



Trabajo práctico 01

Verano 2025

12 de marzo de 2025

Laboratorio de Datos

Grupo: "Pepitas"

Integrante	LU	Correo electrónico
Caporaletti, Gonzalo	707/22	gonzalocaporaletti@gmail.com
Gonzalez Dardik, Micaela Natali	1143/22	micagonzdark@gmail.com



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>

Índice

1. Resumen	2
2. Introducción	2
3. Procesamiento de Datos	2
3.1. Formas Normales	2
3.1.1. Evaluación de 1FN en las bases de datos proporcionadas	3
3.1.2. Evaluación de 2FN y 3FN en las bases de datos proporcionadas	3
3.2. Análisis de calidad de las bases de datos	3
3.2.1. Fuente: centros_culturales.csv	3
3.2.2. Fuente: 2022_padron_oficial_establecimientos_educativos.xlsx	3
3.2.3. Fuente: padron_poblacion.csv	4
3.3. Decisiones Tomadas	4
3.4. Modelado del Diagrama de Entidad-Relación	5
3.5. Modelado Relacional a partir del Diagrama de Entidad-Relación	5
3.6. Extracción de Datos	6
4. Análisis de Datos	7
4.1. Consultas SQL	7
4.1.1. Establecimientos Educativos y Población por Nivel para Cada Departamento	7
4.1.2. Centros Culturales con Capacidad Mayor a 100 Personas por Departamento	7
4.1.3. Centros Culturales, Establecimientos Educativos y Población por Departamento	8
4.1.4. Dominios de Correo Electrónico más Usados por los Centros Culturales	8
4.2. Visualización de Datos	9
4.2.1. Cantidad de centros culturales por provincia	9
4.2.2. Distribución de Establecimientos Educativos por provincia y grupo etario	10
4.2.3. Distribución de EE por departamento, para cada provincia	11
4.2.4. Gráfico de Dispersión que relaciona la cantidad de EE y la cantidad de CC	12
5. Conclusiones	12
A. Anexo	13

1. Resumen

Este proyecto analiza la distribución de **Establecimientos Educativos (EE)** y **Centros Culturales (CC)** en Argentina, con el fin de identificar patrones y determinar, si es que existe, una posible relación entre ambos tipos de instituciones en las distintas regiones del país.

Se recopilaron conjuntos de datos obtenidos de fuentes oficiales y se realizó un análisis **cuantitativo y territorial**, tanto a nivel **departamental** como **provincial**. Se diseñó un **modelo relacional** para manejar las bases de datos necesarias, se llevaron a cabo consultas **SQL** para explorar patrones y se presentaron los resultados mediante **visualizaciones gráficas**.

Los resultados muestran que **no existe una relación proporcional clara entre la cantidad de establecimientos educativos y centros culturales en Argentina**; mientras que los primeros se distribuyen homogéneamente en función de la población escolar, los centros culturales dependen principalmente de la urbanización y concentración económica.

2. Introducción

La educación y la cultura son **pilares fundamentales** para el desarrollo social. Analizar su distribución podría servir para detectar **desigualdades** en el acceso a estos derechos y contribuir a la formulación de **políticas públicas más equitativas**. En Argentina, la distribución de **Establecimientos Educativos (EE)** y **Centros Culturales (CC)** varía ampliamente según la región, lo que podría generar **brechas en las oportunidades** de formación y acceso a la cultura.

El objetivo del trabajo es determinar si la cantidad de **EE y CC** en las provincias argentinas está relacionada y qué factores pueden influir en su distribución. Aunque no se dispone de información **socioeconómica detallada**, se explorarán posibles relaciones a partir de diferencias regionales, considerando variables como la **densidad poblacional** y el **grado de urbanización**.

Para ello, se analizaron bases de datos **oficiales** sobre EE y CC, junto con información poblacional de **censos nacionales**. Comprender su distribución puede contribuir a identificar **desigualdades existentes** y aportar información relevante para la toma de decisiones en políticas públicas.

El análisis comenzará con la **recopilación y exploración** de conjuntos de datos provenientes de **fuentes oficiales**, evaluando su **contenido y estructura**. A continuación, se diseñará un **modelo de base de datos** mediante un **Diagrama Entidad-Relación (DER)** y se construirá un **modelo relacional** que facilite el análisis. Para garantizar la **calidad de los datos**, se aplicarán técnicas de **limpieza y validación** que aseguren su confiabilidad.

Mediante **consultas SQL**, se podrán estudiar temas particulares, y así explorar posibles **relaciones** entre la distribución de los establecimientos y **factores poblacionales y socioeconómicos**. El documento finalizará presentando los **hallazgos obtenidos**, complementados con **visualizaciones gráficas** que permitirán evaluar **tendencias y diferencias** en la cobertura educativa y cultural en el país.

3. Procesamiento de Datos

Para la realización de este trabajo, se utilizaron las siguientes bases de datos:

- **centros_culturales.csv**: Padrón de Centros Culturales de la República Argentina.
- **2022_padron_oficial_establecimientos_educativos.xlsx**: Padrón Oficial de Establecimientos Educativos 2022 de la República Argentina.
- **padron_poblacional.xlsx**: Datos de población por departamento de la República Argentina.

Cabe destacar que estas bases de datos provienen de distintos organismos del Gobierno Nacional, lo que implica que su estructura, códigos y estándares no sean homogéneos.

La primera problemática a abordar fue la **integración y normalización de los datos**. Dado que las bases originales no cumplían con los principios de normalización, se diseñó un **modelo de base de datos** propio, extrayendo y reorganizando la información relevante. A partir de este modelo, se generaron **dataframes** en **Tercera Forma Normal (3FN)**, asegurando una estructura optimizada para el análisis.

Además, se llevó a cabo un proceso de **filtrado y mejora de calidad**, eliminando datos redundantes o irrelevantes. Para esto, se desarrollaron **diagramas y modelos** que proporcionaron una comprensión clara de la información necesaria y facilitaron la creación de una base de datos estructurada y eficiente.

3.1. Formas Normales

Las formas normales establecen reglas para estructurar bases de datos minimizando redundancias y asegurando integridad:

- **Primera Forma Normal (1FN)**: Garantiza datos atómicos y la existencia de una clave primaria.
- **Segunda Forma Normal (2FN)**: Asegura que todos los atributos dependan completamente de una clave.
- **Tercera Forma Normal (3FN)**: Elimina dependencias transitivas para optimizar la estructura.

A continuación, se explica el análisis y la transformación realizada a las bases de datos de Centros Culturales y Establecimientos Educativos para alcanzar la 3FN. Una vez normalizadas las bases, se elaborará un Diagrama de Entidad-Relación (DER), y su respectivo Modelo Relacional para representar la estructura final del sistema.

3.1.1. Evaluación de 1FN en las bases de datos proporcionadas

Ambas bases de datos analizadas violaban **1FN** debido a la falta de atomicidad en ciertos atributos:

- **centros_culturales.csv**: La columna **Mail** almacenaba múltiples correos en una misma celda.
- **establecimientos_educativos.csv**: Columnas como **Modalidad** y **Teléfono** contenían valores separados por `/.º`.

3.1.2. Evaluación de 2FN y 3FN en las bases de datos proporcionadas

Dado que ninguna cumple con **1FN**, y este es un requisito indispensable para alcanzar formas de normalización superiores, queda descartado el cumplimiento de **2FN** y **3FN**.

3.2. Análisis de calidad de las bases de datos

Para evaluar la calidad de los datos, utilizamos la técnica **GQM** (Goal-Question-Metric), que permite definir objetivos, plantear preguntas clave y establecer métricas cuantitativas.

3.2.1. Fuente: centros_culturales.csv

Problema: Inconsistencias en Provincia, Departamento, Localidad

- **Atributo afectado:** Consistencia.
- **Tipo de problema:** Instancia, con posible impacto en el modelo.
- **GQM:**
 - **Objetivo:** Evaluar la estandarización en Provincia, Departamento y Localidad.
 - **Preguntas:**
 - ¿Qué porcentaje de registros presenta variaciones en nombres de Provincia/Departamento/Localidad?
 - **Métricas:**
 - $M1 = 100 \times \frac{\text{Deptos con múltiples nombres}}{\text{Total deptos}} = 100 \times \left(\frac{2}{1067}\right) \approx 0,19\%$.
 - $M2 = 100 \times \frac{\text{Provincias con múltiples nombres}}{\text{Total provincias}} = 0\%$.
 - $M3 = 100 \times \frac{\text{Localidades con múltiples nombres}}{\text{Total localidades}} = 100 \times \left(\frac{19}{1067}\right) \approx 1,8\%$.
- **Corrección:**
 - Establecer un único valor para cada una de las claves. Este valor será el correspondiente a la **moda** de los valores asociados a la clave (el valor que más se repite).
- **Impacto en la calidad:**
 - Luego de aplicar los criterios de corrección, el impacto en la calidad pasa a ser nulo, ya que es posible corregir todos los errores.

3.2.2. Fuente: 2022_padron_oficial_establecimientos_educativos.xlsx

Problema: Celdas vacías en niveles educativos (modalidad “Común”)

- **Atributo afectado:** Completitud.
- **Tipo de problema:** Instancia.
- **GQM:**
 - **Objetivo:** Evaluar y corregir datos faltantes en niveles educativos.
 - **Pregunta:** ¿Cuántas celdas vacías hay en la sección de niveles educativos?
 - **Métrica:**
 - $M1 = \frac{\text{celdas vacías en la columna "Nivel Educativo"}}{\text{total de celdas en la columna "Nivel Educativo"}}$
 - Calculado con el siguiente código:

```

cantidad_total_celdas = ee[columnas].size
cantidad_total_nans = ee[columnas].isna().sum().sum()

proporcion_nans = cantidad_total_nans / cantidad_total_celdas

```

◦ Resultado obtenido: $M_1 = 85.23\%$

- **Corrección:** Reemplazo de valores NULL por s/d.
- **Impacto:** El impacto es inicialmente alto, reducido a nulo tras corrección.

3.2.3. Fuente: padron_poblacion.csv

Problema: Formato y estructura inconsistentes

- **Atributo afectado:** Consistencia.
- **Tipo de problema:** Modelo.
- **GQM:**
 - **Objetivo:** Evaluar la calidad y consistencia de la estructura de datos, asegurando que cada fila contenga información válida y coherente en las columnas correspondientes.
 - **Pregunta:** ¿Cuántas filas contenían valores de área y comuna en la columna `.Edad` antes del proceso de limpieza?

- $M_1 = \frac{\text{filas con valores de área en la columna .Edad}}{\text{total de filas en el dataset}}$
- Calculado con el siguiente código:

```

# Contamos las filas incorrectas
filas_incorrectas = df[df["Edad"].astype(str).str.contains("AREA #")].shape[0]

# Calculamos la proporción de filas incorrectas
proporcion_incorrectas = filas_incorrectas / df.shape[0]

```

◦ Resultado obtenido: $M_1 = 0.95\%$.

- **Corrección:**
 - Se añadieron nuevas columnas para volcar la información que aparecía como título en la columna de `.Edad`; se recorrió el dataset para asignar correctamente esta información que no estaba en la columna correcta.
 - Se estableció un criterio para que la columna `.Edad` contuviera únicamente valores numéricos. Con esto se eliminaron filas incorrectas como aquellas que en la columna Edad decían 'Totales', y era información irrelevante.
- **Impacto:** Inicialmente alto, nulo con la corrección.

3.3. Decisiones Tomadas

- **Comunas de la Ciudad Autónoma de Buenos Aires:** En el análisis, se decidió NO tratar a las comunas de la Ciudad Autónoma de Buenos Aires (C.A.B.A.) como departamentos independientes, sino agruparlas dentro de un único departamento/provincia, que corresponde a la C.A.B.A. Aunque las comunas no son técnicamente departamentos y la propia C.A.B.A. tampoco lo es, se evaluó cómo deberían clasificarse dentro del contexto de la base de datos. En las bases de datos que contenían información sobre centros culturales, los centros de la C.A.B.A. no tenían ninguna distinción clara en cuanto al departamento o comuna a la que pertenecían. Dado que la base de datos de centros culturales solo incluye las coordenadas geográficas, sin especificar departamentos o comunas, se tomó la decisión de agrupar todas las comunas de C.A.B.A. bajo un único identificador, simplificando así la clasificación y manteniendo la coherencia en los datos.
- Se interpretó que el **valor 0 en la capacidad de los Centros culturales** indica que no se tiene dicha información, por lo que este valor fue reemplazado por s/d
- Se decidió que los **Centros culturales con capacidad 's/d'** serían descartados del conteo de Centros culturales con capacidad mayor a 100 personas
- Los **Departamentos que no cuenten con Centros culturales** fueron descartados de la tabla con dominios de correos electrónicos más frecuentes.
- Los **Centros Culturales sin información de mail disponible** fueron descartados de la tabla con dominios de correos electrónicos más frecuentes. Por ende, si un departamento tiene centros culturales, pero dichos centros culturales no tienen información de mail disponible, este departamento queda fuera de la tabla.

3.4. Modelado del Diagrama de Entidad-Relación

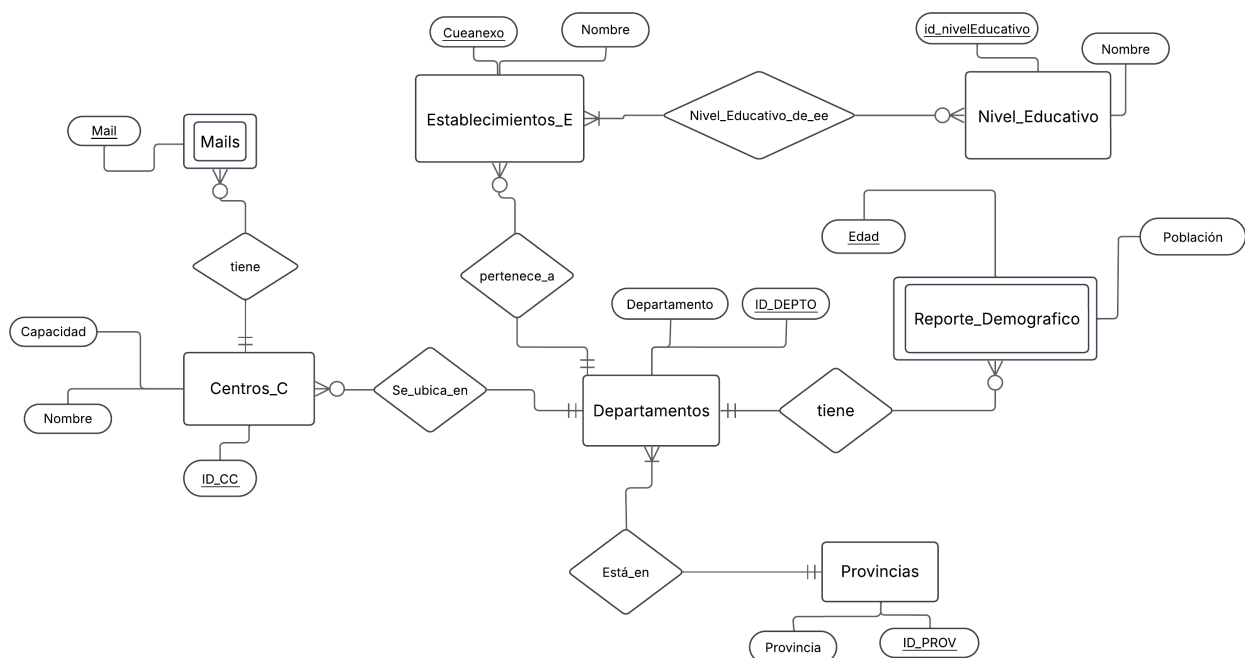


Figura 1: Diagrama de Entidad-Relación

En el diseño del Diagrama Entidad-Relación (DER) (ver Figura 1), identificamos las siguientes entidades: **Centros_C**, **Departamentos**, **Establecimientos_E**, **Mails**, **Nivel_Educativo**, **Nivel_Educativo_de_ee**, **Reporte_Demográfico** y **Provincias**.

A partir de estas entidades, realizamos el correspondiente Modelo Relacional (ver Figura 2).

3.5. Modelado Relacional a partir del Diagrama de Entidad-Relación

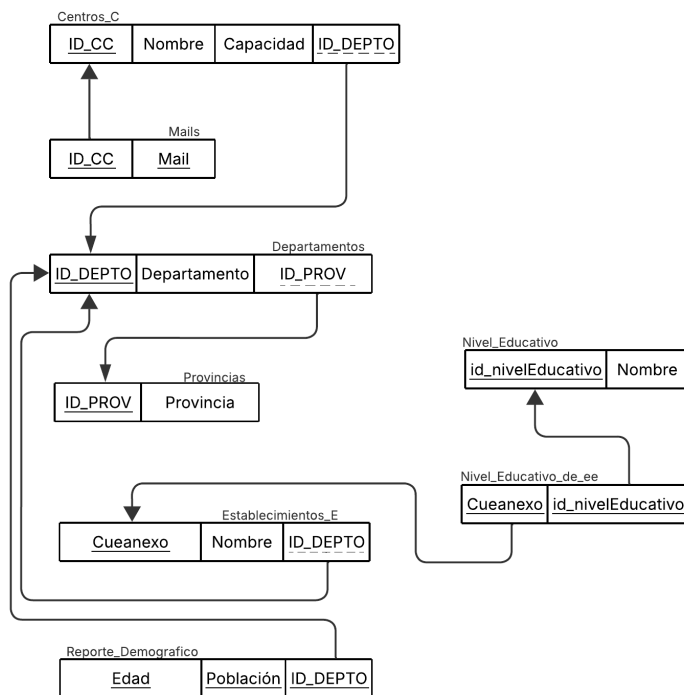


Figura 2: Modelado Relacional

A continuación, se describen los atributos de cada entidad, indentificando sus claves primarias (PK) y claves foraneas (FK):

- **Centros_C:**
 - ID_CC: Identificador único del centro cultural. **(PK)**
 - Nombre: Nombre del centro cultural.
 - Capacidad: Capacidad máxima de personas en el centro cultural.
 - ID_DEPTO: Identificador del departamento al que pertenece el centro cultural. **(FK)**
- **Mails:**
 - ID_CC: Identificador del centro cultural asociado al correo electrónico. **(PK, FK)**
 - Mail: Dirección de correo electrónico del centro cultural. **(PK)**
- **Establecimientos_E:**
 - Cueanexo: Código único del establecimiento educativo. **(PK)**
 - Nombre: Nombre del establecimiento educativo.
 - ID_DEPTO: Identificador del departamento donde se encuentra el establecimiento. **(FK)**
- **Nivel_Educativo:**
 - id_Nivel_Educativo: Identificador único del nivel educativo. **(PK)**
 - Nombre: Nombre del nivel educativo (por ejemplo, primario, secundario, etc.).
- **Nivel_Educativo_de_ee:**
 - Cueanexo: Código único del establecimiento educativo. **(PK, FK)**
 - id_Nivel_Educativo: Identificador del nivel educativo correspondiente al establecimiento. **(PK, FK)**
- **Reporte_Demográfico:**
 - Edad: Rango de edad de la población en el reporte. **(PK)**
 - ID_DEPTO: Identificador del departamento al que corresponde el reporte. **(PK, FK)**
 - Poblacion: Cantidad de personas en el rango de edad especificado en el reporte. **(PK)**
- **Departamentos:**
 - ID_DEPTO: Identificador único del departamento. **(PK)**
 - Departamento: Nombre del departamento.
 - ID_PROV: Identificador de la provincia a la que pertenece el departamento. **(FK)**
- **Provincias:**
 - ID_PROV: Identificador único de la provincia. **(PK)**
 - Provincia: Nombre de la provincia.

3.6. Extracción de Datos

Luego de definir el Diagrama de Entidad-Relación (DER), se extrajeron los datos necesarios de las fuentes de información ya limpias para construir los DataFrames. Solo se conservaron los datos relevantes para los objetivos del trabajo.

En las tres fuentes de datos utilizadas, el manejo de la información sobre los departamentos variaba significativamente: se encontraban diferencias en los nombres de los departamentos, como errores tipográficos o variaciones en la escritura, y en algunos casos, se empleaban claves diferentes para representar el mismo concepto. Así mismo, se podían encontrar departamentos en una base de datos que no estaban en otra. Para resolver estas inconsistencias, se procedió a unificar las claves, los nombres de los departamentos de las tres fuentes y a crear una entidad que reúna la totalidad de los departamentos presentes en las 3 bases de datos.

El archivo `padron_poblacional.xlsx` resultó ser el más completo en cuanto a la información de los departamentos, por lo que se utilizó como referencia principal para la unificación y la creación de los DataFrames.

A continuación, se presenta un listado detallado de las fuentes utilizadas para componer cada DataFrame:

- **Departamentos:** Archivos `padron_poblacion.xlsx` y `2022_padron_oficial_establecimientos_educativos.xlsx` (ambas fueron necesarias por la incompletitud de las mismas).

- **Centros Culturales:** Archivos `centros_culturales.csv` y `padron_poblacion.xlsx`, el cual se utilizó para agregar el `ID_DEPTO`
- **Establecimientos Educativos:** Archivos `2022_padron_oficial_establecimientos_educativos.xlsx` y `padron_poblacion.xlsx`, el cual se utilizó para agregar el `ID_DEPTO`.
- **Nivel Educativo:** Archivo `2022_padron_oficial_establecimientos_educativos.xlsx`.
- **Nivel Educativo de EE:** Archivo `2022_padron_oficial_establecimientos_educativos.xlsx`.
- **Provincias:** Archivo `2022_padron_oficial_establecimientos_educativos.xlsx`.
- **Mails:** Archivo `centros_culturales.csv`.
- **Reporte Demográfico:** Archivo `padron_poblacion.xlsx`.

4. Análisis de Datos

En esta sección se extraen diferentes DataFrames para cada ejercicio, a partir de consultas realizadas con SQL. Se muestran las primeras y últimas filas de cada DataFrame con el fin de ofrecer una muestra representativa de los datos, aunque los DataFrames completos pueden consultarse en el Anexo.

Luego, se presentan visualizaciones que buscan identificar posibles relaciones relevantes para el objetivo del trabajo.

4.1. Consultas SQL

4.1.1. Establecimientos Educativos y Población por Nivel para Cada Departamento

Provincia	Departamento	Jardines	Población Jardín	Primarias	Población Primaria	Secundarios	Población Secundaria
Buenos Aires	La Matanza	378	87162.0	333	225872.0	352	181212.0
Buenos Aires	La Plata	274	29260.0	199	77998.0	219	67326.0
Buenos Aires	Lomas De Zamora	201	28806.0	178	76967.0	205	65257.0
...
Tucumán	Famailla	25	2322.0	23	5489.0	18	4607.0
Tucumán	La Cocha	17	992.0	18	2557.0	15	2282.0
Tucumán	Juan Bautista Alberdi	0	1586.0	0	4124.0	0	3462.0

Tabla 1: Informe por departamento: provincia, cantidad de establecimientos educativos por nivel y población por edad.

(a) Los datos se organizan alfabéticamente por provincia y por la cantidad de escuelas primarias, permitiendo analizar la distribución de los recursos educativos en relación con la población.

El análisis de los datos muestra **desigualdades en la distribución de los establecimientos educativos** en relación con la **población por edad** en los distintos departamentos. En **áreas urbanas** con alta población, la cantidad de escuelas no es proporcionalmente alta, lo que indica que las escuelas en estas zonas albergan a más estudiantes. Esta tendencia se ha observado para casos aislados ¹, pero sugiere que podría repetirse en otros.

Una observación destacada es que algunos departamentos tienen **cero establecimientos educativos** a pesar de contar con población en edad escolar. Por ejemplo, **Juan Martín de Pueyrredón**, en San Luis, tiene **28,900 niños en edad primaria** y no presenta escuelas primarias (lo mismo ocurre en niveles inicial y secundario). Además, este departamento no aparece en el **padrón de establecimientos educativos** ni en el de **centros culturales**, lo que sugiere un **problema en la completitud de los datos**, que podría requerir una **revisión o actualización**.

4.1.2. Centros Culturales con Capacidad Mayor a 100 Personas por Departamento

El análisis muestra una **distribución desigual** de los Centros Culturales con capacidad superior a 100 personas. Como era esperado, se observa que en **áreas urbanas**, con mayores concentraciones de población, existe una mayor cantidad de centros de gran capacidad. En contraste, en departamentos con **baja población** o en zonas rurales, es esperable que no haya centros culturales de esta capacidad, ya que no tendría sentido contar con espacios de gran tamaño para una cantidad reducida de habitantes.

Además, debido a la centralización de los recursos en **Buenos Aires** y otras grandes ciudades, se observa una mayor concentración de centros culturales de gran tamaño. Como ejemplo, si se filtraran los datos por provincia, la suma de centros

¹Las consultas SQL permitieron verificar esto para casos aislados, la visualización facilitará su análisis generalizado.

Provincia	Departamento	Cantidad de CC con cap>100
Buenos Aires	Avellaneda	20
Buenos Aires	La Plata	8
Buenos Aires	Lomas de Zamora	3
...
Tucumán	Simoca	0
Tucumán	Leales	0
Tucumán	Tafi Del Valle	0

Tabla 2: Cantidad de Centros Culturales con Capacidad Mayor a 100 Personas por Departamento

(a) Los datos están organizados alfabéticamente por provincia y de forma descendente según la cantidad de centros culturales.

culturales de gran tamaño, sería considerablemente mayor en las provincias con ciudades grandes, como por ejemplo **Buenos Aires** y **Córdoba**, en comparación con otras provincias, que no posean ciudades grandes.

4.1.3. Centros Culturales, Establecimientos Educativos y Población por Departamento

Provincia	Departamento	Cantidad EE	Cantidad CC	Población Total
Ciudad Autónoma de Buenos Aires	Ciudad Autónoma De Buenos Aires	1782	296	3095454.00
Córdoba	Capital	1136	30	1498060.00
Buenos Aires	La Matanza	977	2	1837168.00
...
San Luis	Libertador General San Martín	0	0	4691.00
Tierra del Fuego, Antártida...	Tolhuin	0	0	6027.00
Tucumán	Juan Bautista Alberdi	0	0	34745.00

Tabla 3: Informe por departamento: provincia, Cantidad de Centros Culturales (CC) y Establecimientos Educativos (EE).

(a) Los datos están organizados alfabéticamente por provincia y de forma descendente según la cantidad de centros culturales.

Este reporte está estrechamente ligado al objetivo del trabajo: se presentan las diferencias a nivel departamental en la relación entre la cantidad de Establecimientos Educativos (EE) y Centros Culturales (CC). En la tabla se visualiza la variación de esta relación entre distintas provincias y departamentos, permitiendo destacar casos con alta presencia de EE pero pocos CC, así como aquellos con una distribución más equilibrada o sin presencia de estos establecimientos.

Si se quisiera, se podría usar esta consulta para fácilmente calcular el promedio de EE y CC por provincia e incluso explorar la posibilidad de una regresión. Sin embargo, dado que entre departamentos de una misma provincia hay grandes diferencias en la cantidad de EE y CC, este enfoque podría generalizar de manera inadecuada esta relación.

4.1.4. Dominios de Correo Electrónico más Usados por los Centros Culturales

Provincia	Departamento	Dominio más Frecuente
Buenos Aires	9 De Julio	hotmail
Buenos Aires	Adolfo Alsina	adolfoalsina
Buenos Aires	Alberti	alberti
...
Tucumán	Cruz Alta	gmail
Tucumán	Lules	gmail
Tucumán	Yerba Buena	gmail

Tabla 4: Dominios de Correo Electrónico más usados por los Centros Culturales

(a) Los datos están organizados alfabéticamente por provincia y por departamento.

Estos datos pueden revelar patrones relacionados con el uso de tecnología en dichas instituciones, ofreciendo así indicios sobre su nivel de modernización y capacidad de adaptación tecnológica. A partir del análisis realizado, se identificó que el dominio más frecuente entre los centros culturales es Gmail, con 110 apariciones, representando un 58,82% del total. Este hallazgo sugiere una predominancia en el uso de servicios gratuitos y masivos de correo electrónico, lo cual podría estar vinculado a factores como accesibilidad, facilidad de uso y familiaridad con dicha plataforma.

4.2. Visualización de Datos

4.2.1. Cantidad de centros culturales por provincia

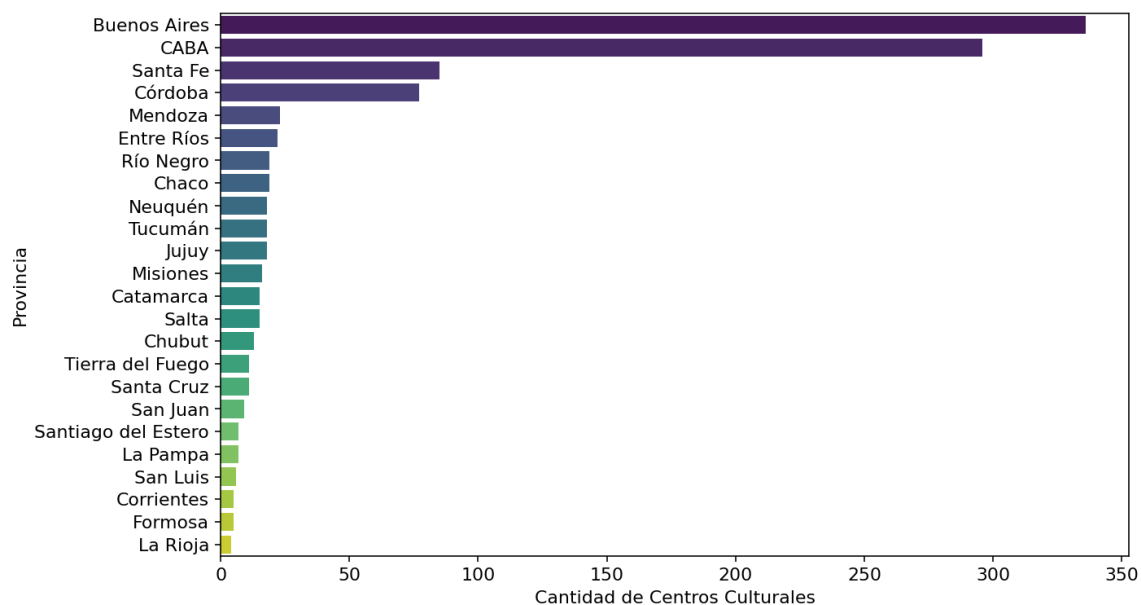


Figura 3: Cantidad de centros culturales por provincia

El análisis de la distribución de centros culturales (CC) en Argentina (ver Figura 3) revela una fuerte concentración en las provincias de Buenos Aires (338), CABA (296) y, en menor medida, Santa Fe (129) y Córdoba (96). A partir de estos datos, se observa claramente un patrón de centralización en áreas con mayor densidad poblacional y actividad económica, destacándose significativamente Buenos Aires y la Ciudad Autónoma de Buenos Aires respecto al resto de las provincias. Este patrón podría reflejar mayores recursos económicos, infraestructura desarrollada y políticas culturales más activas en dichas regiones, lo que podría facilitar la creación y sostenimiento de estos espacios culturales. En contraste, provincias como La Rioja, Formosa y Corrientes presentan la menor cantidad de centros culturales, lo cual podría sugerir limitaciones en infraestructura o políticas públicas menos focalizadas en la promoción cultural.

4.2.2. Distribución de Establecimientos Educativos por provincia y grupo etario

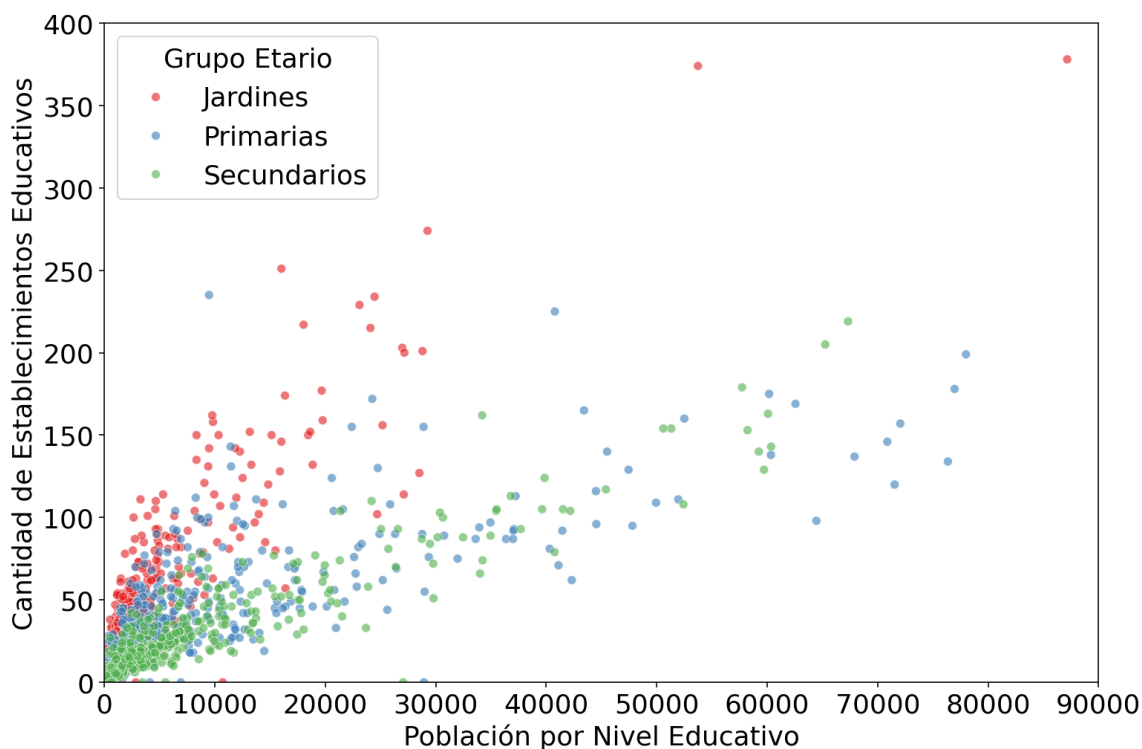


Figura 4: Cantidad de EE de los departamentos en función de la población, separados por nivel educativo y su correspondiente grupo etario

El gráfico presentado (ver Figura 4) ofrece un análisis de la distribución de establecimientos educativos en Argentina en relación con la población en edad escolar, diferenciando cada nivel escolar mediante colores específicos: jardines (rojo), primarias (azul) y secundarias (verde).

En general, se observa una tendencia positiva entre la población por nivel educativo y la cantidad de establecimientos educativos, indicando que a mayor población escolar corresponde, típicamente, una mayor cantidad de establecimientos educativos. No obstante, también se identifican puntos específicos que muestran una menor cantidad de establecimientos en relación con la población, especialmente en niveles educativos secundarios y primarios. Estas desviaciones podrían deberse a factores como la concentración poblacional en grandes establecimientos educativos o la existencia de limitaciones en infraestructura o recursos educativos en ciertas regiones. Esto sugiere que, aunque la tendencia general sea positiva, la distribución no es uniforme.

4.2.3. Distribución de EE por departamento, para cada provincia

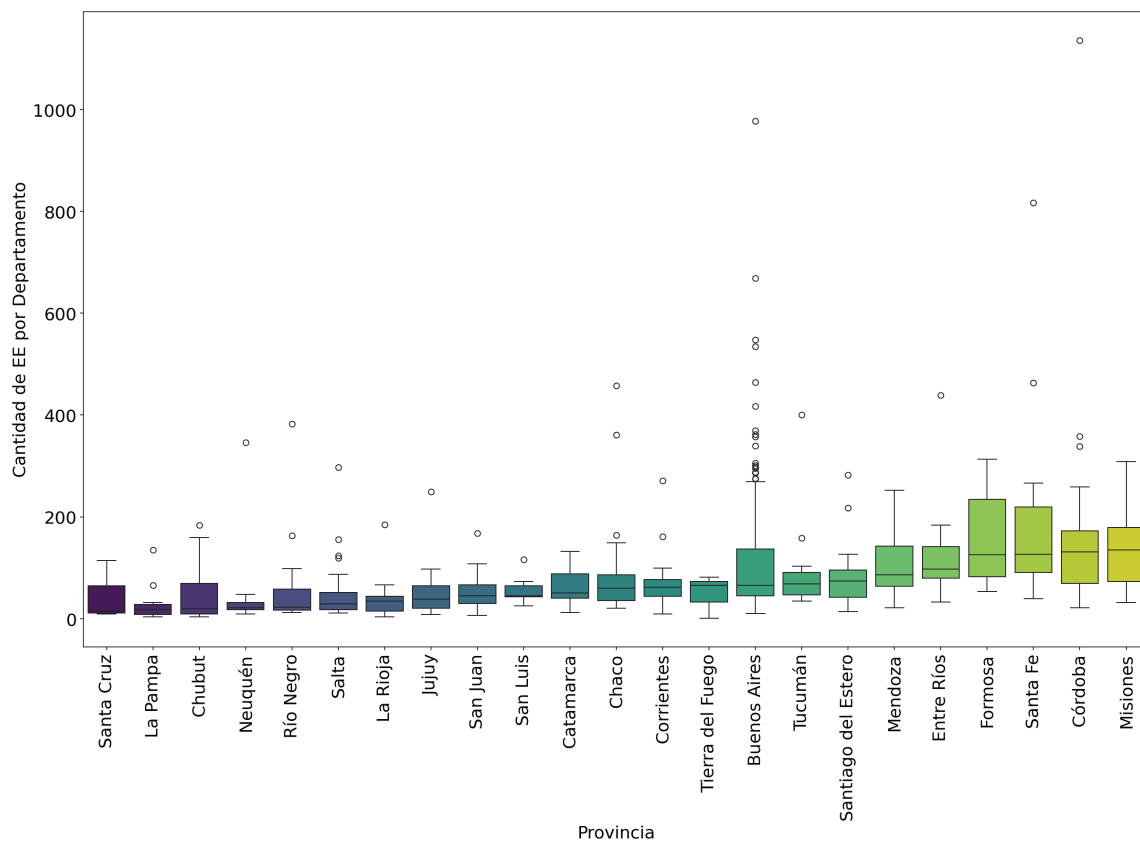


Figura 5: Boxplot por cada provincia, de la cantidad de EE por cada departamento de la provincia. Ordenados por la mediana de cada provincia.

El presente gráfico de cajas (ver Figura 5) muestra la distribución de la cantidad de Establecimientos Educativos (EE) por departamento en cada provincia. Esta visualización permite analizar la variabilidad en la infraestructura educativa dentro de cada jurisdicción, destacando las provincias con una distribución más homogénea frente a aquellas con una marcada desigualdad entre sus departamentos.

Los boxplots han sido ordenados según la mediana de la cantidad de EE por provincia, lo que facilita visualizar de manera estructurada cómo se distribuyen los recursos educativos entre departamentos. La presencia de valores atípicos (outliers) en varias provincias sugiere la existencia de departamentos con una cantidad de EE significativamente mayor que el promedio provincial. Esto podría relacionarse con factores como la concentración poblacional en determinadas áreas urbanas, políticas específicas de infraestructura educativa o diferencias socioeconómicas internas.

Cabe destacar que, la Ciudad Autónoma de Buenos Aires no ha sido incluida en este gráfico debido a que, al ser un único distrito sin subdivisión en departamentos, su representación en este formato no sería adecuada. Además, su alta concentración de establecimientos educativos alteraría la escala del gráfico, dificultando la interpretación comparativa de los datos de otras provincias.

4.2.4. Gráfico de Dispersión que relaciona la cantidad de EE y la cantidad de CC

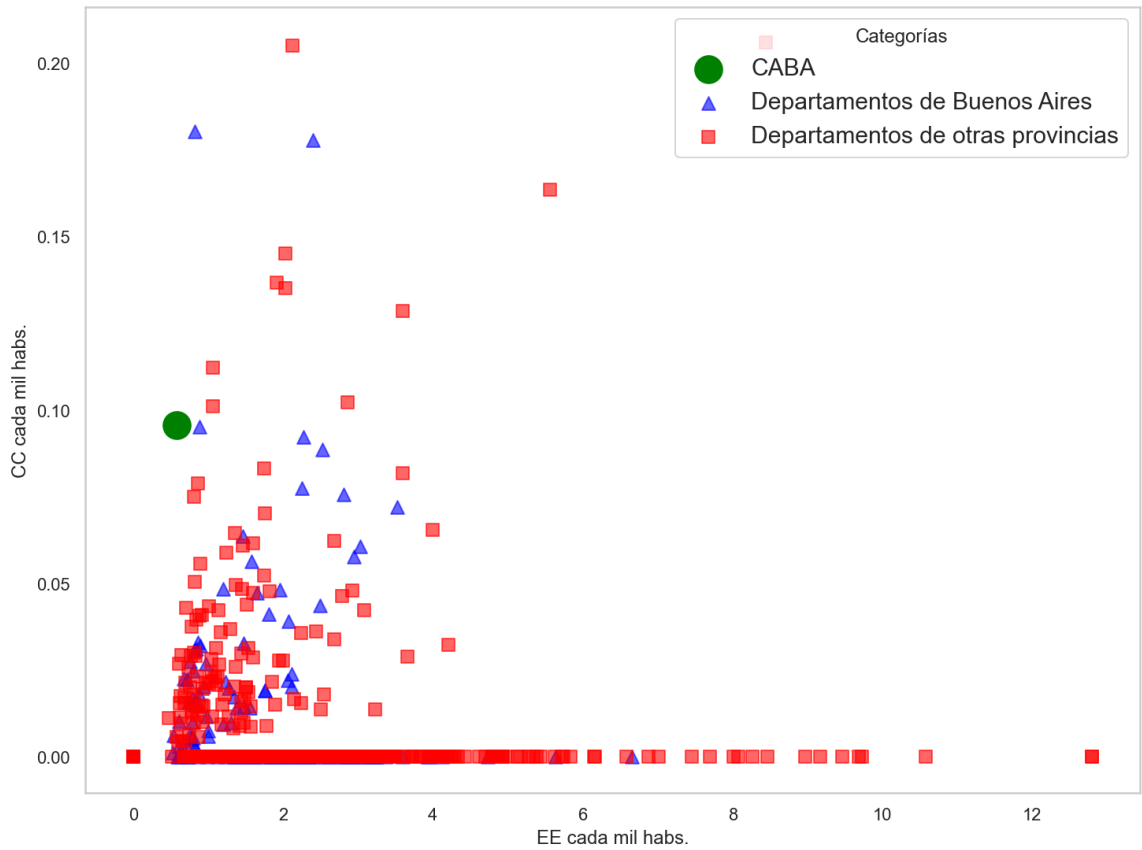


Figura 6: Relación para 3 grupos, entre la cantidad de establecimientos educativos y centros culturales por cada mil habitantes.

(a) Cada grupo se representa con un color, forma y tamaño que se detalla en la leyenda.

En la **Ciudad Autónoma de Buenos Aires** (CABA), se presenta una proporción bastante alta de centros culturales (CC) en relación a la cantidad de establecimientos educativos (EE).

En los departamentos de la **Provincia de Buenos Aires**, las variaciones son más notorias. Se observan varias instancias con proporciones altas de CC en función de EE, pero también hay casos con comportamientos más variados.

Por otro lado, en los departamentos de otras provincias, aunque hay más variaciones debido a la mayor cantidad de departamentos en comparación con los otros grupos, se observa que muchos **no tienen centros culturales registrados**².

La tendencia general sugiere que no parece haber una relación clara entre la cantidad de establecimientos educativos y centros culturales. Aunque en algunos departamentos se observa una mayor oferta cultural, esto no parece depender directamente de la cantidad de establecimientos educativos por cada 1000 habitantes. De hecho, existen varios casos en los que departamentos con una gran cantidad de centros culturales no muestran una relación similar en cuanto a la cantidad de establecimientos educativos. Por lo que no es posible asumir una relación entre ambas variables, ya que no es evidente.

5. Conclusiones

El análisis de la distribución de Establecimientos Educativos (EE) y Centros Culturales (CC) en Argentina muestra que no existe una relación directa ni uniforme entre ambas variables. Si bien ambas instituciones cumplen funciones educativas y culturales, su distribución territorial sigue patrones distintos, que podrían provenir de factores como la densidad poblacional y el grado de urbanización.

Como se observa en el gráfico 6a, que muestra la cantidad de EE y CC por cada mil habitantes en las provincias, la distribución de los EE tiende a ser más homogénea en todo el territorio nacional. En cambio, los CC se concentran mayormente en áreas urbanas, especialmente en la provincia de Buenos Aires y la Ciudad Autónoma de Buenos Aires.

Este fenómeno responde a que los EE están diseñados para garantizar el acceso a la educación básica en todo el país, lo que exige una distribución equitativa en función de la población en edad escolar. En contraste, los CC dependen de una mayor concentración de público y demanda, lo que favorece su sostenibilidad en entornos urbanos.

A nivel provincial, se observan casos como Santiago del Estero y Formosa, donde la cantidad de EE por mil habitantes es elevada, pero la presencia de CC sigue siendo baja. Esto sugiere que, si bien la infraestructura educativa en estas provincias

²Esto se representa como una recta de puntos en el valor $y=0$

está bien distribuida, la oferta cultural es más limitada. En el otro extremo, CABA y Buenos Aires concentran la mayor cantidad de CC del país, lo que refuerza la idea de que su distribución responde más a la demanda urbana que a un criterio poblacional uniforme.

En conclusión, no hay una relación proporcional entre la cantidad de EE y CC en las provincias argentinas. Mientras que la educación formal se distribuye de manera más equitativa en función de las necesidades poblacionales, la infraestructura cultural tiende a concentrarse en áreas urbanas.

A. Anexo

Para facilitar la legibilidad del informe y la visualización de los datos, se presentaron exclusivamente las primeras y últimas 3 filas de cada reporte. Los reportes completos se encuentran en formato .csv dentro de la carpeta **TablasModelo** adjunta a la carpeta principal de la entrega.