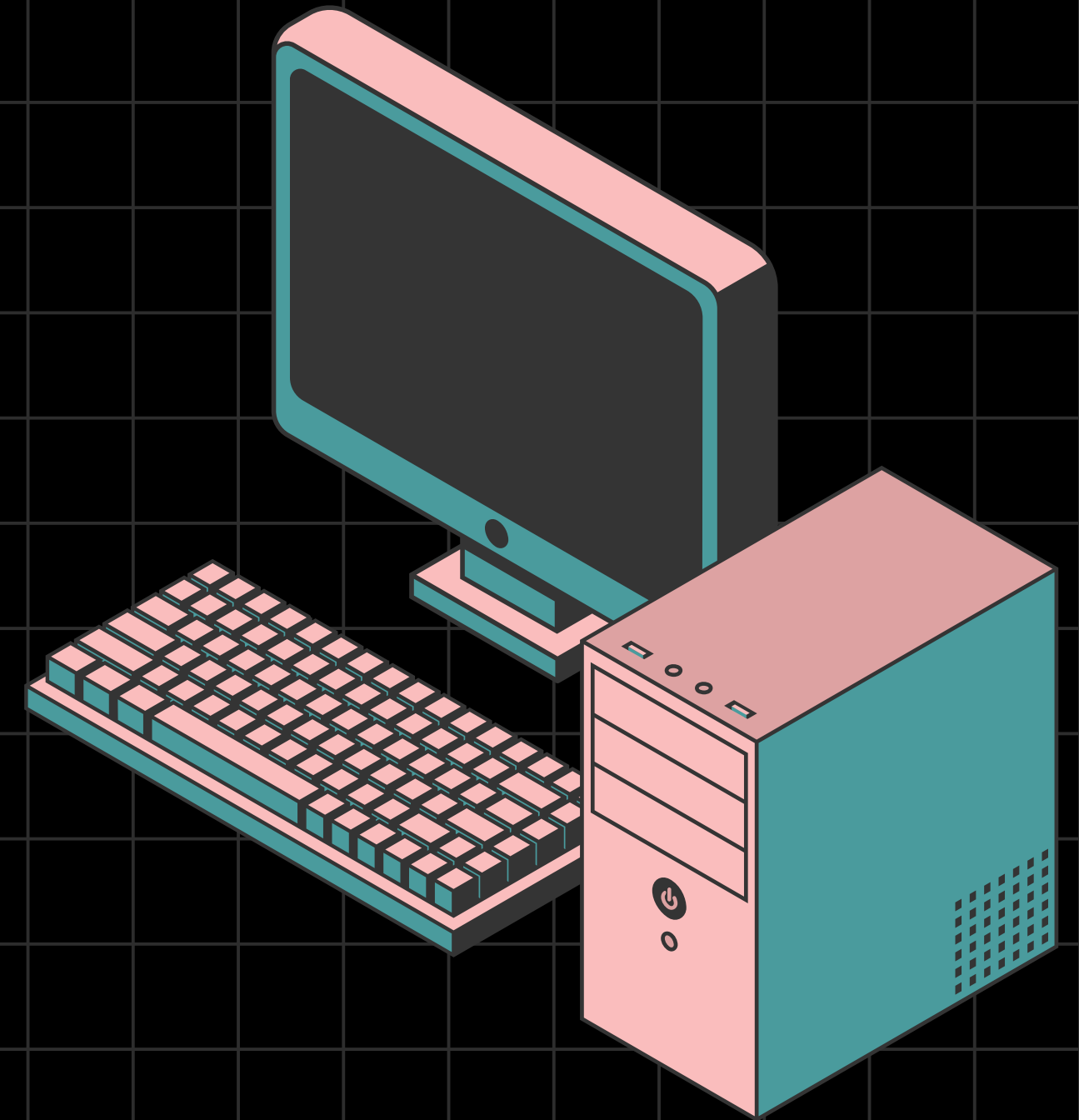


# ETL NACIMIENTOS

Barrionuevo Imanol, Broilo Mateo, Correa Valentín, Díaz Gabriel, Gambino Tomás,  
Gómez Ferrero Andrés, Letona Mateo y Wursten Santiago



# DataSets



## Nacimientos en Argentina por departamento (2012 - 2022)

Frecuencia de personas con DNI nacidas en Argentina entre 2012 y 2022 por año de nacimiento, según datos extraídos en junio de 2023.

[DESCARGAR](#)[CONSULTAR EN EL  
PORTAL DE ORIGEN](#)

### Campos de este recurso

Título de la columna	Tipo de dato	Descripción	Unidad de medida
anio	Número entero (integer)	Año de nacimiento	
provincia_nombre	Texto (string)	Provincia de residencia en el DNI de las personas identificadas.	
provincia_id	Número entero (integer)	Código designado por el Instituto Nacional de Estadística y Censos para cada departamento. CABA se considera como un único departamento.	
departamento_nombre	Texto (string)	Departamento de residencia en el DNI de las personas identificadas.	
departamento_id	Número entero (integer)	Código de departamento designado por el Instituto Nacional de Estadística y Censos	
nacimientos	Número entero (integer)	Cantidad de personas identificadas con DNI	Personas

# Departamentos

```
"categoria","centroide_lat","centroide_lon","fuente","id","nombre","nombre_completo","provincia_id","provincia_interseccion","provincia_nombre"
"Partido",-37.9646159068483,-60.2482821323384,"IGN","06014","Adolfo Gonzales Chaves","Partido de Adolfo Gonzales Chaves","06",0.0125815079288771,
"Departamento",-31.2902983724746,-58.2378696898985,"IGN","30015","Concordia","Departamento Concordia","30",0.0414681664516468,"Entre Ríos"
"Departamento",-30.9924267798293,-58.8918600477175,"IGN","30035","Federal","Departamento Federal","30",0.0657426173552007,"Entre Ríos"
"Departamento",-33.0223490410459,-58.7858051710614,"IGN","30056","Gualeguaychú","Departamento Gualeguaychú","30",0.0970757725759829,"Entre Ríos"
"Departamento",-33.6249961546235,-58.9338958542555,"IGN","30063","Islas del Ibicuy","Departamento Islas del Ibicuy","30",0.0641372927357608,"Entre Ríos"
"Departamento",-32.7839424597941,-60.2180700308372,"IGN","30105","Victoria","Departamento Victoria","30",0.08535097955251,"Entre Ríos"
```

# Provincias

```
"categoria","centroide_lat","centroide_lon","fuente","id","iso_id","iso_nombre","nombre","nombre_completo"
"Ciudad Autónoma",-34.6144420654301,-58.4458763250916,"IGN","02","AR-C","Ciudad Autónoma de Buenos Aires","Ciudad Aut
"Provincia",-38.6419828626673,-70.1198972237318,"IGN","58","AR-Q","Neuquén","Neuquén","Provincia del Neuquén"
"Provincia",-33.7611035381154,-66.0252312714021,"IGN","74","AR-D","San Luis","San Luis","Provincia de San Luis"
"Provincia",-30.7088227091528,-60.9506872769706,"IGN","82","AR-S","Santa Fe","Santa Fe","Provincia de Santa Fe"
"Provincia",-29.6849372775783,-67.1817575814487,"IGN","46","AR-F","La Rioja","La Rioja","Provincia de La Rioja"
"Provincia",-27.3359537960762,-66.9478972451295,"IGN","10","AR-K","Catamarca","Catamarca","Provincia de Catamarca"
"Provincia",-26.948283501723,-65.3647655803683,"IGN","90","AR-T","Tucumán","Tucumán","Provincia de Tucumán"
"Provincia",-26.3869871835867,-60.765116260356,"IGN","22","AR-H","Chaco","Chaco","Provincia del Chaco"
"Provincia",-24.8950871761481,-59.9321901121647,"IGN","34","AR-P","Formosa","Formosa","Provincia de Formosa"
```

---

# Configuración Inicial con Docker



# Archivo .env.db

```
#Definimos cada variable
DATABASE_HOST=db
DATABASE_PORT=5432
DATABASE_NAME=postgres
DATABASE_USER=postgres
DATABASE_PASSWORD=postgres
POSTGRES_INITDB_ARGS="--auth-host=scram-sha-256 --auth-local=trust"
# Configuración para inicializar postgres
POSTGRES_PASSWORD=${DATABASE_PASSWORD}
PGUSER=${DATABASE_USER}
# Configuración para inicializar pgadmin
PGADMIN_DEFAULT_EMAIL=postgres@postgresql.com
PGADMIN_DEFAULT_PASSWORD=${DATABASE_PASSWORD}
# Configuración para inicializar superset
SUPERSET_SECRET_KEY=your_secret_key_here
```

# Archivo docker-compose.yml

```
networks:
  net:
    external: false

volumes:
  postgres-db:
    external: false

services:
  db:
    image: postgres:alpine
    env_file:
      - .env.db
    restart: unless-stopped
    environment:
      - POSTGRES_INITDB_ARGS=--auth-host=md5 --auth-local=trust
    healthcheck:
      # Prueba de salud para el contenedor
      test: [ "CMD-SHELL", "pg_isready" ]
      interval: 10s
      timeout: 2s
      retries: 5
    ports:
      - 5432:5432
    volumes:
      - postgres-db:/var/lib/postgresql/data
      - ./scripts:/docker-entrypoint-initdb.d # Scripts que se ejecutan solamente al iniciar por primera vez el servidor
      - ./datos:/datos
    networks:
      - net
```

# Archivo docker-compose.yml

```
pgadmin:
  image: dpage/pgadmin4
  restart: unless-stopped
  env_file:
    - .env.db
  ports:
    - 5050:80
  depends_on:
    db:
      condition: service_healthy
  networks:
    - net
```



# Archivo docker-compose.yml

```
superset:
  image: apache/superset:4.0.0
  restart: unless-stopped
  env_file:
    - .env.db
  ports:
    - 8088:8088
  depends_on:
    db:
      condition: service_healthy # Aguarda hasta que la base de datos este funcionando correctamente
  networks:
    - net
```

# Archivo init.sh

```
#!/bin/bash
echo "Inicializamos el usuario de superset"
docker compose exec -it superset superset fab create-admin \
    --username admin \
    --firstname Superset \
    --lastname Admin \
    --email admin@superset.com \
    --password admin
echo "Migramos la base de datos"
docker compose exec -it superset superset db upgrade
echo "Seteamos los Roles"
docker compose exec -it superset superset init
```

# Armado del Script



```
CREATE TEMPORARY TABLE temp_nacimiento (  
    provincia_id TEXT,  
    provincia_nombre VARCHAR,  
    departamento_id TEXT,  
    departamento_nombre VARCHAR,  
    anio INTEGER,  
    nacimientos_cantidad INTEGER,  
    poblacion_total INTEGER,  
    tbn FLOAT  
);
```

```
CREATE TEMPORARY TABLE temp_departamentos (  
    categoria VARCHAR,  
    centroide_lat FLOAT,  
    centroide_lon FLOAT,  
    fuente VARCHAR,  
    id VARCHAR,  
    nombre VARCHAR,  
    nombre_completo VARCHAR,  
    provincia_id VARCHAR,  
    provincia_interseccion FLOAT,  
    provincia_nombre VARCHAR  
);
```

```
CREATE TEMPORARY TABLE provincias_temp (  
    categoria VARCHAR,  
    centroide_lat FLOAT,  
    centroide_lon FLOAT,  
    fuente VARCHAR,  
    id VARCHAR,  
    iso_id VARCHAR,  
    iso_nombre VARCHAR,  
    nombre VARCHAR,  
    nombre_completo VARCHAR  
);
```

```
CREATE TABLE public.nacimiento (  
    id serial,  
    departamento_id INTEGER,  
    anio INTEGER,  
    cantidad_nacimientos INTEGER,  
    poblacion_total INTEGER,  
    tbn FLOAT  
);
```

```
CREATE TABLE public.departamento (  
    id BIGINT,  
    nombre VARCHAR,  
    nombre_completo VARCHAR,  
    centroide_lat FLOAT,  
    centroide_lon FLOAT,  
    categoria VARCHAR,  
    provincia_id BIGINT  
);
```

```
CREATE TABLE public.provincia (  
    id BIGINT,  
    nombre VARCHAR,  
    nombre_completo VARCHAR,  
    centroide_lat FLOAT,  
    centroide_lon FLOAT,  
    categoria VARCHAR  
);
```

```
/*
Agrego las restricciones de clave primaria y foránea a las tablas
*/
ALTER TABLE public.nacimiento
ADD CONSTRAINT pk_nacimiento PRIMARY KEY (id);

ALTER TABLE public.provincia
ADD CONSTRAINT pk_provincia PRIMARY KEY (id);

ALTER TABLE public.departamento
ADD CONSTRAINT pk_departamento PRIMARY KEY (id);

ALTER TABLE public.departamento
ADD CONSTRAINT fk_departamento_provincia FOREIGN KEY (provincia_id) REFERENCES provincia (id);

ALTER TABLE public.nacimiento
ADD CONSTRAINT fk_nacimiento_departamento FOREIGN KEY (departamento_id) REFERENCES departamento (id);
```

```
COPY provincias_temp
FROM '/datos/provincias.csv' DELIMITER ',' CSV HEADER;
```

```
INSERT INTO
    public.provincia (
        id,
        nombre,
        nombre_completo,
        centroide_lat,
        centroide_lon,
        categoria
    )
SELECT
    id::INTEGER,
    nombre,
    nombre_completo,
    centroide_lat,
    centroide_lon,
    categoria
FROM provincias_temp;
```

```
COPY temp_departamentos
FROM '/datos/departamentos.csv' DELIMITER ',' CSV HEADER;
```

```
INSERT INTO
    public.departamento (
        id,
        nombre,
        nombre_completo,
        centroide_lat,
        centroide_lon,
        categoria,
        provincia_id
    )
SELECT
    id::INTEGER,
    nombre,
    nombre_completo,
    centroide_lat,
    centroide_lon,
    categoria,
    provincia_id::INTEGER
FROM temp_departamentos;
```

```
INSERT INTO
    public.nacimiento (
        departamento_id,
        anio,
        cantidad_nacimientos,
        poblacion_total,
        tbn
    )
SELECT
    CASE
        WHEN departamento_id = 'NA'
        AND provincia_id = 'NA' THEN 9998 /* Información de la provincia "La Rioja" */
        WHEN departamento_id = 'NA'
        AND provincia_id::INTEGER = 2 THEN 9999 /* Información de la provincia "Buenos Aires" */
        ELSE departamento_id::INTEGER
    END AS departamento_id,
    anio,
    nacimientos_cantidad,
    poblacion_total,
    tbn
FROM temp_nacimiento;
```



# Casos Excepcionales



```

INSERT INTO
  public.departamento (id, nombre, provincia_id)
VALUES (
  9999,
  'Ciudad Autónoma de Buenos Aires',
  2
),
(
  9998,
  'No definido en La Rioja',
  46
);

```

```

1  "provincia_id","provincia_nombre","departamento_id","departamento_nombre","anio","nacimientos_cantidad","poblacion_total",
701  2,"Ciudad Autónoma de Buenos Aires",NA,"Ciudad Autónoma de Buenos Aires",2013,44892,3044076,14.75
702  2,"Ciudad Autónoma de Buenos Aires",NA,"Ciudad Autónoma de Buenos Aires",2014,44892,3044076,14.75

```

```

1  "provincia_id","provincia_nombre","departamento_id","departamento_nombre","anio","nacimientos_cantidad","poblacion_total",
312  NA,"La Rioja",NA,NA,2012,159,9927,16.02

```

```
SELECT
CASE
    WHEN departamento_id = 'NA'
    AND provincia_id = 'NA' THEN 9998 /* Información de la provincia "La Rioja" */
    WHEN departamento_id = 'NA'
    AND provincia_id::INTEGER = 2 THEN 9999 /* Información de la provincia "Buenos Aires" */
    ELSE departamento_id::INTEGER
END AS departamento_id,
anio.
```

# Esquema Public Formas Normales



## ESQUEMA PUBLIC

```
▷ Run | ◻ Select  
CREATE TABLE public.nacimiento (  
    id serial,  
    departamento_id INTEGER,  
    anio INTEGER,  
    cantidad_nacimientos INTEGER,  
    poblacion_total INTEGER,  
    tbn FLOAT  
);
```

# FORMAS NORMALES

Descartamos algunas propiedades del DataSet para poder cumplir con las formas normales.

"nombre\_departamento" → "id\_departamento" → "nacimiento"

~~"nombre\_departamento"~~ → **Tabla aparte**

"provincia" → "departamento" → "nacimiento"

~~"provincia"~~ → **Tabla departamento**

# Ejecución del Script



```
$ docker compose up -d
```

```
[+] Running 5/5
```

✓ Network postgres-etl_net	Created
✓ Volume "postgres-etl_postgres-db"	Created
✓ Container postgres-etl-db-1	Healthy
✓ Container postgres-etl-superset-1	Started
✓ Container postgres-etl-pgadmin-1	Started



- `$ cat scripts/000_initdb.sql | docker exec -i postgres-etl-db-1 psql`  
`DROP TABLE`  
`DROP TABLE`  
`DROP TABLE`  
`CREATE TABLE`  
`CREATE TABLE`  
`CREATE TABLE`  
`ALTER TABLE`


```
$ . init.sh
Inicializamos el usuario de superset
logging was configured successfully
2025-05-15 23:03:06,126:INFO:superset.utils.logging_configurator:logging was configured successfully
```

```
PS C:\Users\Usuario> docker inspect postgres-etl-db-1
```

```
"IPAddress": "172.19.0.2",
```

localhost:8088/login/

Importar favoritos | Escribe el siguiente...

Superset

Sign In

Enter your login and password below:

USERNAME:

admin

PASSWORD:

admin

SIGN IN



## Settings

## Security

## List Users

## List Roles

## Action Log

## Row Level Security

## Data

## Database Connections

X

STEP 1 OF 3

### Select a database to connect



## PostgreSQL



Presto



MySQL



## SQLite

Or choose from a list of other databases we support:

## Connect a database



STEP 2 OF 3

### Enter the required PostgreSQL credentials

Need help? [Learn more about connecting to PostgreSQL..](#)

HOST \* 

PORT \*

DATABASE NAME \*

Copy the name of the database you are trying to connect to.

USERNAME \*

PASSWORD



DISPLAY NAME \*

BACK

CONNECT


## Databases

NAME	EXPOSE IN SQL LAB	AQE	MODIFIED BY
<input type="text" value="Type a value"/>	<input type="text" value="Select or type a value"/>	<input type="text" value="Select or type a value"/>	<input type="text" value="Select or type a value"/>

Name ▾	Backend	AQE ▾	DML ▾	CSV upload ▾	Expose in SQL Lab ▾	Last modified ▾
PostgreSQL	postgresql	×	×	×	✓	now

SQL ▾

- SQL Lab
- Saved Queries
- Query History

 **Superset**

[Dashboards](#) [Charts](#) [Datasets](#) [SQL ▾](#)

⋮ Untitled Query 1 ● ×

+

DATABASE

postgresql PostgreSQL ▾

SCHEMA

public ▾ ↺

SEE TABLE SCHEMA

Select table or type to search tables ▾ ↺

! 1

SELECT ...

RUN

LIMIT: 1 000 ▾

SAVE ▾

RESULTS

QUERY HISTORY

# Queries y Resultados





```

1 SELECT anio AS "Año", cantidad_nacimientos as "Cantidad Nacimiento", departamento.nombre AS "Departamento"
2 FROM nacimiento
3 INNER JOIN departamento ON (nacimiento.departamento_id = departamento.id)
4 INNER JOIN provincia ON (departamento.provincia_id = provincia.id)
5 WHERE provincia.nombre = 'Córdoba' AND departamento.nombre = 'General San Martín'
6 ORDER BY nacimiento.cantidad_nacimientos DESC
7 LIMIT 3
8 ;|

```

```
SELECT anio AS "Año", cantidad_nacimientos as "Cantidad Nacimiento", departamento.nombre AS "Departame...
```

Año ↕	Cantidad Nacimiento ↕	Departamento ↕
2014	2497	General San Martín
2016	2446	General San Martín
2013	2402	General San Martín

Años con mayor cantidad de nacimientos en el departamento General San Martín

```
1 SELECT departamento.nombre AS "Departamento", cantidad_nacimientos as "Cantidad Nacimiento", anio AS "Año"
2 FROM nacimiento
3 INNER JOIN departamento ON (nacimiento.departamento_id = departamento.id)
4 WHERE anio = 2022
5 ORDER BY nacimiento.cantidad_nacimientos DESC
6 LIMIT 3
7 ;|
```

```
SELECT departamento.nombre AS "Departamento", cantidad_nacimientos as "Cantidad Nacimiento", anio AS "Año"
```

Departamento ⚡	Cantidad Nacimiento ⚡	Año ⚡
Ciudad Autónoma de Buenos Aires	23397	2022
La Matanza	15729	2022
Rosario	14896	2022

Departamentos con mayor cantidad de nacimientos en el año 2022

```
1 SELECT departamento.nombre AS "Departamento", cantidad_nacimientos as "Cantidad Nacimiento", anio AS "Año"
2 FROM nacimiento
3 INNER JOIN departamento ON (nacimiento.departamento_id = departamento.id)
4 WHERE anio = 2022
5 ORDER BY nacimiento.cantidad_nacimientos
6 LIMIT 3
7 ;|
```

```
SELECT departamento.nombre AS "Departamento", cantidad_nacimientos as "Cantidad Nacimiento", anio AS "...
```

Departamento ↕	Cantidad Nacimiento ↕	Año ↕
Limay Mahuida	0	2022
Lihuel Calel	1	2022
Mártires	2	2022

Departamentos con menor cantidad de nacimientos en el año 2022

```
1 SELECT provincia.nombre AS "Provincia", SUM(nacimiento.cantidad_nacimientos) AS "Total Nacimientos 2012-2022"
2 FROM public.nacimiento
3 INNER JOIN public.departamento ON (nacimiento.departamento_id = departamento.id)
4 INNER JOIN public.provincia ON (departamento.provincia_id = provincia.id)
5 GROUP BY provincia.nombre
6 ORDER BY "Total Nacimientos 2012-2022" DESC
7 ;
```

```
SELECT provincia.nombre AS "Provincia", SUM(nacimiento.cantidad_nacimientos) AS "Total Nacimientos 2012-2022"
```

Provincia	Total Nacimientos 2012-2022
Buenos Aires	2713829
Córdoba	577149
Santa Fe	548428
Ciudad Autónoma de Buenos Aires	402492



Provincias con mayor cantidad de nacimientos entre los años 2012-2022

```

1 SELECT provincia.nombre AS "Provincia", SUM(nacimiento.cantidad_nacimientos) AS "Total Nacimientos 2012-2022"
2 FROM public.nacimiento
3 INNER JOIN public.departamento ON (nacimiento.departamento_id = departamento.id)
4 INNER JOIN public.provincia ON (departamento.provincia_id = provincia.id)
5 GROUP BY provincia.nombre
6 ORDER BY "Total Nacimientos 2012-2022"
7 ;

```

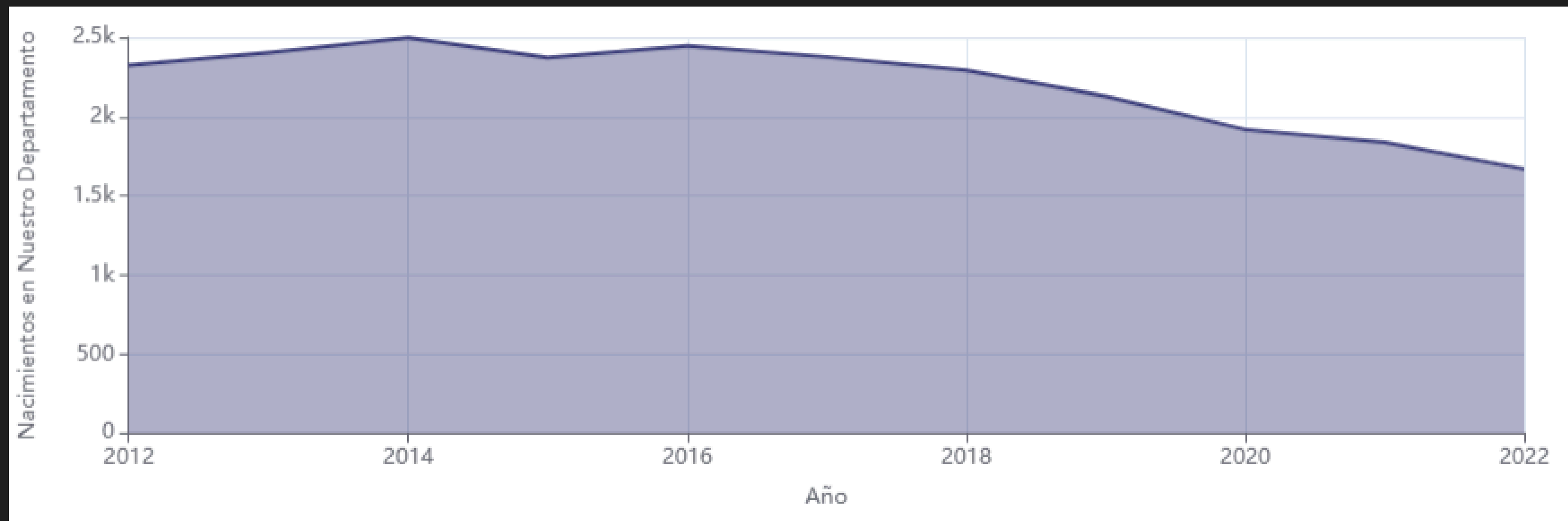
```
SELECT provincia.nombre AS "Provincia", SUM(nacimiento.cantidad_nacimientos) AS "Total Nacimientos 2012-2022"
```

Provincia 	Total Nacimientos 2012-2022 
Tierra del Fuego, Antártida e Islas del Atlántico Sur	27214
La Pampa	51902
Santa Cruz	56864
La Rioja	61520

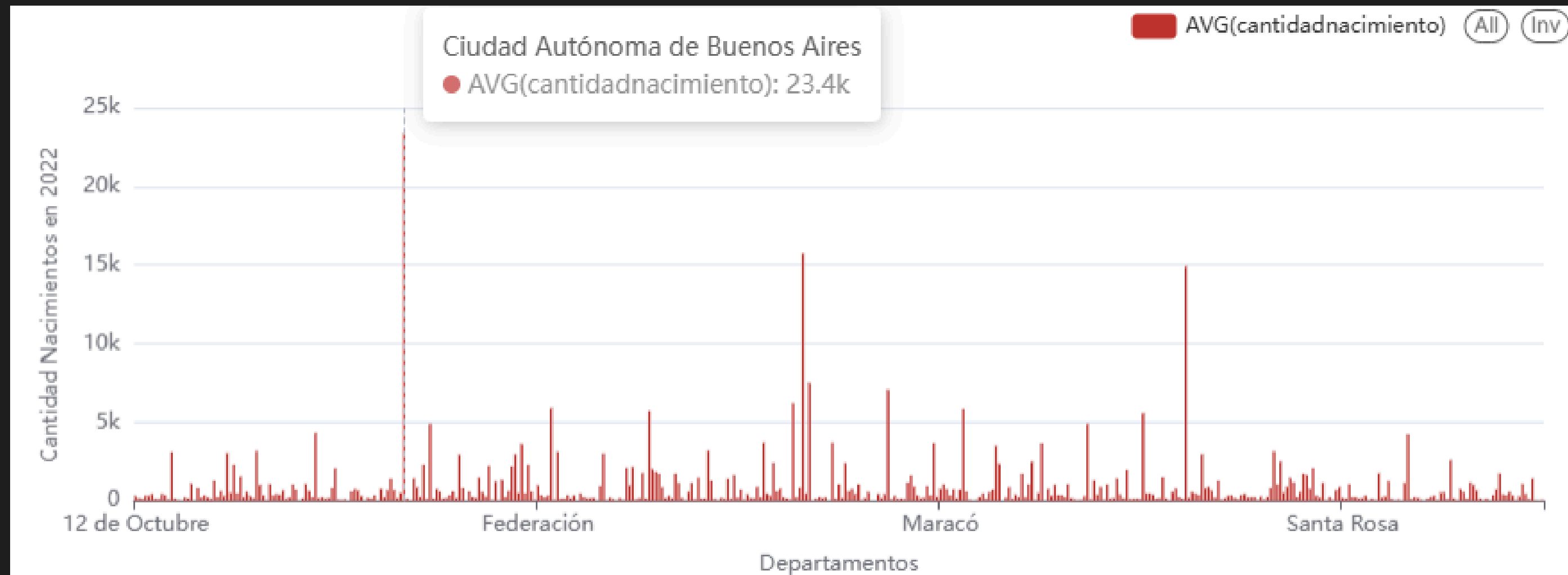
Provincias con menor cantidad de nacimientos entre los años 2012-2022

# Gráficos con Superset



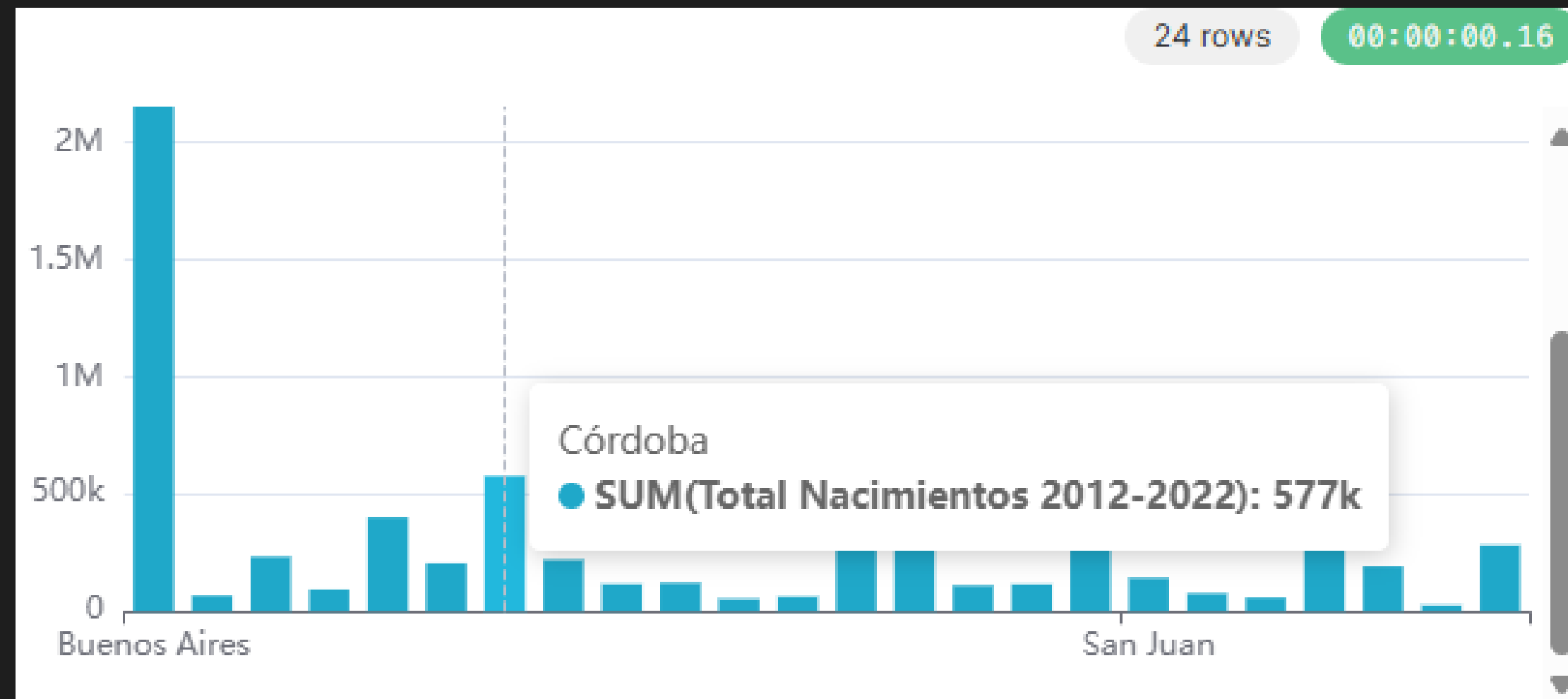


Cantidad de nacimientos por año en el departamento General San Martín



Cantidad de nacimientos por departamento en el año 2022





Cantidad de nacimientos por provincia entre los años 2012-2022