



# Trabajo Práctico 01

## Análisis de Datos

Laboratorio de Datos

### Grupo Ctrl

<u>Integrante</u>	<u>LU</u>
Arango, Joaquín	342/24
Cardinale, Dante	593/24
Herrero, Lucas	179/24



# Resumen

---

El informe a continuación aborda la posible relación entre la cantidad de Bibliotecas Populares y Establecimientos Educativos en los departamentos del país.

A partir de bases de datos oficiales correspondientes al año 2022, se desarrolló un proceso completo de integración, limpieza y transformación de datos, resolviendo inconsistencias, formatos heterogéneos y ambigüedades semánticas.

Se construyó un modelo relacional coherente, acompañado por consultas SQL y visualizaciones que permitieron explorar los vínculos entre ambas instituciones, y apreciar diversas peculiaridades.

Como conclusión, se detectaron patrones que sugieren una correlación entre la presencia de EE y BP, especialmente en zonas con mayor densidad poblacional.

## Introducción

---

Este informe procura identificar una conexión entre la cantidad de Bibliotecas Populares (de ahora en más, BP o bibliotecas) y la cantidad de Establecimientos Educativos (o, abreviadamente, EE) en los distintos departamentos del país. Como fuentes de información tomaremos las bases de datos 'bibliotecas-populares', '2022\_padron\_oficial\_establecimientos\_educativos' y 'padron\_poblacion', brindadas por el gobierno, tomando 2022 como año de referencia. No obstante, estas tablas serán solo el punto de partida, ya que es indispensable la realización de una serie de transformaciones (observando que cuentan con datos ajenos a nuestro objeto de análisis; así como datos de gran relevancia pero inconsistentes entre las diversas tablas, semánticamente confusos o erróneos), y una reorganización de las mismas, preservando su información, pero de forma más legible y concisa. En el proceso para llevarlo a cabo, utilizaremos herramientas como el modelado conceptual de los datos (DER), la representación de los datos mediante el Modelo Relacional, el empleo de consultas SQL, métodos de visualización de los datos y más recursos que aprendimos a lo largo de la materia. De esta forma, realizaremos un análisis competente en relación con el cumplimiento del objetivo establecido.

## Procesamiento de Datos

---

### Revisión calidad de datos

En esta subsección, nos enfocamos en el análisis de la calidad de datos de las bases de datos: Establecimientos Educativos y Bibliotecas Populares.

Algunos de los problemas de calidad de datos que posee el esquema EE son:

- Valores faltantes en mail, C.P., Código de Área, Teléfono y Ámbito; y gran variedad de valores vacíos en la columna Servicios Complementarios (que adicionalmente afecta la semántica de la BD pues no se llega a comprender qué es un servicio complementario). En general, el problema daña la 'Compleitud' de los datos en la BD y se puede clasificar como problema del 'Modelo de datos' pues el daño hacia la calidad de los datos deriva de una deficiencia en el diseño del esquema que define cómo se organizan y relacionan los datos, así como la ausencia de información por la inexistencia de tales valores.
- Valores Imposibles, como en la columna Teléfono donde hay 0; o hay un Código de área que es -3777511302, lo que afecta el atributo 'Correctitud/Corrección' de los datos, y se puede clasificar al problema como de 'Instancia' pues son registros individuales los que se ven afectados, con incidencias enmarcadas en formatos inadecuados/errores tipográficos.

A raíz de esto, para cuantificar cuán grave es el problema de valores faltantes en la columna Teléfono, utilizamos la técnica GQM.

Nuestro objetivo es analizar la 'Compleitud' del dato Teléfono asociado a cada Establecimiento Educativo.

**Goal:** el dato correspondiente al Teléfono de cada EE esté completo.

**Question:** ¿Qué tan significativa es la cantidad de EE que no poseen número de teléfono en relación a la totalidad de EE?

**Metric:**  $\frac{\#EE-sin-teléfono}{\#EE-totales}$

Para obtener el valor de esta métrica, programamos la siguiente función:

```
def porcentaje_ee_sin_telefono(df):  
    cant_ee_sin_telefono = 0  
    cant_ee_totales = len(df)  
    for fila in df.itertuples():  
        if (fila.Teléfono == 'N/D' or fila.Teléfono == 'S/D'):  
            cant_ee_sin_telefono +=1  
    return cant_ee_sin_telefono/cant_ee_totales
```

*Código ilustrativo a modo de ejemplo*

El resultado obtenido es alrededor de 3e-4, lo que en porcentaje equivale a un 0.03%. Es decir solo 20 de entre más de 63600 EE no tienen un teléfono registrado.

El problema puede haber surgido de la inexistencia de un número de teléfono de la institución o la pérdida de esa información entre tantos registros.

Cualquiera sea el caso, el impacto es muy bajo, por lo que no consideramos la necesidad de un accionar/implementación de medidas para contrarrestarlo.

En cuanto al esquema BP, los problemas de calidad que se destacan son:

- Columnas enteras con valores NULL (subcategoría, web, informacion\_adicional) lo que afecta el atributo 'Relevancia' (pues, el hecho de que se encuentran completamente vacíos denota insignificancia) y otras que presentan valores faltantes en gran cantidad (mail, teléfono, etc.), afectando de esta forma la 'Compleitud' de los datos en la BD. Éstos son problemas del 'Modelo de Datos' por la imposibilidad/inexistencia de los datos que se desean almacenar, produciendo una deficiencia en el diseño del esquema.
- Pares Latitud-Longitud duplicadas, incluso para bibliotecas que se hallan en distintas localidades (entre ellas, Bib.Pop. Dr. Manuel Belgrano y Bib.Pop. Villa de las Rosas), lo que implica una falla en la 'Correctitud/Corrección' de los datos. Éste es un problema de 'Instancia' pues son errores que afectan registros individuales con incidencias enmarcadas en valores duplicados.
- Localidades distintas con el mismo cod\_localidad (Wilde-Avellaneda, Berazategui-Plátanos, etc.), por lo cual la 'Consistencia' del dato se ve perjudicada. Nuevamente, se trata de un problema de 'Instancia' pues deriva de errores que afectan registros individuales.
- Columnas enteras que poseen un mismo valor repetido como "categoría: Biblioteca Popular", y "tipo\_latitud\_longitud: Exacta". En este último caso, como hemos visto pares latitud-longitud duplicadas, esto impacta en la 'Confiabilidad' de los datos, pues no es cierta la exactitud mencionada. En general, las columnas enteras con mismo valor es un problema de 'Modelo de datos' ya que deriva de una deficiencia en el diseño de la BD.

Para cuantificar cuán grave es el problema de columnas cuyos valores son, en su totalidad, NULL en la fuente de datos Bibliotecas Populares, utilizamos la técnica GQM.

Llámesse a estas columnas 'columnas vacías'; y al resto 'columnas no vacías'.

**Goal:** cada columna de nuestro esquema bibliotecas\_populares sea una 'columna no vacía'

**Question:** ¿cuál es la proporción de 'columnas vacías' en nuestro esquema?

**Metric:**  $\frac{\#columnas-vacías*100}{\#columnas-totales}$

Para obtener el valor de ésta métrica, diseñamos el siguiente programa en Python:

```
def proporcion_columnas_nulas(df):
    cant_columnas_totales = len(df.columns)
    cant_columnas_vacias = 0
    for columna in df.columns:
        todos_nulos = True
        for elemento in df[columna]:
            if pd.notna(elemento):
                todos_nulos = False
                break

        if (todos_nulos):
            cant_columnas_vacias += 1
    return (cant_columnas_vacias * 100)/cant_columnas_totales
```

*Código ilustrativo a modo de ejemplo*

El resultado en nuestro caso es de un 20%. Es decir que 5 de nuestras 25 columnas son 'columnas vacías'.

La causa la asociamos a que los datos buscados no aplican a la relación en cuestión, lo que genera que ninguna tupla posea un valor no nulo en esos campos.

Es por eso que para mejorar la calidad de la base de datos, la propuesta es la eliminación de dichas columnas. Es decir, que se mantengan sólo aquellas en las que se ha proporcionado cierto valor para alguna de las tuplas. De esta forma, es muy significativa la reducción de almacenamiento que se logra, así como también mejora el atributo de Relevancia de lo almacenado.

### Revisión formas normales

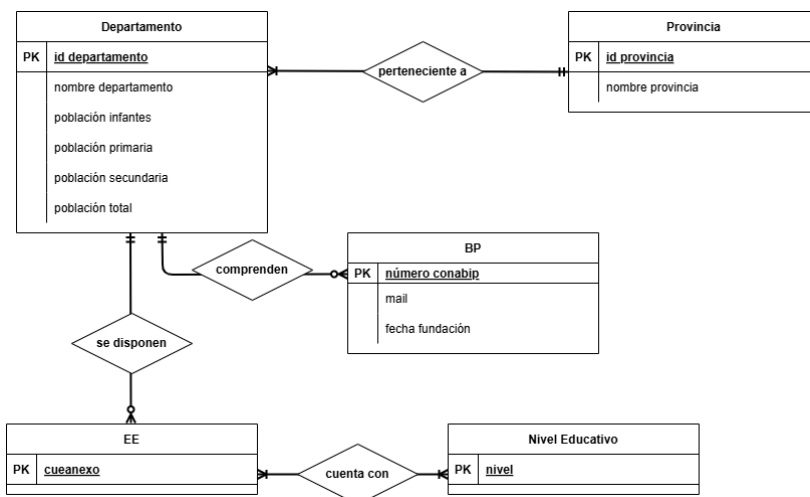
En esta subsección, nos enfocamos en el análisis de las formas normales de las fuentes de datos 'Establecimientos Educativos' y 'Bibliotecas Populares'.

Ninguna consigue alcanzar la 1FN. Las razones son:

- **EE:** no alcanza la 1FN pues tiene relaciones dentro de relaciones (o se puede pensar como que presenta relaciones como valores de atributos dentro de tuplas), además de que no posee valores atómicos en todas sus columnas (por ejemplo, en Teléfono hay filas que poseen más de uno).
- **BP:** cumple con la ausencia de relaciones dentro de relaciones, pero no con la indivisibilidad/simplicidad de los elementos del dominio de cada atributo (por ejemplo, en mail hemos hallado una tupla que tiene cargado dos correos), y es por ello que no alcanza la 1FN.

### Modelado conceptual de los datos: DER

En esta subsección nos proponemos realizar el Modelo DER a partir de las [Decisiones Tomadas](#).

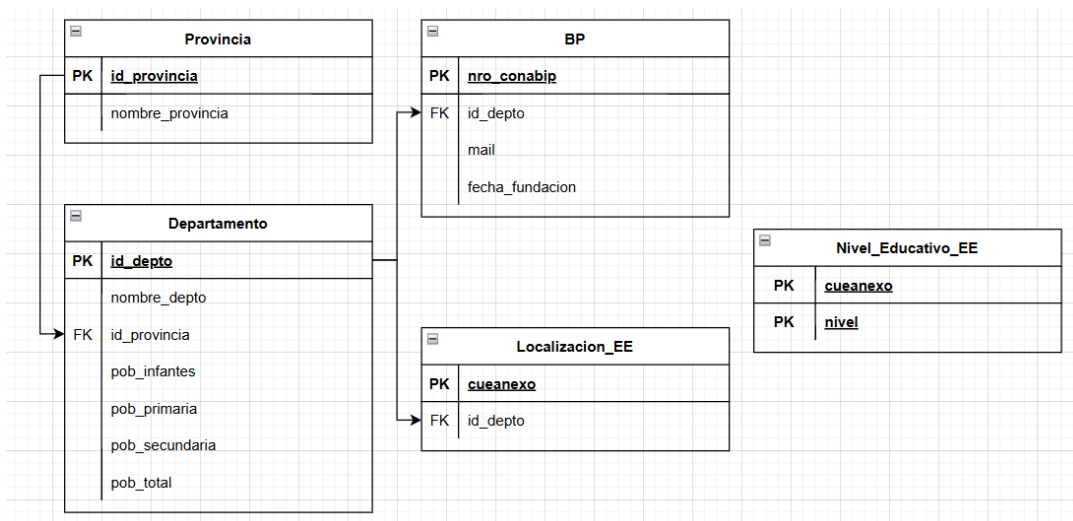


Para su armado, tuvimos en cuenta lo siguiente:

- **Departamento**: como Primary Key seleccionamos el id de cada departamento, pues se trata de su identificador único. Cada Departamento pertenece a una, y solo una, Provincia. Adicionalmente, los departamentos pueden contar con varios o ningún EE y BP. En cuanto a los datos poblacionales, al tener en claro los grupos etarios que nos interesan, sumado a que lo que distingue una población de otra es meramente su procedencia (el departamento), nos resulta conveniente que sean atributos del departamento, y no así una entidad aparte.
- **Provincia**: Como el id de cada una es único, lo tomamos como clave primaria. Una provincia puede llegar a tener muchos departamentos, y siempre tiene al menos uno. A la entidad le agregamos también el atributo 'nombre provincia'.
- **BP**: Cada biblioteca tiene un único numero\_conabip que lo identifica, por lo cual esta es nuestra PK. Luego, una BP solo corresponde a uno y solo un departamento; y cuenta con un mail y una fecha de fundación. Una aclaración importante a realizar es que se han detectado dos bibliotecas distintas, y pertenecientes a distintos departamentos, que poseen un mismo correo. Esto genera que el atributo 'mail' no sea viable para la identificación única de las bibliotecas.
- **EE**: cuentan con un identificador llamado cueanexo, que es único para cada establecimiento. Los establecimientos pertenecen a exactamente un departamento; y pueden tener múltiples niveles educativos, pero siempre al menos uno.
- **Nivel Educativo**: la clave es el nivel, el cual puede ser 'jardín', 'primario' o 'secundario'. Cada nivel educativo puede darse en uno o más establecimientos, y no hay establecimiento que no tenga alguno de estos niveles.

### Modelo Relacional:

En esta subsección nos proponemos realizar el Modelo Relacional a partir del DER y las [Decisiones Tomadas](#).



Para representar las relaciones establecidas en el modelo conceptual, llamamos Foreign Key (FK) a aquellos atributos que nos permiten relacionar dos esquemas. Las flechas en el diagrama ayudan a observar la procedencia de estas FK.

Como podemos observar, no creamos una tabla propia para EE ni Nivel Educativo debido a que poseen un solo atributo. En su lugar, decidimos establecer los esquemas:

- **Localización\_EE**: representa la relación de EE con Departamento. La PK es el cueanexo del establecimiento, ya que éste pertenece exclusivamente a un departamento (y necesariamente a uno).

- Nivel\_Educativo\_EE: se plasma la relación entre EE y Nivel, tomando como PK el cueanexo del establecimiento y el nivel educativo referido (no basta solo con el identificador del establecimiento pues éste puede contar con varios niveles).
- El resto del esquema coincide con lo realizado en el DER.

Es importante tener en cuenta que en cada relación, la PK es la única que determina el resto de los atributos. Es decir, en cada tabla, la única dependencia funcional existente es la que posee cada atributo (lado derecho de la dependencia) a con su clave (lado izquierdo de la dependencia).

El no poseer relaciones anidadas, y tener atomicidad en los valores de cada atributo, se cumple con la 1FN; luego, como en cada relación la clave es única (o no hay atributos no primos en el caso de Nivel\_Educativo\_EE), se logra también la 2FN; y finalmente, teniendo en cuenta que las únicas dependencias funcionales existentes son las de los atributos no primos a con su PK, no se presentan dependencias transitivas, lo que desemboca en el cumplimiento de la 3FN.

### Proceso de importación de datos

Para la generación de las tablas asociadas al Modelo Relacional, se tuvieron que implementar diversas estrategias de recolección, procesamiento y filtrado de datos mediante programas producidos en lenguaje python, uso de funciones de las librerías numpy y pandas, y desarrollo de consultas SQL a través de la librería duckdb.

Cabe destacar que para la obtención y utilización de los datos de la fuente Padrón Población se debieron tomar medidas adicionales, como la realización de un reformato completo del esquema, que funcionó como punto de partida para luego poder analizar la extracción de la información contenida. Esto se debe a que la BD presentaba varias tablas separadas, donde solo se hallaban los datos poblacionales y no así los códigos de los departamentos ni sus nombres que, por el contrario, se encontraban celdas por encima.

Es entonces, que en línea con nuestro objetivo, se determinó:

1. Extraer, del archivo original de Bibliotecas Populares, solo la información correspondiente al nro\_conabip, id\_departamento, mail y fecha\_fundacion; pues observamos que son éstos los atributos que poseen cierta competencia en relación a nuestros objetivos.
2. Recompilar, del archivo original Padrón Oficial de Establecimientos Educativos, solo la información correspondiente a Cueanexo, Código de Localidad de los establecimientos educativos de Modalidad Común, así como los niveles educativos que cada uno presentaba (ver [Decisiones Tomadas](#) para profundizar acerca de la significancia de 'nivel educativo', y el uso de los códigos de localidad).
3. Sustraer, del archivo original Padrón Población, solo la información correspondiente al Área (código) de cada departamento, su nombre y la información poblacional (infantil, primaria, secundaria y total; véase [Decisiones Tomadas](#) para mayor claridad).

## **Decisiones Tomadas**

Para conseguir relaciones con semántica clara, almacenamiento mínimo (que, a la vez, no permitan anomalías de inserción, deleción y modificación), conservación de dependencias funcionales, cumplimiento de la 3FN y con buena calidad de datos (enfaticando en la relevancia, consistencia y corrección de los datos), hemos optado por:

- En el archivo original Padrón Oficial de Establecimientos Educativos, la información respecto a los distintos niveles educativos se agrupó de la siguiente forma:
  - 'Nivel inicial - Jardín maternal' y 'Nivel inicial - Jardín de infantes' como nivel 'jardín'.
  - 'Secundario' y 'Secundario – INET' como nivel 'secundario'.
- Se definieron los siguientes grupos etarios a partir de la información poblacional presente en el archivo original Padrón Población:

- Infantil: población de entre 3 y 5 años (bordes incluidos).
  - Primaria: población de entre 6 y 12 años (bordes incluidos).
  - Secundaria: población de entre 13 y 17 años (bordes incluidos).
  - Total: cantidad general de habitantes de dicho departamento.
- Las comunas de Ciudad de Buenos Aires no se han considerado como departamentos. La información poblacional correspondiente a las mismas, así como los datos de localización de sus BP y EE, ha sido agrupada indistintamente como 'Ciudad Autónoma de Buenos Aires'. A raíz de esto, los id que eran asignados a las comunas en los archivos de Establecimientos Educativos y Padrón Poblacional, fueron sobrescritos como '2000', que es aquel con el que se hace referencia a CABA en el archivo Bibliotecas Populares.
  - En el atributo mail de la tabla BP, se detectó un mail escrito dos veces. Éste fue corregido para el cumplimiento de la atomicidad del atributo.
  - Siguiendo con los problemas hallados en relación a los mails de las bibliotecas, se corrigió un dominio presente en uno de los mails. Éste era '@educar' (dominio inválido, inexistente; de formato incorrecto) en vez de '@educ.ar' (perteneciente al portal educativo oficial de Argentina).
  - Los identificadores de departamento de cada establecimiento educativo fueron hallados mediante los códigos de localidad presentes en la base de datos original, pues se pudo notar una correspondencia entre las primeras cinco cifras de éste código y el id del departamento. No obstante, en el caso de la Ciudad y Provincia de Buenos Aires, son las cuatro primeras, tras descontar la primer cifra, ya que en estas provincias se observa la presencia de un 0 en el inicio.
  - De manera casi análoga, existe correspondencia entre el id del departamento y la provincia a la cual pertenece. Las dos primeras cifras de cada id representan la provincia de la cual son parte. Sin embargo, en caso de la Ciudad y Provincia de Buenos Aires donde, a diferencia del resto, el id de los departamentos es un número menor a 10000, el identificador de la provincia es únicamente la primer cifra.

## Análisis de Datos

En esta sección empezaremos mostrando los resultados de las consultas SQL propuestas, describiendo qué representa cada una de ellas. Vale aclarar que estos resultados no se presentarán en su versión completa, sino que se mostrarán las primeras siete filas a modo representativo. Para visualizar de forma completa los reportes, véase el [Anexo](#); donde se encuentran en formato xlsx .

1)

Provincia	Departamento	Jardines	Poblacion_Jardin	Primarias	Poblacion_Primaria	Secundarios	Poblacion_Secundaria
Buenos Aires	La Matanza	325	87162	333	225872	335	181212
Buenos Aires	La Plata	232	29260	199	77998	206	67326
Buenos Aires	Lomas de Zamora	162	28806	178	76967	191	65257
Buenos Aires	General Pueyrredón	178	23100	169	62565	170	57730
Buenos Aires	Quilmes	162	26975	146	70881	153	60085
Buenos Aires	Almirante Brown	133	25188	137	67913	145	58227
Buenos Aires	Moreno	118	28521	134	76357	134	60359

Para cada departamento se detalló la provincia a la cual pertenece, la cantidad de establecimientos educativos, segmentados por nivel, que se encuentran en el departamento; así como también la información correspondiente a la población de individuos en edad de escolarización de los distintos niveles educativos.



La información resultante fue ordenada alfabéticamente por provincia, y dentro de cada una, el orden se estableció en base a la cantidad de escuelas primarias en forma descendente.

2)

Provincia	Departamento	Bibliotecas_fundadas_desde_1950
Buenos Aires	La Plata	15
Buenos Aires	La Matanza	15
Buenos Aires	Moreno	13
Buenos Aires	Bahía Blanca	12
Buenos Aires	Tigre	12
Buenos Aires	Tandil	10
Buenos Aires	General San Martín	10

Para cada departamento se detalla su nombre, la provincia de la cual es parte y la cantidad de bibliotecas populares que fueron fundadas en el departamento posterior al año 1950.

El reporte fue ordenado alfabéticamente por provincia y, dentro de cada provincia, de forma descendente en base a la cantidad de observaciones de bibliotecas con dicha característica.

3)

Provincia	Departamento	Cant_EE	Cant_BP	Poblacion_Total
Ciudad Autónoma de Buenos Aires	Ciudad de Buenos Aires	1567	43	3095454
Córdoba	Capital	1054	20	1498060
Buenos Aires	La Matanza	957	16	1837168
Santa Fe	Rosario	742	39	1337958
Buenos Aires	La Plata	621	33	756074
Buenos Aires	Lomas de Zamora	520	12	685644
Buenos Aires	General Pueyrredón	500	6	660569

Para cada departamento, se detalló su nombre, provincia de pertenencia, cantidad de establecimientos educativos de modalidad común que posee, la cantidad de bibliotecas populares que dispone y su población total.

El reporte está ordenado en base a la cantidad de establecimientos (primero) y bibliotecas populares (después) en forma decreciente, nombre de provincia en forma ascendente, y finalmente por nombre del departamento, también en forma ascendente.



4)

Provincia	Departamento	Dominio más frecuente en BP
Buenos Aires	Adolfo Alsina	hotmail.com
Buenos Aires	Adolfo Gonzales Chaves	yahoo.com.ar
Buenos Aires	Alberti	hotmail.com
Buenos Aires	Almirante Brown	yahoo.com.ar
Buenos Aires	Avellaneda	yahoo.com.ar
Buenos Aires	Ayacucho	none
Buenos Aires	Azul	yahoo.com.ar

Para cada departamento, se detalla su nombre, provincia de pertenencia y el dominio de mail de mayor frecuencia en las bibliotecas populares que se encuentran en el departamento.

En los casos en que el departamento no posea bibliotecas populares, o las bibliotecas del departamento no posean mails cargados; entonces el resultado será 'none'. Basta con que exista alguna biblioteca popular del departamento que posea mail cargado para que éste no sea el caso.

### Visualización de datos

En esta subsección se encuentran los gráficos solicitados así como el análisis/evaluación que realizamos de los mismos.

1.

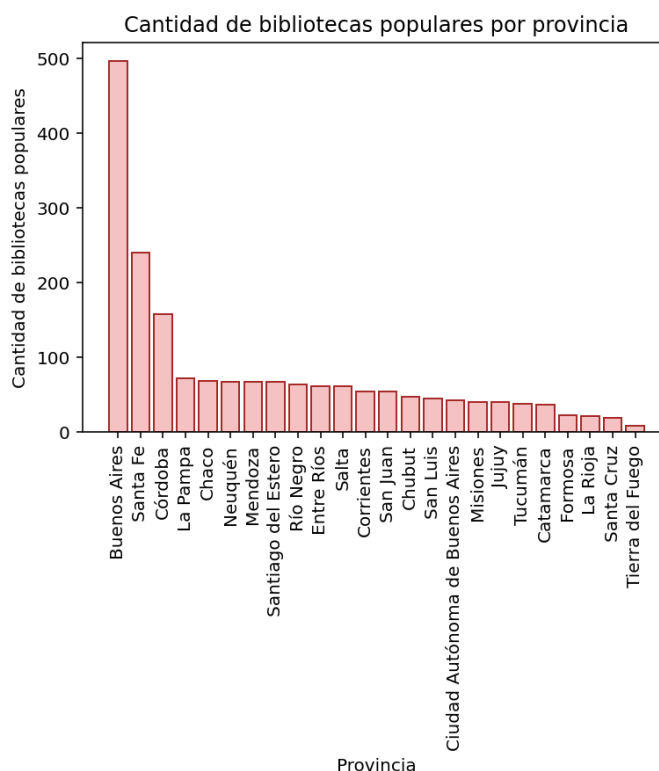


Figura I. Gráfico de la cantidad de BP por cada provincia argentina.

Para esta visualización optamos por un GRÁFICO DE BARRAS, ya que éste permite representar una variable numérica, que por lo general describe una cantidad, en nuestro caso, cantidad de BP; así como variables categóricas (una barra por categoría), las cuales en la [Figura I](#) son las provincias. La ventaja de este gráfico es que su formato nos facilita la comparación de las cantidades de BP en cada provincia, logrando así destacar sin dificultad a Buenos Aires con mayor abundancia de bibliotecas, seguida, aunque con amplia diferencia, de Santa Fe y Córdoba. Exceptuando algunos casos puntuales, la mayoría de las demás provincias presentan cifras relativamente similares entre sí, y se distingue que hay cantidades exactamente iguales entre varias de ellas. Un detalle no menor es que hay bibliotecas populares en todas las provincias.

2.

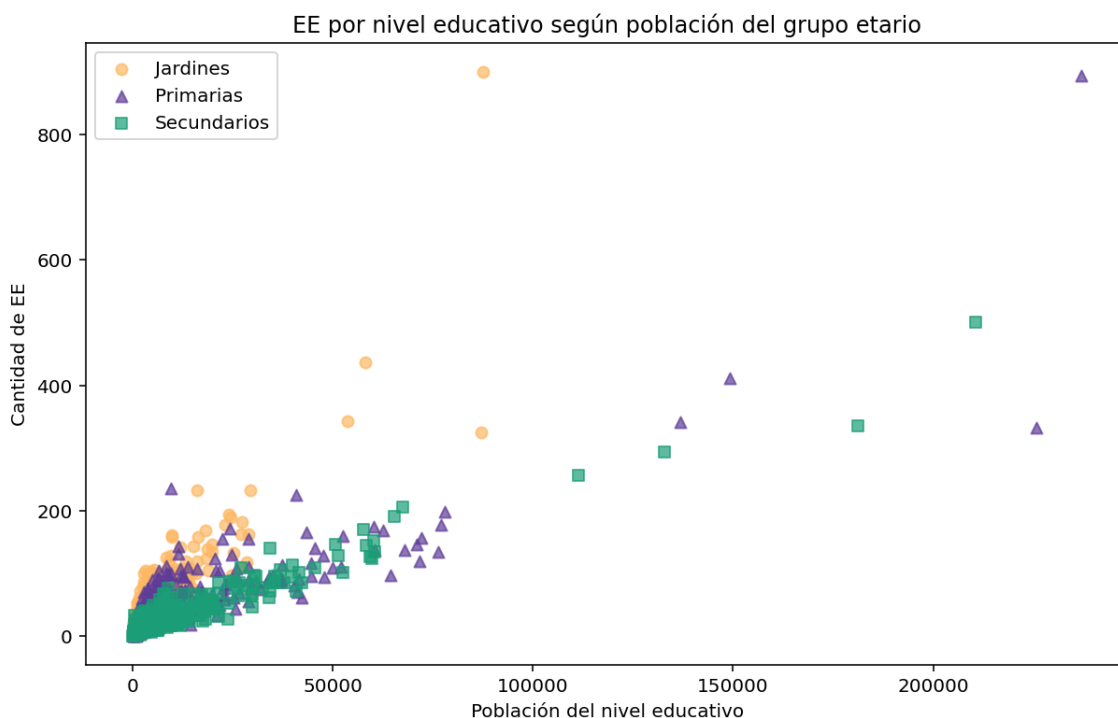


Figura II. SCATTERPLOT de la cantidad de EE (separados por nivel educativo) de los departamentos en función de la población del grupo etario.

La [Figura II](#) muestra la relación entre la cantidad de Establecimientos Educativos (EE) y la población de cada provincia, diferenciando los datos según el nivel educativo y su grupo etario correspondiente. El uso de un gráfico de dispersión (SCATTERPLOT) permite visualizar la distribución y comportamiento de la cantidad de EE de cada nivel educativo en función de la población.

Es observable que para poblaciones menores a 100000 habitantes, la cantidad de establecimientos educativos no suele (en la gran mayoría de los casos, con pocas excepciones) superar los 200 en los distintos niveles. El nivel jardín es el que más excepciones de ésta índole presenta.

Por otra parte, se puede analizar que hay gran cantidad de departamentos que poseen menos de un EE cada 250/300 individuos del grupo etario correspondiente; pudiéndose calcular además casos que esa cifra llega incluso a un establecimiento cada aproximadamente 400 personas de la población; siendo las primarias y las secundarias las que en mayor medida presentan estas métricas.

Aunque, a la vez, son identificables varios puntos que indican que hay departamentos que cuentan con al menos un establecimiento educativo cada 100 habitantes (llegando a ser incluso menor la cifra), lo que marca una asimetría entre los distintos departamentos del país en lo referente a este ámbito.

3.

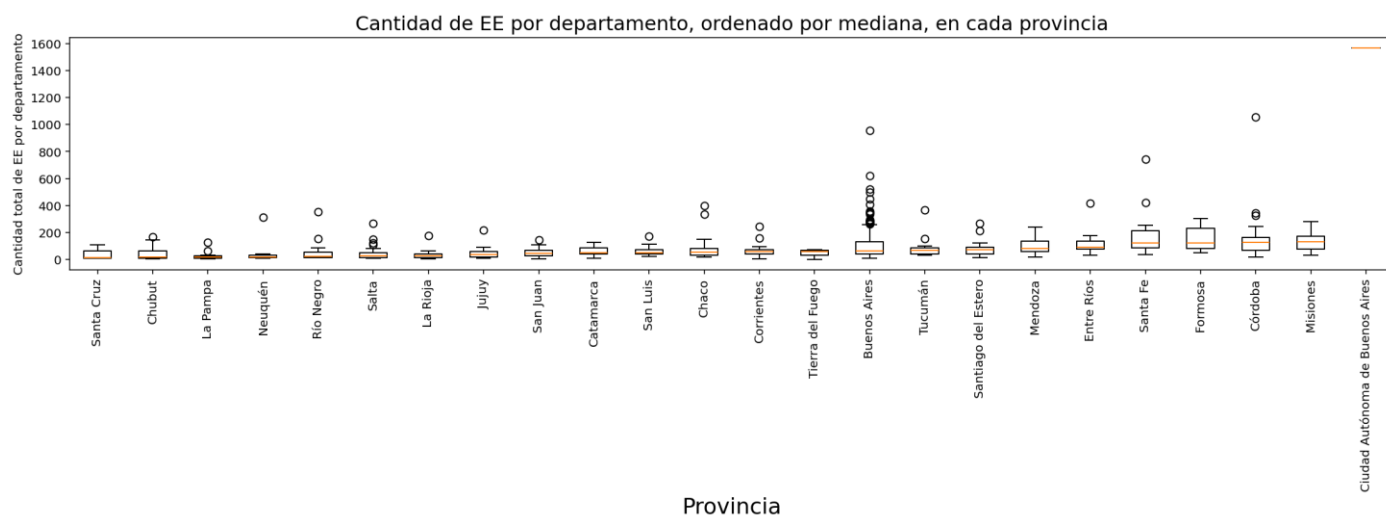


Figura III. BOXPLOT de la cantidad de EE por cada departamento de las provincias argentinas.

La [Figura III](#) muestra la cantidad de EE en los departamentos agrupados por sus correspondientes provincias, ordenado en función de la mediana de EE de los departamentos en la provincia. Lo primero que hay que mencionar sobre este gráfico es el caso Ciudad de Buenos Aires: al haber agrupado todas las comunas, no se destacan departamentos en la ciudad, por lo que toda la información resulta encapsulada en la mediana. Cabe destacar que si bien ésta presenta una gran diferencia en comparación al resto, si se supone una distribución uniforme en las comunas, entonces la mediana rondaría entre 100 y 110 EE, y por lo tanto no lograría distanciarse del resto de las provincias, donde la mediana es relativamente similar (al menos lo es desde la mitad de la tabla). Por otra parte, son varias las provincias que, por dentro, presentan cantidades similares de EE en sus departamentos, ya que suelen ser muy puntuales los casos de “datos extremos”/outliers; aunque la presencia de los outliers sí son importantes en la Provincia de Buenos Aires, lo que implica que es la que mayor disparidad entre departamentos presenta, seguida de Córdoba.

4.

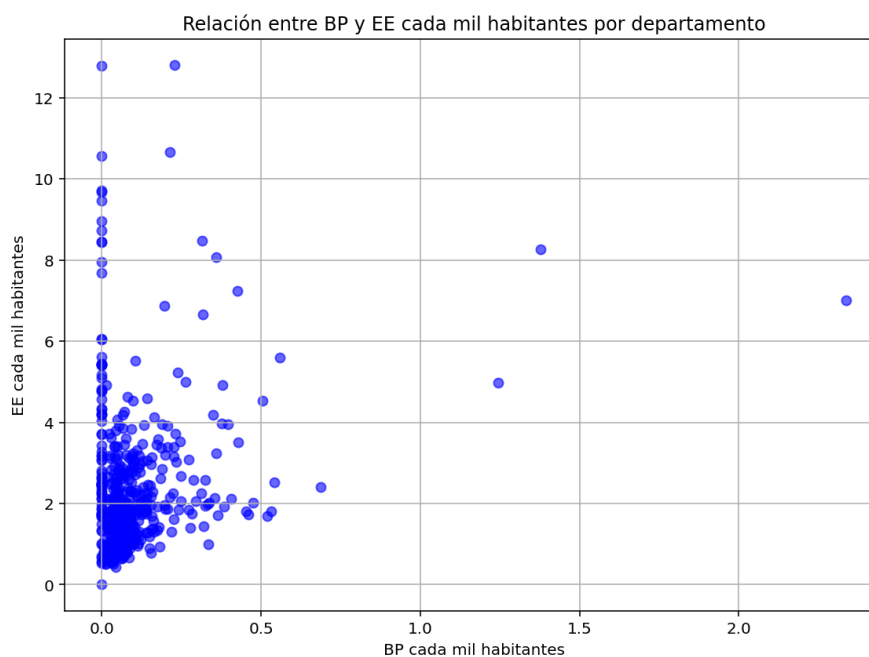


Figura IV. Gráfico de la relación entre la cantidad de BP y EE cada mil habitantes por departamento.

La [Figura IV](#) muestra la relación entre la cantidad de Bibliotecas Populares y Establecimientos Educativos, cada mil habitantes, por departamento.

El gráfico de dispersión permite observar cómo se distribuyen la cantidad de BP y de EE; logrando así ver que, por lo general, los valores se concentran en la zona baja del eje X (pocas BP por mil habitantes) y también en valores bajos del eje Y (pocos EE cada mil habitantes).

Sin embargo, hay varios departamentos que presentan una importante cantidad de establecimientos educativos, siendo inusualmente alta (al menos, por lo observado en el gráfico) en relación con su población. Esto nos sugiere que la relación entre ambos tipos de instituciones varía considerablemente entre departamentos, al haber aquellos que poseen gran densidad de establecimientos educativos pero nula cantidad de bibliotecas populares; y otros que presentan cierta paridad/sintonía en sus cantidades (aun siendo bajas).

Aunque, por otra parte, vale destacar que en la figura no se distinguen departamentos que presenten bajos niveles de establecimientos educativos pero altos de bibliotecas populares. Este último punto, entonces, es afín a la generalidad de las áreas observadas.

## Conclusiones

---

Habiéndose realizado todos los procesos de recolección, procesamiento y limpieza de datos; y tras haber implementado técnicas de análisis de los datos, con el respaldo de las visualizaciones gráficas producidas, hemos determinado que se puede dar fe de una relación entre las cantidades de establecimientos educativos y bibliotecas populares.

La detección de varios departamentos con gran disposición de establecimientos educativos pero bajos (o nulos) de bibliotecas populares y, la identificación de otros de mayor armonía; nos dan a entender que no existe fenómeno que determine, a partir de una dada cantidad de EE, cuántas BP dispone el departamento. Dicho en otras palabras, las irregularidades presentes en los análisis nos dan a entender que la cantidad de establecimientos no es una magnitud que fije o precise el número de bibliotecas.

No obstante, analizando la problemática de manera inversa, tras observarse la suba del número de BP, hemos distinguido cierta afinidad entre volúmenes, ya que los números de establecimientos también se notaron elevados. Esto sugiere que, si bien no es cierto que si hay pocas BP entonces habrán pocos EE; sí se puede determinar que a mayor presencia de bibliotecas populares, se puede esperar un incremento en la cifra de establecimientos educativos.

Además, al incorporar la variable poblacional en el análisis, se evidenció que muchos de los departamentos con mayor proporción de instituciones (ya sean EE o BP), también poseen una población notablemente más numerosa, lo cual podría actuar como un factor que impulsa la creación de estos espacios.

En definitiva, hemos podido alcanzar nuestro objetivo de caracterizar el vínculo entre estas dos instituciones, llegando a la conclusión de la existencia de cierta repercusión sobre la cantidad de establecimientos educativos de un departamento en función de la proporción de bibliotecas populares en el mismo.

## **ANEXO**

En este módulo se encuentran las tablas expuestas en la primer parte de la sección de Análisis de Datos, en un único archivo de Excel, en el que cada hoja de cálculo del mismo representa un reporte, donde el orden de las hojas respeta la secuencia en la que fueron mostrados los reportes en el informe. Adicionalmente, a cada hoja se le asignó un nombre declarativo para identificar sin esfuerzo el ejercicio del cual es resultado.

Esto mismo lo dejamos disponible en Google Sheets para permitir el acceso de manera online (mismo material, la decisión de proporcionar distintos formatos apunta a la facilitación de acceso, acorde a las comodidades y necesidades del lector).



TP01-CTRL-REPORTE  
S.xlsx

<https://docs.google.com/spreadsheets/d/1UMnhUSRVk0TA4hYuMVGk3ft0zTWpY2AOVWroy89wLqE/edit?usp=sharing>

## **BIBLIOGRAFÍA**

- Establecimientos Educativos (EE). Padrón Oficial de Establecimientos Educativos del año 2022. Disponible en: <https://www.argentina.gob.ar/educacion/evaluacion-e-informacion-educativa/padron-oficial-de-establecimientos-educativos>
- Bibliotecas Populares (BP). Padrón de Bibliotecas Populares. [https://datos.gob.ar/dataset/cultura-mapa-cultural-espacios-culturales/archivo/cultura\\_01c6c048-dbeb-44e0-8efa-6944f73715d7](https://datos.gob.ar/dataset/cultura-mapa-cultural-espacios-culturales/archivo/cultura_01c6c048-dbeb-44e0-8efa-6944f73715d7)
- Población. Datos de población por Departamento. Se pueden obtener de los datos del censo de 2022, sección Estructura por edad de la población. Está disponible en: <https://www.indec.gob.ar/indec/web/Nivel4-Tema-2-41-165> . Para descargar el archivo xlsx, seguir este enlace: <https://docs.google.com/spreadsheets/d/1M9EXOxmu3orgsMa06-wdx7rHQhgxEAvb/edit?gid=1552365536#gid=1552365536>
- Material teórico y práctico correspondiente a la materia Laboratorio de datos, FCEyN, UBA.