



Informe:

“Análisis relacional de establecimientos educativos y establecimientos productivos de Argentina en el año 2022”

Materia: Laboratorio de Datos

Comisión: Cotik

Cuatrimestre: 2do. Cuatrimestre 2025

Grupo: NotNull

Integrantes:

- Alvarez, Valentina Jazmín
- Eker, Elif Sena
- Sanchez, Araceli

Entrega: 22-10-2025





Resumen

El presente informe aborda el objetivo de determinar el vínculo entre la oferta educativa y el desarrollo productivo en Argentina, utilizando datos públicos de Establecimientos Educativos, Establecimientos Productivos y Población del año 2022.

La metodología se centró en la superación de la fragmentación y los problemas de calidad de datos (ambigüedad, incompletitud y falta de atomicidad) presentes en las fuentes originales. Para ello, se diseñó un Modelo Relacional normalizado en Tercera Forma Normal (3FN), integrando las entidades Departamento, Personas, Establecimientos Educativos y Establecimientos Productivos. A partir de este modelo, se procesaron y limpian los datos, realizando agrupaciones y selecciones de atributos relevantes para el análisis. ✓

Los resultados del análisis, presentados en reportes SQL y gráficos, confirman una fuerte correlación positiva entre la cantidad de Establecimientos Educativos (EE) y la Población, así como entre la cantidad de EE y la Cantidad Total de Empleados por departamento. Se observó que departamentos con mayor empleo total (como La Matanza en Buenos Aires) también presentan la mayor cantidad de EE. No obstante, el análisis de EE y Empleados cada mil habitantes mostró una gran dispersión, con valores atípicos como CABA y Tierra del Fuego. Adicionalmente, se identificó una marcada brecha de género en el ámbito productivo, con actividades que concentran casi la totalidad del empleo femenino y otras que prácticamente lo excluyen.

En conclusión, el estudio establece que la oferta educativa se ajusta al tamaño poblacional y productivo de las regiones.





Introducción

En el presente informe abordamos como objetivo central, a partir de conocimientos adquiridos en la materia, conocer si existe un vínculo entre la educación y el desarrollo de actividades productivas en Argentina. Para ello, utilizamos fuentes de datos públicos nacionales sobre Establecimientos Educativos, Establecimientos Productivos y Población del año 2022.

La problemática principal a la cual nos enfrentamos es la fragmentación de la información que se necesita, donde las fuentes de datos poseen defectos en la calidad de datos y diseño. Por lo tanto, se realizó un trabajo de lectura y entendimiento, normalización, limpieza y adaptación de los datos para luego realizar el análisis necesario en función del objetivo del trabajo.

Como se mencionó anteriormente, existían problemas de calidad y diseño de las fuentes utilizadas, por consiguiente fueron tomadas diversas decisiones sobre qué datos conservar y cómo organizarlos para optimizar el manejo y análisis posterior. Por ende las fuentes de datos se trabajaron a partir de dataframes creados con la librería *Pandas* y se realizaron consultas *SQL* para filtrar la información relevante.

Además, se elaboraron reportes y gráficos con dichos dataframes para visualizar las relaciones entre las instituciones educativas y productivas en función de la población de cada región para dar respuesta al objetivo principal del trabajo.





Procesamiento de datos

Análisis de formas normales

Se comenzó la tarea de procesamiento de datos analizando en qué formas normales se encontraban las tablas de Establecimientos Productivos y Establecimientos Educativos:

- Teniendo en cuenta que la primer forma normal solo permite valores atómicos¹ como dominio de un atributo, en la tabla de Establecimientos educativos los atributos como “Teléfono” y “Mail”, en muchos casos poseen más de un valor. Por lo tanto podríamos decir que no se encuentra en ninguna de las demás formas normales.
- La tabla de Establecimientos Productivos se encuentra en primera formal normal, sin embargo no se encuentra en segunda debido a que existe atributos con dependencias parciales como por ejemplo “provincia” con “provincia_id” y “departamento” con “id_departamento”. En consecuencia, podríamos decir que no se encuentra en tercera forma normal.

Calidad de datos

A partir la lectura y consultas SQL de las tablas mencionadas en el apartado anterior se identificaron problemas de calidad de datos:

- En una primer instancia de visualización rápida de la tabla de Establecimientos Productivos se observó que había nombres de departamento repetidos en diversas provincias, para comprobar si afectaba la consistencia de calidad de datos, siendo un problema de instancia y modelo, realizamos consultas SQL² con el objetivo (Goal) de determinar el porcentaje de tuplas afectadas por la ambigüedad en los nombres de los departamentos. La pregunta (Question) que se utilizó fue “¿Cuál es el porcentaje de nombres de departamentos iguales en provincias diferentes para el año 2022?” y la métrica (Metric) para responder la pregunta es el porcentaje de nombres de departamentos ambiguos:

$$\left(\frac{\text{Departamentos repetidos en provincias diferentes}}{\text{Total de departamentos}} \right) * 100$$

El resultado de esta métrica fue 8,91%, por lo tanto casi 9 de cada 100 nombres de departamentos en la tabla son ambiguos y en consecuencia, la consistencia se ve afectada. Se podría mejorar el modelo si solo se permitiera ingresar un nombre repetido de departamento junto con una referencia a la provincia. Por ejemplo, no permitir ingresar “Capital”, más bien especificar “Capital-Córdoba”.

Por otra parte, se identificó un problema en la relevancia de los datos cuando analizamos el atributo “Clae2” debido a que no aporta información nueva a la tabla porque deriva de otro atributo (“Clae6”)

- Asimismo, en la tabla de Establecimientos Educativos se observa a simple vista un gran problema de completitud en los atributos que hacen referencia a modalidades

¹ Valores simples e indivisibles.

² Se encuentran en el archivo .py llamado “tp” en la carpeta “TP1”



educativas como “Adultos”, “Artística”, “Especial”, “Hospitalaria”, “Intercultural” y “Encierro” debido a poseen en su mayoría valores Nan. Es importante resaltar que los valores que están completados en esas columnas son 1, por lo tanto la ausencia de los establecimientos educativos de la tabla deberá ser presentada con el valor 0 por convención y puede derivar en un problema de confiabilidad por falta de datos. Estos tipos de problemas son de instancia por hacer referencia a la ausencia de valores.

Con objetivo de cuantificar la completitud de oferta educativa diferente a la modalidad común, la pregunta que se utilizó fue “¿Cuál es el porcentaje de tuplas que carecen de valores en las columnas de modalidades educativas diferentes a común?” y la métrica empleada fue el porcentaje de filas con valores nulos de dichos atributos:

$$\left(\frac{\text{Filas con valores Nan en los atributos mencionados}}{\text{Total de filas}} \right) * 100$$

El resultado de esta métrica fue de un 96,27% , infiriendo que casi la totalidad de los establecimientos no poseen datos en alguna de las modalidades “Adultos”, “Artística”, “Especial”, “Hospitalaria”, “Intercultural” y “Encierro”. Llegando a la conclusión de que existe un problema de completitud en la tabla, pudiendo mejorar completando las columnas con 0 y 1, o creando una columna donde los valores sean los nombres de la modalidad del establecimiento.

Diagrama Entidad-Relación

Para modelar de manera conceptual los datos necesarios para resolver los problemas planteados posteriormente, se realizó un Diagrama Entidad-Relación que permite visualizar las entidades, sus atributos, claves y relaciones entre las bases de datos de Establecimientos Educativos, Establecimientos Productivos, Población.

En el diagrama se ilustran en rectángulos las entidades “Departamento”, “Establecimientos productivos”, “Establecimientos educativos” y “Personas”, en rombos las relaciones entre ellas con sus cardinalidades y en óvalos los atributos de cada entidad, con las claves subrayadas.

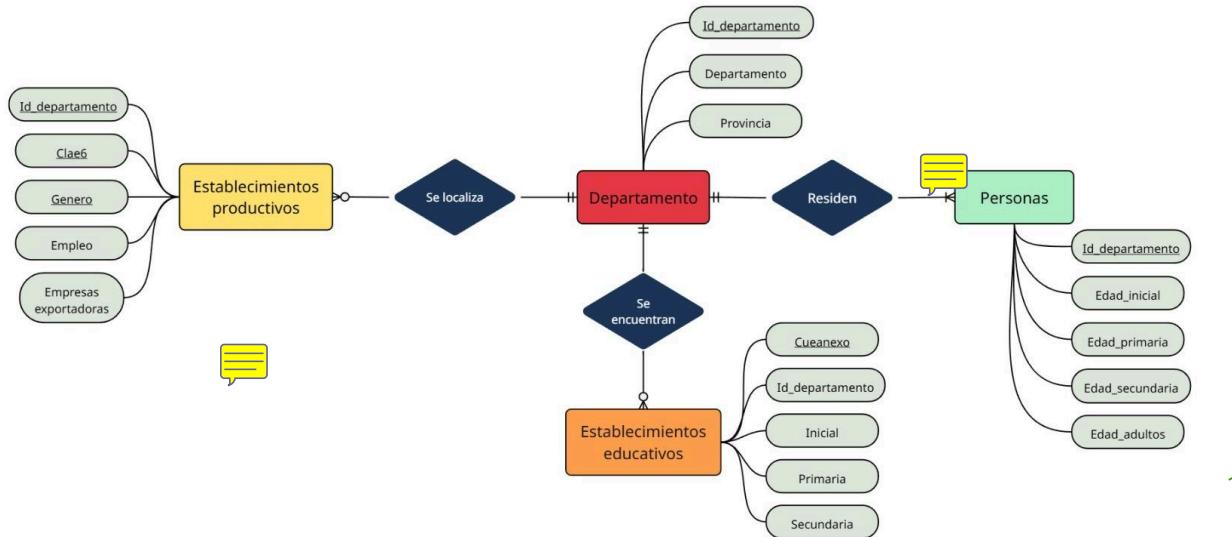


Figura 1. Diagrama Entidad-Relación resultante de la integración de las bases de datos proporcionadas.

Modelo Relacional

Partiendo del DER elaborado en el apartado anterior, se realizó el modelo relacional, donde cada entidad posee su esquema:

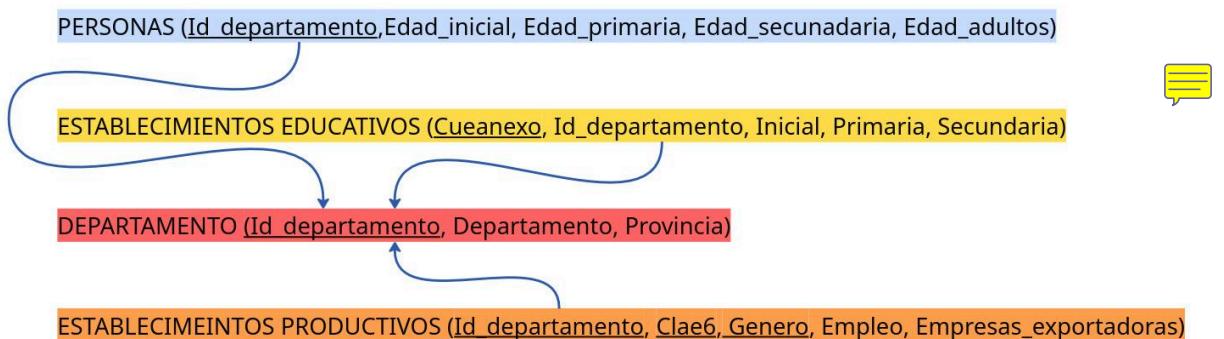


Figura 2. Modelo relacional del Diagrama Entidad-Relación de la Imagen 1.

Como todas las relaciones son uno a muchos con “Departamento”, siendo esta la entidad en la lado de uno, las relaciones “Personas”, “Establecimientos educativos” y “Establecimientos productivos” incluyen la clave primaria de la relación “Departamento” (“Id_departamento”) como clave foránea indicado con flechas.

Cada esquema está en las tres formas normales debido a que se cumple que todos los atributos son atómicos, cada atributo que no pertenece a la clave primaria depende de la clave primaria en su totalidad y no existen dependencias funcionales transitivas, con el objetivo de minimizar la redundancia y anomalías de inserción, delección y modificación.



- El esquema de “Personas” posee como PK³ “Id_departamento”, la DF⁴ presente se basa en que “Id_departamento” determina “Edad_inicial”, “Edad_primaria”, “Edad_secundaria” y “Edad_adultos”, y la FK⁵ como se mencionó anteriormente es “Id_departamento” porque muchas personas pertenecen a un mismo departamento. Esta relación cumple las tres formas normales debido a que todos los atributos poseen valores atómicos, la PK es un único atributo lo cual conlleva dependencias funcionales totales sobre la clave y no existen dependencias funcionales transitivas porque los atributos que hacen referencias a los grupos etarios no dependen de otro atributo primo intermedio, solo de la PK.
- El esquema de “Establecimientos Educativos” posee como PK “Cueanexo”, la DF presente se basa en que “Cueanexo” determina a qué departamento pertenece el establecimiento, y si ofrece nivel inicial, primaria y/o secundaria, y la FK como se mencionó anteriormente es “Id_departamento” porque muchas establecimientos educativos pertenecen a un mismo departamento. Esta relación cumple las tres formas normales por las mismas razones que el esquema anterior. ✓
- Por otro lado, el esquema de “Departamento” posee como PK “Id_departamento”, la DF presente se basa en que “Id_departamento” determina el nombre y provincia de cada departamento. Esta relación cumple las tres formas normales por las mismas razones mencionadas. ✓

DataFrames

Para generar los dataframes que se necesitan, en primer lugar se exportaron los archivos originales de las bases de datos proporcionadas y luego se “limpio” cada tabla en función de lo que servirá para realizar el trabajo.

Para el dataframe de Establecimientos Productivos original se seleccionaron las tuplas del “anio” 2022 y únicamente las columnas “in_departamentos”, “departamento”, “provincia”, “clae6”, “genero”, “Empleo”, “Establecimientos” y “Empresas_exportadoras” que posteriormente se renombraron para sean consistentes con el esquema del modelo relacional. ✓

Para el dataframe de Departamento se utilizó la tabla de establecimientos Productivos y se recopilaron los “in_departamentos”, “departamento” y “provincia” sin contar repeticiones, renombrando nuevamente los atributos para que coincidan con el modelo relacional.

Por otro lado, en el dataframe de Establecimientos Educativos original, se omitieron las primeras filas y columnas como “Sector”, “Ámbito”, “Nombre”, entre otras porque no tenían información relevante. Luego se unificaron los establecimientos que tenían “Nivel inicial - Jardín maternal” y “Nivel inicial - Jardín de infantes” en una nueva columna llamada “Inicial”, además de “Secundario”, “Secundario - INET”, “SNU” y “SNU - INET” en una columna llamada “Secundaria”, para luego eliminar aquellas columnas que agrupe. En este caso el atributo que debería ser “Id_departamento” se llamaba “Codigo de localidad”, por lo tanto se cambió el nombre y además se agregan ceros cuando tenía menos dígitos y se realizó un for para cambiar los “Id_departamento” de las “Comunas” de CABA para que sea ✓

³ Clave primaria.

⁴ Dependencia funcional.

⁵ Clave foránea.



Laboratorio de Datos
Trabajo Práctico 01
2do. Cuatrimestre - 2025



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

consistente con el dataframe de Departamento. Por último, solo se seleccionaron los atributos “Id_departamento”, “Primaria”, “Inicial” y “Secundaria”.

Para el dataframe de Población se realizó un for que guarde el inicio, final e id de cada “mini tabla” del archivo, para posteriormente crear un diccionario que contenga el “Id_departamento” como clave y una lista con la suma de personas de cada rango etario por nivel educativo como valor. Luego se creó un dataframe a partir de ese diccionario.

✓



Decisiones tomadas

Las fuentes de datos proporcionadas para la realización del trabajo incluían diversos atributos que eran innecesarios para el análisis en función del objetivo planteado. Por consecuencia tomamos las siguientes decisiones:

- En la fuente de establecimientos productivos fueron eliminados atributos como “año” debido a que sus valores se reducían a 2021 y 2022, como la fuente de establecimientos educativos y de población contenía datos únicamente del año 2022 se consideró conveniente realizar el análisis con las tuplas que se refieren a dicho año y remover las que hacían referencia al 2021.

Por otro lado, se eliminaron atributos como “departamento” y “provincia” en vista de que con tener la información del “id_departamento” es suficiente por ser clave foránea de Departamento⁶, además de ser un atributo más adecuado que el nombre del departamento por su unicidad. Asimismo decidimos eliminar los atributos “Clae2” y “letra” porque eran irrelevantes teniendo en cuenta lo pedido en el trabajo y por ejemplo, la “Clae2” está contenida en la “Clae6” siendo los dígitos posteriores al cuarto, contados desde el final de la “Clae6”.

- En la fuente de datos de establecimientos educativos se eliminó la primera fila que cumplía una función de categorización de los atributos. Sin embargo en esta fuente también existían varios atributos como “Jurisdicción”, “Nombre”, “Sector”, “Ambito”, “Domicilio”, “C. P.”, “Codigo de area”, “Teléfono”, “Departamento”, “Localidad”, y “Mail” que son importantes para describir un establecimiento educativo en sí mismo, aunque no eran datos de utilidad en este caso.

Teniendo en cuenta que en las consignas del trabajo nos piden únicamente analizar los establecimientos educativos de modalidad común, se eliminaron los atributos que corresponden a los distintos tipos de modalidades (“Común”, “Especial”, “Adultos”, “Artística”, “Hospitalaria”, “Intercultural”, “Encierro” y “Servicios complementarios”) y las tuplas de establecimientos que no pertenecían a la modalidad a analizar. Se conservaron los atributos que denotaban la división de niveles educativos de “Común” y se fusionaron por rango etario para optimizar el análisis. En particular, se condensó los establecimientos que ofrecen “Nivel inicial- Jardín maternal” y “Nivel inicial - Jardín de Infantes” (que en muchos casos eran los mismos) porque se consideró innecesario tener una división de rango etario de 0 a 2 años y de 3 a 5 años aproximadamente. De igual forma con “Secundario”, “Secundario - INET”, “SNU” y “SNU- INET” que se refieran a distintos tipos de educación secundaria y su relación con el INET (Instituto Nacional de Educación Tecnológica), por lo tanto fusionamos los 4 atributos en “Secundaria” que posee un rango etario aproximado de 11 a 18 años⁷.

- La fuente de datos de población estaba estructurada de una forma muy impráctica a nuestros fines, por lo tanto dividimos las edades en grupos etarios de interés siendo

⁶ Tabla creada especialmente para guardar los datos de los identificadores de cada departamento, su nombre y a qué provincia pertenecen.

⁷ Se utilizó dicho rango etario debido a que la edad de inicio depende de la duración de la escuela primaria (6 o 7 años), mientras que la duración total de la secundaria puede ser de 5 o 6 años según la provincia.



Laboratorio de Datos
Trabajo Práctico 01
2do. Cuatrimestre - 2025



DEPARTAMENTO
DE COMPUTACION

Facultad de Ciencias Exactas y Naturales - UBA

“Edad jardín”, “Edad primaria”, “Edad secundaria” y “Edad adultos” los nuevos atributos. Se eliminaron los atributos que hacían referencia a los porcentajes y el total de la población debido que es información que se puede obtener a partir de otros atributos.



Análisis de datos

Reportes realizados con SQL

Debido a que todos los reportes poseen muchas filas, solo serán presentados de forma ilustrativa. Todos los reportes se encuentran como archivo csv en la carpeta "TablasModelo" y su nombre está especificados en los epígrafes a continuación:

- Consigna 1: En el reporte se registra la información de la cantidad de jardines, primarias y secundarios, la cantidad de población de cada rango etario correspondiente a los niveles mencionados por departamento de cada provincia. Además tenemos la información ordenada ascendente por el nombre de la provincia y luego dentro de provincias decreciente por cantidad en "Primarias".

De esta forma, se observa fácilmente cual es el departamento de cada provincia que tiene mayor cantidad de escuelas primarias. En el caso de Buenos Aires es La Matanza, la cual tiene 1186 primarias, teniendo 225872 personas en ese grupo etario. Además podemos observar que hay una relación entre la cantidad de escuelas que hay por cada nivel y la cantidad de población del grupo etario que asiste. Por lo tanto, la cantidad de jardines, primarias y secundarios aumenta con la población que puede asistir (en función de su edad).

Provincia	Departamento	Jardines	Población Jardín	Primarias	Población Primaria	Secundarios	Población Secundaria
La Matanza		325	157034	1186	225872	355	181212
La Plata		232	52075	832	77998	254	67326
General Pueyrredón		178	41427	664	62565	205	57730
Lomas de Zamora		162	50825	647	76967	219	65257
Quilmes		162	47353	542	70881	167	60085
Almirante Brown		133	43762	489	67913	152	58227
Moreno		118	50791	428	76357	143	60359

Figura 3. Fracción ilustrativa del reporte "tabla_poblacion_educacion.csv"

- Consigna 2: En el presente reporte se informa la cantidad de empleados por departamento y la provincia a la cual pertenecen. Los datos se ordenaron primero de forma alfabética y luego de manera descendente por la cantidad de empleados.

De esta manera podemos observar el departamento que mayor cantidad de empleados tiene por provincia. Si miramos el caso de Buenos Aires, La Matanza es el departamento con mayor cantidad de empleados.

Si comparamos los datos de los dos primeros reportes podemos ver que hay una relación con la cantidad de empleados y la cantidad de establecimientos educativos, en vista de que por ejemplo La Matanza es el departamento de Buenos Aires con la mayor cantidad de empleados y la mayor cantidad de jardines, primarias y secundarias.

Provincia	Departamento	Cantidad total de empleados en 2022
Buenos Aires	La Matanza	127962.0
Buenos Aires	Vicente López	120347.0
Buenos Aires	General Pueyrredón	118660.0
Buenos Aires	San Isidro	98372.0
Buenos Aires	La Plata	89699.0
Buenos Aires	Tigre	83661.0



Figura 4. Fracción ilustrativa del reporte “empleo_por_departamento.csv”

- Consigna 3: En este reporte se tiene la cantidad de empresas exportadoras que emplean mujeres, la cantidad de establecimientos empresas exportadoras y la población que hay por departamento de cada provincia. El primer criterio de ordenamiento es la cantidad de empresas exportadoras que emplean mujeres de manera descendiente. Luego la cantidad de empresas exportadoras descendiente, nombre de provincia ascendente y nombre de departamento ascendente con ese orden.

Por ejemplo podemos ver que Córdoba Capital es el lugar donde tienen mayor cantidad de empresas exportadoras que emplean mujeres con una suma de 1296. Al tener en cuenta el caso de Buenos Aires y comparar con los datos de los dos anteriores reportes, se observa que La Matanza es el departamento con mayor cantidad de empresas exportadoras que emplean mujeres teniendo 864.

Provincia	Departamento	Cant_Expo_Mujeres	Cant_EE	Población
Cordoba	Capital	12960	1400	1498060
Santa Fe	Rosario	14250	1248	1337958
Buenos Aires	La Matanza	8640	1186	1837168
Buenos Aires	La Plata	2330	832	756074
Santa Fe	La Capital	3370	728	568259
Buenos Aires	General Pueyrredón	7930	664	660569
Buenos Aires	Lomas de Zamora	3290	647	685644
Entre Ríos	Paraná	2750	591	388716

Figura 5. Fracción ilustrativa del reporte “tabla_edu_export_pob.csv”

- Consigna 4: El reporte denota la cantidad de empleados por rubro y departamento que superan el promedio de cantidad de empleados por provincia, además de indicar la Clae3 y la provincia. Por ejemplo podemos ver que en La Matanza en el rubro con Clae3 472 hay 1963 empleados.



Provincia	Departamento	CLAE3	Cant. Empleados
Entre Ríos	Uruguay	464	75.0
Entre Ríos	Uruguay	614	1.0
Entre Ríos	Uruguay	641	79.0
Entre Ríos	Uruguay	900	3.0
Entre Ríos	Uruguay	960	38.0
Formosa	Formosa	14	590.0
Formosa	Formosa	16	7.0
Formosa	Formosa	462	4.0

Figura 6. Fracción ilustrativa del reporte “departamentos_empleo_rubro_principal.csv”

Gráficos

- Consigna 1: Para analizar la cantidad de empleados por provincia en el año 2022, se elaboró el siguiente gráfico (Fig. 7). En él se puede observar una clara diferencia en la cantidad de empleados en la provincia de Buenos Aires, que cuenta con más de 2.000.000 de trabajadores, superando casi cuatro veces a Santa Fe, la provincia con mayor cantidad de empleados si se excluye a la Ciudad Autónoma de Buenos Aires (CABA). Esto evidencia que Buenos Aires y CABA concentran la mayor parte del empleo del país, superando ampliamente al resto de las provincias.

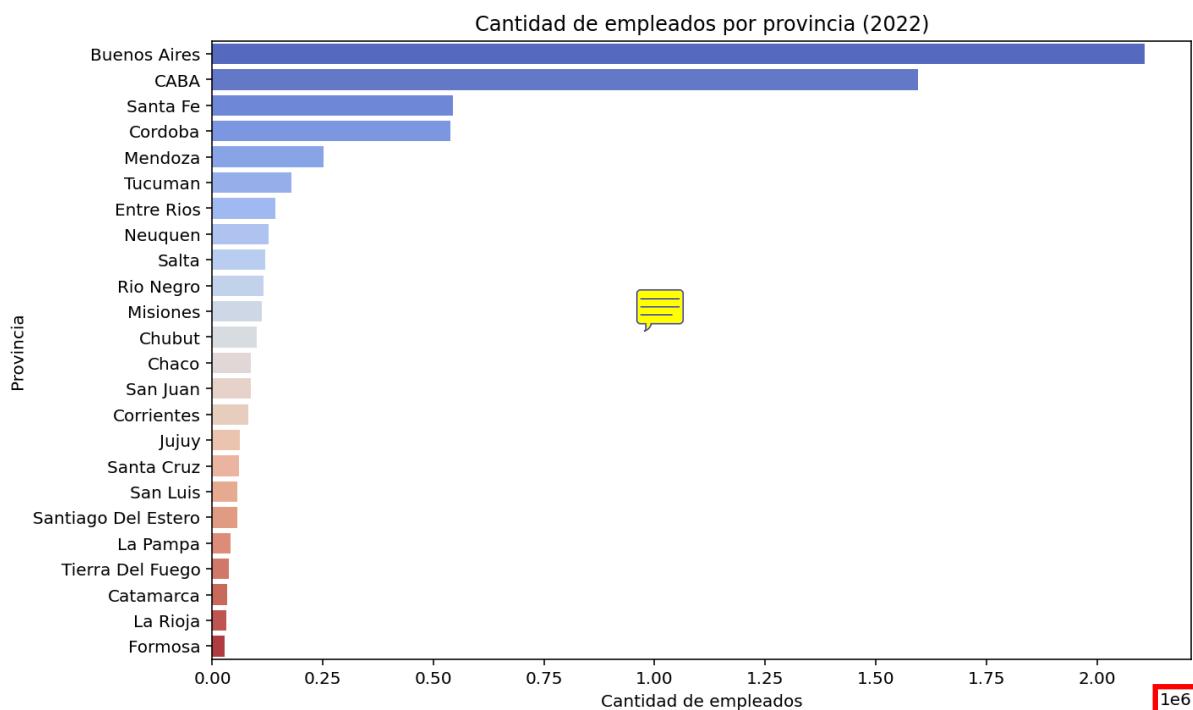


Figura 7. Cantidad de empleados por provincia (2022).

- Consigna 2: Para representar la relación entre la cantidad de establecimientos educativos (EE) y la población por departamento, se elaboró un gráfico de dispersión



en escala logarítmica(Fig. 8). En el eje X se muestra la población y en el eje Y la cantidad de EE. Cada punto representa un departamento, y los colores diferencian el nivel educativo y su correspondiente grupo etario: Jardín (Azul), Primaria (Naranja), Secundaria (Verde).Se empleó una escala logarítmica en ambos ejes para facilitar la visualización de los datos.

Se observa una tendencia creciente: a medida que aumenta la población, también lo hace la cantidad de establecimientos educativos. Este comportamiento es consistente para los tres niveles, aunque con distinta magnitud. El nivel primario presenta la mayor cantidad de establecimientos, seguido por el secundario y finalmente el inicial(ver Fig. 12). El uso de la escala logarítmica en ambos ejes permite visualizar de manera más clara tanto los departamentos con baja población como aquellos con gran densidad poblacional, manteniendo la proporcionalidad de la relación. En conjunto, el gráfico refleja cómo la oferta educativa se ajusta al tamaño poblacional de los distintos departamentos.

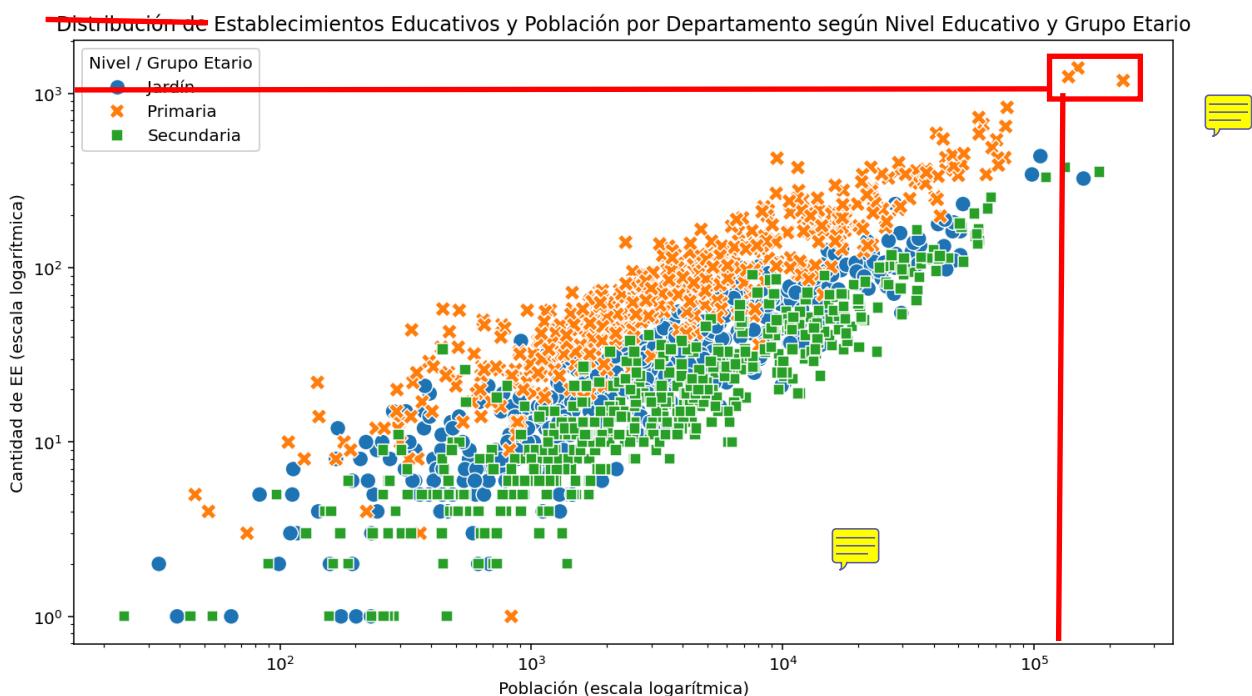


Figura 8.Distribución de establecimientos educativos por población y nivel educativo. Ambos ejes están expresados en escala logarítmica.

- Consigna 3: En el siguiente gráfico se representa la distribución de la cantidad de establecimientos educativos (EE) por departamento dentro de cada provincia (Fig. 9). Cada boxplot muestra la mediana, los cuartiles y la dispersión de los valores, permitiendo comparar la variabilidad interna entre provincias. Las provincias fueron ordenadas de menor a mayor según la mediana de cantidad de EE por departamento. Esto facilita identificar aquellas con mayor concentración de



establecimientos educativos en promedio. Se observa que provincias como CABA, Santa Fe y Córdoba presentan las mayores medianas y una amplia dispersión, lo que indica una alta concentración de establecimientos en varios de sus departamentos. En contraste, provincias como Santa Cruz, La Pampa y Chubut presentan menores medianas y menor dispersión, reflejando una distribución más homogénea y con menos cantidad total de establecimientos por departamento

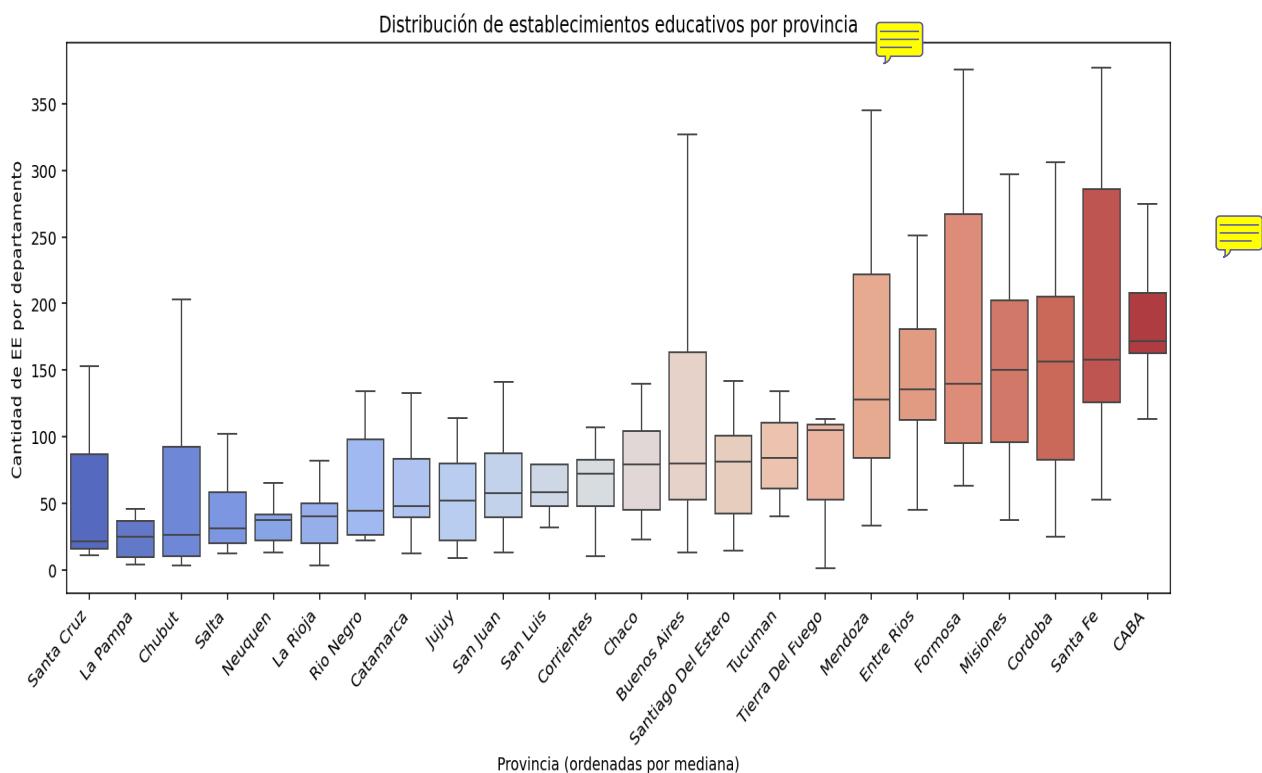


Figura 9. Distribución de la cantidad de establecimientos educativos por departamento en cada provincia (ordenadas por mediana)

- Consigna 4 : Se buscó relacionar la cantidad de personas empleadas y la cantidad de establecimientos educativos (EE) cada mil habitantes en el año 2022, por departamento y diferenciadas por provincia (Fig 10). En el eje X se observa la cantidad de empleados cada mil habitantes, mientras que en el eje Y se muestra la cantidad de EE cada mil habitantes. Los distintos colores identifican las provincias a las que pertenece cada departamento. Se empleó una escala logarítmica en ambos ejes para facilitar la visualización de los datos, dado que las variables presentan una alta dispersión entre departamentos.

A partir del gráfico se observa que la mayoría de los departamentos se concentran en una zona intermedia, lo que indica valores moderados tanto en empleo como en oferta educativa por habitante, también se destacan algunos valores atípicos que se alejan del conjunto principal, como por ejemplo, CABA con un elevado número de empleados por habitantes, esto puede ser porque representa una estructura económica altamente concentrada. También está Tierra del Fuego con menor EE por habitantes o Santa Cruz con mayor EE por habitantes.

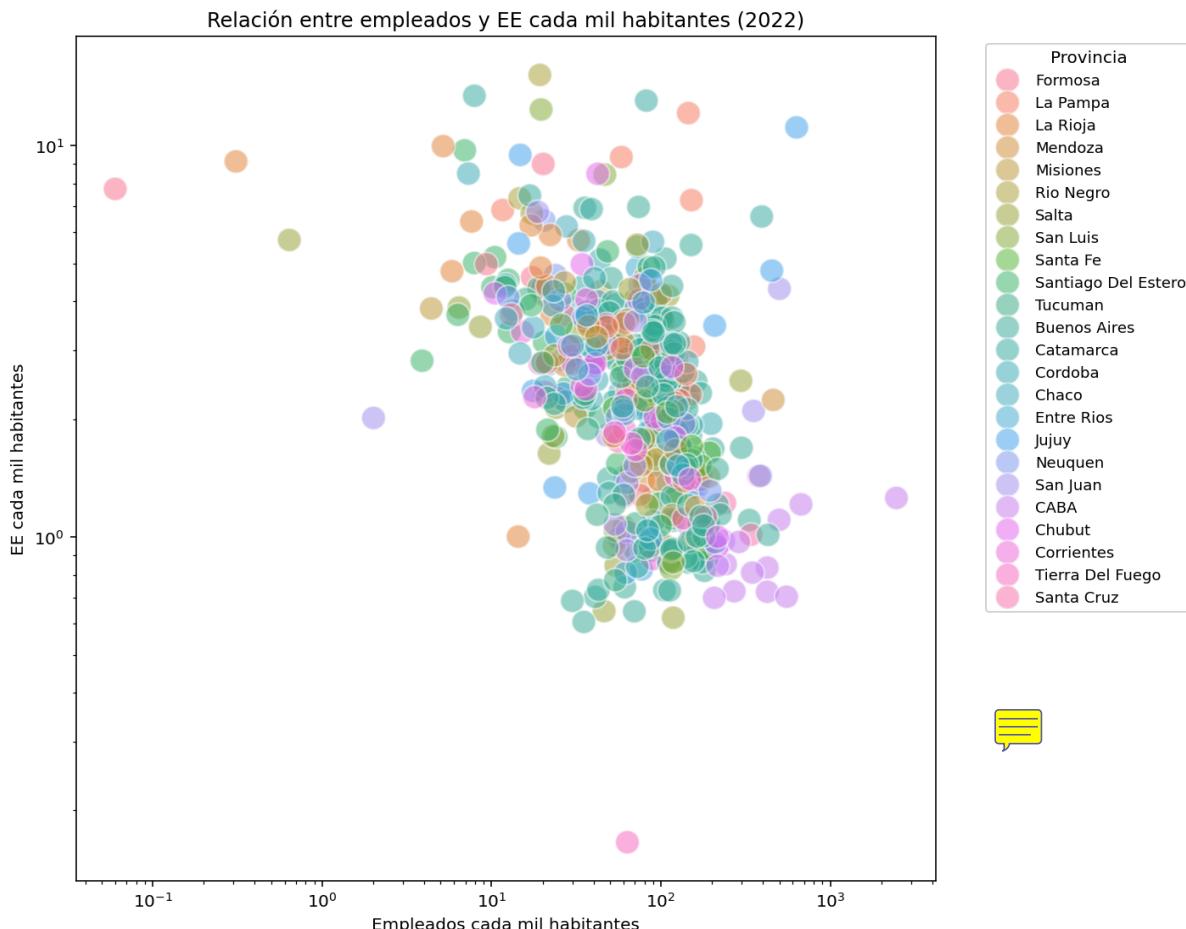


Figura 10. Relación entre la cantidad de empleados y de establecimientos educativos (EE) cada mil habitantes en 2022, por departamento y provincia. Cada punto representa un departamento; el color identifica la provincia correspondiente. Ambos ejes están expresados en escala logarítmica

- Consigna 5: Se graficaron los 5 establecimientos productivos por código de actividad (clae6) con mayor y menor proporción de empleo femenino para el 2022(Fig. 11). Las actividades representadas en color salmón muestran valores muy altos, cercanos o superiores al 85–95%. Esto indica que en esos sectores el empleo consta con una presencia dominante de mujeres. Las barras celestes muestran proporciones muy bajas, cercanas a cero. Esto sugiere que esas actividades están con una participación femenina mínima. La línea horizontal negra marca el promedio general de empleo femenino (27,63%). Esta queda muy por debajo de las actividades del Top 5 y muy por encima de las del Bottom 5, mostrando una gran brecha de género entre sectores.
Esto refleja una segmentación laboral por género: algunas actividades concentran a la mayoría de las mujeres empleadas, mientras que otras prácticamente las excluyen.

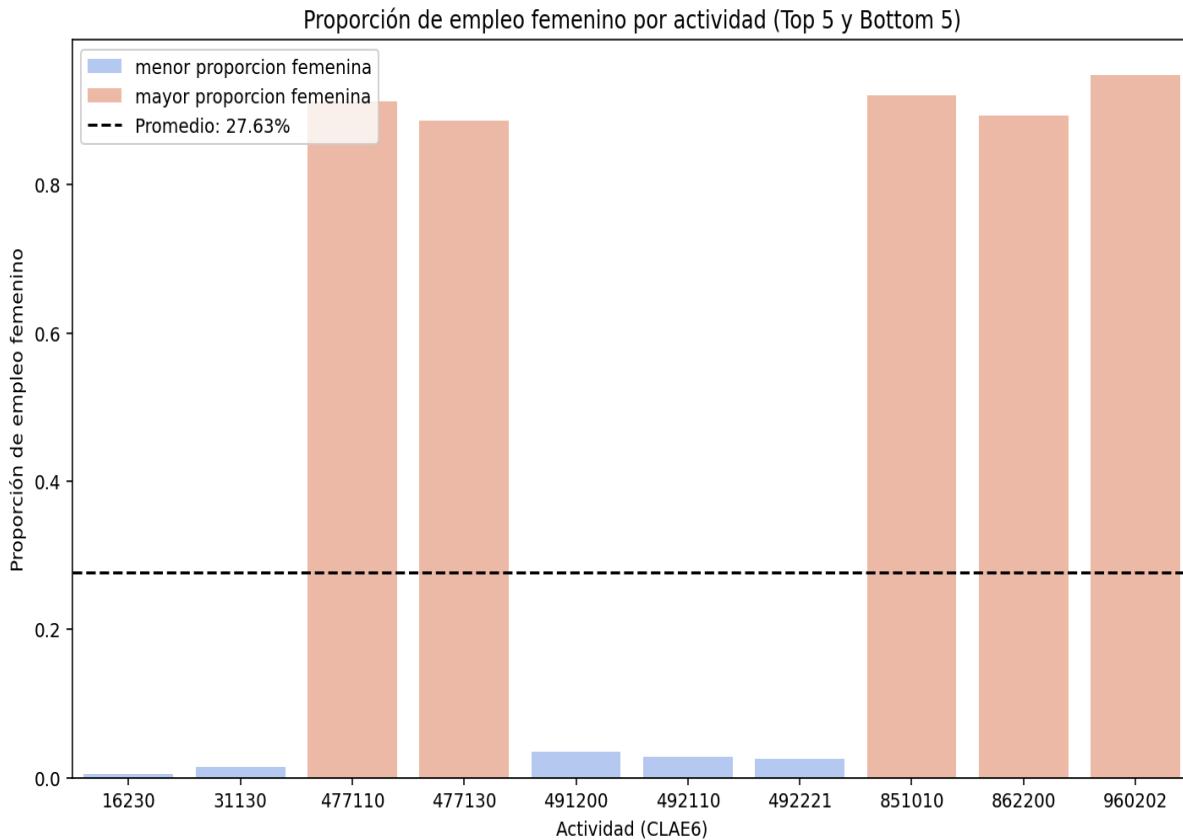


Figura 11. Distribución del empleo femenino por actividad económica (CLAE6) en Argentina para el año 2022. Se destacan las cinco actividades con mayor y menor proporción de trabajadoras mujeres, junto con la línea de referencia que indica el promedio general de empleo femenino (27,63%).





Conclusiones

El objetivo central de este informe, que era conocer si existe un vínculo entre la educación y el desarrollo de actividades productivas en Argentina, se ha respondido afirmativamente a partir del análisis de los datos del año 2022.

Se encontró una clara y fuerte correlación entre la cantidad de Establecimientos Educativos y la población por departamento, manteniéndose esta tendencia para los niveles Inicial, Primario y Secundario. De igual manera, se verificó una relación entre la cantidad de empleados y la oferta educativa; los departamentos con mayor cantidad de empleo (como La Matanza en Buenos Aires) son también aquellos que concentran la mayor cantidad de instituciones educativas. Esto sugiere que la infraestructura educativa se ajusta al tamaño poblacional y a la concentración de la actividad económica de cada región.

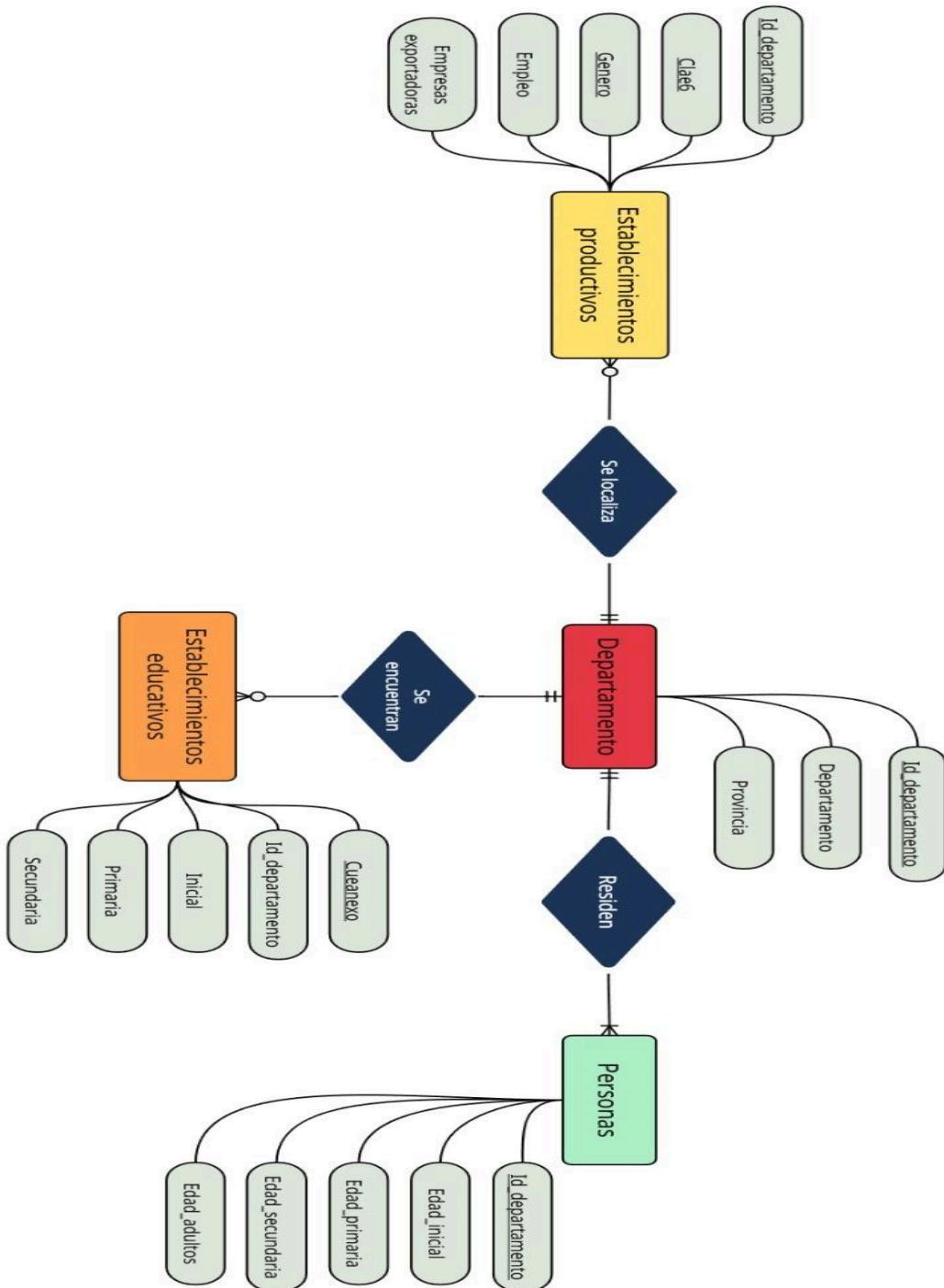
El análisis de la distribución de establecimientos educativos por provincia mostró que CABA, Santa Fe y Córdoba presentan las mayores medianas y una alta dispersión, indicando una fuerte concentración de establecimientos en sus principales departamentos. Esto es consistente con la distribución del empleo, donde Buenos Aires y CABA concentran la mayor parte del total de trabajadores del país.

En resumen, el trabajo demuestra que la oferta de educación común en Argentina se encuentra ligada a la densidad poblacional y al desarrollo productivo a nivel departamental.



Anexo

Debido a que la Imagen 1 no se ve muy clara, la adjuntamos nuevamente de forma apaisada para mejor lectura de la misma



Se adjunta la relación de la Figura 8 con la diferencia de la separación de cada nivel educativo para mejor visualización, debido a que en dicha figura hay datos poco apreciables por su superposición



Laboratorio de Datos
Trabajo Práctico 01
2do. Cuatrimestre - 2025



DEPARTAMENTO
DE COMPUTACION
Facultad de Ciencias Exactas y Naturales - UBA

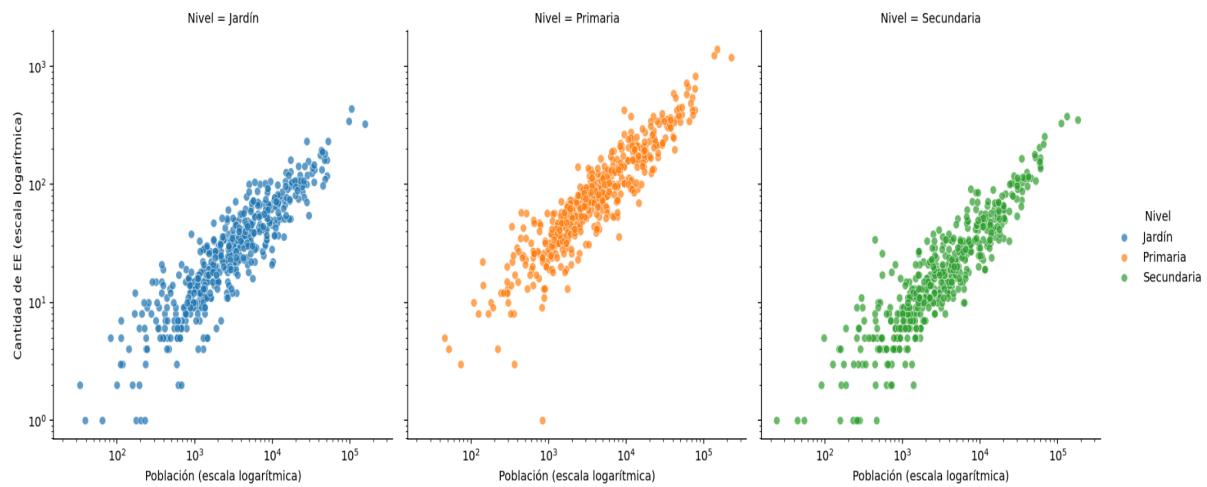


Figura 12. Relación entre la cantidad de establecimientos educativos y la población, separadas por el nivel educativo (2022).