

# **Relación entre los Establecimientos Educativos y los Centros Culturales de la República Argentina**

Evaluación: Trabajo Práctico 01

Materia: Laboratorio de datos - FCEyN - UBA

Autor : Kamala Jimeno Leiton, Francisco Peix y Carolina Cuiña

Fecha : 11/02/2025 - 23/02/2025



## Resumen

Este trabajo, tiene el propósito de analizar tres distintas fuentes de datos de la República Argentina, correspondientes al año 2022, las cuales no cumplen con ciertos atributos de calidad y pautas de diseño que deberían tener. Para poder responder a la pregunta de si existe cierta relación entre la cantidad de establecimientos educativos en cada una de las provincias, y la cantidad de centros culturales.

Mediante las herramientas aprendidas en clase, buscamos la forma de poder obtener información relevante de las mismas, con el fin de poder trabajar con los datasets de manera cómoda y así poder dar un diagnóstico de lo obtenido y observado.

Además, para poder complementar y perfeccionar nuestros resultados hallados, realizamos diversos gráficos, de distinto tipo, con el fin de poder entender y utilizar el contenido más eficientemente que al hacerlo mirando simplemente los datos crudos, pudiendo así detectar una relación entre los datos que estas fuentes presentan; y con ello poder elaborar ciertas hipótesis acerca de lo observado.

Finalmente, arribamos a varias conclusiones como que en todo departamento de la República Argentina se presenta una cantidad de Establecimientos Educativos superior a la de Centros Culturales; y que dichas cantidades se relacionan entre sí. Además de que se ven influenciadas en cada área según el número de habitantes que esta posea.

## Introducción

El problema principal que se busca resolver es determinar, si existe, algún tipo de relación entre la cantidad de establecimientos educativos en cada una de las provincias, y la cantidad de centros culturales.

Nuestro objetivo general es encontrar una respuesta al problema y fundamentos que la acompañen, mediante la revisión y análisis en profundidad de los datos obtenidos, partiendo de la hipótesis de que la cantidad de centros culturales dependen de la cantidad de establecimientos educativos por departamento.

Para llevar a cabo esto, contamos con tres fuentes de datos brindadas asociadas a dicho asunto, las cuales pertenecen al mismo año (2022):

- **Padrón Oficial de Establecimientos Educativos** [1]
- **Padrón de Centros Culturales** [2]
- **Datos de población por Departamento (padrón de población)** [3]

En primer lugar, deberemos enfocarnos en reformar dichos datasets, los cuales poseen abundantes errores de calidad, para poder tratar con ellos de una manera rápida y efectiva. En este punto, además, deberemos crearnos distintos dataFrames (distintos esquemas) para poder *normalizar* los datos; así, logramos minimizar la redundancia y además suprimir anomalías de inserción, deleción y modificación.

Una vez hecho esto, nos enfocaremos en examinar en profundidad los datos, para poder así obtener abundante información relevante; y, con ella, poder determinar ciertas hipótesis con respecto a nuestro objetivo general.

Una parte muy importante que nos ayuda en esta etapa es la visualización de datos mediante gráficos, pues estos nos brindan mayor capacidad para detectar patrones, anomalías y relaciones entre variables que al hacerlo mirando simplemente los datos crudos.

Finalmente, luego de un profundo trabajo, expondremos nuestras propias deducciones acerca de lo estudiado, justificando así la respuesta hallada a la pregunta establecida.

## Procesamiento de Datos

En primer lugar, podemos observar que, por un lado, el dataset asociado a Centros culturales se encuentra en **primera forma normal (1FN)**, pues se cumple que el dominio de sus atributos sean valores atómicos (considerando que los datos pertenecientes a la columna 'Fuente' son valores atómicos ya que hacen referencia a una única ruta del navegador); pero, como no todos los atributos no primos dependen funcionalmente de manera completa sobre todas las claves de la tabla, se puede asegurar que la misma no se encuentra en segunda forma normal y, por consiguiente, tampoco en tercera. Un ejemplo para poder demostrar esto es, con respecto a los atributos : "Latitud", "Longitud" y "LatitudLongitudTipo". Este último depende de "Latitud" y de "Longitud", pero no depende de, por ejemplo "Cod\_Loc", que es una CK (Candidate Key).

Con respecto a la fuente de Establecimientos Educativos, en cambio, pudimos contemplar que, en el atributo "Telefono", se presentan, en ciertas instancias legales, la presencia de valores no atómicos; así, por lo tanto, podemos determinar que esta tabla **no** se encuentra ni siquiera en primer forma normal y, por lo tanto, tampoco en segunda ni tercera.

## Calidad De Datos

En los datasets presentes, surgen abundantes problemas de calidad, entre los que se pueden destacar los siguientes:

- **Datos nulos en distintos campos y difícil accesibilidad a los datos:**

Para la tabla Padrón de Centros Culturales, en las columnas "Observación" y "InfoAdicional", en todas las filas se tiene valores nulos. Esto es un problema en el modelo. En este caso el atributo de calidad que se ve afectado es la **completitud**;

### GQM

*Goal (Objetivo)* Que los datos correspondientes a Observación de cada centro estén completos.

*Question (Pregunta)* ¿Cuál es la proporción de centros que tienen el dato correspondiente a Observación vacío?

*Metric (Métrica)* Cantidad de registros con campo Observación vacío en tabla Centros Culturales, es decir,

$$\frac{\text{Cantidad de registros de centro con campo Observación vacío}}{\text{Cantidad total de registros de centro}}$$

Asimismo, un caso similar puede ocurrir en la fuente de datos Padrón de Población, donde sus datos resultan poco accesibles, pues, aunque parece cómodo para poder visualizarlo, no tiene la estructura para poder ser analizado fácilmente. Describe un problema de modelo ocasionado por un problema de software a causa de su interfaz

poco amigable. En este caso el atributo de calidad que se ve afectado es el de **disponibilidad**, pues resulta difícil la accesibilidad de los datos a analizar.

#### **GQM**

*Goal (Objetivo):* Que los datos correspondientes al atributo “Casos” contengan información pertinente a la columna (evitando nulos, repetición del nombre o datos no relevantes).

*Question (Pregunta):* ¿Cuál es la proporción de la cantidad de valores nulos de ‘Casos’ respecto al total de registros en Padron Poblacion ?

*Metric (Métrica):* 
$$\frac{\text{Cantidad de valores nulos de 'Casos'}}{\text{total de registros en Padron Poblacion}}$$

- **Datos inestables y faltantes:**

Nuevamente, en la fuente de datos Padrón de Centros Culturales, podemos detectar que, con los datos asociados a la Ciudad Autónoma de Buenos Aires, se presenta una inconsistencia con respecto a sus campos “ID\_DEPTO” y “Departamento”. Esto se debe a que la información que debería ser correspondiente a estos campos no se encuentra disponible dentro de la fuente (falta de datos) y, por defecto, se estableció colocar, en el campo “Departamento” a la Jurisdicción nuevamente, es decir “Ciudad Autónoma de Buenos Aires” (Aclaración: sería una mejor opción reemplazar el título de la columna “Provincia” por “Jurisdicción”, pues al colocar este también incluye a CABA, es por esto que mencionamos dicha palabra, pero, en la tabla actual, hacemos referencia a la columna “Provincia”) . Esto provoca que sea muy complejo poder analizar la información asociada a la misma en estos casos (falta de información relevante).

Dicho problema está asociado a la instancia, debido a que los datos no han sido almacenados con la precisión necesaria. En este caso, el atributo que se ve afectado es la **confiabilidad**:

#### **GQM**

*Goal (Objetivo)* El dato correspondiente a Departamento no posea el valor de “Ciudad Autónoma de Buenos Aires”, sino que posea la información real asociada a la ubicación en donde se encuentra, y con ello también sea modificado el id de Departamento

*Question (Pregunta)* ¿Cuál es la proporción de Departamentos que poseen el valor de “Ciudad Autónoma de Buenos Aires”?

*Metric (Métrica)* Proporción de Departamentos que poseen el nombre de departamento “Ciudad Autónoma de Buenos Aires” en la tabla Centros Culturales, es decir,

$$\frac{\text{Cantidad de Departamentos que poseen el nombre de departamento "Ciudad Autónoma de Buenos Aires"}}{\text{Cantidad total de registros de centro}}$$

- **Valores no estandarizados:**

Con respecto a el dataset llamado Padrón Oficial de Establecimientos Educativos, en el atributo correspondiente al teléfono (“Telefono”) se puede apreciar que estos no poseen una sistemática en común, por ejemplo, uno de los valores dados fue “4783-0889 INT19”, mientras que otro de ellos era “4960508 - 4962173”, pudiendo provocar posibles confusiones a futuro al momento de generar un análisis con respecto a estos (falta de un formato común para los valores).

Este corresponde a un problema tanto de modelo como de instancia (problema asociado a procesos), pues no solo no se estandariza el tipo de valor que tiene que ingresar; sino que también los datos ingresados en distintas instancias son de distinta forma. En este caso, el atributo de calidad que se ve afectado es la **consistencia**, pues no existe un tipo de valor estándar en dicho atributo.

### **GQM**

*Goal (Objetivo)* El dato correspondiente a ‘Telefono’ de cada establecimiento educativo esté con el mismo formato. Establecemos que el formato utilizado sea el mayoritariamente ingresado.

*Question (Pregunta)* ¿Cuál es la proporción de Establecimientos Educativos que tienen el dato correspondiente a ‘Telefono’ con un formato distinto al establecido como estándar?

*Metric (Métrica)* Cantidad de valores de ‘Telefono’ con distinto formato al establecido como estándar en tabla Establecimientos Educativos, es decir,

Cantidad de valores de ‘Telefono’ con distinto formato al establecido  
Cantidad total de registros de Establecimientos Educativos

## DER

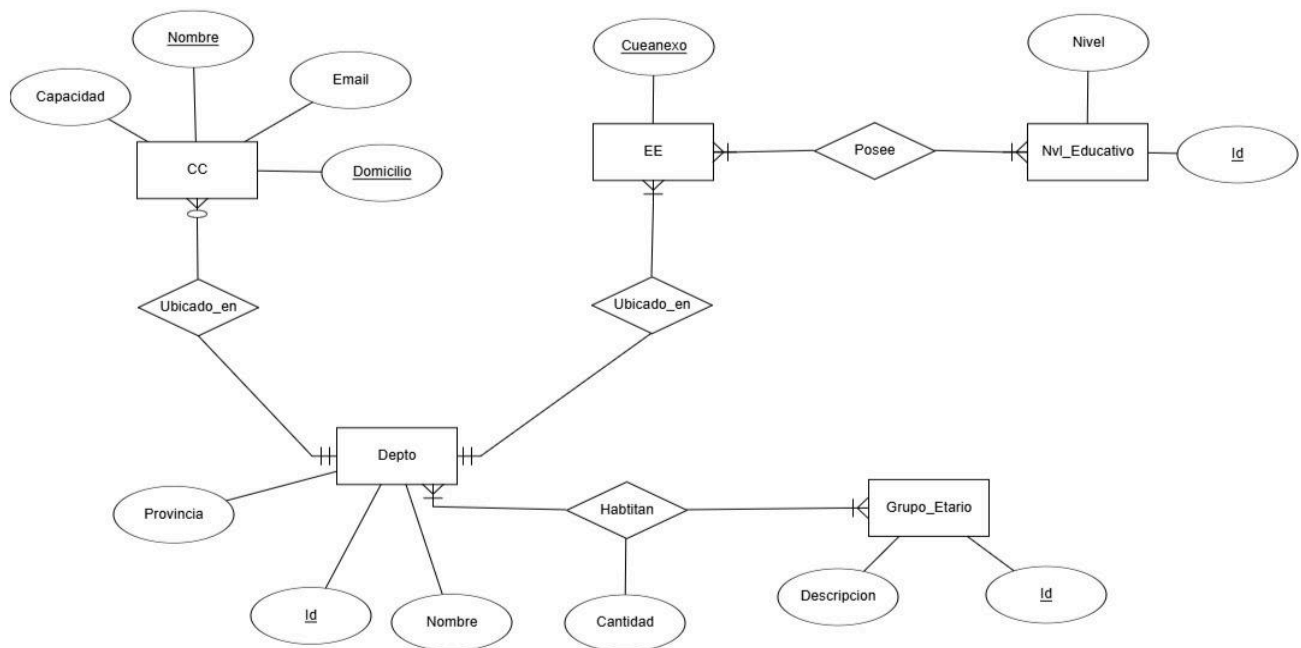


Diagrama-Entidad-Relación

## Modelo Relacional - Mapeo desde DER

Centros\_Culturales ( NombreC, Domicilio, Capacidad, Email, *ID\_Depto*)  
 Departamento (ID\_Depto , Provincia, NombreD)  
 Establecimientos\_Educativos (Cueanexo, *ID\_Depto*)  
 Grupo\_Etario (ID\_Grupo\_Etario, DescripciónG)  
 Habitan\_En (*ID\_Grupo\_Etario*, *ID\_Depto*, Cantidad)  
 Nivel\_Educativo(ID\_Nvl\_Educativo, DescripciónN)  
 Nivel\_Establecimiento (Cueanexo, *ID\_Nvl\_Educativo*)

Aclaraciones sobre notación:

- Las claves primarias se encuentran subrayadas .
- Las foreign keys se encuentran en *cursiva*.

## Dependencias funcionales - conjunto minimal

Aplicando las reglas de inferencia y aplicando el algoritmo de búsqueda de un cubrimiento minimal F para un conjunto de DFs, conseguimos como resultado el siguiente conjunto:

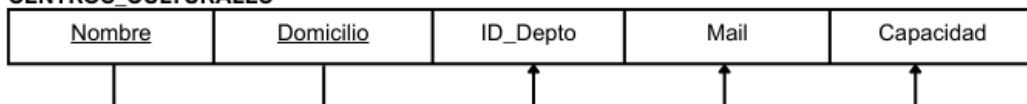
DFs = { {NombreC, Domicilio} → {Capacidad, Email, ID\_Depto},  
 ID\_Depto → {Provincia, NombreD},  
 Cueanexo → ID\_Depto,  
 ID\_Grupo\_Etario → DescripciónG,  
 {ID\_Depto, ID\_Grupo\_Etario} → Cantidad,  
 ID\_Nvl\_Educativo → DescripciónN }

## Esquemas correspondientes al modelo relacional del DER

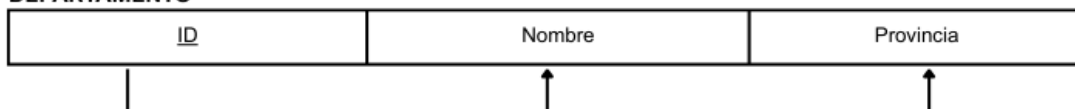
Al armar nuestros esquemas correspondientes al modelo relacional del DER, debimos procurar que estos se encuentren en tercera forma normal (3FN). Para ello, nos centramos en que se verifiquen las siguientes condiciones:

- En el dominio de nuestros atributos se encuentran únicamente valores atómicos, lo que hace que sea **1FN**.
- Todo atributo no primo A de R (siendo R cualquier esquema de los planteados anteriormente) depende completamente (de manera funcional) de todas las claves de R, por lo que está en **2FN**
- Para toda dependencia funcional no trivial  $X \rightarrow A$  de R, se cumple alguna de las siguientes condiciones:
  - X es SK de R
  - A es atributo primo de Rlo cual cumple con la definición de **3FN**.

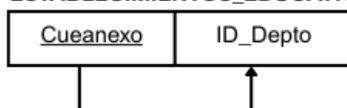
### CENTROS CULTURALES



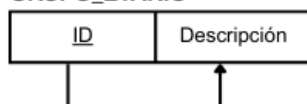
### DEPARTAMENTO



### ESTABLECIMIENTOS\_EDUCATIVOS



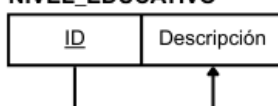
### GRUPO\_ETARIO



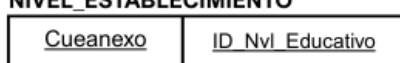
### HABITAN\_EN



### NIVEL\_EDUCATIVO



### NIVEL\_ESTABLECIMIENTO





A continuación, se presentan nuestros esquemas con los datos ya cargados:

#### CENTROS\_CULTURALES

Nombre	Domicilio	ID_Depto	Mail	Capacidad
Oihoy Casa Abierta	14 De Julio 426	02000	oihoy	0
Carpe Diem	15 De Noviembre De 1889 1774	02000	nan	0
Casa Cultural Victor Jara	24 De Noviembre 2273	02000	gmail	80
...	...	...	...	...

#### DEPARTAMENTO

ID	Nombre	Provincia
02000	Ciudad de Buenos Aires	Ciudad de Buenos Aires
06007	Adolfo Alsina	Buenos Aires
...	...	...

#### ESTABLECIMIENTOS\_EDUCATIVOS

Cueanexo	ID_Depto
020000100	02000
020000200	02000
...	...

#### GRUPO\_ETARIO

ID	Descripción
1	0 a 5 años
2	6 a 12 años
3	13 a 18 años
4	19 a 26 años
5	27 a 59 años
6	60 a 110 años

#### HABITAN\_EN

ID_Depto	Area	ID_Grupo_Etario	Casos
02000	Ciudad de Buenos Aires	1	155314
02000	Ciudad de Buenos Aires	2	201989
02000	Ciudad de Buenos Aires	3	245665
02000	Ciudad de Buenos Aires	4	342658
02000	Ciudad de Buenos Aires	5	1448872
02000	Ciudad de Buenos Aires	6	700956
06007	Adolfo Alsina	1	1292
...		...	...

#### NIVEL\_EDUCATIVO

ID	Descripcion
1	Jardín
2	Primaria
3	Secundaria

#### NIVEL\_ESTABLECIMIENTO

Cueanexo	ID_NIVEL
020000100	1
020000100	2
020000100	3
020000300	1
...	...

Estos esquemas fueron completados con la información proveniente de las fuentes de datos mencionadas anteriormente, en el apartado **Introducción**.

Al trabajar con estos esquemas, notamos la importancia de que estos se encuentren en 3FN, pues es por ello que evitamos la pérdida de datos fundamentales; mientras a su vez también eludimos la creación de tuplas espúreas.

## Decisiones Tomadas

Para facilitar nuestro análisis, tomamos una serie de decisiones cruciales.

En primer lugar, decidimos cambiar el formato de presentación del archivo Padrón de Población para evitar nulos innecesarios y facilitar el acceso a los datos. Para esto, seleccionamos las columnas que contenían información de utilidad para nuestro análisis, y volvimos a plantear la presentación, teniendo como resultado una tabla con las columnas: 'Edad'; 'Casos'; 'Area'; y 'ID\_DEPTO'. Además, decidimos descartar la sección 'resumen' la cual contenía la suma de los 'Casos' por edad de todas las áreas, y, como nosotros debíamos enfocarnos en la información de cada departamento, no nos servía para responder a nuestro objetivo.

Debimos también corregir datos de departamentos que no eran compatibles entre las tablas. Decidimos analizar Ciudad de Buenos Aires como un único departamento con código de identificación y nombre únicos y propios. Esto se debe a que el archivo Padrón de Centros Culturales presentaba a Ciudad de Buenos Aires como un único departamento, en contraste a las otras dos fuentes en las cuales se encontraba dividida en diversas comunas. Además, homogeneizamos el código de identificación de los departamentos pertenecientes a Tierra del Fuego, los cuales tampoco coincidían entre las fuentes.

Para poder unificar las fuentes dadas, logramos identificar la presencia de una conexión entre estas tres mediante el código de la localidad, que son 5 dígitos:

- **Padrón Oficial de Establecimientos Educativos:** en la columna denominada "Código de localidad", los primeros 5 dígitos de cada valor ingresado
- **Padrón de Centros Culturales:** en la columna denominada "ID\_DEPTO", el código, será "0" + los 4 dígitos de cada valor ingresado (en caso de que el valor tenga 4 dígitos); o los 5 dígitos que posea el valor si es que tiene dicha cantidad.
- **Datos de población por Departamento (padrón de población):** el código de "Área" determinado para cada registro, cuya longitud son 5 dígitos.

Además tuvimos el compromiso de tomar la partición de los grupos etarios para facilitar el análisis de la población por provincia. En esta ocasión las particiones se realizaron con respecto a los distintos niveles educativos: tomando el grupo 1 de 0 a 5 años; grupo 2 de 6 a 12; grupo 3 de 13 a 18; grupo 4 de 19 a 26; grupo 5 de 27 a 59; y grupo 6, de 60 a 110 . Esta decisión nos favoreció a la hora de graficar, pues al coincidir la edad con el nivel no debimos hacer cambios excedentes.

Dentro de la tabla de Centros Culturales hemos tomado ciertas decisiones para homogeneizar los datos recibidos. Consideramos como mail a aquellos que son del estilo "algo@dominio.tld" , ya que algunos datos no coinciden con el formato general. Además, contamos sólo aquellas direcciones cuyos dominios poseen 4 o más caracteres, para obviar los valores nulos o aquellos que significan sin dirección (s/d) pues los centros que se pierden luego del filtro no son una cantidad significativa.

En cuanto a la tabla de Establecimientos Educativos, decidimos tomar maternas y jardines como un solo nivel -nivel inicial - , al igual que Secundario, Secundario INET, SNU y SNU -

INET - nivel Secundario-. Agregando a lo anterior, los datos utilizados para analizar fueron extraídos únicamente de los establecimientos clasificados con modalidad común.

Finalmente, para sumar a nuestros análisis, hemos hecho una vasta investigación sobre la Ley 26.206 de Educación Nacional y utilizamos gráficos provenientes de la página oficial del INDEC, la cual fue provista como fuente principal por nuestros profesores.

## Análisis de Datos

### Consultas en SQL

A partir de los esquemas presentados en la sección “Procesamiento de Datos”, hemos realizado una serie de reportes a través de consultas de SQL los cuales nos ayudaron a encontrar una relación entre las fuentes y, posteriormente, a graficar.

En primer lugar, presentamos para cada departamento informando su provincia, cantidad de Establecimientos Educativos de cada nivel, considerando solamente la modalidad común, y la cantidad de habitantes por edad según los niveles educativos:

#### RESULTADO\_1:

Provincia	Departamento	Jardines	Población_Jardín	Primarias	Población_Primaria	Secundarios	Población_Secundaria
Chaco	12 de Octubre	23	2262	31	2450	15	2724
Chaco	1° de Mayo	8	1360	12	1577	7	1495
Chaco	2 de Abril	8	745	17	783	5	880
Misiones	25 de Mayo	45	3369	68	3671	36	3903
...	...	...	...	...	...	...	...

Al analizar esta consulta observamos que hay un departamento que no contiene Establecimientos Educativos que cumplan con la condición de modalidad común, este es Tolhuin, ubicado en Tierra del Fuego. Con excepción de este, todos los departamentos restantes los cuales pertenecen a esta tabla tienen, por lo menos, 1 establecimiento educativo en todos los niveles.

Analizando más en profundidad, notamos un contraste entre la cantidad de establecimientos educativos con la población que está en el rango de edad correspondiente. En anexo [\[1\]](#) se encuentran 2 tablas en las cuales se puede observar que, aunque la población secundaria tiene la mayor cantidad de habitantes, es la que menos establecimientos educativos posee. Para visualizar la tabla completa en cuestión, se deberá revisar, dentro de los archivos adjuntos, aquel cuyo nombre sea “RESULTADO\_1.csv”.

Luego, producimos otra consulta la cual para cada departamento debimos informar la provincia y la cantidad de CC con capacidad mayor a 100 personas:

## RESULTADO\_2:

Provincia	Departamento	Cantidad de CC con cap >100
Buenos Aires	Avellaneda	20
Buenos Aires	La Plata	8
Buenos Aires	Lomas de Zamora	3
...	...	...

Dentro de los departamentos con mayor cantidad de CC con capacidad superior a 100 sacando la Ciudad de Buenos Aires se encuentran: Avellaneda, Buenos Aires con 20 ; La Plata, Buenos Aires con 8 ; Capital, Santa Fe con 6 ; y Rosario, Santa Fe con 6.

Esto podría significar que estos departamentos son los que más apuestan a la cultura. Notamos también que hay una gran diferencia entre la cantidad de centros culturales en Avellaneda con respecto a los demás registros. Por otro lado, la *moda* de la cantidad de centros es el 1 ya que aparece 34 veces sobre los 56 departamentos que contienen centros.

Con respecto a las provincias, las que más centros tienen son Buenos Aires, Santa Fe y Córdoba. Sin tener en cuenta a la Ciudad de Buenos Aires la cual contiene 78 centros en total (sumando todas sus comunas).

La tabla completa con los datos incorporados puede encontrarse en el archivo “RESULTADO\_2.csv”, incluido en los materiales adjuntos.

Asimismo, construimos otra consulta en la cual, para cada departamento, indicamos su provincia, cantidad de CC, cantidad de EE (de modalidad común) y población total:

## RESULTADO\_3

Provincia	Departamento	Cant_EE	Cant_CC	Poblacion Total
Ciudad de Buenos Aires	Ciudad de Buenos Aires	1782	296	3.09545e+06
Córdoba	Capital	1136	30	1.49806e+06
Buenos Aires	La Matanza	977	2	1.83717e+06
...	...	...	...	...

Observamos que, sacando a la Ciudad de Buenos Aires, los departamentos que cuentan con más cantidad de EE son: Capital, Córdoba ; La Matanza, Buenos Aires ; Rosario, Santa Fe ; y La Plata, Buenos Aires.

Por otro lado, los que más CC tienen son: La Plata, Buenos Aires ; Avellaneda, Buenos Aires ; Rosario, Santa Fe ; y Capital, Córdoba.

Asimismo, los que más población tienen son: La Matanza, Buenos Aires ; Capital, Córdoba ; Rosario, Santa Fe ; y La Plata, Buenos Aires.

Lo cual demuestra que, en la mayoría de los casos, son los mismos departamentos los que poseen mayores cantidades en los respectivos elementos a considerar. Así, se podría

proponer que, a medida que aumentan la cantidad de establecimientos educativos aumentan la cantidad de centros culturales. En otras palabras, mientras más se apuesta a la educación se tiene más en cuenta el fomento a la cultura. Asimismo, podemos decir que la densidad de población por departamento influye en la cantidad de las instituciones.

También es importante mencionar que se pudo apreciar que todos los departamentos, excepto Tolhuin, ubicado en Tierra del Fuego, tienen establecimientos educativos; aunque no todos tienen centros culturales.

Se recomienda revisar el archivo “*RESULTADO\_3.csv*” para acceder a la versión completa de la tabla mencionada.

Por último, armamos una consulta la cual, para cada departamento, indica provincia y qué dominios de mail se usan con mayor frecuencia para los centros culturales:

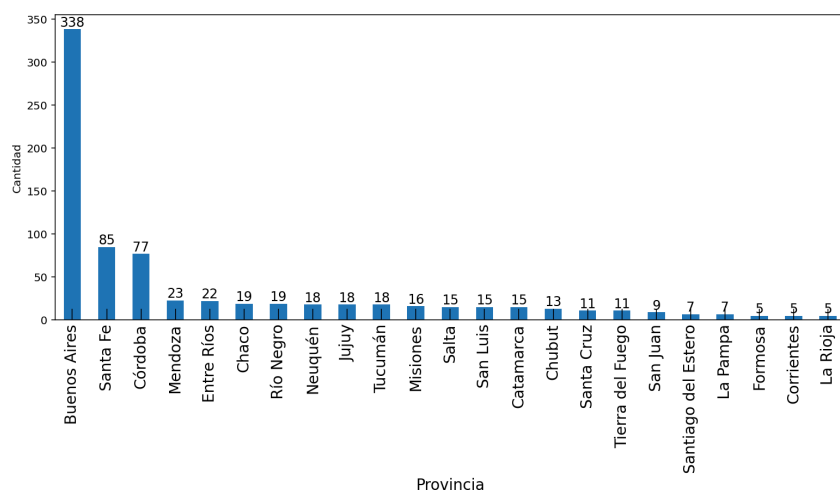
#### RESULTADO\_4

Provincia	Departamento	Dominio más frecuente en CC
Buenos Aires	9 de Julio	hotmail
Buenos Aires	Adolfo Alsina	adolfoalsina
...	...	...

Aquí se puede apreciar que muchos Centros Culturales no usan correo electrónico como contacto, pues en muchos de los departamentos el valor más frecuente en su respectiva celda suele ser “s/d”, “-”, o un valor nulo; y que dentro de los que contienen dirección, el dominio más utilizado es “gmail”.

La tabla completa se ha guardado en el archivo “*RESULTADO\_4*”, el cual se encuentra disponible entre los documentos anexos.

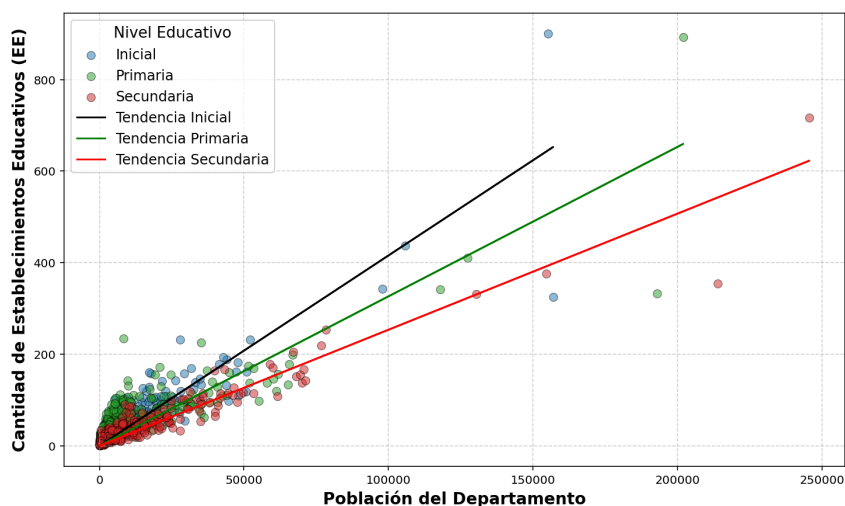
## Herramientas de visualización



### *Cantidad de centros culturales por provincia*

Este gráfico muestra la cantidad de centros culturales por provincia. Observamos una gran diferencia entre Buenos Aires y el resto, aunque Santa Fe y Córdoba también son significativas.

Desde nuestro punto de vista hay una relación entre la cantidad de centros culturales con la densidad de la población ya que Buenos Aires, Córdoba, Santa Fe y Mendoza son de las provincias con más flujo de personas en nuestro país como se muestra en el gráfico anexoado[2] .

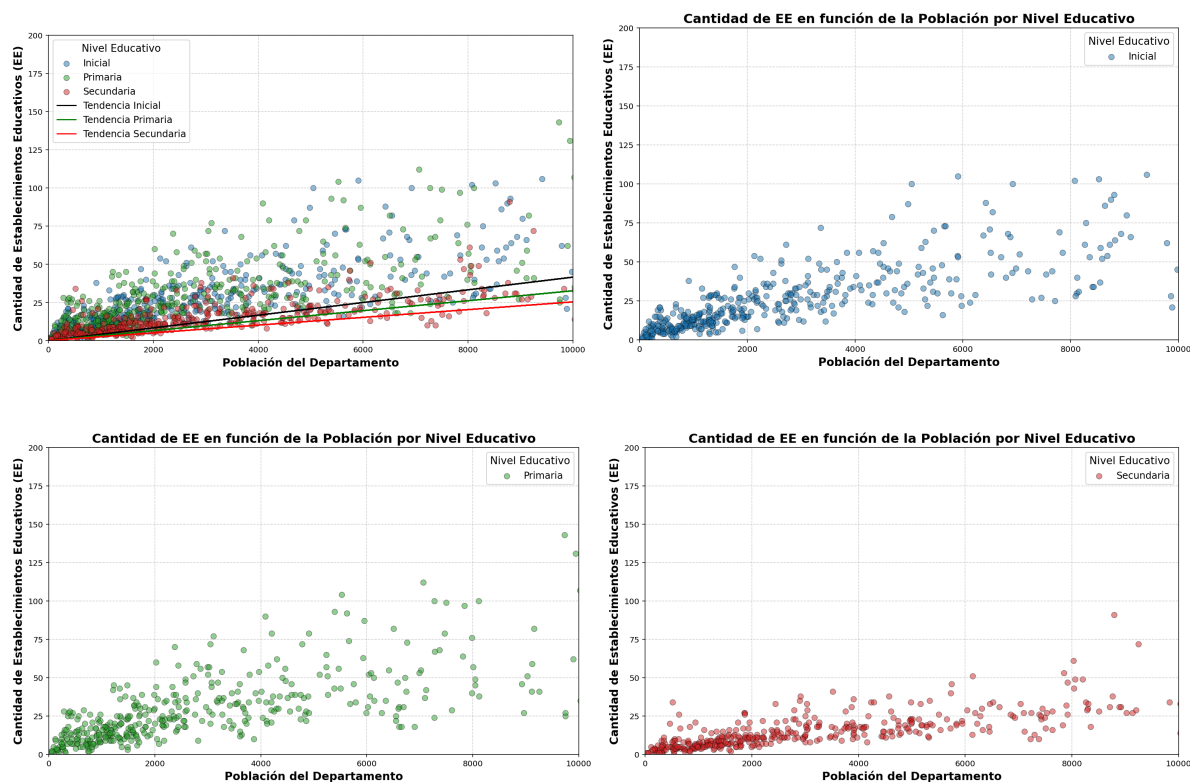


### *Cantidad de Establecimientos Educativos en función de su población separado por niveles sin acotar*

En este gráfico, se puede apreciar la cantidad de Establecimientos Educativos de departamentos en función de su población separado por niveles y su correspondiente grupo etario. En nuestro caso, optamos por separar únicamente por nivel educativo dado que el grupo etario 1 corresponde a la edad en que las personas van al jardín (0 a 5 años), el 2, al que van al nivel primario (6 a 12 años) y el 3 al que van al secundario (13 a 18 años).



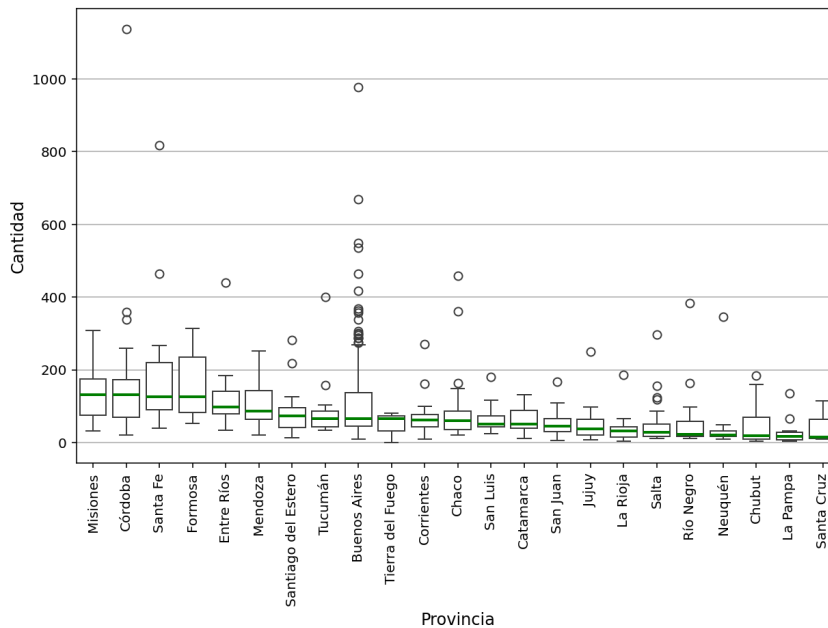
Además, a continuación, realizamos un gráfico aparte, en donde acotamos la población a 10000 y la cantidad de EE por 200 para ver con más claridad la dispersión de los puntos en la zona donde se encuentra mayor concentración de los mismos.



1. Cantidad de Establecimientos Educativos en función de su población separado por niveles acotado
2. Cantidad de Establecimientos Educativos de Nivel Inicial en función de la población del departamento
3. Cantidad de Establecimientos Educativos de Nivel Primario en función de la población del departamento
4. Cantidad de Establecimientos Educativos de Nivel Secundario en función de la población del departamento

Con estos datos podemos ver que la pendiente del Nivel Educativo “Secundaria” es menor al de “Primaria” e “Inicial”, lo que indica que hay proporcionalmente menos EE de secundaria respecto a la población en comparación con los otros niveles. La pendiente que hace referencia a el Nivel Inicial es mayor a la del Nivel Primario y Secundario. Esto indica que, a medida que crece la población, el número de EE de Nivel Inicial aumenta con más velocidad en comparación con el primario y secundario.

Por otro lado, hay más EE de Nivel Inicial y Primario que de Secundario. Nosotros consideramos que esta situación se podría dar por la inasistencia a los secundarios respecto a los otros niveles, puesto que, en este grupo etario, se pueden encontrar habitantes que se encuentran en edad legal para trabajar, lo que puede ser un factor importante a considerar para que se produzca este fenómeno (decrecimiento de la cantidad de habitantes que se centran en terminar este nivel educativo). Esto se puede observar claramente en el gráfico anexoado [\[3\]](#).



*Boxplot por cada provincia, de la cantidad de EE por cada departamento de la provincia ordenados por la mediana.*

En este gráfico, se presenta de manera clara la comparación de la cantidad de EE acumulada en cada departamento, por provincia.

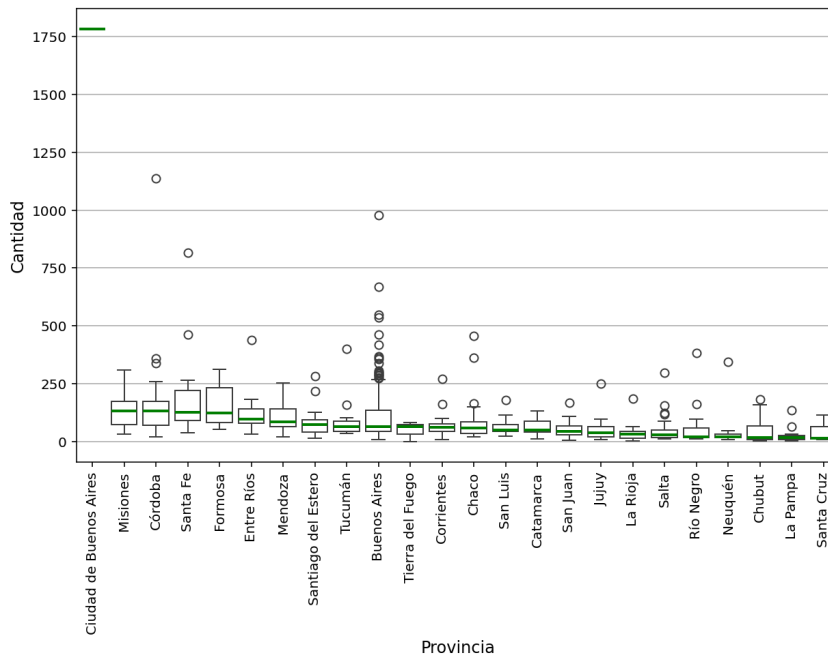
Las provincias con mayor *mediana* por departamento son Misiones, Córdoba, Santa Fe y Formosa. Mientras que las menores son Neuquén, Chubut, La Pampa y Santa Cruz.

Buenos Aires presenta una alta dispersión de EE entre sus departamentos debido a que hay una amplia brecha de cantidad entre unos y otros.

Por otro lado, La Pampa y Neuquén tienen una distribución de los EE más compacta, lo que indica que la cantidad de EE es más homogénea entre sus departamentos, y es por esto que su caja asociada es más plana.

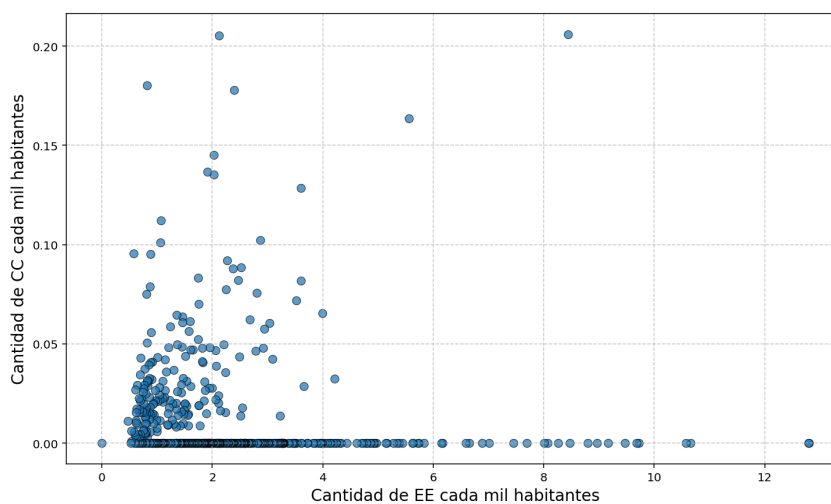
Se presentan abundantes outliers (puntos fuera de los bigotes), principalmente en Misiones, Santa Fe y Buenos Aires. Esto indica que existen departamentos con muchos más establecimientos educativos que la mayoría de los otros dentro de su respectiva provincia.

Desde nuestro punto de vista, estos puntos pueden estar asociados a que, en esos departamentos, hay una gran aglomeración de gente; y, por lo tanto, hay más EE. Por ejemplo, La Matanza, cuenta con 977 establecimientos, mientras que, Tres Lomas, que también forma parte de Buenos Aires, cuenta con 19 establecimientos; pero en este viven 9090 personas, en tanto que en La Matanza hay 1.83717e+06 habitantes.



*Boxplot por cada provincia, de la cantidad de EE por cada departamento de la provincia ordenados por la mediana incluyendo la Ciudad de Buenos Aires.*

Ahora, teniendo en cuenta la Ciudad de Buenos Aires, se puede observar que la mediana de esta es muy alta en comparación a la de las otras provincias. Es importante recordar que su caja es compacta porque unificamos todas las comunas de la ciudad, teniendo así un único valor total. Es por esto que el gráfico no refleja la dispersión por comunas.

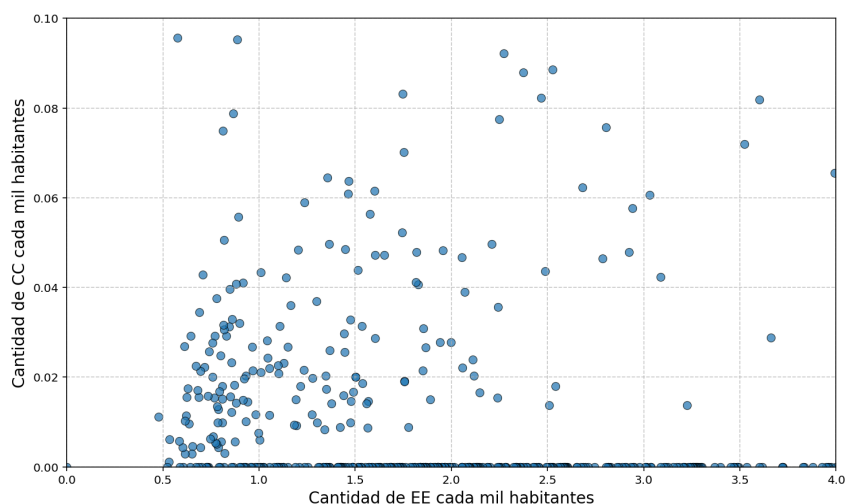


*Relación entre la cantidad de CC cada mil habitantes y de EE cada mil habitantes sin acotar*

En este gráfico se observa la proporción de EE y CC cada mil habitantes, por departamento, en donde cada punto hace referencia a un departamento en particular. Aquí se aprecia que es usual que en los departamentos no haya presencia de centros culturales; o que la proporción de cada mil habitantes de esta sea muy escasa; mientras que la de Establecimientos Educativos resulta ser superior, pero en muy pocos casos es excesivamente alta (pocos outliers). Dos de ellos pueden ser, en Formosa, El Alto y Ancastí, los cuales su proporción es de 12.8 y 12.79, respectivamente.

Resulta también interesante agregar que se presentan algunos puntos en los cuales la proporción de CC de cada mil habitantes es considerablemente grande a lo que la mayoría de los puntos poseen, entre los que se destacan Minas, ubicada en la provincia de Córdoba; o también Tilcara, localizada en Jujuy, cuyas proporción son en ambas de 0.20.

Para analizar un poco más en detalle este esquema, recurrimos a acotar los límites de los ejes, puesto que vimos que la mayoría de los puntos (departamentos) se encontraban entre unas proporciones similares. A continuación, se presenta el gráfico con sus ejes acotados.



#### *Relación entre la cantidad de CC cada mil habitantes y de EE cada mil habitantes acotado*

En este gráfico notamos que en todos los casos hay mayor cantidad de establecimientos educativos que de centros culturales, es interesante resaltar que la variable correspondiente a la cantidad de CC cada mil habitantes se mueve de 0.0 a 0.10, mientras que la que se corresponde a la cantidad de EE cada mil habitantes se encuentra entre 0.0 y 4.0 lo que marca una gran diferencia entre los límites elegidos.

Además, se puede observar que la mayoría de los puntos se concentran en valores bajos de centros culturales y establecimientos educativos lo que sugiere que en la mayoría de los casos ambas variables son bajas en relación con la población.

También notamos la existencia de valores atípicos que pueden estar influenciados por políticas culturales específicas o una alta concentración poblacional en ciertas regiones, y, continuando con la misma deducción, existen algunos departamentos con valores medios de EE y una cantidad de CC relativamente alta, lo que indica posibles diferencias en la planificación cultural de cada departamento.

Para profundizar esta investigación, graficamos la relación entre la cantidad de CC y de EE por departamento, lo cual nos ayudó a responder a nuestro objetivo general. Dichos gráficos se encuentran anexados [\[4\]](#). Al armarlos, observamos claramente la relación entre la cantidad de centros culturales con respecto a la de establecimientos educativos, sin tener en cuenta la densidad de población.

Aquí volvemos a notar que hay puntos atípicos donde hay muchos centros culturales con respecto a otros departamentos con valores similares en la otra variable. Lo que justifica nuestra propuesta de que la cantidad de centros culturales depende del fomento y la importancia que le da cada departamento.

Además, los departamentos que cuentan con más centros culturales son los que también cuentan con muchos establecimientos educativos. Lo que sostiene la idea que los departamentos que le dan más importancia a la educación también le otorgan más entidad y relevancia a la cultura.

Dado a que los datos que estamos trabajando son actuales (2022) a nuestro parecer estos departamentos deberían ser un ejemplo para los otros departamentos del país, así este comportamiento dejaría de ser un caso atípico.

Luego de realizar las consultas y gráficos, notamos una diferencia muy fuerte entre la cantidad de Establecimientos Educativos con respecto a la cantidad de Centros culturales. Buscando la causa de este fenómeno, encontramos que existe una ley, la 26.206 de Educación Nacional, la cual presentamos en el anexo [\[5\]](#) . Haciendo hincapié en el artículo 1 y 4 de la misma, como la educación es un derecho, se explica la razón por la cual existe, por lo menos, un Establecimiento Educativo por departamento, siendo esta la causa de la proporción entre la cantidad de establecimientos educativos y la población.

Con respecto a Tolhuin, este contiene instituciones educativas de modalidad distinta a la considerada en el análisis (modalidad común), es por eso que no se encuentran datos sobre sus instituciones.

Por otro lado, no encontramos una Ley Nacional que regule los Centros Culturales, lo que explica la gran dispersión respecto de su número, creación y pertenencia. Esto indica que la cantidad de estos centros depende de la importancia que en cada provincia se les otorgue.

## Conclusión

Luego de haber realizado este vasto informe con el fin de hallar la existencia de algún tipo de relación entre la cantidad de establecimientos educativos en cada una de las provincias, y la cantidad de centros culturales, hemos encontrado una serie de conclusiones que nos parecen importantes reportar:

- Hallamos que todo departamento de la República Argentina posee, al menos, un Establecimiento Educativo; pero no siempre un Centro Cultural.
- En cada departamento se presenta una mayor cantidad de Establecimientos Educativos que Centros Culturales.
- La cantidad de establecimientos educativos es proporcional a la densidad de la población por departamento.
- A medida que aumenta la población por provincia, también aumentan los centros culturales.
- Los departamentos más desarrollados son los que tienen más población, establecimientos educativos y centros culturales.

Asimismo, hemos logrado realizar ciertas deducciones a partir de nuestra investigación que responden de manera implícita el por qué de nuestros resultados, agregando información relevante para entender los números de Centros Culturales y Establecimientos Educativos:

- Existen casos atípicos de departamentos donde le otorgan más entidad y relevancia a la cultura. La cantidad de centros culturales depende no solo de la población y de la cantidad de establecimientos educativos, sino también del fomento y la importancia que le otorga cada departamento a estos.
- Hay provincias cuya distribución de departamentos respecto a la cantidad de establecimientos educativos es homogénea.
- A pesar del hecho de que hay más habitantes dentro del rango etario correspondiente al nivel educativo 3 (nivel secundario) en contraste a los demás niveles, encontramos abundantes jardines y primarias en comparación a la cantidad de secundarias.
- Al haber una ley que ampara el derecho a la educación, se verifica la existencia de por lo menos un establecimiento educativo por departamento accesible a los habitantes. Mientras que como los centros culturales no tienen una ley nacional que los regule su existencia depende de la importancia que cada región le otorgue.

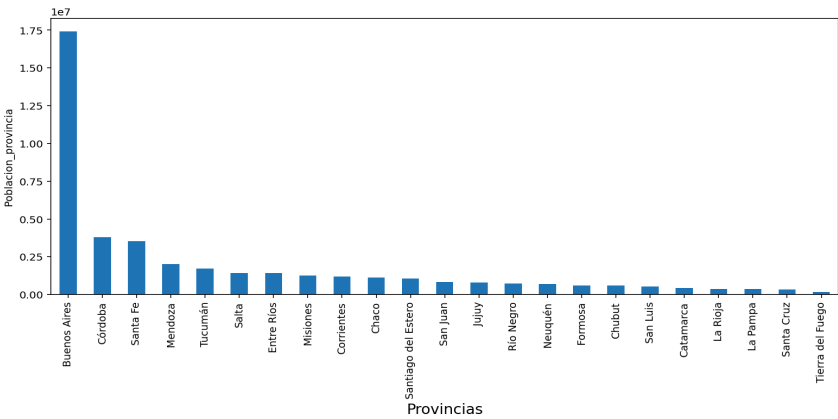
1. Ministerio de Capital Humano, área de Educación, Información y Evaluación Educativa. (2022). *Establecimientos Educativos (EE)*. Padrón Oficial De Establecimientos Educativos 2022. Retrieved February 20, 2025, from <https://www.argentina.gob.ar/educacion/evaluacion-e-informacion-educativa/padronoficial-de-establecimientos-educativos>
2. Sistema de Información Cultural de la Argentina - Ministerio de Cultura de la Nación. (2022, September 9). *Centros culturales (2022)*. Retrieved February 20, 2025, from [https://datos.gob.ar/dataset/cultura-mapa-cultural-espacios-culturales/archivo/cultura\\_0e9a431c-b4f7-455b-aa1a-f419b5740900](https://datos.gob.ar/dataset/cultura-mapa-cultural-espacios-culturales/archivo/cultura_0e9a431c-b4f7-455b-aa1a-f419b5740900)
3. Instituto Nacional de Estadística y Censos [INDEC]. (2022). *Datos de población por departamento*. Censo 2022. Retrieved February 20, 2025, from <https://www.indec.gob.ar/indec/web/Nivel4-Tema-2-41-165>  
De esta fuente hemos utilizado el recorte de datos proveído por nuestros profesores e investigando, hemos encontrado gráficos que nos ayudaron a llegar a nuestra conclusión final.
4. Ministerio de Justicia de La Nación [InfoLEG - información legislativa]. (2006, December 27). *LEY DE EDUCACIÓN NACIONAL*. Disposiciones Generales. Sistema Educativo Nacional. Educación De Gestión Privada. Docentes Y Su Formación. Políticas De Promoción De La Igualdad Educativa. Calidad De La Educación. Educación, Nuevas Tecnologías Y Medios De Educación. Educación a Distancia Y No Formal. Gobierno Y Administración. Cumplimiento De Los Objetivos De La Ley. Disposiciones Transitorias Y Complementarias. Retrieved February 21, 2025, from <https://servicios.infoleg.gob.ar/infolegInternet/anexos/120000-124999/123542/texact.htm>

[1]

Poblacion_Jardin	Poblacion_Primaria	Poblacion_Secundaria
3.52702e+06	4.36574e+06	5.02582e+06

Jardines	Primarias	Secundarias
21159	22872	16098

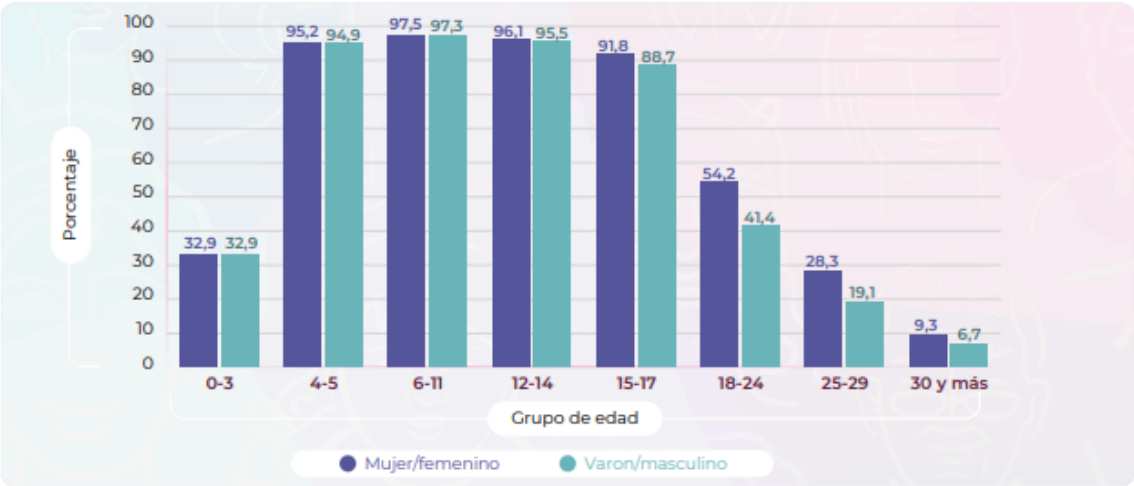
[2]



Población por provincias.

[3]

Gráfico 4. Población en viviendas particulares que asiste a un establecimiento educativo, por sexo registrado al nacer, según grupo de edad. Total del país. Año 2022



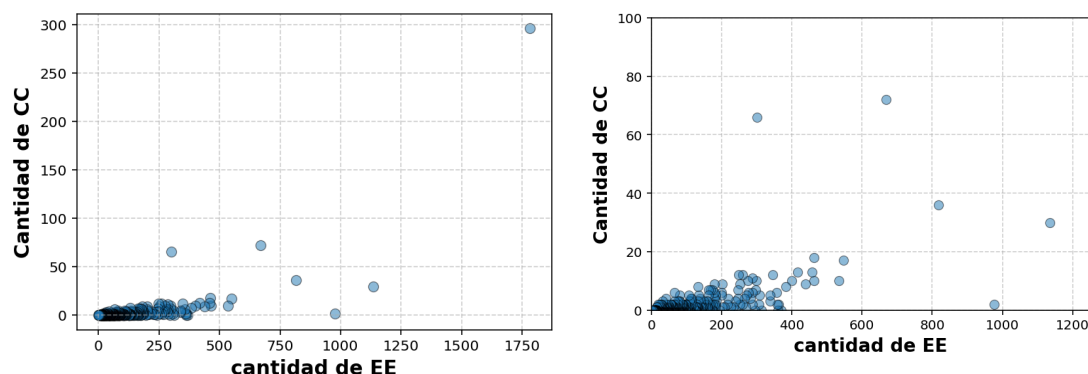
**Nota:** de acuerdo con la evaluación de calidad y consistencia de los resultados definitivos, y para cumplir con los estándares de calidad estadística requeridos por el INDEC, la categoría X de las respuestas a la pregunta por sexo registrado al nacer se distribuye entre las categorías Mujer/femenino y Varón/masculino.

**Fuente:** INDEC, Censo Nacional de Población, Hogares y Viviendas 2022. Resultados definitivos.

retirado de la página del censo publicada por el INDEC



[4]



1. Relación entre la cantidad de CC y de EE por departamento sin acotar
2. Relación entre la cantidad de CC y de EE por departamento acotado

[5]

## LEY DE EDUCACIÓN NACIONAL

“ARTÍCULO 1º — La presente ley regula el ejercicio del derecho de enseñar y aprender consagrado por el artículo 14 de la Constitución Nacional y los tratados internacionales incorporados a ella, conforme con las atribuciones conferidas al Honorable Congreso de la Nación en el artículo 75, incisos 17, 18 y 19, y de acuerdo con los principios que allí se establecen y los que en esta ley se determinan.

ARTÍCULO 2º — La educación y el conocimiento son un bien público y un derecho personal y social, garantizados por el Estado.

ARTÍCULO 3º — La educación es una prioridad nacional y se constituye en política de Estado para construir una sociedad justa, reafirmar la soberanía e identidad nacional, profundizar el ejercicio de la ciudadanía democrática, respetar los derechos humanos y libertades fundamentales y fortalecer el desarrollo económico-social de la Nación.

ARTÍCULO 4º — El Estado nacional, las provincias y la Ciudad Autónoma de Buenos Aires tienen la responsabilidad principal e indelegable de proveer una educación integral, permanente y de calidad para todos/as los/as habitantes de la Nación, garantizando la igualdad, gratuidad y equidad en el ejercicio de este derecho, con la participación de las organizaciones sociales y las familias.

ARTÍCULO 5º — El Estado nacional fija la política educativa y controla su cumplimiento con la finalidad de consolidar la unidad nacional, respetando las particularidades provinciales y locales.

ARTÍCULO 6º — El Estado garantiza el ejercicio del derecho constitucional de enseñar y aprender. Son responsables de las acciones educativas el Estado nacional, las provincias y la Ciudad Autónoma de Buenos Aires, en los términos fijados por el artículo 4º de esta ley; los municipios, las confesiones religiosas reconocidas oficialmente y las organizaciones de la sociedad; y la familia, como agente natural y primario.

ARTÍCULO 7° — El Estado garantiza el acceso de todos/as los/as ciudadanos/as a la información y al conocimiento como instrumentos centrales de la participación en un proceso de desarrollo con crecimiento económico y justicia social.

ARTÍCULO 8° — La educación brindará las oportunidades necesarias para desarrollar y fortalecer la formación integral de las personas a lo largo de toda la vida y promover en cada educando/a la capacidad de definir su proyecto de vida, basado en los valores de libertad, paz, solidaridad, igualdad, respeto a la diversidad, justicia, responsabilidad y bien común.

ARTÍCULO 9° — El Estado garantiza el financiamiento del Sistema Educativo Nacional conforme a las previsiones de la presente ley. Cumplidas las metas de financiamiento establecidas en la Ley N° 26.075, el presupuesto consolidado del Estado nacional, las provincias y la Ciudad Autónoma de Buenos Aires destinado exclusivamente a educación, no será inferior al SEIS POR CIENTO (6%) del Producto Interno Bruto (PIB).

ARTÍCULO 10. — El Estado nacional no suscribirá tratados bilaterales o multilaterales de libre comercio que impliquen concebir la educación como un servicio lucrativo o alienten cualquier forma de mercantilización de la educación pública.[...]"