**Trabajo Práctico 1 - Laboratorio de Datos**

**Escuelas y Bibliotecas**

**Fecha:** 23 de mayo de 2025

**Integrantes:** Barrios, Bruno; Mur, Santiago; Operti, Bruno

**Resumen**

Este trabajo tiene como objetivo analizar si existe una relación entre la cantidad de establecimientos educativos y las bibliotecas populares por departamento en Argentina. Se utilizaron fuentes de datos oficiales para establecimientos educativos, bibliotecas populares y datos poblacionales, se realizó limpieza y normalización de los datos, se construyó un modelo relacional en 3FN, se ejecutaron consultas SQL relevantes, y se elaboraron visualizaciones que permiten observar y reflexionar sobre las posibles correlaciones entre estas variables.

**Introducción**

La problemática que se aborda es identificar patrones o relaciones entre dos instituciones fundamentales para el desarrollo educativo y cultural de las comunidades: las escuelas y las bibliotecas. El objetivo general es establecer si existe correlación entre su distribución territorial. Para esto se realizarán tareas de exploración, limpieza, modelado, consulta y visualización sobre los datos provistos. A lo largo del informe se documentará el procesamiento, las decisiones metodológicas, los resultados obtenidos y las conclusiones alcanzadas.

**Procesamiento de Datos**

**Formas normales originales:**

* **Establecimientos Educativos:** no cumple 1FN en el campo "Mail" (valores no atómicos).
* **Bibliotecas Populares:** contiene campos nulos permanentes como "subcategoria".

**Calidad de Datos:**

* **Establecimientos Educativos:** atributo "Mail" contiene múltiples correos en una misma celda. Se afecta la disponibilidad y consistencia. Métrica: 26.51% de los registros no son atómicos. Posible solución: extraer sólo el primer correo o normalizar en otra tabla.
* **Bibliotecas Populares:** campo "subcategoria" completamente nulo. Afecta completitud y relevancia. Posible solución: eliminar el campo.

**DER y Modelo Relacional:**  
Se diseñó un modelo de datos en 3FN con las siguientes tablas:

* departamentos (id\_depto PK, nombre\_depto, provincia)
* establecimientos (id\_establecimiento PK, nombre, id\_depto FK, ambito, modalidad\_comun, ...)
* niveles\_establecimientos (id\_establecimiento FK, id\_nivel FK)
* niveles (id\_nivel PK, descripcion)
* nivel\_educativo\_por\_departamento (id\_depto FK, id\_nivel FK, cantidad)
* bibliotecas (id\_biblioteca PK, id\_depto FK, nombre, dominio\_email, año\_fundacion)

**Importación de datos:**  
Los datos limpios se cargaron a pandas DataFrames con los mismos nombres que las tablas del modelo.

**Decisiones Tomadas**

* Se eliminó el campo "subcategoria" de Bibliotecas Populares.
* Se mantuvo solo el primer correo en "Mail" de Establecimientos.
* Se integraron los datos poblacionales al nivel educativo según departamento para mejor comparación.

**Análisis de Datos y Consultas SQL**

**Consulta i:** Relación EE - población por nivel.

* Datos ordenados por provincia y cantidad de escuelas primarias.
* Permite observar si la infraestructura acompaña la cantidad de niños y jóvenes.

**Consulta ii:** BP fundadas desde 1950.

* Permite observar desarrollo cultural reciente. Predominan provincias con grandes urbes.

**Consulta iii:** Cantidad de EE, BP y población total.

* Da una vista integral de la dotación institucional de cada departamento.

**Consulta iv:** Dominio de mail más frecuente en BP.

* gmail y hotmail son los más frecuentes, reflejando tecnologías de uso común.

**Visualizaciones**

1. **Cantidad de BP por provincia:**
   * Buenos Aires lidera ampliamente. Posible asociación con tamaño poblacional.
2. **Cantidad de EE por nivel vs población:**
   * El gráfico muestra si la infraestructura acompaña el volumen demográfico etario.
3. **Boxplot de EE por departamento por provincia:**
   * Provincias con alta dispersión presentan mayor desigualdad territorial.
4. **Relación EE y BP cada mil habitantes:**
   * Algunos departamentos tienen buena cobertura de ambas instituciones, otros revelan desigualdades.

**Conclusiones**

Del análisis se desprende que en general existe cierta correlación positiva entre cantidad de EE y BP por departamento, aunque no es estricta. Existen casos donde el número de BP es bajo pese a alta cantidad de EE o población, lo cual podría deberse a políticas provinciales, infraestructura o demanda cultural local. El modelo relacional permitió estructurar los datos para consultas eficientes, y las visualizaciones facilitaron la detección de patrones. Como extensión, podría explorarse si la presencia de BP impacta en indicadores educativos o culturales.

G