

# Laboratorio de Datos

## Trabajo Práctico 01:

### Educación y empleo

---

23 de Octubre de 2025

#### Grupo 1

Integrante	LU	Correo electrónico
Hildt, Lautaro	1166/24	<a href="mailto:lautarohildt43@gmail.com">lautarohildt43@gmail.com</a>
Fisanotti, Juan Bautista	1146/24	<a href="mailto:bautifisanotti@gmail.com">bautifisanotti@gmail.com</a>
Smirnoff, Manuel	144/24	<a href="mailto:manumax04@gmail.com">manumax04@gmail.com</a>



#### Facultad de Ciencias Exactas y Naturales

Universidad de Buenos Aires

Ciudad Universitaria - (Pabellón I/Planta Baja)

Intendente Güiraldes 2610 - C1428EGA

Ciudad Autónoma de Buenos Aires - Rep. Argentina

Tel/Fax: (++54 +11) 4576-3300

<http://www.exactas.uba.ar>

→ **Resumen:**

En este trabajo nos centramos en analizar la relación entre la existencia de instituciones educativas y el desarrollo de actividades productivas en los departamentos de Argentina, utilizando los datos oficiales del año 2022. La problemática que nos planteamos fue ver si existe una relación entre la cantidad de establecimientos educativos y el número de personas empleadas en cada departamento, también ver si hay más personas empleadas en departamentos con mayor cantidad de población. También quisimos identificar los departamentos con mayor proporción de empleo respecto a su población total y calcular si efectivamente en zonas donde hay mayor población de gente joven hay mayor cantidad de establecimientos educativos.

Para poder cumplir estos objetivos construimos un modelo de datos relacional que integra las tres fuentes principales, Población, Establecimientos Educativos Y Establecimientos Productivos. Este proceso requirió de una limpieza y normalización de las tablas ya que cada fuente usaba distintos formatos para referirse al mismo departamento. Una vez que juntamos los identificadores departamentales y tuvimos el mismo para todos, diseñamos tablas en Tercera Normal que permitieron organizar la información de manera consistente y facilitar los cruces necesarios, también definimos categorías de población por rangos etarios que creímos relevantes para el análisis educativo.

A partir de las consultas realizadas, los resultados que obtuvimos muestran que los departamentos con mayor población tienden a tener también un mayor número de empleos aunque la intensidad de esto varía según la provincia y la región, ya que pudimos ver una clara diferencia entre la provincia de Buenos Aires, Santa Fe y Córdoba con las demás provincias, evidenciando la desigualdad de territorio y población. Por otro lado la distribución de establecimientos educativos por departamentos en cada provincia muestra una gran heterogeneidad, mientras que en algunas provincias hay una oferta de empleo más equilibrada, en otras se pueden ver fuertes dispersidades entre departamentos.

→ **Introducción**

**Problemática a resolver:**

Definir si, tras un análisis y manipulación de datos correspondiente, es posible encontrar o establecer una relación entre el número de Establecimientos Educativos y la cantidad de puestos de trabajo en las provincias del país. Con este objetivo tendremos en cuenta la distribución o densidad tanto de la población como de los establecimientos nombrados anteriormente.

**Objetivo general:**

Llevar a cabo un posible diseño de base de datos, un procesamiento de nuestros datos originales para construir el modelo diseñado, y un posterior análisis de nuestro modelo de datos que nos ayude a determinar la existencia o no de la relación buscada.

### Actividades a realizar:

Nuestra “hoja de ruta” para llevar a cabo nuestro objetivo consistió en:

- Descargar y hacer un primer análisis a nuestras fuentes de datos originales mencionadas anteriormente: analizar estructuras, información disponible y posibles conclusiones a partir de esta.
- Diseñar un DER (Diagrama de Entidad-Relación) que represente cómo va a verse nuestro modelo de datos, y cómo nuestras entidades (junto con sus atributos) van a relacionarse dentro de este.
- A partir del DER, diseñar un Modelo Relacional que lo represente correctamente.
- Realizar un análisis de calidad de nuestras bases de datos originales, utilizando el método GQM.
- Llevar a cabo una limpieza a nuestros datos originales, obteniendo así nuestras entidades del modelo relacional junto con sus atributos.
- Desarrollar el análisis de nuestros datos limpios a través de consultas SQL, y visualizaciones de la información obtenida. Evaluar la conclusión obtenida a través de nuestro desarrollo.

### Acerca de este informe:

A continuación en nuestro documento, desarrollaremos acerca de:

- El procesamiento llevado a cabo sobre las bases de datos originales
- Las decisiones tomadas durante la ejecución de nuestra hoja de ruta
- El análisis sobre nuestro modelo de datos
- La conclusión obtenida a partir de este proceso.

#### → Procesamiento de datos

##### Formas normales de nuestras fuentes de datos

Tras analizar los datos originales, podemos ver que una tabla se encontraban en la Primera Forma Normal (1FN), ya que todos sus atributos contienen valores atómicos sin repetición, pero la otra no, ya que tenía atributos que no eran atómicos .

**Establecimientos educativos:** esta tabla no está en 1FN, ya que Mail no es un atributo atómico. Hay registros que tienen más de un Mail en esa columna.

**Establecimientos Productivos:** esta tabla está en 1FN ya que todos sus atributos son atómicos pero no está en 2FN ya que si tomamos como PK los atributos in\_departamentos y clae6, por ejemplo el atributo departamento está relacionado sólo con un elemento de la PK que sería in\_departamentos, no con toda la PK.

##### Calidad de Datos

###### • Establecimientos Educativos

Comenzamos con la base de datos de Establecimientos Educativos. En un primer vistazo, logramos identificar que: para el atributo “Mail”, contamos con numerosos registros que contienen más de una dirección de mail para un único establecimiento, delimitados por caracteres como “,”, “;” o “/”. De este modo, estamos violando el principio de atomicidad de los datos, es decir, nuestros datos no están compuestos por una única unidad atómica e indivisible de información, sino por varias piezas de información.

Pero este atributo no es un atributo que nos interese para el análisis que queremos hacer. Por lo tanto hicimos foco en atributos que nos interesen para el análisis como los códigos de las localidades. De estos códigos a nosotros nos interesan los primeros 5 caracteres ya que son los que coinciden con el código del Departamento al que corresponden y es lo que nos permite relacionar esta tabla con las demás.

El primer problema con el que nos encontramos es que estaban como tipo de dato numérico, por lo tanto si un código comenzaba con el carácter cero, ese código tenía un carácter menos que los demás. Esto nos pasó con los códigos de la provincia de Buenos Aires que son los que comienzan con ese carácter. El segundo problema con el que nos encontramos fue que para los departamentos de la Ciudad de Buenos Aires los primeros 5 dígitos estaban mal, por lo tanto no correspondían con los id de los departamentos de la Ciudad de Buenos Aires de las otras tablas.

Los atributos de la calidad de esta base de datos afectados serían:

- **Corrección:** no cumple este atributo ya que no representa la situación real al tener códigos de departamento incorrectos, generando no solo un fallo al distribuir la información, sino también un problema de comparación de datos, para cualquier sistema que dependa de estos códigos.
- **Confiabilidad:** si los errores son frecuentes, se pone en juego la confianza de cualquier análisis que se haga en la base de datos.

Planteamos entonces el análisis GQM sobre el atributo "Código de Localidad" de la base de datos de Establecimientos Educativos.

### Análisis GQM

- **Goal:** el atributo "Código de Localidad" de la base de datos original de establecimientos educativos debería ser correcto.
- **Question:** ¿Qué proporción de los registros del atributo "Código de Localidad" contienen datos erróneos?
- **Metric - Definimos:**

*Ntotal : cantidad de registros del atributo "Código de Localidad"*

*Nerróneos : cantidad de registros del atributo "Código de Localidad" con valores erróneos.*

*%erróneos = (Nerróneos/Ntotal) \* 100 : porcentaje de registros con datos erróneos*

*Ntotal = 63662*

*Nerróneos = 2776*

*%erróneos = (2776 / 63662) = 4.36%*

Podemos ver de este modo que el 4.36% de los registros de Código de Localidad presentan datos erróneos, por lo que impiden el uso instantáneo del dato para su manipulación y brindan inconsistencia sobre lo que representan. Podría tratarse de un problema de actualización de código a medida que pasó el tiempo.

La mejor solución a este problema sería actualizar los códigos de departamento de las comunas 1 a 15 con sus respectivos valores verdaderos, para evitar problemas de comparación e información incorrecta.

- **Establecimientos Productivos:**

Para la base de datos de Establecimientos Productivos, podemos notar que en muchas ocasiones la cantidad de expresas exportadoras representada supera a la cantidad de establecimientos, lo que es una contradicción.

Podemos decir que este fallo en la calidad afecta al atributo de **Consistencia** al tener muchas filas una contradicción lógica, dado que la cantidad de establecimientos totales no puede ser superada por un subconjunto de estos establecimientos. Es un problema de Instancia, porque los datos no tienen la precisión necesaria.

Por otro lado afecta al atributo de **Confiabilidad** ya que si la cantidad de contradicciones es grande no refleja la realidad y cualquier análisis que se haga con estos datos.

### Análisis GQM

- **Goal:** La cantidad de establecimientos productivos debería ser un número mayor o igual a la cantidad de empresas exportadoras.
- **Question:** ¿Cuántas contradicciones entre la cantidad de establecimientos y la cantidad de empresas exportadoras se presentan?
- **Metric** - Definimos:

*Ntotal : cantidad de registros de los atributos establecimientos y cantidad de empresas exportadoras*

*Ncontradicciones : cantidad de registros contradictorios entre estos dos atributos*

$\%contradicciones = (Ncontradicciones/Ntotal) * 100$  : porcentaje de contradicciones

$\%contradicciones = (15543 / 316697) * 100 = 4.9\%$

De este modo podemos ver que el 4,9% de los registros de la tabla presenta contradicciones en la lógica de cantidad de establecimientos y cantidad de empresas exportadora. Por lo tanto generan un problema de consistencia y confiabilidad hacia los datos.

Este problema puede solucionarse solo reemplazando esos datos por los datos que reflejan la realidad.

### Diagrama Entidad-Relación (DER)

En esta sección se presenta el Diagrama Entidad Relación (DER) diseñado a partir de las cuatro fuentes de datos previamente procesadas: establecimientos educativos, establecimientos productivos y padrón de población. A partir del análisis y la limpieza de datos, se identificaron las entidades relevantes, sus atributos, y las relaciones necesarias para representar adecuadamente la información. El diseño busca normalizar los datos y reflejar de forma precisa las asociaciones entre departamentos, establecimientos productivos, establecimientos educativos y población. A continuación, se describe brevemente cada entidad, sus claves primarias y relaciones, y se incluye el diagrama correspondiente.

#### Entidades:

- **Departamento:** Unidad geográfica y administrativa básica.

*Atributos:* Departamento\_Id, Departamento, Provincia

*Relación:* Cada departamento puede tener cero o muchos establecimientos educativos y productivos pero tiene si o si tiene una y una única población.

- **Establecimiento Productivo:** Unidad física donde se realizan actividades económicas para producir bienes o servicios.

*Atributos:* Clae6, Departamento\_Id, Departamento, Provincia, Genero, Cant\_Employados, Cant\_Establecimientos, Cant\_Empresas\_Exportadoras

*Relación:* Cada establecimiento productivo registra su actividad en uno y un único departamento y corresponde a una y una única actividad económica.

- **Establecimiento Educativo:** Institución que brinda enseñanza en una o más modalidades.

*Atributos:* CueAnexo, Nombre, Sector, Departamento\_Id, Departamento, Provincia, Nivel.

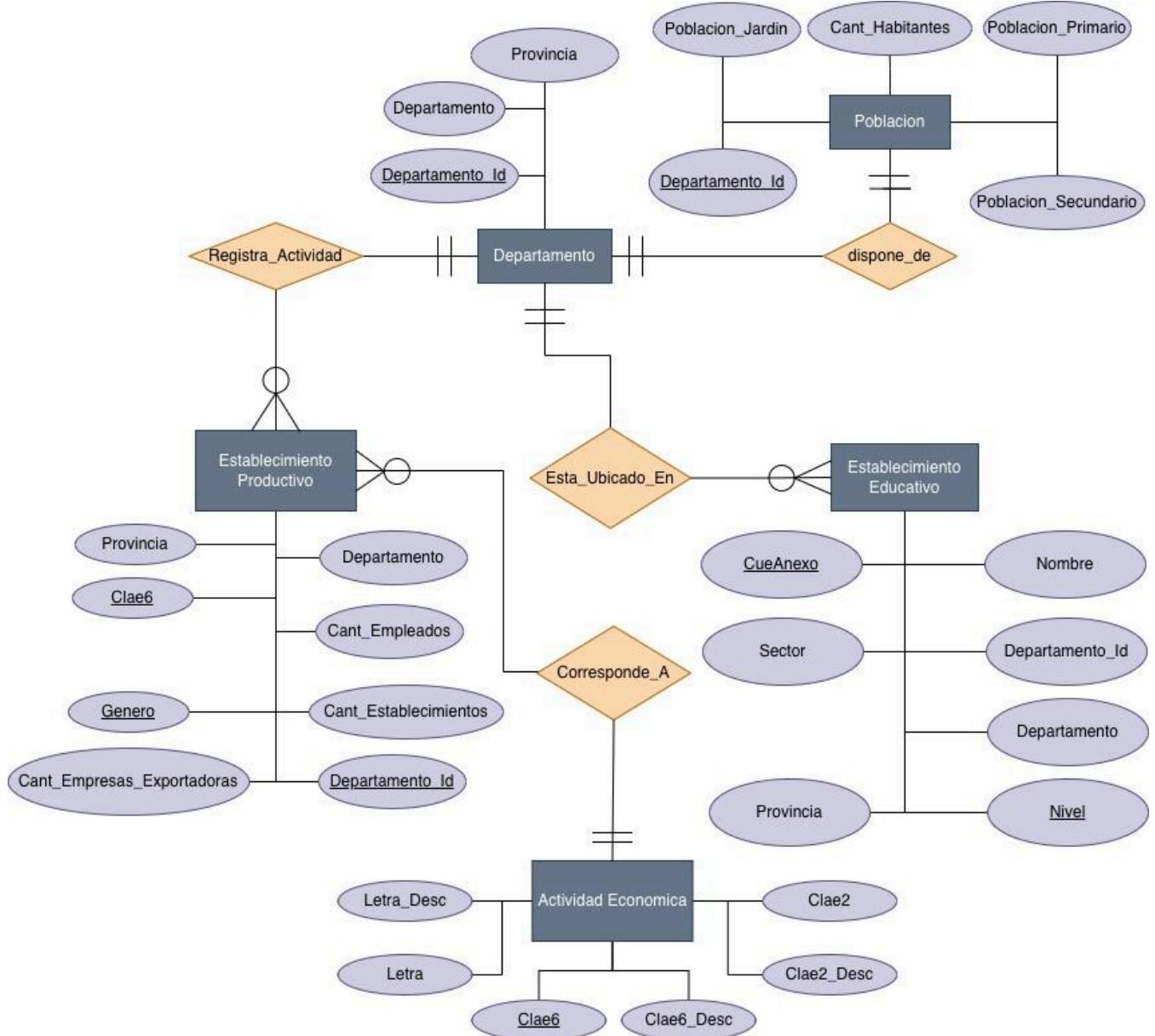
*Relación:* Cada establecimiento está ubicado en un único departamento. Puede estar vinculado a uno o más niveles educativos.

- **Actividad Económica:** Representa una actividad específica a la que se dedica una determinada empresa.  
*Atributos:* Clae6, Clae6\_Desc, Clae2, Clae2\_Desc, Letra, Letra\_Desc.  
*Relación:* Cada actividad económica corresponde a cero o muchas empresas.
  - **Población:** Representa la cantidad de personas que viven en cada departamento.  
*Atributos:* Departamento\_Id, Cant\_Habitantes, Poblacion\_Jardin, Poblacion\_Primario, Poblacion\_Secundario.  
*Relación:* Cada población vive en uno y un solo departamento.
- 

#### Relaciones:

- **Registra Actividad:** Cada establecimiento productivo registra su actividad en uno y sólo un departamento pero un departamento puede tener cero o muchas empresas registradas en el.
- **Dispone de:** Cada departamento dispone de una y sólo una población y una población dispone de uno y sólo un departamento.
- **Está ubicado en:** Cada establecimiento educativo está ubicado en uno y sólo un departamento pero un departamento puede ser ubicación de cero o muchos establecimientos educativos.
- **Corresponde a:** Un establecimiento productivo corresponde a una y sólo una actividad pero una actividad económica puede ser correspondida por cero o muchas empresas.

A continuación, nuestro **Diagrama de Entidad-Relación**:



Basándonos en el DER, construimos nuestro modelo relacional. En este, representamos las entidades y sus relaciones mediante tablas, de acuerdo al modelo de base de datos realizado. En cada relación contamos con *primary keys* (PK) y *foreign keys* (FK), vinculadas mutuamente. Sumado a esto tenemos *dependencias funcionales* (DF), las cuales describen las dependencias de los atributos con respecto a las claves. Al idear el modelo relacional, buscamos que todas las relaciones se encuentren en la **Tercera Forma Normal** (3FN). De este modo, no tendremos redundancias en la estructura. La única tabla que no pasamos a 3FN es la tabla de actividad económica ya que es solo para tener una referencia sobre la actividad económica de cada establecimiento productivo y por ese motivo no era necesario pasarlala a 3FN.

Las relaciones luego de pasarlas a la tercera forma normal quedaron de la siguiente manera:

- **Departamentos\_Info( Departamento\_Id, Departamento, Provincia, Poblacion\_Jardin, Poblacion\_Primario, Poblacion\_Secundario, Cant\_Total\_Habitantes )**

En esta tabla la PK es Departamento\_Id y no tiene FK ya que no depende de ninguna otra tabla.

- **Establecimientos\_Productivos( Clae6, Departamento\_Id, Genero, Cant\_Empleados, Cant\_Establecimientos, Cant\_Empresas\_Exportadoras )**

En esta tabla la PK es la clave compuesta por los atributos Clae6, Departamento\_Id y Genero. Y las FK serían Departamento\_Id y Clae6.

- **Establecimientos\_Educativos( Cueanexo, Establecimiento\_Nombre, Sector, Departamento\_Id )**

En esta tabla la PK es Cueanexo y la única FK es Departamento\_Id.

- **Niveles\_Educativos ( Cueanexo, Nivel )**

En esta tabla la PK es la compuesta por sus dos atributos Cueanexo y Nivel. Y a su vez la FK es Cueanexo.

- **Actividades\_Economicas( Clae6, Clae6\_Desc, Clae2, Clae2\_Desc, Letra, Letra\_Desc )**

En esta tabla Clae6 es PK y a la vez también es FK ya que relaciona esta tabla con la tabla de establecimientos productivos.

### Importación y limpieza de datos

Para generar los datasets o tablas asociadas al modelo relacional, desarrollamos distintos scripts de Python que se adapten a las distintas necesidades de limpieza de datos correspondientes a cada base de datos original. Procedimos de la siguiente forma:

- **Carga de datasets originales:** haciendo uso de la biblioteca Pandas, importamos los distintos archivos CSV y Excel correspondientes a las tablas de información originales (Establecimientos Educativos, Establecimientos Productivos, Población por departamento y Actividades Económicas).
- **Limpieza y normalización:** nos aseguramos de no tener elementos duplicados, filas incompletas y datos erróneos. Priorizamos crear dataframes provisarios para no modificar los de las tablas originales. Renombramos columnas para unificar nomenclaturas para PKs y FKs.

### Decisiones tomadas

Durante el proceso de limpieza de datos y generación de entidades y relaciones, tomamos las decisiones de:

- Eliminar columnas vacías e irrelevantes
- Corregir valores erróneos
- Los datos referenciados con la provincia como Ciudad de Buenos Aires los transformamos a Buenos Aires ya que CABA no es una provincia.

## Laboratorio de Datos - 2025

Grupo: “*Grupo 1*” - Hildt, Fisanotti, Smirnoff

- Para determinar cantidad de personas por nivel educativo, tomamos:
  - ◆ 3 a 5 años → Nivel inicial (Jardín)
  - ◆ 6 a 12 años → Nivel Primario
  - ◆ 13 a 17 años → Nivel Secundario

### → Análisis de Datos:

#### Reporte i: Cantidad de EE por nivel educativo y población por grupo etario por departamento.

Provincia	Departamento	Jardines	Población Jardín	Primarias	Población Primaria	Secundarias	Población Secundaria
Buenos Aires	La Matanza	325	87162	333	225872	335	151594
Buenos Aires	La Plata	215	29260	199	77998	206	54517
Buenos Aires	Lomas de Zamora	160	28806	178	76967	191	54431
Buenos Aires	General Pueyrredón	177	23100	169	62565	170	47412
Buenos Aires	Quilmes	160	26975	146	70881	153	50346
Buenos Aires	Almirante Brown	132	25188	137	67913	145	48665
Buenos Aires	Moreno	118	28521	134	76357	134	50667
Buenos Aires	Merlo	110	27105	120	71541	124	49991
Buenos Aires	Lanús	118	16028	116	44506	114	33067
Buenos Aires	Tigre	121	18454	111	51961	109	37885
Buenos Aires	Pilar	105	18864	109	49944	102	34708
Buenos Aires	Florencio Varela	97	24783	98	64455	102	43778
Buenos Aires	General San Martín	99	15897	96	44539	95	32905

Esta tabla se encuentra en la carpeta de reportes con el nombre de *reporte\_i*. Este es un recorte de la misma.

Podemos ver que: contamos con mayor cantidad de establecimientos de determinado nivel, mientras mayor cantidad de personas de su rango etario correspondiente se encuentren en su respectivo departamento.

#### Consulta ii: Cantidad de empleados totales por departamento para el año 2022

Provincia	Departamento	Cantidad total de empleados en 2022
Buenos Aires	Comuna 1	1.07108e+06
Buenos Aires	Comuna 4	303452
Buenos Aires	Comuna 14	263598
Buenos Aires	La Matanza	254636
Buenos Aires	Vicente López	238423
Buenos Aires	General Pueyrredón	227681
Buenos Aires	Comuna 13	215462
Buenos Aires	San Isidro	193094
Buenos Aires	Comuna 3	189615
Buenos Aires	Comuna 2	174547
Buenos Aires	Tigre	164487
Buenos Aires	Comuna 15	163905
Buenos Aires	General San Martín	159485
Buenos Aires	La Plata	158748
Buenos Aires	Quilmes	137894

## Laboratorio de Datos - 2025

Grupo: “*Grupo 1*” - Hildt, Fisanotti, Smirnoff

Esta tabla se encuentra en la carpeta de Reportes con el nombre reporte\_ii. Este es un recorte de la misma.

Notamos que los departamentos que tienen mayor cantidad de empleos en 2022 son de la provincia de Buenos Aires, lo cual dice mucho sobre lo centralizado que está nuestro país.

### **Consulta iii: Cantidad de empresas exportadoras que emplean mujeres, cantidad de EE y población total por departamento**

Departamento_Id	Provincia	Departamento	Cantidad de empresas exportadoras con mujeres	Cantidad de EE	Población Total
14014	Cordoba	Capital	2560	1400	1498060
82084	Santa Fe	Rosario	2836	1248	1337958
06427	Buenos Aires	La Matanza	1749	1186	1837168
06441	Buenos Aires	La Plata	447	832	756074
22140	Chaco	San Fernando	383	810	413764
82063	Santa Fe	La Capital	700	728	568259
06357	Buenos Aires	General Pueyrredón	1544	664	660569
06490	Buenos Aires	Lomas de Zamora	644	647	685644
30084	Entre Ríos	Paraná	544	591	388716
90084	Tucuman	Capital	585	586	587289
62042	Río Negro	General Roca	989	548	379627
06658	Buenos Aires	Quilmes	820	542	631774
06028	Buenos Aires	Almirante Brown	508	489	583209
58035	Neuquén	Confluencia	1074	452	467217
54028	Misiones	Capital	290	445	391498

Esta tabla se encuentra en la carpeta de Reportes con el nombre reporte\_iii. Este es un recorte de la misma.

Notamos que, los departamentos de provincias mas pobladas (Bs As, CABA, Cordoba) cuentan con mayor cantidad de establecimientos educativos y a su vez mayor cantidad de empresas exportadoras que emplean a mujeres

### **Consulta iv: Departamentos y (Clae3 que más empleo genera en ese departamento) con mayor cantidad de empleados que el promedio de la provincia.**

Nombre	Departamento	CLAE3	Cant_Employados
Buenos Aires	Almirante Brown	851	8376
Buenos Aires	Avellaneda	851	6483
Buenos Aires	Bahía Blanca	471	3810
Buenos Aires	Berazategui	851	6190
Buenos Aires	Comuna 1	641	81801
Buenos Aires	Comuna 10	801	8120
Buenos Aires	Comuna 11	410	4995
Buenos Aires	Comuna 12	620	8482
Buenos Aires	Comuna 13	869	12896
Buenos Aires	Comuna 14	511	12793
Buenos Aires	Comuna 15	812	5470
Buenos Aires	Comuna 2	869	12951

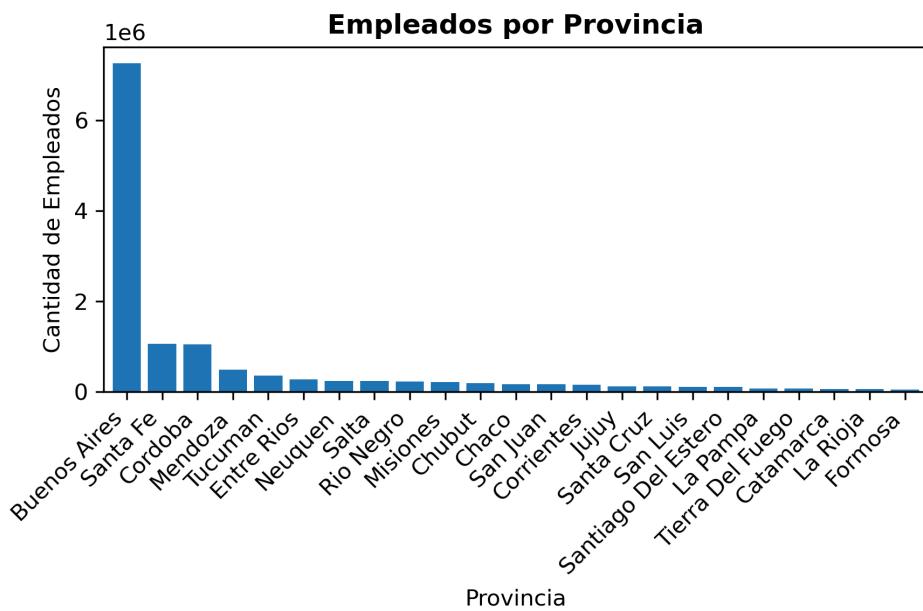
Esta tabla se encuentra en la carpeta de Reportes con el nombre reporte\_iv. Este es un recorte de la misma.

Con este análisis, notamos que la cantidad de departamentos que generan más empleos que el promedio son pocos en relación a la cantidad total de departamentos. Lo que nos deja ver que hay pocos departamentos que generan gran cantidad de empleo comparado con el resto.

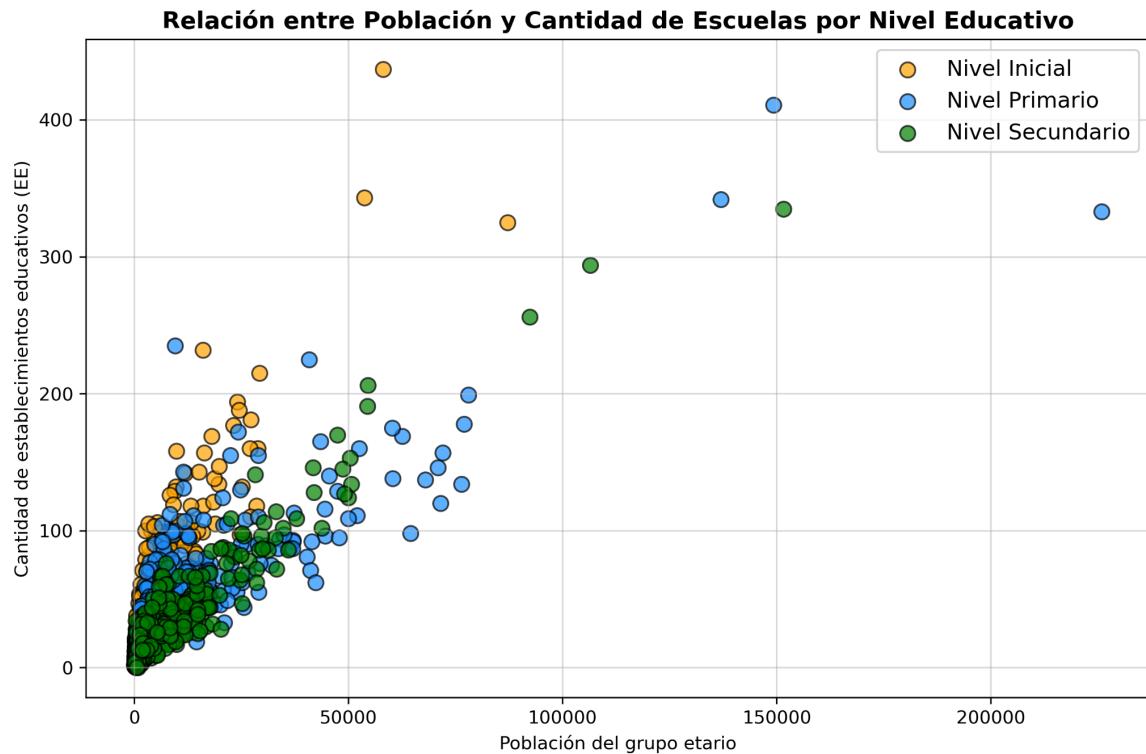
→ Análisis de Visualizaciones:

i) Cantidad de empleados por provincia.

El gráfico de barras muestra la cantidad total de empleados por provincia. Se observa que **Buenos Aires** lidera ampliamente, seguida por **Córdoba** y **Santa Fe**. Esta distribución parece correlacionarse con los niveles de urbanización y concentración poblacional. En contrapartida, las provincias del norte y sur del país presentan una menor cantidad de empleados, lo cual puede atribuirse a factores geográficos, demográficos o económicos que dificultan la creación o el sostenimiento de establecimientos productivos.

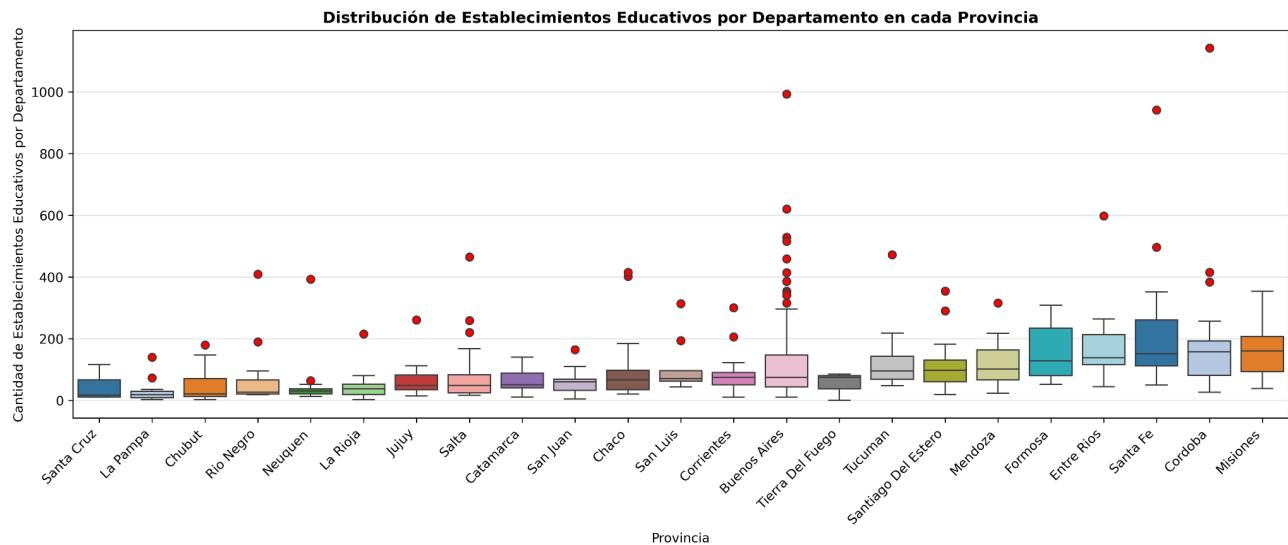


ii) Cantidad de EE de los departamentos en función de la población, separando por nivel educativo y su correspondiente grupo etario.



Este gráfico de dispersión representa la relación entre la población correspondiente a cada grupo etario y la cantidad de establecimientos educativos comunes por nivel (jardín de infantes, primario y secundario). Se observa una **correlación esperada**: cuanto mayor es la población en edad correspondiente, mayor es la cantidad de EE. Sin embargo, se identifican ciertos departamentos con mayor cantidad de EE que lo esperado, lo cual podría deberse a criterios de distribución territorial que priorizan el acceso en zonas rurales o alejadas, independientemente de la densidad poblacional.

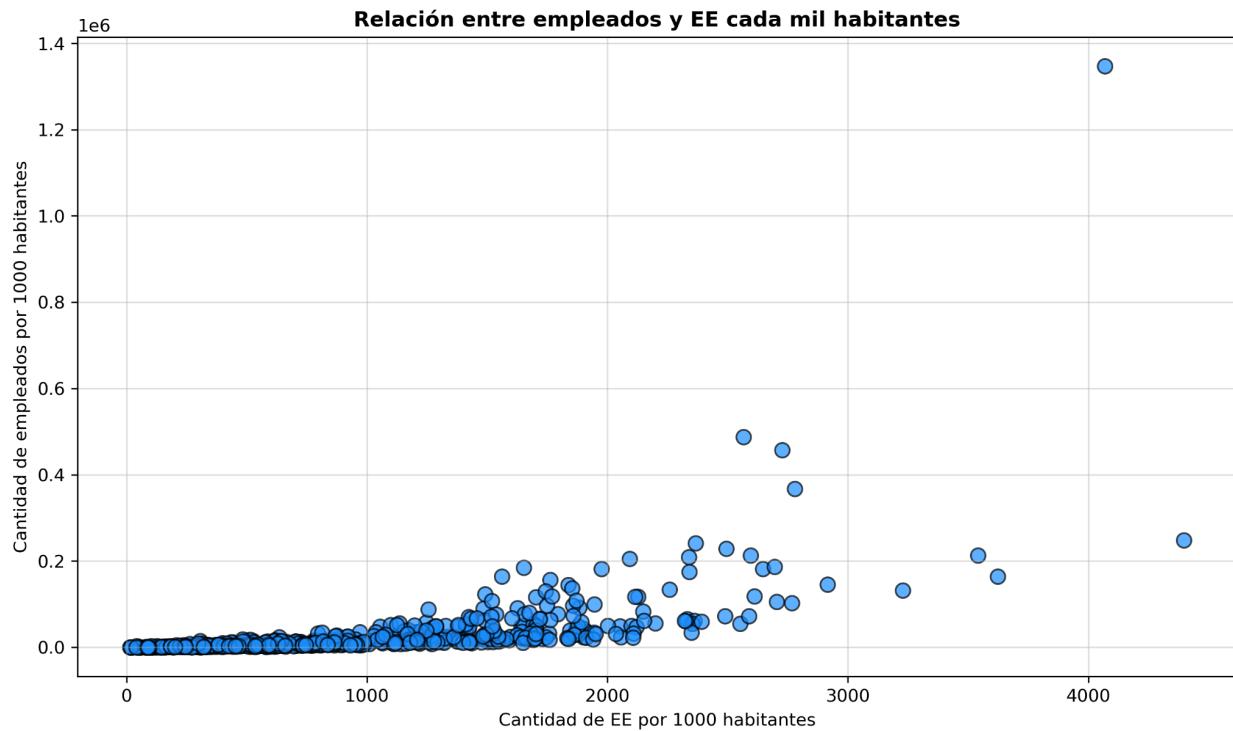
### iii) Distribución de EE por departamento en cada provincia.



El boxplot muestra la **variabilidad de EE entre departamentos** dentro de cada provincia. El ordenamiento por mediana permite observar cuáles provincias tienen mayor "centralidad" en su distribución. **Buenos Aires**, por ejemplo, presenta una dispersión muy amplia, lo cual refleja su gran heterogeneidad territorial. Por otro lado, provincias como **La Pampa** o **Formosa** muestran una distribución más homogénea, con menor diferencia entre departamentos. Esto puede estar relacionado con políticas educativas más uniformes o con una estructura territorial menos fragmentada.

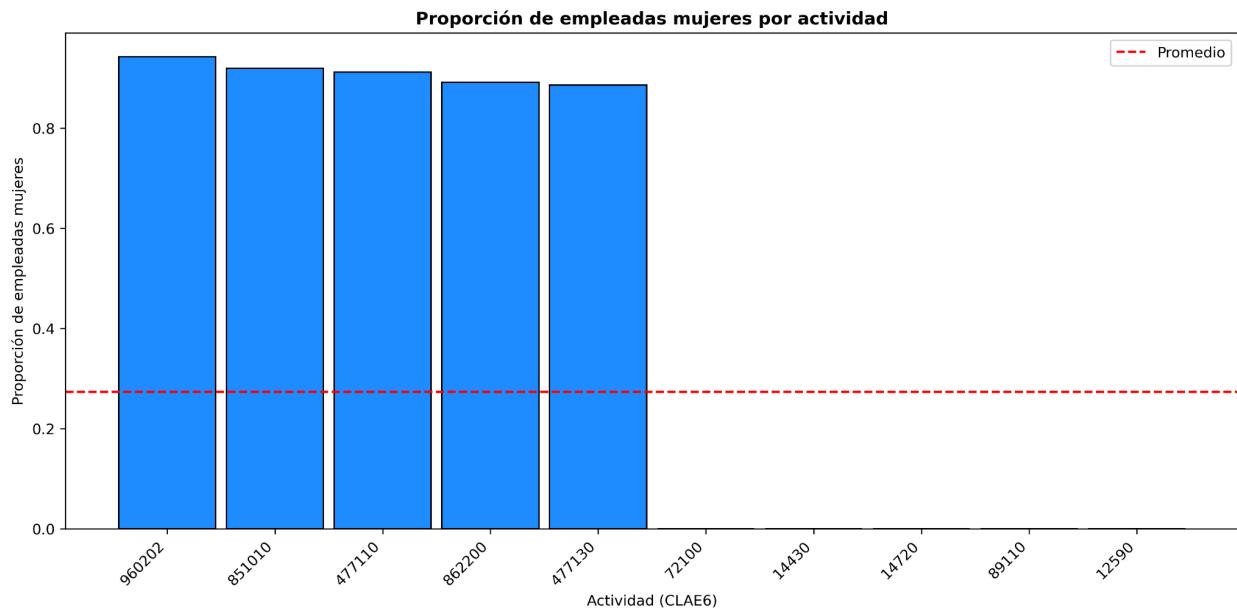
Relación entre la cantidad de empleados cada mil habitantes (para 2022) y de EE cada mil habitantes por departamento

**iv) Relación entre cantidad de empleados y EE por cada 1000 habitantes.**



Este gráfico explora la relación entre la cantidad de Establecimientos Educativos comunes y la cantidad de empleados en cada departamento **normalizada por población**. Se observa una relación débil entre cantidad de empleados y cantidad de establecimientos educativos por habitante. También podemos observar una gran concentración de departamentos con valores bajos en ambos atributos y algunos casos atípicos con alto nivel de empleo. Esto demuestra una disparidad en el desarrollo económico y educativo entre departamentos.

v) Las 5 actividades con mayor y menor proporción de empleadas mujeres.



Este grafico de barras muestra las 5 actividades economicas con mayor y menor participacion de mujeres con un linea de referencia que muestra el promedio general de empleadas femeninas. Lo que podemos ver es que las actividades con mayor participacion de mujeres suelen estar relacionadas con servicios sociales, educacion y salud, mientras que las de menor participacion corresponden a sectores industriales y extraccion de minerales en minas.

## Conclusiones

Nosotros nos propusimos explorar si existe una relación entre la cantidad de establecimientos educativos y la cantidad de empleados en las distintas provincias del país. A partir del análisis de las tablas generadas mediante consultas SQL y los gráficos construidos sobre ellas, pudimos obtener algunas conclusiones relevantes.

En primer lugar, observamos que la cantidad de empleos por provincia guarda una relación con la cantidad de habitantes, como se logra evidenciar en el gráfico "Cantidad empleados por provincia". Además, al analizar la relación entre EE y empleados cada mil habitantes, notamos que no existe una correspondencia lineal entre ambas variables. Incluso, en algunos casos, los departamentos con mayor cantidad de EE cada mil habitantes no cuentan con muchos empleados cada mil habitantes. Este dato resulta especialmente llamativo, ya que revela que no hay una relación clara entre las dos variables.

En conclusión, no se encontró una relación clara entre la cantidad de establecimientos educativos y la de empleos a nivel departamental. Ambos indicadores parecen variar de forma independiente entre ellos. En todo caso, podría afirmarse que la cantidad de EE está más influenciada por la estructura administrativa y urbana de cada provincia, como se aprecia en el gráfico "Cantidad de EE de los departamentos en función de la población, separando por nivel educativo y su correspondiente grupo etario".