

Laboratorio de Datos - Instituto de Cálculo - Comisión 2

Trabajo Práctico 1

El informe se debe entregar por campus (solapa "Evaluación") en formato Rmd junto con el PDF generado (lo debe subir un representante del grupo, ambos archivos con el nombre de "Grupo_X", siendo X el número de grupo. Posteriormente, debe entregarse la versión impresa del PDF. Considere que la extensión del informe no debe superar las 12 carillas.

Fecha límite de entrega por campus: lunes 5/5/2025 a las 23:50 hs

Fecha límite de entrega impresa: martes 6/5/2025 a las 14:00 hs

Importante: leer todo el enunciado del TP antes de comenzar a resolverlo.

Enunciado

En este trabajo práctico interesa analizar datos relativos a la oferta de establecimientos educativos en la Provincia de Buenos Aires (PBA) y sus características.

En particular interesa responder las siguientes preguntas:

- A. ¿Cómo es la distribución de cantidad de establecimientos educativos en la PBA por categoría, área y ámbito?
- B. ¿Existe alguna relación entre la matrícula y la cantidad de secciones de un establecimiento. ¿Qué ocurre si se analiza discriminando por modalidad, área, sector, etc.?
- C. ¿Cuántos estudiantes y cuántas secciones tienen en promedio los establecimientos de una misma dependencia? ¿Cómo es la variabilidad? En especial interesa analizar las dependencias que tengan más de 500 establecimientos bajo su órbita, mientras que el resto puede ser analizado en conjunto como una categoría "otro" (en referencia a otro tipo de dependencia).
- D. ¿Cómo se caracteriza la oferta de turnos para cada uno de los niveles educativos? ¿Varía según área, sector y el tipo de dependencia?

Notas:

Modalidad: modalidad educativa (ejemplo: educación especial, común, etc.)

Área: refiere a si es conurbano o interior

Sector: Ámbito de gestión del establecimiento (ejemplo: privado, estatal, etc.)

Dependencia: Jurisdicción de la que dependen normativamente (oficial, municipal, privada, nacional, u otros organismos).

Ámbito: Refiere a si es ámbito urbano, rural disperso, rural agrupado, o itinerante.

Categoría: Clasificación según pautas de organización y otras características de los niveles educativos (existen establecimientos de primera, segunda, y tercera categoría).

Secciones: Cantidad de grupos escolares organizados (pueden pensarlo como cantidad de cursos por escuela).

Matrícula: Cantidad de alumnos registrados en una unidad educativa

Turnos: Jornadas escolares durante las cuales opera el establecimiento (ejemplo: mañana, tarde, vespertino, etc.)

Nivel: Nivel educativo (ejemplo: inicial, primario, secundario, etc.)

Dependencia: Jurisdicción de la que dependen normativamente (ejemplo: municipal, nacional, etc.)

Parte A

Ejercicio 1

Diseñar un DER en función de los datos requeridos para responder las preguntas clave previamente establecidas, identificando entidades fuertes, débiles (en caso de existir), relaciones y cardinalidad de las mismas. Incorporar los atributos necesarios, subrayando aquellos que sean identificatorios. Identificar el grado de participación de una entidad en una relación (si tienen más de una opción explicitar los supuestos).

Para elaborar el DER, sugerimos asumir lo siguiente:

- Categoría, sector, área, nivel, dependencia y ámbito pueden ser considerados atributos simples de un establecimiento.
- Un establecimiento tiene un solo tipo de modalidad (Ej, “educación común”).

Un establecimiento puede ofrecer varios turnos, cada uno de ellos con un rango horario único definido.

Ejercicio 2

Mapear al modelo relacional el DER propuesto en el Ejercicio 1.

Parte B

Para abordar las preguntas de interés (o al menos algunas de ellas), se utilizarán datos provenientes de [datos abiertos del Gobierno de la Provincia de Buenos Aires](https://datos.abiertos.del.gobierno.de.la.provincia.de.buenos.aires), específicamente el conjunto de datos denominado “Establecimientos educativos por modalidad y nivel”, disponibles en el siguiente enlace: <https://catalogo.datos.gba.gob.ar/dataset/establecimientos-educativos>

En dicho sitio encontrarán el dataset en distintos formatos, así como un archivo titulado “**Detalle ...**”, que resulta fundamental para comprender la estructura y contenido del conjunto de datos. Además, se recomienda complementar esta información consultando otras fuentes que permitan interpretar adecuadamente el significado de las variables.

Ejercicio 3

Responder las preguntas de interés A, B y C, utilizando las herramientas trabajadas en clase, ampliando su aplicación en caso de ser necesario. La presentación de los resultados debe incluir tablas y/o gráficos, según corresponda a la naturaleza de cada análisis.

Parte C

Ejercicio 4

¿Cómo procederían para poder responder la Pregunta “D”? ¿Consideran que existe una forma más adecuada de ofrecer los datos que mediante un único esquema (tabla)? Fundamenten su respuesta y relacionarla con el DER o el mapeo al Modelo Relacional propuesto en la Parte A.

Ejercicio 5

Para una segunda etapa del proceso de análisis de datos, es fundamental contar con la variable “categoría” lo más completa posible. Aplicar la metodología **Goal-Question-Metric (GQM)** para evaluar el atributo de *completitud* de esta variable.

Sugerencias para la elaboración del informe

A continuación se presentan algunas pautas para guiar el análisis de datos y la respuesta a las preguntas de interés. Estas sugerencias no son excluyentes: pueden incorporarse otras herramientas o enfoques que se consideren pertinentes. Es importante tener en cuenta que todas las decisiones metodológicas deben ser justificables y coherentes con los objetivos del análisis.

Pautas generales:

- Descripción del dataset: Incluir una caracterización general del conjunto de datos: cantidad de observaciones, número de variables, tipo de variables, etc.
- Identificación de variables relevantes: Determinar qué variables son pertinentes para responder a las preguntas planteadas.
- Análisis de datos faltantes: Evaluar la presencia de datos faltantes, tanto en términos generales como por variable. ¡No olvidar verificar cómo fueron codificados!
- Tablas resumen: Elaborar tablas que incluyan medidas de tendencia central, dispersión, frecuencias absolutas y relativas, según corresponda a la naturaleza de cada variable. Las tablas deben estar numeradas (por ejemplo, *Tabla 1*, *Tabla 2*, etc.), correctamente estructuradas y acompañadas de un epígrafe claro y descriptivo.
- Visualización de datos: Cuando corresponda, incluir gráficos. Estos deben estar numerados (por ejemplo, *Figura 1*, *Figura 2*, etc.), ser legibles, contar con títulos y rótulos adecuados, y un epígrafe explicativo.

Estructura sugerida del informe:

Aunque hay distintas formas válidas de organizar un informe, se sugiere la siguiente estructura básica:

1. Carátula
2. Introducción
3. Metodología
4. Resultados
5. Discusión
6. Referencias bibliográficas

Criterios de evaluación del trabajo práctico:

- Pertinencia del contenido respecto a lo trabajado en clase: el abordaje debe estar en línea con los contenidos y herramientas vistos en clase, pudiendo profundizar los mismos.
- Calidad y claridad del informe: Se valorará la capacidad para comunicar de forma precisa, coherente y estructurada los resultados y conclusiones del análisis.
- Adecuación al formato requerido: El informe debe respetar las pautas de formato especificadas (estructura, extensión, estilo, etc.).
- Correcta ejecución técnica y presentación del archivo .Rmd: Se evaluará que el archivo se ejecute sin errores, y que el código esté organizado y con comentarios necesarios que faciliten su comprensión.