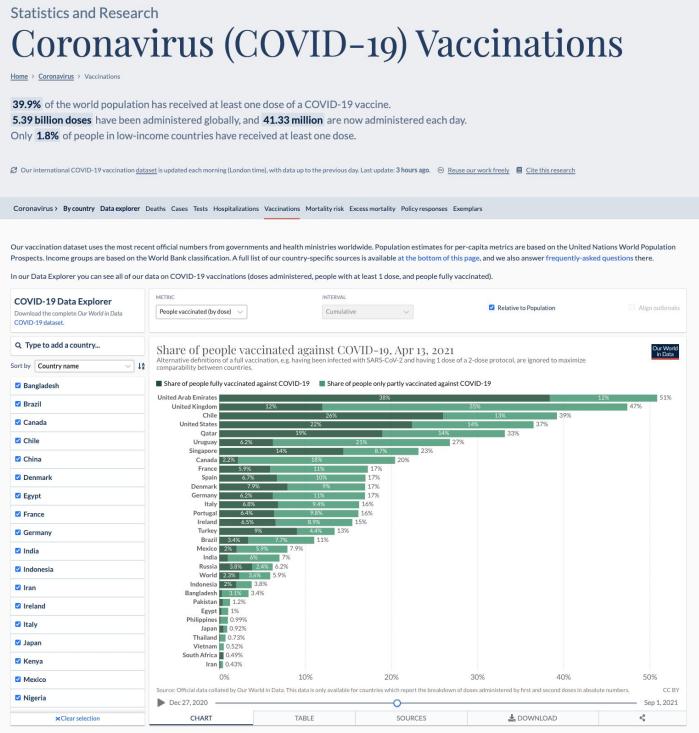
Está claro que no hay una única respuesta, pero podemos ponernos de acuerdo en algunos puntos. El de la [provincia de Buenos Aires](#) parece bastante completo, pero carece de la cantidad acumulada de vacunados; dato importante a tener en cuenta a la hora de enfrentar nuevas cepas. 🤔 Si lo están viendo desde una computadora traten de ingresar desde el celular ¿Cambió su opinión? Desde un dispositivo móvil su navegación parece un poco engorrosa , ¿cierto?

El del [Ministerio de Sanidad del Gobierno de España](#) es bien completo pero un tanto confuso, cuesta entender qué nos está tratando de contar. Muchos links y frases sueltas parecen marear.

El [Monitor de Vacunación](#) es simple, quizás por demás. Lo vemos en el siguiente ejercicio. ➡

SOLUCIÓN PROPUESTA

Actividad N° 3 - Tableros III

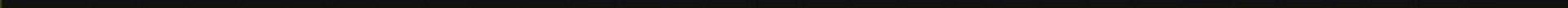


No te enojes, pero acá tampoco hay una única solución. 🤷‍♂️ No creemos que nadie tenga el tablero perfecto, pero hay ejemplos muy buenos. Por ejemplo el sitio *Our World in Data* se especializa en el análisis de datos y publicación de información. El [tablero de vacunaciones](#) (mira solamente el primer gráfico) es muy completo, posee infinidad de opciones, filtros, posibilidad de ver valores acumulados y una estética simple y efectiva. Pruébalo, juega un rato. 🎈



DATA ANALYTICS

Clase 12 - Gráficos eficientes



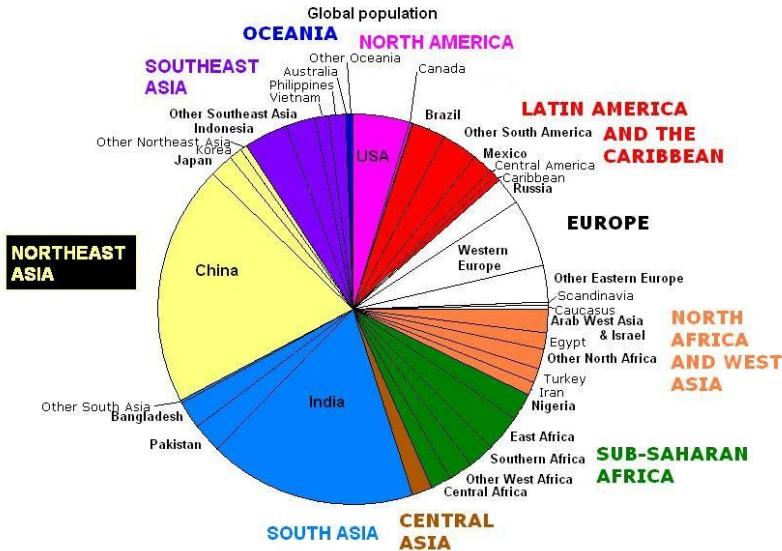
ACTIVIDADES





ERRORES EN GRÁFICOS I

Actividad N° 1



En este capítulo exploraremos distintos gráficos, en su mayoría con errores y, trataremos de encontrar una manera de corregirlos. Empezaremos con el gráfico de torta de la izquierda ¿Qué te parece? ¿Le corregirías algo? Si es así ¿Qué corregirías? Y ¿Por qué?



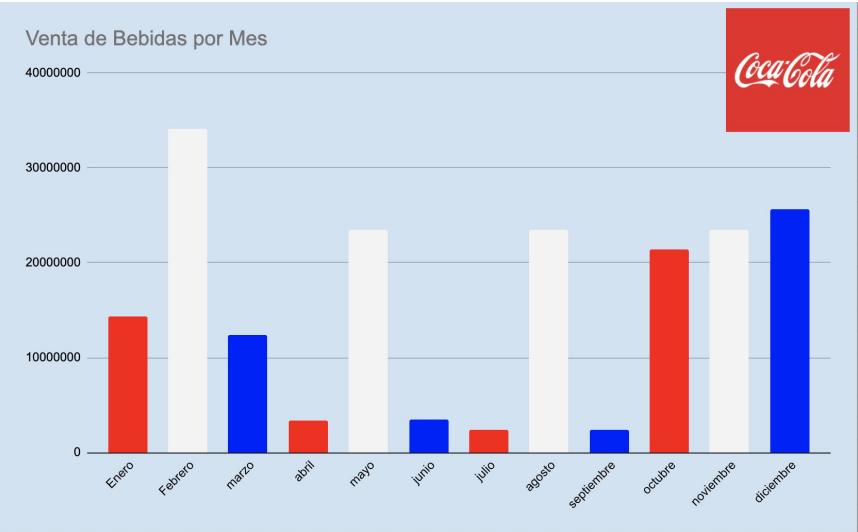
Notas actividad 1:

Recuerda siempre, si un gráfico hay que explicarlo o mirarlo mucho para entenderlo es porque hay algo malo en él.



ERRORES EN GRÁFICOS II

Actividad N° 2



Este viene con trampa. 🤪

Se trata de un gráfico diseñado para la compañía Coca Cola que contiene información sobre ventas de bebidas mensual ¿Qué te parece? Registra todo lo que consideres importante de modificar, sumar o restar. Puede haber más de un error.



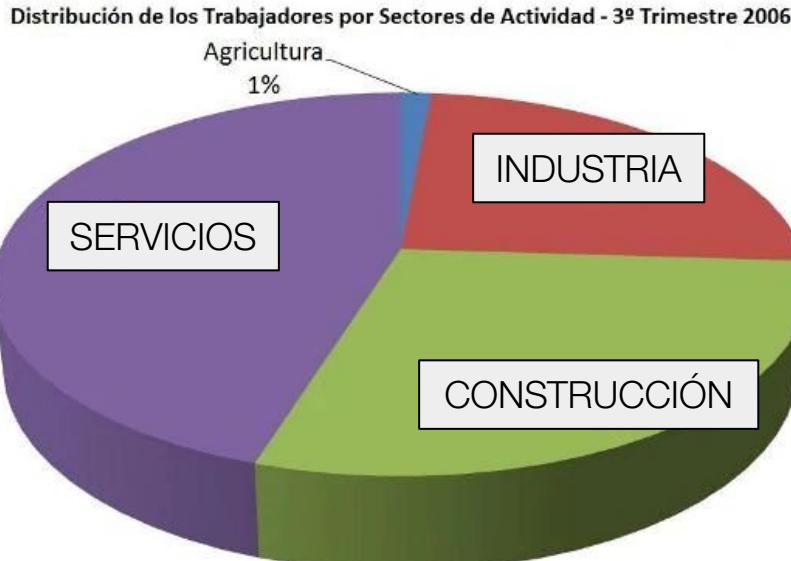
Notas actividad 2:

Una presentación de calidad cuida hasta el más mínimo detalle. Lo creen que nadie verá es justo donde todos pondrán sus miradas. 😊



ERRORES EN GRÁFICOS III

Actividad N° 3



Otro problema es la perspectiva. 😎
Fíjate en este gráfico y de nuevo, ¿qué te parece? ¿Reorganizarías algo? ¿El diseño 3D encaja? ¿Cómo ordenarías las categorías según su tamaño?



Notas actividad 3:

Perspectiva, colores, fuente, unidades, todos los detalles deben cuidarse. Estar atento a todo es esencial. 😯

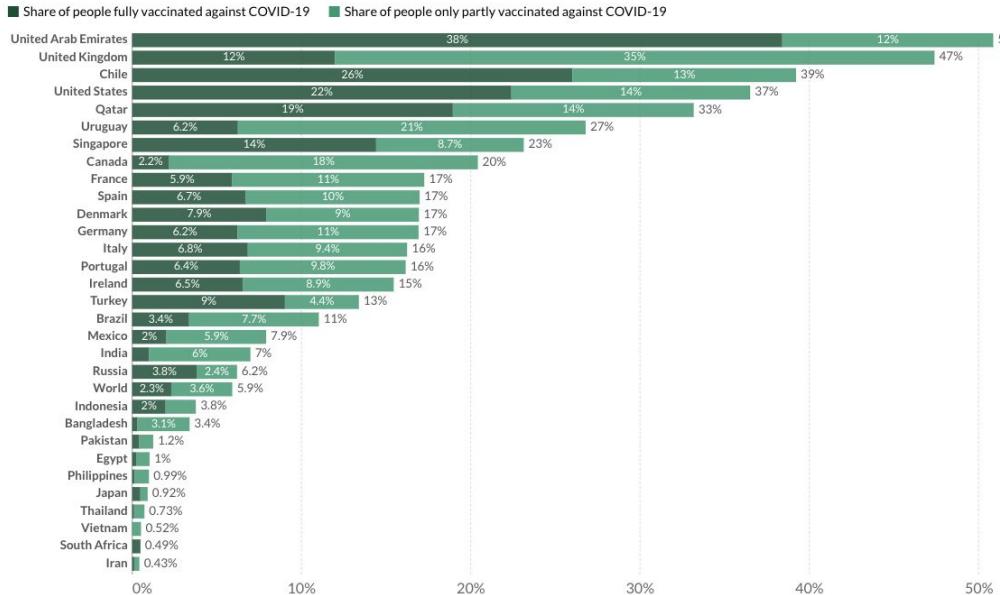
RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - *Errores en gráficos I*

Share of people vaccinated against COVID-19, Apr 13, 2021
Alternative definitions of a full vaccination, e.g. having been infected with SARS-CoV-2 and having 1 dose of a 2-dose protocol, are ignored to maximize comparability between countries.



Source: Official data collated by Our World in Data. This data is only available for countries which report the breakdown of doses administered by first and second doses in absolute numbers.

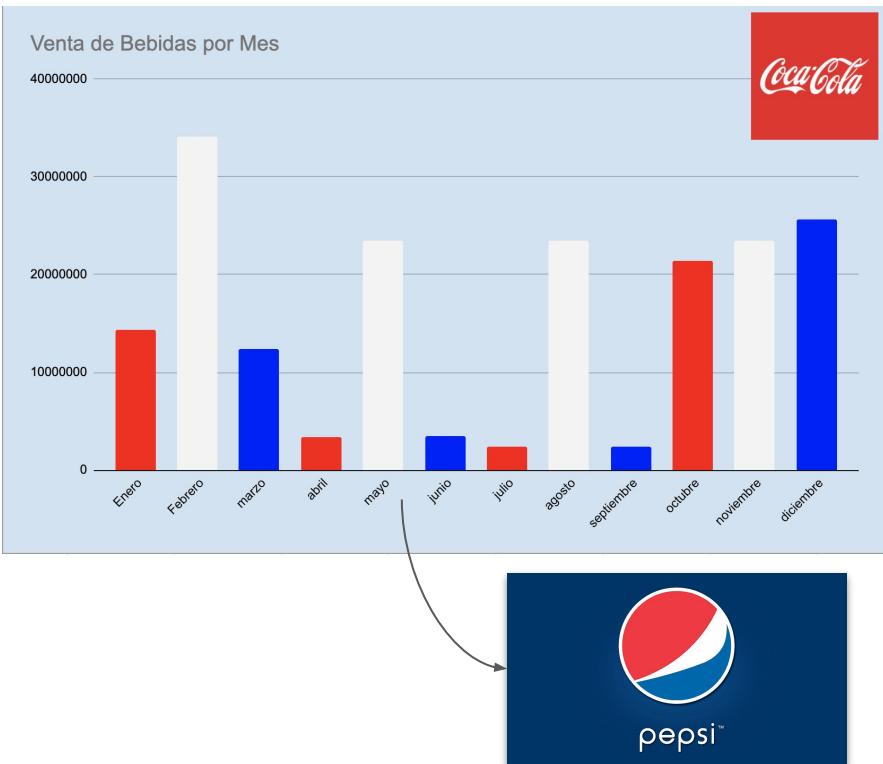
Our World
in Data

CC BY

Re-usando el gráfico de un ejercicio anterior, mostramos una posible solución. 🙌 El gráfico de tortas anterior era claramente difícil de leer. Cuando son muchas las categorías es ideal usar un **gráfico de barras horizontales** o bien agrupar categorías en una genérica del estilo “OTROS”. ✓

SOLUCIÓN PROPUESTA

Actividad N° 2 - Errores en gráficos II



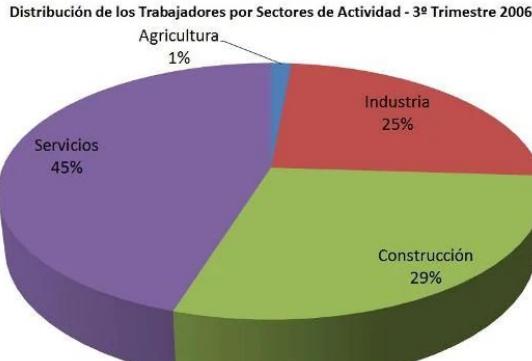
Acá hay al menos dos errores.

El primero, es que no está especificada la unidad de medida de las ventas. No se sabe si son dólares, pesos, litros, botellas o latas lo que se vende.

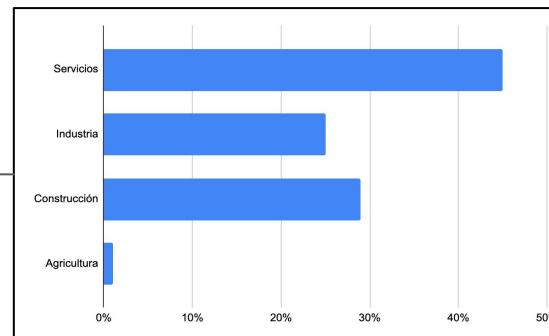
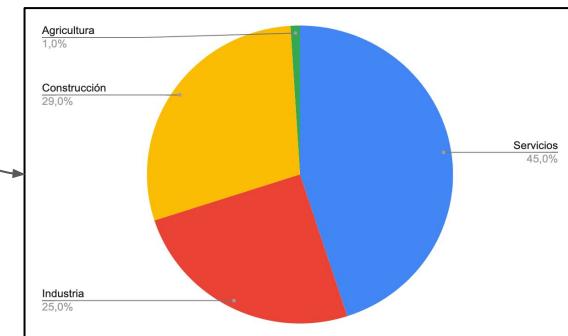
El segundo, más delicado y peligroso, es haber usado unos colores para los gráficos muy similares a los del principal competidor. Cada empresa tiene su manual de marca, búscalos y úsalos.

SOLUCIÓN PROPUESTA

Actividad N° 3 - Errores en gráficos III



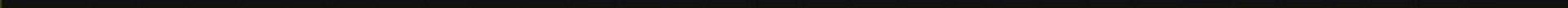
En el gráfico de torta que vimos inicialmente, 🧐 nota que SERVICIOS es casi el doble de grande de INDUSTRIA y un 50% más grande que CONSTRUCCIÓN. A la vez CONSTRUCCIÓN es mayor que INDUSTRIA pero, ¿se nota eso claramente? Fíjate los gráficos alternativos y compara ¿Cuáles tienen mejor lectura y permiten un entendimiento más rápido?





DATA ANALYTICS

Clase 13 - Introducción a Power BI



ACTIVIDADES





POWER BI - CARGA DE DATOS I

Actividad N° 1

 gob.pe | Plataforma Nacional de Datos Abiertos

Datos Abiertos

Marco de Gobernanza de Datos del Estado Peruano está constituido por instrumentos técnicos y normativos que establecen los requisitos mínimos que las entidades de la Administración Pública deben implementar conforme a su contexto legal, tecnológico y estratégico para asegurar un nivel básico y aceptable para la recopilación, procesamiento, publicación, almacenamiento y apertura de los datos que administre.

DATATON 2021

Home / Datasets / Población Peru

Ver Revisores

Licencia
Open Data Commons Attribution License

Población Peru

Dataton 2021 Salud

Contiene el número de personas según proyección INEI 2021, a nivel departamento, provincia, distrito, sexo y grupos de edad.

Otros accesos

Empezamos a trabajar con Power BI. Vamos a descargar de la Plataforma Nacional de Datos Abiertos de Perú el archivo que detalla la población de dicho país. Luego debemos importarlo en Power BI y, desde allí, analizar su contenido ¿Notas algo extraño? 😊



Notas actividad 1:

Casi todas las fuentes de datos cuentan con un diccionario. Este explica 🧑 qué significa cada dato que compone el dataset.



POWER BI - CARGA DE DATOS II

Actividad N° 2

 gob.pe | Plataforma Nacional de Datos Abiertos

Datos Abiertos

Marco de Gobernanza de Datos del Estado Peruano está constituido por instrumentos técnicos y normativos que establecen los requisitos mínimos que las entidades de la Administración Pública deben implementar conforme a su contexto legal, tecnológico y estratégico para asegurar un nivel básico y aceptable para la recopilación, procesamiento, publicación, almacenamiento y apertura de los datos que administre.

DATATON 2021

Home / Datasets / Población Peru

Ver Revisores

Licencia
Open Data Commons Attribution License


Otros accesos

Población Peru

Dataton 2021 Salud

Contiene el número de personas según proyección INEI 2021, a nivel departamento, provincia, distrito, sexo y grupos de edad.

Seguimos trabajando con el archivo que detalla la población de Perú. Identifica las medidas y las dimensiones ¿Ha sido todo bien generado? O ¿Power BI ha hecho suposiciones incorrectas? 😳



Notas actividad 2:

Las herramientas son inteligentes. Aún así no debemos dar nada por supuesto. Cada carga de datos, cada transformación debe ser controlada. Nada debe ser dejado librado al azar. 💪



POWER BI - CARGA DE DATOS III

Actividad N° 3

The screenshot shows the 'gob.pe' website's data portal. At the top, there's a red header with the logo and the text 'Plataforma Nacional de Datos Abiertos'. Below it, a grey sidebar on the left has 'Datos Abiertos' and a 'DATATON 2021' button. The main content area shows a dataset titled 'Población Peru'. It includes a 'Licencia' section with 'Open Data Commons Attribution License' and 'DATATON 2021 Salud'. A note says it contains population numbers by projection from INEI 2021, at the departmental, provincial, district, sex, and age group levels. There are 'Ver' and 'Revisores' buttons, and a 'Otros accesos' section.

Ahora uniremos el archivo que detalla la población de Perú con el que detalla los datos de cada provincia, el cual puedes descargar acá. Luego de descargarlo, impórtalo a Power BI y relaciona ambos archivos entre sí.



Notas actividad 3:

Cada fuente de datos cuenta con diccionario y detalles sobre cómo se enlazan los datos entre sí. Cómo analistas debemos estar al tanto de esos detalles.



POWER BI - MODIFICACIÓN DE TABLAS

Actividad N° 4

The screenshot shows the gob.pe platform interface. At the top, there's a red header with the gob.pe logo and "Plataforma Nacional de Datos Abiertos". Below it, a sidebar on the left lists "Datos Abiertos". The main content area displays the "Población Peru" dataset. It includes a brief description of the data, a "DATATON 2021" button, navigation links ("Home", "Datasets", "Población Peru"), and a "Ver" button. A "Licencia" section specifies the "Open Data Commons Attribution License". Below the main content, there's a "Otros accesos" section.

En Power BI toma una de las dos tablas cargadas y cámbiale el nombre a TABLA_PERU. Luego genera una nueva columna insertando la fecha del día en cada una de las filas.

Notas actividad 4:

El entorno de trabajo de Power BI es muy similar a las demás herramientas de Microsoft Office. Si has manejado Excel, las funciones de Power BI te resultarán similares. 😊





POWER BI - GALERÍA

Actividad N° 5

The screenshot shows the Microsoft Power BI Community Data Stories Gallery. At the top, there's a navigation bar with links like Overview, Products, Pricing, Solutions, Partners, Resources, and Community. Below the navigation is a yellow header bar with the text "This board". The main content area is titled "Microsoft Power BI Community > Galleries > Data Stories Gallery". It features a "Featured" section with three cards: one showing a dashboard with charts and a "29%" metric, another showing a soccer ball with a hand icon, and a third showing a chart titled "2018 AMERICAN ASSOCIATION Most Cases Involve Unintentional Deaths". Below the featured section are buttons for "Top Kudos", "Recently Posted", and "All Data Stories".

Power BI posee una galería propia donde cada usuario sube distintos modelos de tableros. Busca la galería en Google y navégala. Encuentra algún diseño que te resulte interesante e identifica algunos de sus componentes.



Notas actividad 5:

Las galerías sirven para obtener inspiración y buenas prácticas para nuestros diseños. Es un gran lugar donde encontrar ejemplos valiosos.

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Carga de datos I

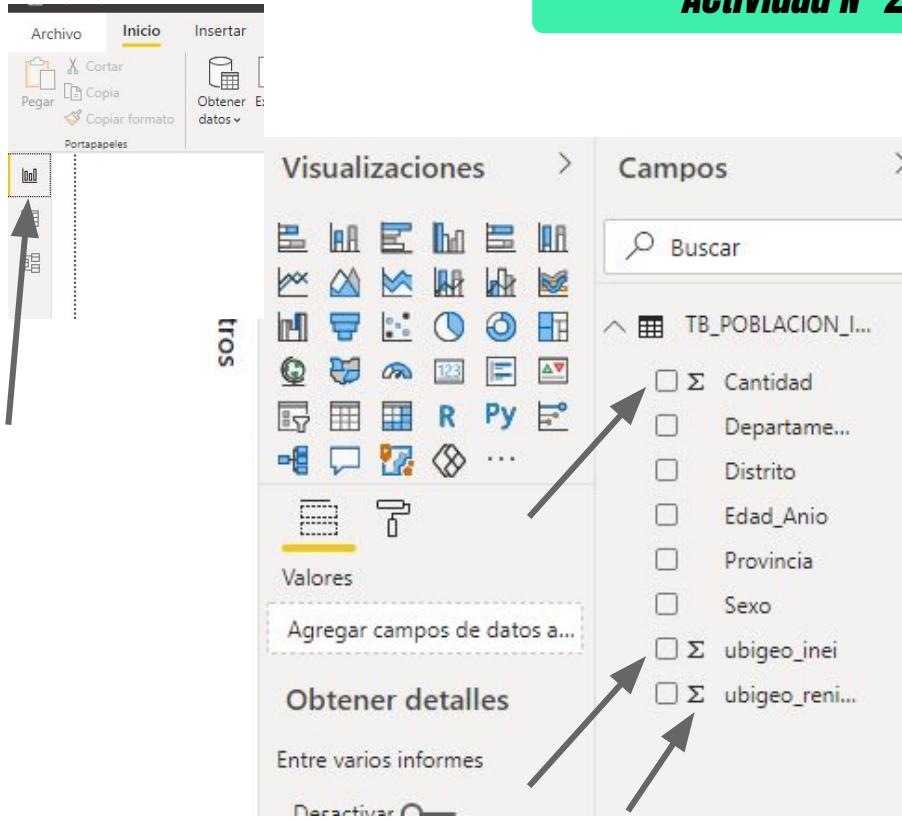
ubigeo_reniec	ubigeo_inei	Departamento	Provincia	Distrito	Edad_Año	Sexo	Cantidad
140129	150139	LIMA	LIMA	SANTA ROSA	10	M	169
140129	150139	LIMA	LIMA	SANTA ROSA	0	M	155
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	9	M	2147
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	8	M	2256
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	80 +	M	6347
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	7	M	2140
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	75-79	M	4885
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	70-74	M	6419
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	6	M	2150
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	65-69	M	8464
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	60-64	M	10315
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	5	M	2194
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	55-59	M	12545
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	50-54	M	13702
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	4	M	2219
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	45-49	M	14138
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	40-44	M	15368
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	3	M	2124
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	35-39	M	15719
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	30-34	M	14467
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	2	M	2012
140130	150140	LIMA	LIMA	SANTIAGO DE SURCO	25-29	M	15073

Presionando sobre el ícono resaltado de la izquierda, podrás hacer un análisis visual de los datos recién importados. Es importante revisar que los datos hayan sido correctamente cargados. 

Ahora miremos el campo EDAD-ANIO. Hay mezcla de datos numéricos con datos de rango y hasta con símbolos. La edad, que podríamos pensar es un dato numérico, en este conjunto de datos no lo es. 😬 Incluso en el diccionario de datos está definido como un campo VARCHAR de 5, es decir, caracteres variables de una longitud máxima de 5.

SOLUCIÓN PROPUESTA

Actividad N° 2 - Carga de datos II



Presionando sobre el ícono resaltado de la izquierda podrás acceder al panel de “Campos”.

Una vez allí, notarás que con el signo de sumatoria se marcan 3 medidas (Cantidad, Ubigeo_inei y Ubigeo_renec). El resto son todas dimensiones. Ahora bien, “Cantidad” es una medida, pero las dos “Ubigeo” corresponden a campos clave, son dimensiones pero Power BI detecta números en su contenido e interpreta que son medidas. Debemos cambiar esto antes de proseguir con nuestro trabajo.

SOLUCIÓN PROPUESTA

Actividad N° 3 - Carga de datos III

The screenshot shows the Microsoft Power BI ribbon at the top with various tabs like Portapapeles, Cortar, Copia, Obtener datos, Conjuntos de datos de Power BI, SQL Server, Especificar Dataverse, Orígenes recientes, Transformar datos, Actualizar consultas, and Administrar relaciones. Below the ribbon, two tables are selected: 'TB_POBLACION_INEI' and 'TB_UBIGEOS'. A large arrow points from the 'Relaciones' tab in the ribbon down to the 'Relaciones' button in the ribbon bar below. The main area is the 'Administrador de relaciones' (Relationship Manager) dialog box. It shows two tables: 'TB_POBLACION_INEI' and 'TB_UBIGEOS'. Under 'TB_POBLACION_INEI', the columns 'Cantida...', 'Departamento...', 'Distrito...', 'Edad_Anio...', 'Provincia...', 'Sexo...', 'ubigeo_inei', and 'ubigeo_reniec' are listed. Under 'TB_UBIGEOS', the columns 'ISO_3166_2...', 'latitud', 'longitud', 'macroregion_inei', 'macroregion_minsa', 'provincia...', 'provincia_inei', 'region', 'superficie', 'ubigeo_inei', and 'ubigeo_reniec' are listed. A relationship is being defined between 'ubigeo_inei' in 'TB_POBLACION_INEI' and 'ubigeo_reniec' in 'TB_UBIGEOS', indicated by a line connecting the two columns with a star symbol at each end. The 'Relaciones' dialog box has buttons for 'Nuevo...', 'Detección automática...', 'Editar...', 'Eliminar', and 'Cerrar'.

Presionando sobre el ícono “Modelo” podemos Administrar Relaciones. Esto nos permite relacionar ambos conjuntos de datos cargados. Debemos seleccionar el campo que une una tabla con la otra (clave foránea) de forma tal que se genere la relación. Una vez hecho esto el diagrama mostrará la conexión creada. Siempre revisen el diccionario de datos para saber cómo unir los datos.

SOLUCIÓN PROPUESTA

Actividad N° 4 - Modificación de tablas

The screenshot shows the Power BI Report Designer interface. On the left, there's a preview area with the text "Creación de objetos visuales con sus datos" and "Seleccione o arrastre campos desde el panel Campos hasta el lienzo del informe." Below this is a small icon illustrating dragging fields from the Fields pane to the report canvas. On the right, the Fields pane is open, showing a tree structure with "TB_POBLACION_INEI" expanded. A context menu is open over the table name, with several options highlighted by red arrows:

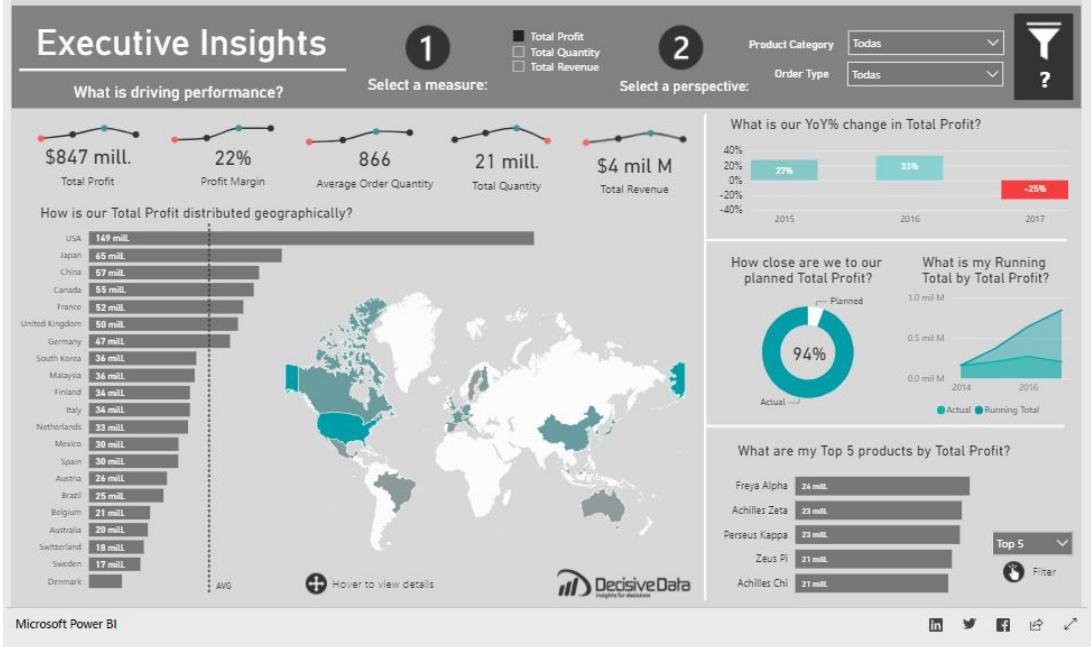
- Nueva medida
- Nueva columna
- Nueva medida rápida
- Actualizar datos
- Editar consulta
- Actualización incremental
- Administrar agregaciones
- Cambiar nombre
- Eliminar
- Ocultar
- Marcar como tabla de fechas ▾
- Ver ocultos
- Mostrar todos
- Contraer todo
- Expandir todo

Above the Fields pane, the ribbon menu is visible with tabs like "Formato", "Propiedades", "Ordenar", "Grupos", "Relaciones", "Cálculos", and "Visualizaciones".

Haciendo clic en el botón derecho sobre el nombre de la tabla aparecerá el menú donde podremos cambiar el nombre de la misma y agregar una columna. Al crear una nueva columna deberemos ingresar la fórmula TODAY() para generar la fecha del día.

SOLUCIÓN PROPUESTA

Actividad N° 5 - Galería



Hay miles de ejemplos disponibles, no obstante, el que seleccionamos para mostrarte está muy completo porque presenta casi todos los componentes vistos: mapas, gráficos, indicadores, filtros. Adicional nota la paleta de colores definida, siempre siguen la misma gama. La fuente se respeta, se mantienen los tamaños. Hay armonía entre todas las partes del tablero.



DATA ANALYTICS

Clase 14 - Transformación de datos

ACTIVIDADES





POWER BI - TRANSFORMACIÓN DE DATOS I

Actividad N° 1



Sigamos trabajando con la Plataforma Nacional de Datos Abiertos de Perú, precisamente con [el archivo que detalla la población de dicho país.](#) Vamos a transformar la columna Sexo de la tabla POBLACION_INEI: en lugar de M o F, deberá decir 1 o 0.



Notas actividad 1:

Poder transformar datos dentro de la herramienta de BI nos ahorrará mucho tiempo. Pero recuerda que los datos no son modificados en su origen, sólo en la herramienta.



POWER BI - TRANSFORMACIÓN DE DATOS II

Actividad N° 2



Continuando con la columna SEXO modificada en la actividad anterior, ahora cambiaremos el tipo de dato a **Número Entero**.

1 2
3 4



Notas actividad 2:

Siempre hay problemas con los tipos de datos. Debemos verificarlo siempre y corregirlo en los casos que sea necesario o conveniente.



POWER BI - TRANSFORMACIÓN DE DATOS III

Actividad N° 3



En esta ocasión agregaremos una nueva columna a la tabla UBIGEOS. Deberá llamarse PAIS y su contenido debe ser Perú, dado que toda la tabla tiene datos de ese país.



Notas actividad 3:

En muchas ocasiones dividir una columna en dos o tres nos ayudará a resolver problemas más fácilmente. Es importante pensar “fuera de la caja”. 📦



POWER BI - TRANSFORMACIÓN DE DATOS IV

Actividad N° 4



Una vez hecho lo anterior, modificaremos una columna de la tabla UBIGEOS. La columna SUPERFICIE contiene valores nulos. Los reemplazaremos con 0.



Notas actividad 4:

La detección de valores nulos es una de las tareas más importantes. Detectar dónde están y decidir si se dejan o se reemplazan no es una cuestión menor. 😊



POWER BI - TRANSFORMACIÓN DE DATOS V

Actividad N° 5



Ahora usaremos nuevamente la columna SUPERFICIE. Crea una nueva columna que se llame DOBLE_SUPERFICIE que contenga el resultado de multiplicar cada valor de superficie por dos. 😊



Notas actividad 5:

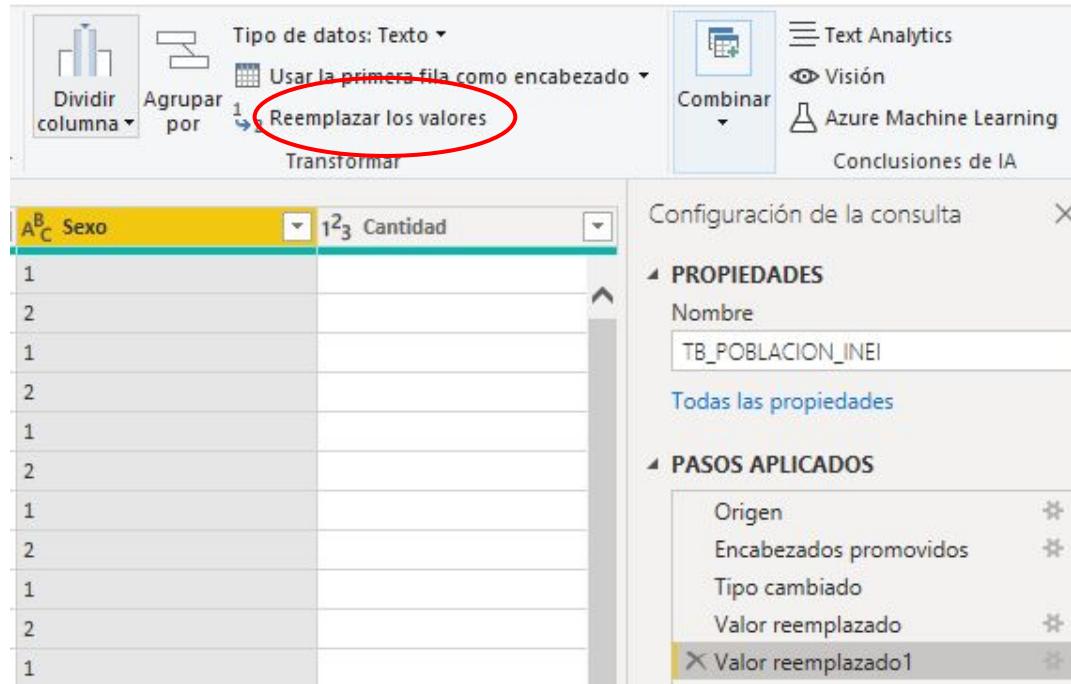
La posibilidad de insertar fórmulas en Power BI nos permite realizar cálculos a la hora de generar tableros, sin depender únicamente de SQL. 😊

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Transformación de datos I



The screenshot shows the Power BI Data Editor interface. At the top, the ribbon has the 'TRANSFORMAR' tab selected. In the 'Acciones' (Actions) bar, there are several icons: 'Dividir columna', 'Agrupar por', 'Tipo de datos: Texto', 'Usar la primera fila como encabezado', and a circled 'Reemplazar los valores' (Replace values) button. To the right of the actions bar is a 'Combinar' (Combine) button followed by a dropdown menu with options: 'Text Analytics', 'Visión', 'Azure Machine Learning', and 'Conclusiones de IA'. Below the actions bar is a table preview showing columns 'Sexo' and 'Cantidad'. On the right side of the screen, there is a 'Configuración de la consulta' (Query configuration) pane. It contains sections for 'PROPIEDADES' (Properties) where 'Nombre' is set to 'TB_POBLACION_INEI', and 'PASOS APLICADOS' (Applied steps) which include 'Origen', 'Encabezados promovidos', 'Tipo cambiado', 'Valor reemplazado', and a highlighted step 'Valor reemplazado'.

Dentro del menú TRANSFORMAR DATOS podremos encontrar el botón REEMPLAZAR LOS VALORES. Allí podremos cambiar los datos fácilmente. 😊

SOLUCIÓN PROPUESTA

Actividad N° 2 - Transformación de datos II

The screenshot shows the 'Transform' tab of the Power BI Data Editor. A red oval highlights the 'Tipo de datos: Número entero' button in the top toolbar. Below it, the 'Sexo' column is selected, and its data type is being changed from Text to Integer. The 'Cantidad' column is also visible. The data in the 'Sexo' column consists of the values 1 and 2.

Sexo	Cantidad
1	
2	
1	
2	
1	
2	
1	

En la misma solapa del ejercicio anterior se puede cambiar el **tipo de dato**, de Texto a Numérico. Nota como la alineación del contenido de las celdas cambió de izquierda a derecha. 😊

SOLUCIÓN PROPUESTA

Actividad N° 3 - Transformación de datos III

Agregar columna

Vista Herramientas Ayuda

Categoría función Personalizada Duplicar columna

Formato

Columna condicional Columna de índice

Combinar columnas Trigonometría

Fecha Hora Duración

Text Analytics Visión Conclus

Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna País

Fórmula de columna personalizada ○

= "Perú"

Columnas disponibles

- ubigeo_reniec
- ubigeo_inei
- Departamento
- Provincia
- Distrito
- Edad_Año
- Sexo

<< Insertar

Información sobre fórmulas de Power Query

✓ No se han detectado errores de sintaxis.

Aceptar Cancelar

	123 superficie	123
1	154	-77.8728 Perú
1	26	-77.7108 Perú
1	357	-78.0197 Perú
1	57	-77.7008 Perú
1	143	-77.7375 Perú
1	279	-77.8542 Perú
1	181	-77.6286 Perú
	3041	-6.1064
1	49	-6.1736
1	380	-6.4847
)	373	-6.7075

SOLUCIÓN PROPUESTA

Actividad N° 4 - Transformación de datos IV

A screenshot of a data transformation interface. At the top, there are four columns labeled '1.2 superficie', '1.2 altitud', '1.2 latitud', and '1.2 longit'. The first column, '1.2 superficie', is highlighted with a yellow background and has a red circle drawn around it. A context menu is open over this column, listing various options: Copiar, Quitar, Quitar otras columnas, Duplicar columna, Agregar columna a partir de los ejemplos..., Quitar duplicados, Quitar errores, Cambiar tipo, Transformar, Reemplazar los valores..., Reemplazar errores..., Agrupar por..., Rellenar, Anulación de dinamización de columnas, Anulación de dinamización de otras columnas, Anular dinamización de las columnas seleccionadas únicamente, Cambiar nombre..., Mover, and Rastrear desagrupando datos. The 'Reemplazar los valores...' option is highlighted with a gray background.

Seleccionando la columna SUPERFICIE, botón derecho, aparece el menú de la izquierda. En la opción REEMPLAZAR VALORES podremos cambiar los NULL por 0. 😊

Reemplazar los valores

Reemplace un valor con otro de las columnas seleccionadas.

Valor que buscar

null

Reemplazar con

0

SOLUCIÓN PROPUESTA

Actividad N° 5 - Transformación de datos V

Columna personalizada

Agregue una columna que se calcula a partir de otras columnas.

Nuevo nombre de columna

Doble_Superficie

Fórmula de columna personalizada ⓘ

= [superficie] * 2

Información sobre fórmulas de Power Query

Columnas disponibles

macreregion_minsa
iso_3166_2
fips
superficie
altura
latitud
longitud

<< Insertar

Aceptar

Cancelar

Desde el botón COLUMNAS PERSONALIZADA pueden crear una columna con los valores solicitados. Desde la ventana COLUMNAS DISPONIBLES pueden insertar SUPERFICIE y luego, en el campo de fórmula, aplicar la multiplicación por dos. ✨

✓ No se han detectado errores de sintaxis.



DATA ANALYTICS

Clase 15 - Dimensiones y medidas

ACTIVIDADES





POWER BI - TRANSFORMACIONES AVANZADAS I

Actividad N° 1



En esta ocasión usaremos un conjunto de datos referente a [Proyectos de Investigación del Ministerio de Ciencia y Tecnología de Colombia](#). En primera instancia crearemos una nueva MEDIDA que cuente la cantidad total de proyectos distintos existentes.



Notas actividad 1:

Antes de poner manos a la obra deberás SIEMPRE analizar los datos. Apurarse solo produce errores. Toma el tiempo necesario para revisar los datos.





POWER BI - TRANSFORMACIONES AVANZADAS II

Actividad N° 2



Usando el mismo conjunto de datos, inserta un gráfico donde se muestre el total de proyectos y el detalle de proyectos por año. Utiliza la medida creada en el ejercicio anterior. 



Notas actividad 2:

Crear medidas nos permite amasar los datos sin necesidad de usar SQL. Recuerden que los datos originales no son modificados con lo que hagan en Power BI. 



POWER BI - TRANSFORMACIONES AVANZADAS III

Actividad N° 3



Usando el mismo conjunto de datos, crearemos una tabla calendario tomando como fecha de referencia FECHA_REGISTRO.



Notas actividad 3:

La creación de la tabla calendario nos permite manejar las jerarquías de fechas con mayor claridad.



POWER BI - TRANSFORMACIONES AVANZADAS IV

Actividad N° 4



Ahora crearemos nuestro perfil en Power BI. De esta manera nuestro trabajo podrá ser subido a la web. 💪💻



Notas actividad 4:

Recuerden que en el mundo actual todo el trabajo es colaborativo. 🤝 Saber trabajar en la nube es fundamental.



POWER BI - TRANSFORMACIONES AVANZADAS V

Actividad N° 5



Una vez creado el perfil podremos subir el proyecto a nuestro espacio de trabajo en la web. 🚀 Publica tu proyecto en la web y verifica que se pueda acceder desde un navegador (Chrome, Edge, Firefox, Safari, etc).



Notas actividad 5:

Siempre verifica que tu trabajo puede ser visto correctamente desde la web. Una buena práctica es ingresar en modo incógnito y ver si funciona correctamente. 🌐

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Transformaciones avanzadas I

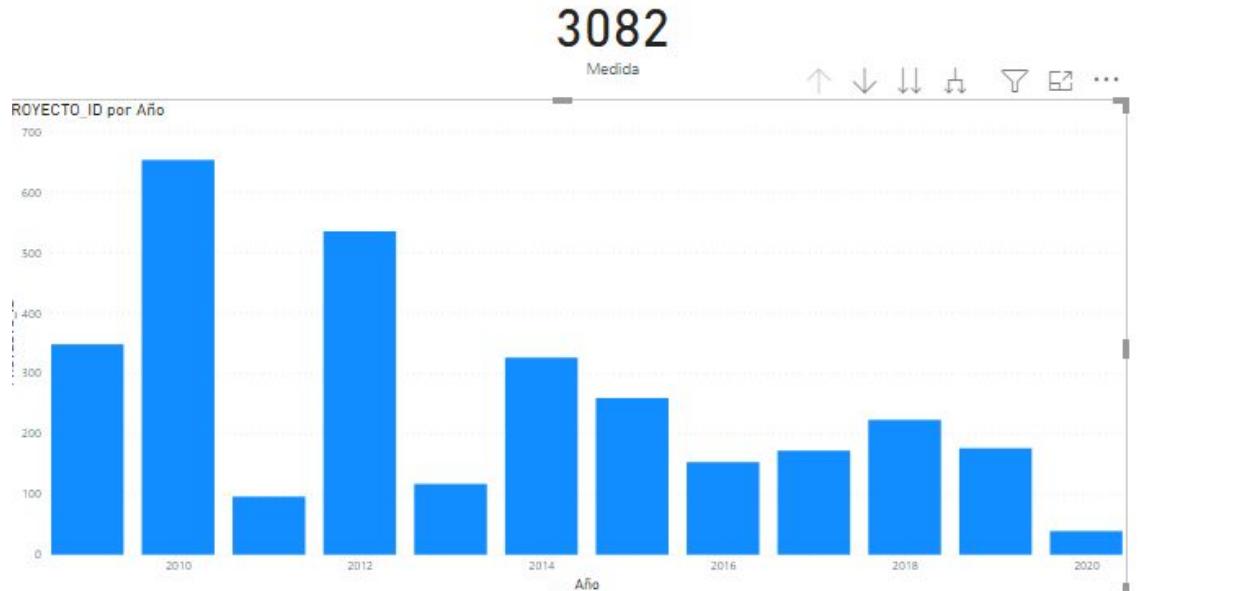
The screenshot shows the Power BI desktop interface. At the top, there's a ribbon with tabs like 'Archivo inicial', 'Proyectos_de_inves...', 'Formato' (with dropdowns for '\$', '%', and 'Automática...'), and 'Cálculos'. Below the ribbon, there are four tabs: 'Estructura', 'Formato', 'Propiedades', and 'Cálculos'. A message in a blue box says: 'Autorrecuperación contiene algunos archivos recuperados que no se han abierto.' In the 'Cálculos' tab, a formula is defined: '1 Medida = DISTINCTCOUNT(Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009[PROYECTO_ID])'. The main area displays a data grid with columns: PROYECTO_ID, CONVOCATORIA_ID, ANO_CONVOCATORIA, DESC_CONVOCATORIA, CODIGO_PROYECTO, and TITULO. The data includes rows for various projects from 2010, such as '508-2010 BANCO DE PROYECTOS ESTANDARIZACION Y ACF' and '521-2010 Banco de Proyectos de Investigacion Cientifica o'.

PROYECTO_ID	CONVOCATORIA_ID	ANO_CONVOCATORIA	DESC_CONVOCATORIA	CODIGO_PROYECTO	TITULO
37759	11160	2010	508-2010 BANCO DE PROYECTOS ESTANDARIZACION Y ACF	222850827518	Implementacion de l...
37761	11160	2010	508-2010 BANCO DE PROYECTOS ESTANDARIZACION Y ACF	222850827520	Estandarizacion de p...
37739	11160	2010	508-2010 BANCO DE PROYECTOS ESTANDARIZACION Y ACF	123750827498	Consolidacion de un...
37399	11160	2010	508-2010 BANCO DE PROYECTOS ESTANDARIZACION Y ACF	110150827155	Migracion global y es...
38773	11168	2010	521-2010 Banco de Proyectos de Investigacion Cientifica o	120452128532	Busqueda de estrate...
38782	11168	2010	521-2010 Banco de Proyectos de Investigacion Cientifica o	120352128541	Corroboracion de la i...
38788	11168	2010	521-2010 Banco de Proyectos de Investigacion Cientifica o	120352128547	ValorizaciÂn de resu...

Recuerda que las fórmulas de Power BI son bastante similares a las de Excel. Sus nombres son bastante elocuentes y, además, son muy simples de usar.

SOLUCIÓN PROPUESTA

Actividad N° 2 - Transformaciones avanzadas II



Filtros

The Power BI interface showing various filters and settings. The "Filtros" section includes a large icon for "FECHA_REGISTRO" and smaller icons for "Año", "Trimestre", "Mes", and "Día". The "Eje" section shows the "FECHA_REGISTRO" field selected. The "Leyenda" section has a placeholder "Agregar campos de datos aquí". The "Valores" section shows the "PROYECTO_ID" field.

SOLUCIÓN PROPUESTA

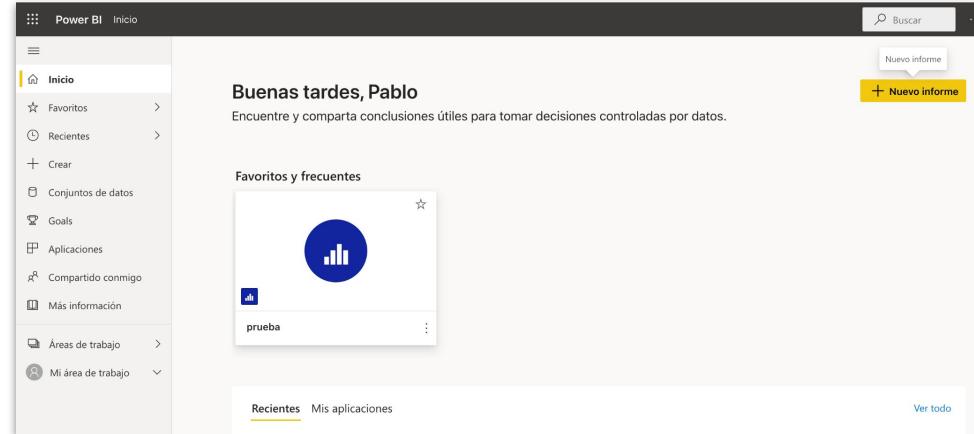
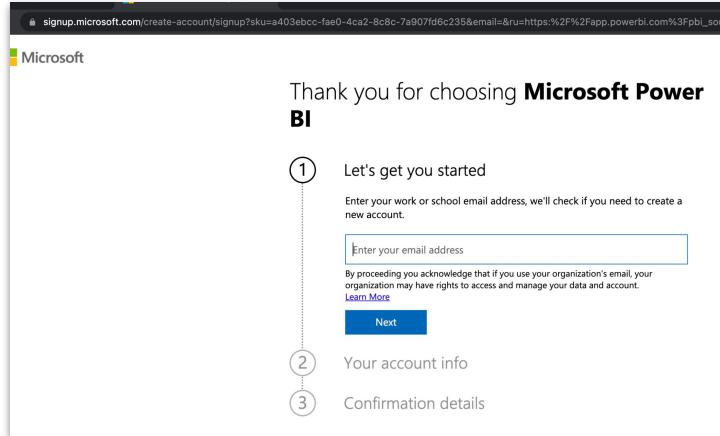
Actividad N° 3 - Transformaciones avanzadas III

The screenshot shows the Microsoft Power BI Data Editor interface. On the left, there is a preview grid of data with columns: FECHA_REGISTRO, Anio, Mes, and Trimestre. The 'Anio' column contains the value '2010'. The 'Mes' and 'Trimestre' columns also contain values (10 and 4 respectively). The 'FECHA_REGISTRO' column shows dates from 'viernes, 15 de octubre de 2010' repeated multiple times. At the top, a formula bar displays: '1 Anio = YEAR('Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2 (2)'[FECHA_REGISTRO])'. To the right, there are tabs for Estructura, Formato, Propiedades, Ordenar, Grupos, Relaciones, Cálculos, and Campos. The 'Campos' pane is open, showing a search bar and a list of fields from the source table, with 'Anio' being the currently selected field.

Nota cómo las fórmulas son bastante simples (se llaman MONTH, QUARTER y YEAR). El nombre de la tabla cambiará dependiendo de cómo se hayan denominado cada uno. Como el campo de las fórmulas posee texto predictivo, si empiezas a escribir te brindará sugerencias.

SOLUCIÓN PROPUESTA

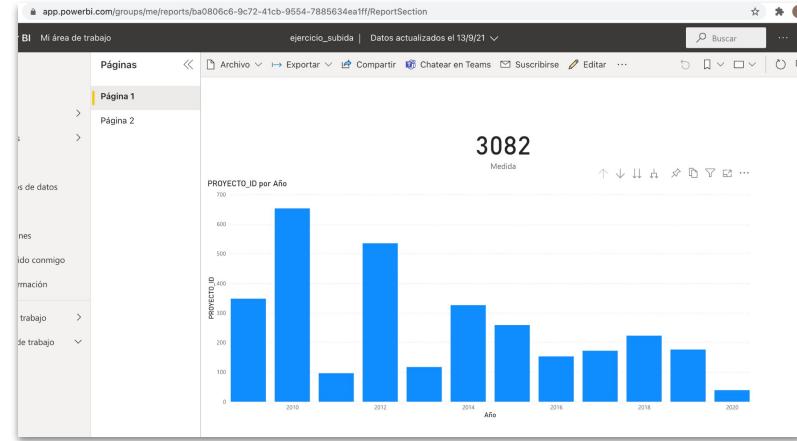
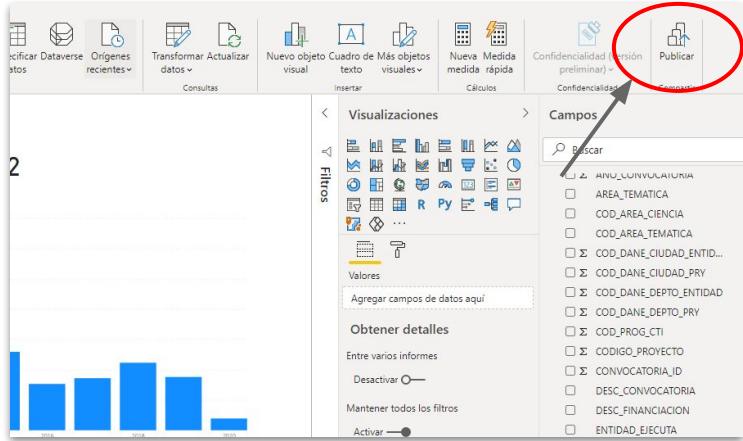
Actividad N° 4 - Transformaciones avanzadas IV



Primero deberás completar el registro (captura de la izquierda)👉. Luego podrás ingresar a la plataforma de Power BI, tal como se observa en la captura de la derecha.👉

SOLUCIÓN PROPUESTA

Actividad N° 5 - Transformaciones avanzadas V

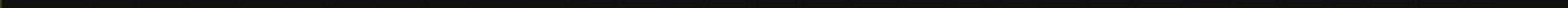


Simplemente presionando el botón PUBLICAR y completando unos simples datos (captura de la izquierda), el proyecto quedará disponible en la web (captura de la derecha).



DATA ANALYTICS

Clase 16 - Gráficos y segmentaciones



ACTIVIDADES





POWER BI - GRÁFICOS I

Actividad N° 1



Seguimos usando los datos de [Proyectos de Investigación del Ministerio de Ciencia y Tecnología de Colombia](#). Necesitamos crear un gráfico 📈 donde muestre a qué entidades ejecutivas le fueron aprobados más proyectos (usar el campo ENTIDAD_EJECUTA).



Notas actividad 1:

Como en este caso, a veces las definiciones son poco claras e implican conocer los datos. Siempre se recomienda preguntar 🎤 para evitar ambigüedades y trabajar en vano.



POWER BI - GRÁFICOS II

Actividad N° 2



Continuando con el ejercicio N° 1, + crea una nueva hoja, agrégale un título a la hoja, repite el mismo gráfico pero agrégale un título y etiquetas de datos. Desactiva el título del eje X.



Notas actividad 2:

En los detalles está la excelencia. 🏆 No dejar nada librado al azar ni a la interpretación del usuario. Todo debe estar bien claro.



POWER BI - GRÁFICOS III

Actividad N° 3



Agreguemos una nueva hoja e insertemos un mapa del tipo ArcGIS Maps. Para la **ubicación** usar el campo NME_DEPTO_ENTIDAD, para el **color** recuento del campo PROYECTO_ID. De esta forma veremos en un mapa la cantidad de proyectos por departamento.



Notas actividad 3:

Los mapas son una muy buena manera de mostrar gráficamente la distribución de las variables analizadas. Hay distintos tipos, cada uno con una función particular.



POWER BI - GRÁFICOS IV

Actividad N° 4



Otra vez, agreguemos una nueva hoja e insertemos otro mapa del tipo ArcGIS Maps y un gráfico de barras horizontales. Ambos gráficos deben mostrar la cantidad de proyectos por Departamento (campos PROYECTO_ID y NME_DEPTO_ENTIDAD). Haz clic sobre una de las barras o uno de los departamentos y ve que sucede. ☺



Notas actividad 4:

Power BI tiene muchas funciones automáticas que brindan muchas soluciones pero ¡Ojo! Hay que conocerlas para saber cómo se comporta la herramienta. !



POWER BI - GRÁFICOS V

Actividad N° 5



Ahora genera tres gráficos en una nueva hoja: uno de barras, otro de anillos y un Treemap. Los tres con recuento de proyectos por ciudad (PROYECTO_ID y NME_CIUDAD_ENTIDAD).

👉 Desactiva la interacción entre uno de los 3 gráficos.

Notas actividad 5:

Power BI tiene muchas herramientas. Es muy difícil conocerlas todas, así que no desesperen. La práctica hace al maestro.

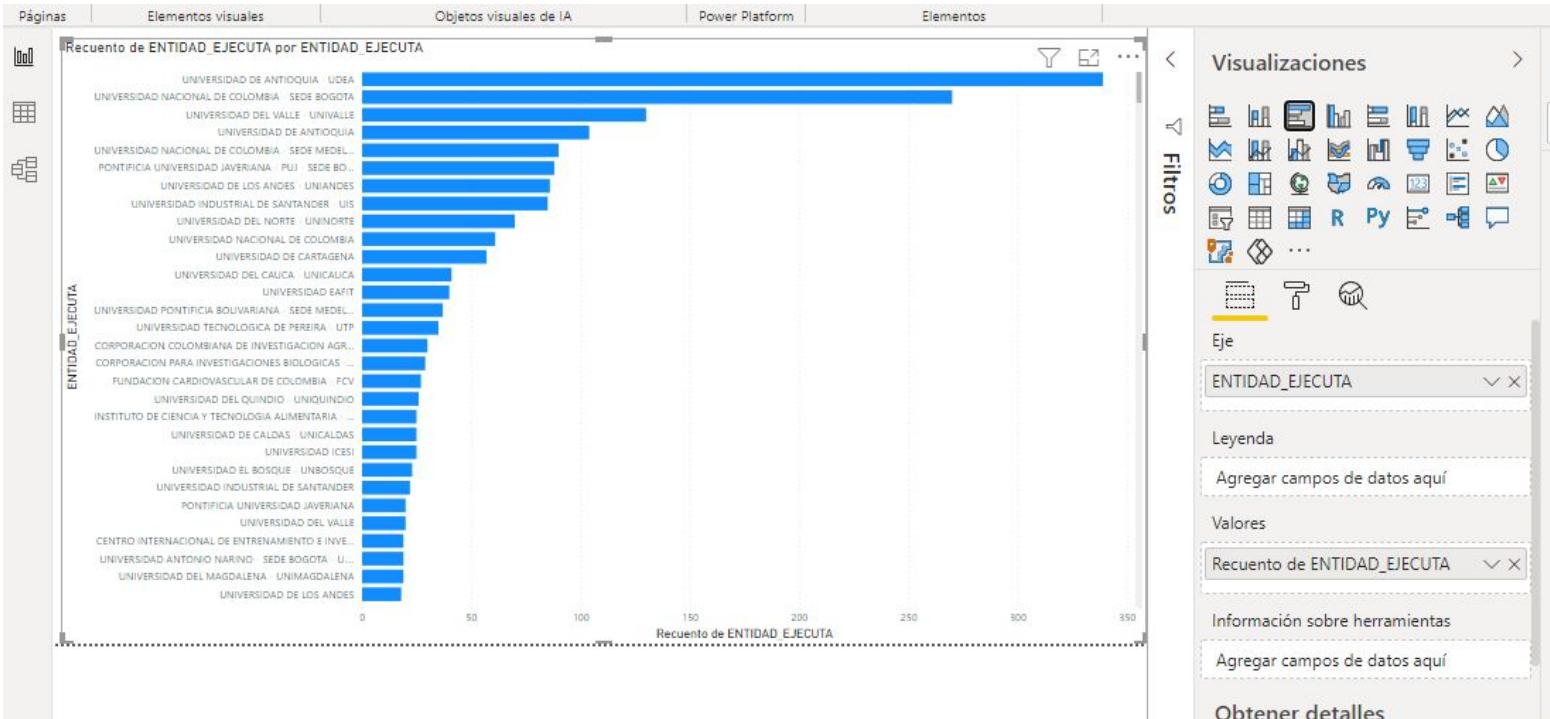


RESOLUCIONES



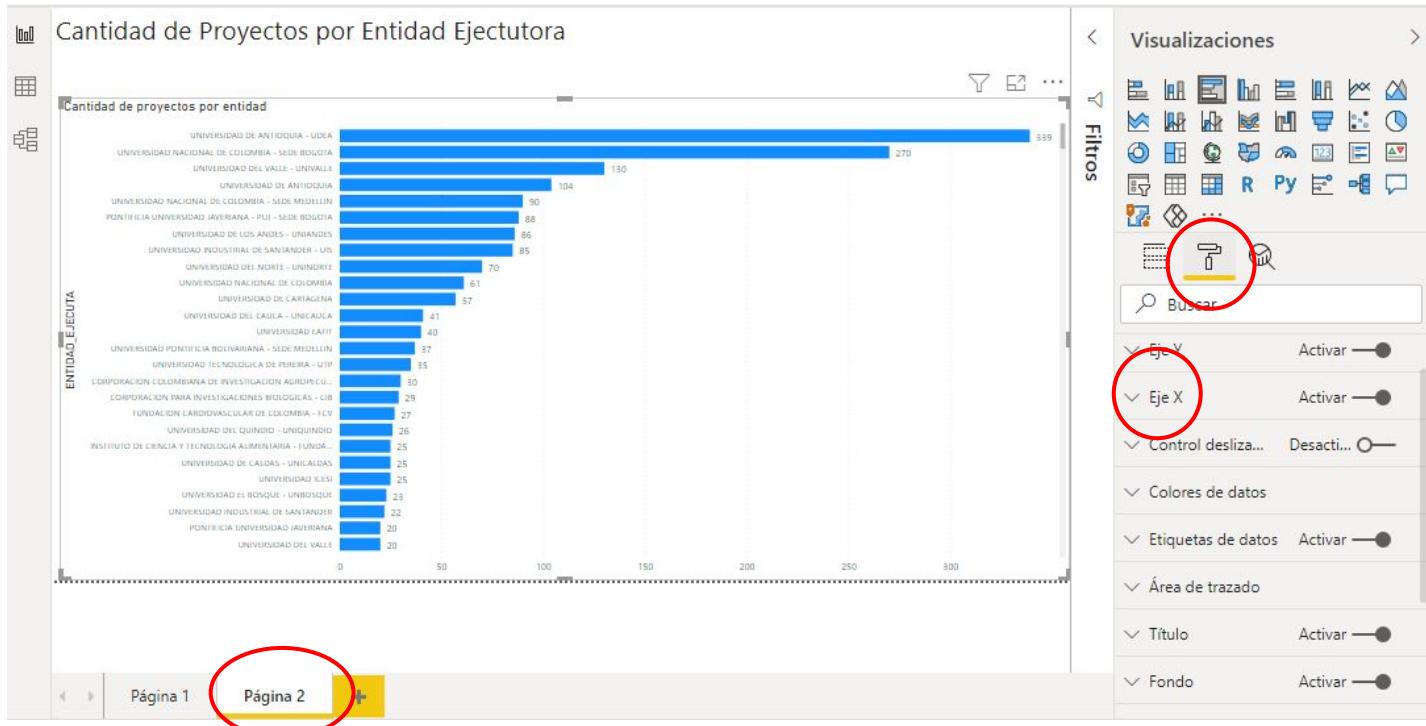
SOLUCIÓN PROPUESTA

Actividad N° 1 - Power BI - Gráficos /



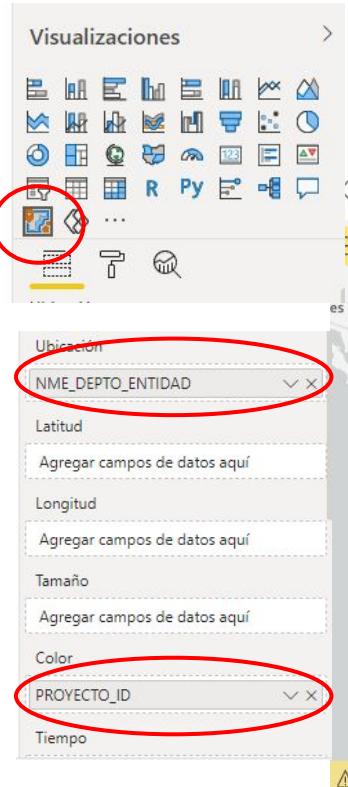
SOLUCIÓN PROPUESTA

Actividad N° 2 - Power BI - Gráficos II

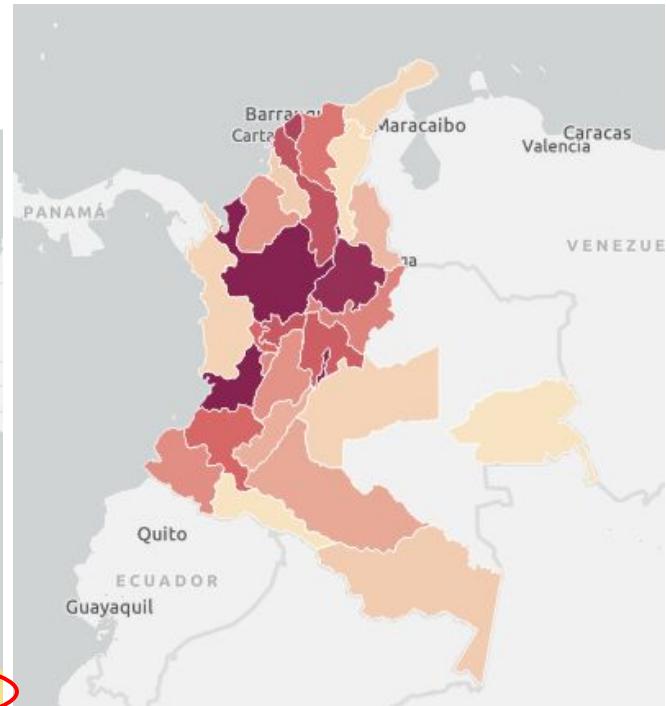


SOLUCIÓN PROPUESTA

Actividad N° 3 - Power BI - Gráficos III

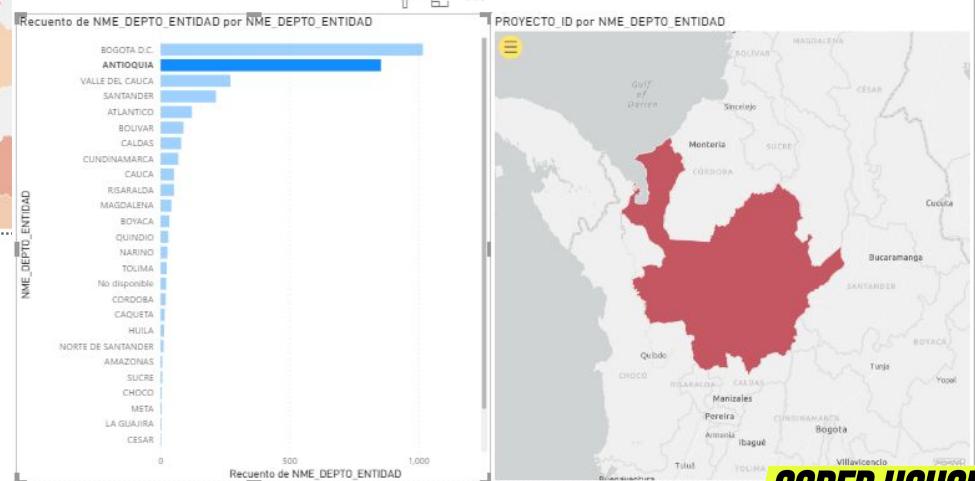
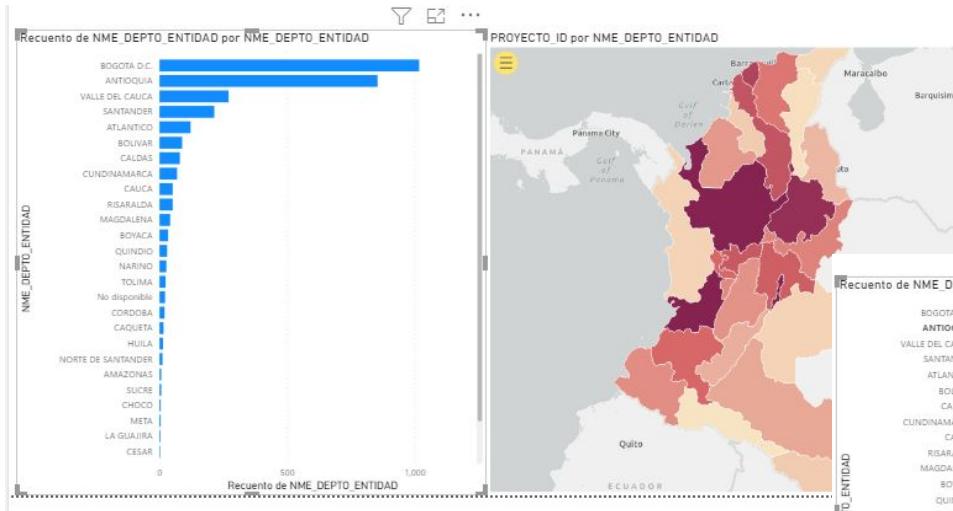


The screenshot shows the 'Ubicación' settings for a map visualization. The 'Ubicación' dropdown is set to 'NME_DEPTO_ENTIDAD'. A red circle highlights this dropdown. Below it, the 'Color' dropdown is set to 'PROYECTO_ID', also highlighted with a red circle. A callout box titled 'Tipo de ubicación' provides details on how locations are represented (as borders) and which country they are in (Colombia). It also includes options for 'Departments' and 'Coincidencia más cercana' (selected). A red circle highlights the 'Coincidencia más cercana' radio button. At the bottom, a warning message is circled in red: '⚠ La precisión de las ubicaciones se puede mejorar. Vaya a Tipo de ubicación para obtener más información.' (⚠ Location precision can be improved. Go to Location type for more information.)



SOLUCIÓN PROPUESTA

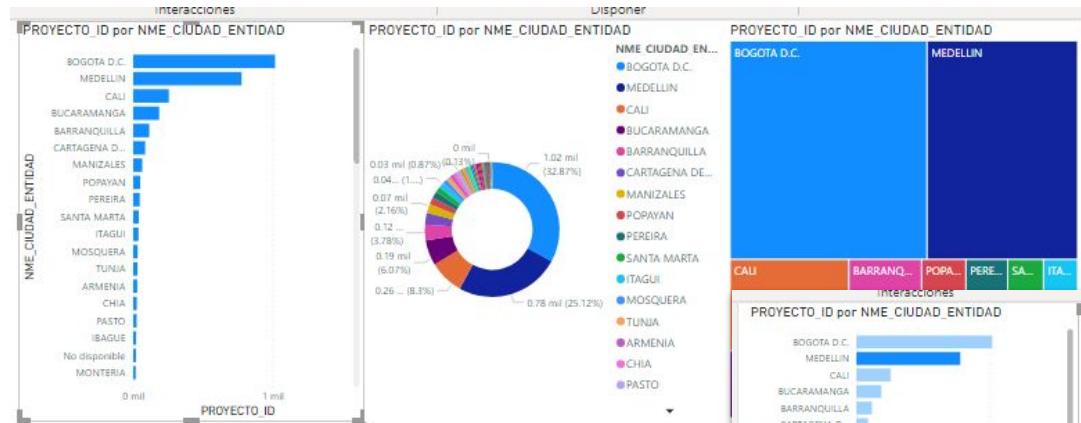
Actividad N° 4 - Power BI - Gráficos IV



Noten como seleccionando una barra solo se sombra ese departamento en el mapa.
Se llama INTERACCIÓN. 😊

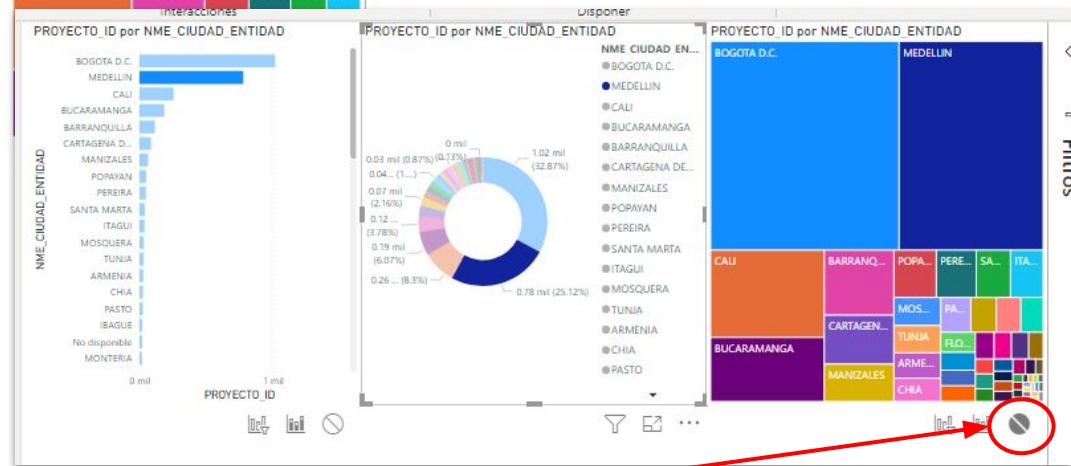
SOLUCIÓN PROPUESTA

Actividad N° 5 - Power BI - Gráficos V



Vean como la INTERACCIÓN se ha desactivado en el tercer gráfico (solapa FORMATO, EDITAR INTERACCIONES).

Nota como la INTERACCIÓN se mantiene entre el primer y segundo gráfico, no así en el tercero. 😞



CODER HOUSE

DATA ANALYTICS

Clase 17 - Medidas avanzadas

ACTIVIDADES





POWER BI - CÁLCULOS AVANZADOS I

Actividad N° 1



Seguimos usando los datos de [Proyectos de Investigación del Ministerio de Ciencia y Tecnología de Colombia](#). Vamos a crear una medida que muestre la cantidad de proyectos de la ciudad de BOGOTA D.C.

Mostrar los resultados en una tarjeta.



Notas actividad 1:

Una de las mayores fortalezas de Power BI son las fórmulas. Aprender a usarlas nos otorgará muchísimos beneficios y facilitará las tareas.



POWER BI - CÁLCULOS AVANZADOS II

Actividad N° 2



Ahora, continuando con los datos del [Proyectos de Investigación del Ministerio de Ciencia y Tecnología de Colombia](#), vamos a crear una medida que muestre la cantidad de proyectos de la ciudad de BOGOTA D.C. para el año 2020. 😊

Mostrar los resultados en una tarjeta.



Notas actividad 2:

Recuerda que la mayoría de los errores son fáciles de resolver: una coma o un paréntesis faltante ocasionan gran parte de las fallas en las fórmulas. 🛠



POWER BI - CÁLCULOS AVANZADOS III

Actividad N° 3



Crea una variable que devuelva el porcentaje de monto financiado sobre el monto total (columnas MONTO_TOTAL_AP y MONTO_FINANCIADO_AP). De nuevo, inserta el resultado en una tarjeta. 



Notas actividad 3:

Las variables son parte esencial de un entorno de programación. En cualquier lenguaje son el alma del código, por lo que usarlas es una obligación. 



POWER BI - CÁLCULOS AVANZADOS IV

Actividad N° 4



Crear un parámetro (what if) numérico, entero, que vaya de 1 a 30. Denominarlo “Parámetro_Ejemplo”, y ubicarlo en la misma hoja que el resultado del ejercicio anterior (porcentaje). 😊



Notas actividad 4:

Power BI, tal como hemos dicho, ofrece miles de herramientas distintas. Son muchos los caminos que nos llevan a un mismo lugar.





POWER BI - CÁLCULOS AVANZADOS V

Actividad N° 5



En esta oportunidad crea una nueva medida que tome el parámetro del [ejercicio anterior](#) y el porcentaje del [ejercicio 3](#). La medida debe devolver el resultado de dividir el porcentaje por el parámetro. Insertarlo en una tarjeta. 



Notas actividad 5:

Si las consignas son confusas léelas de nuevo. Y otra vez. Y otra. Y si aún así quedan dudas, consulta con el usuario. Las consignas confusas son más habituales de lo deseado. 

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Power BI - Cálculos avanzados I

The screenshot shows the Power BI interface with the following details:

- Estructura (Structure) tab:** Contains DAX code:

```
1 Proyectos_Bogota = CALCULATE(COUNT  
(Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009[PROYECTO_ID]),  
Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009[NOMBRE_CIUDAD_ENTIDAD]  
= "BOGOTA D.C.")
```
- Formato (Format) tab:** Shows a large number **1018** with the label **Proyectos_Bogota** below it.
- Propiedades (Properties) tab:** Shows filter settings:
 - Filtros de este objeto visual ...
 - Proyectos_Bogota es (todos)
 - Agregar campos de datos ...
- Cálculos (Calculated Column) tab:** Shows filter settings:
 - Filtros de esta página ...
 - Agregar campos de datos ...
- Bottom navigation bar:** Shows the path **ID por Año** and the **CODER HOUSE** logo.

SOLUCIÓN PROPUESTA

Actividad N° 2 - Power BI - Cálculos avanzados II

The screenshot shows the Power BI Data View interface. On the left, there is a code editor window displaying DAX code:

```
1 Proyectos_Bogota_Año_Anterior = CALCULATE(
2     COUNT(Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009
[PROYECTO_ID]),
3     Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009
[NME_CIUDAD_ENTIDAD] = "BOGOTA D.C.",
4     YEAR(Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009
[FECHA_APROBACION]) = 2020)
```

On the right, there is a visual representation of the data flow. It consists of a large rectangle containing two numbers: 1018 and 41. Above the rectangle is a funnel icon with three dots next to it. To the right of the rectangle is a vertical column of four items:

- Proyectos_Bogota_Año es (todos)
- Agregar campos de d
- Filtros de esta página
- Agregar campos de d

At the bottom right of the visual area, the text "CODER HOUSE" is displayed.

SOLUCIÓN PROPUESTA

Actividad N° 3 - Power BI - Cálculos avanzados III

```
1 porcentaje =  
2     var monto_financiado = SUM  
(Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009  
[MONTO_FINANCIADO_AP])  
3     var monto_total = sum  
(Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009[MONTO_TOTAL_AP])  
4  
5 return round(monto_financiado / monto_total * 100,2)
```

51.69

porcentaje

Filtros de todas las páginas

Agregar campos de datos

SOLUCIÓN PROPUESTA

Actividad N° 4 - Power BI - Cálculos avanzados IV

The screenshot shows the Power BI Data Model interface. A parameter named "Parámetro_Ejemplo" is defined with the formula `GENERATESERIES(0, 30, 1)`. Below the parameter, a visual displays the value **51.69** with the label "porcentaje".

Nota como los parámetros también generan una fórmula. Por ende, podrían ser creados mediante el botón PARAMETRO NUEVO (barra MODELADO) o bien mediante una fórmula. 🧑

SOLUCIÓN PROPUESTA

Actividad N° 5 - Power BI - Cálculos avanzados V

Estructura | Formato | Propiedades

✓ 1 porcentaje2 = [porcentaje] / 'Parámetro_Ejemplo'[Valor Parámetro_Ejemplo]

The screenshot shows the Power BI interface with three main sections: Estructura, Formato, and Propiedades. In the Estructura section, there is a parameter control labeled 'Parámetro Ejemplo' with the value '10'. Below it, a calculated field 'porcentaje2' is displayed with the value '5.17'. In the bottom section, another calculated field 'porcentaje' is shown with the value '51.69'. Red circles highlight the parameter value '10', the calculated field value '5.17', and the final result '51.69'.



Al deslizar la barra del parámetro se debe modificar el *porcentaje2*. En la imagen se muestra el 10% sobre 51,69.

CODER HOUSE

DATA ANALYTICS

Clase 18 - Workshop Power BI

Sin actividades 😊



DATA ANALYTICS

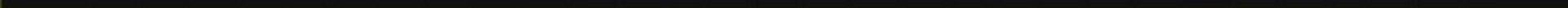
Clase 19 - Charla de un especialista visualización de datos

Sin actividades 😊



DATA ANALYTICS

Clase 20 - Medidas avanzadas



ACTIVIDADES





POWER BI - TENDENCIA CENTRAL I

Actividad N° 1



Seguiremos usando los datos de [Proyectos de Investigación del Ministerio de Ciencia y Tecnología de Colombia](#). Vamos a crear un mini tablero. Calcula y muestra la media de MONTO TOTAL AP. 



Notas actividad 1:

A la hora de tener una consigna es importante saber cómo resolverla y cómo mostrarla. Ambas tareas son fundamentales. Un buen resultado pero mal mostrado, no sirve. 



POWER BI - TENDENCIA CENTRAL II

Actividad N° 2



Calcula y muestra la media de MONTO TOTAL AP para el año 2020 y agrégala a la misma hoja donde realizaste el ejercicio anterior.



Notas actividad 2:

No se trata solo de hacer cuentas, sino también de usar la imaginación. Muchas veces ni la persona que pregunta sabe cómo quiere la respuesta.



POWER BI - TENDENCIA CENTRAL III

Actividad N° 3



En la misma hoja, ahora agrega un gráfico que pueda mostrar los resultados anteriores. Quizás ninguno se ajuste completamente pero puedes buscar el que más se adapte. 



Notas actividad 3:

Power BI tiene una comunidad que genera herramientas para graficar y mejorar nuestros tableros. Muchos de ellos son sumamente útiles. Pero cuidado, no todos están aprobados por Microsoft. 



POWER BI - TENDENCIA CENTRAL IV

Actividad N° 4



Agregar en la misma hoja un nuevo gráfico que muestre la moda de la ciudad (NME_CIUDAD_ENTIDAD). 🏙️



Notas actividad 4:

Estadística, programación, diseño. Todas esas materias se mezclan en el escritorio de un analista de datos. 🎨



POWER BI - TENDENCIA CENTRAL IV

Actividad N° 5



Por último, inserta un histograma del monto total (MONTO_TOTAL_AP). Acomoda la hoja para que el tablero quede armado. Recuerda ponerle un título. 😊



Notas actividad 5:

Recuerden que si no encuentran el gráfico que necesitan pueden importarlo desde el menú *Visualizaciones*, y obtener más objetos visuales (triple puntos suspensivos). 😊👍

RESOLUCIONES



SOLUCIÓN PROPUESTA

Actividad N° 1 - Tendencia central I

The screenshot shows a software interface with a toolbar at the top and a code editor window below it. The toolbar has four tabs: 'Estructura' (Structure), 'Formato' (Format), 'Propiedades' (Properties), and 'Cálculos' (Calculations). The code editor window contains the following code:

```
1 media_monto = AVERAGE  
([Proyectos_de_Investigaci_n_e_Innovaci_n_evaluados_y_aprobados_desde_el_a_o_2009[MONTO_TOTAL_AP]])
```

Below the code editor, the word 'Media' is displayed. In the results panel, the value '590.77 mill.' is shown, followed by the label 'media_monto'.

SOLUCIÓN PROPUESTA

Actividad N° 2 - Tendencia central II

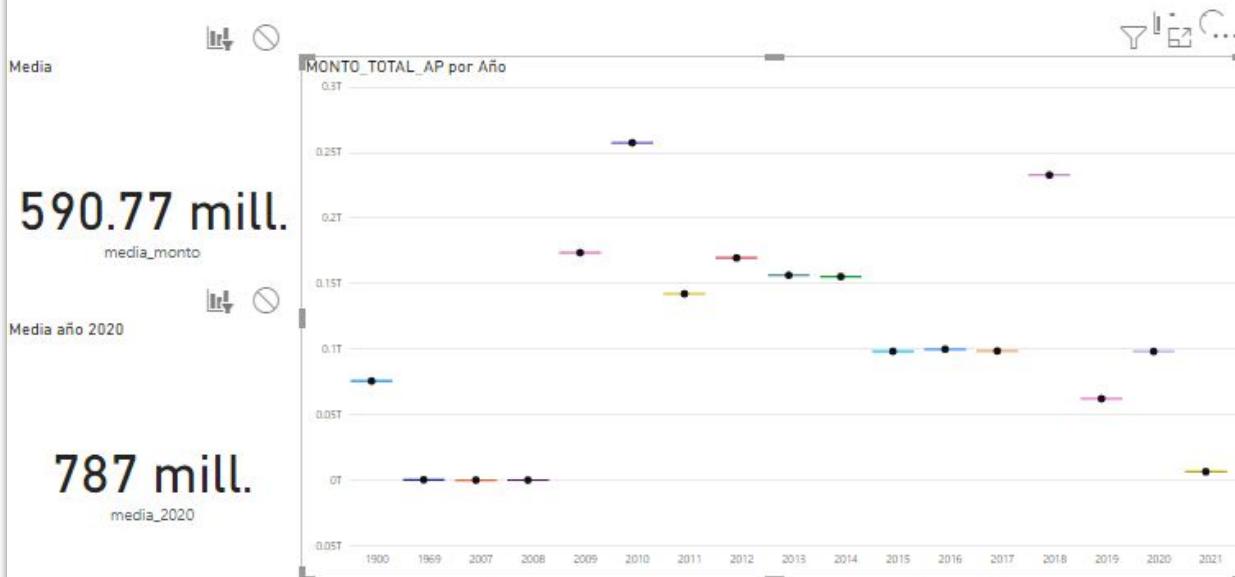
Medidas de Tendencia Central



SOLUCIÓN PROPUESTA

Actividad N° 3 - Tendencia central III

Medidas de Tendencia Central



SOLUCIÓN PROPUESTA

Actividad N° 4 - Tendencia central IV

Medidas de Tendencia Central

Media

590.77
mill.

media_monto

Media año 2020

787 m...

media_2020

MONTO_TOTAL_AP por Año

0.3T

0.25T

0.2T

0.15T

0.1T

0.05T



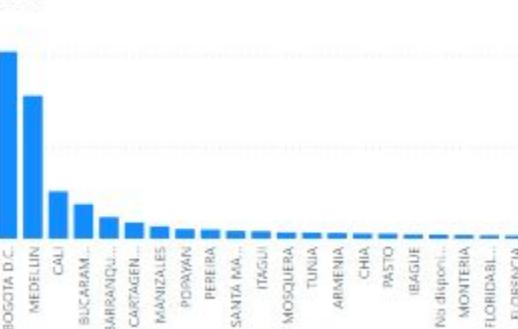
Ciudad - Moda

Re cuento de NME_CIUDAD..

1,000

500

0



SOLUCIÓN PROPUESTA

Actividad N° 5 - Tendencia central V

Medidas de Tendencia Central

Media

590.77
mill.

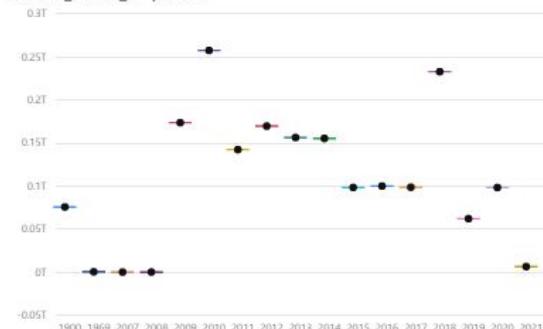
media_monto

Media año 2020

787 mill.

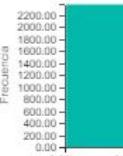
media_2020

MONTO_TOTAL_AP por Año



MONTO_TOTAL_AP

Frecuencia



CODER HOUSE

DATA ANALYTICS

Clase 21 - Tips and Tricks

Sin actividades 😊

CODER HOUSE

DATA ANALYTICS

Clase 22 - Proyecto Final

Sin actividades 😊



DATA ANALYTICS

Clase 23 - Presentación de Proyecto Final

Sin actividades 😊

¿Are you ready for this?

FINAL DE LOS EJERCICIOS

Si llegaste hasta aquí y lograste realizar cada uno de los ejercicios propuestos, déjanos decirte que

You rock! 🤘

Esperamos que toda la información brindada haya sido de mucho provecho y que haya agregando valor a tu desarrollo profesional. No obstante **nos gustaría conocer tu opinión**, por lo que te invitamos a que nos compartas tu experiencia sobre el booklet por  [acá](#). 

De ahora en adelante el compromiso es contigo para mantener estas buenas prácticas e ir por todo lo que sabes que puedes lograr, siempre dando lo mejor de ti en cada proyecto.

¡Muchos éxitos en lo que sea que te propongas! 🙌