



# Práctica Nro. 5

## Análisis de Datos y Visualización

**Publicación: 27/10/2025**  
**Finalización: 03/11/2025**

### PARTE I. Datos abiertos

1. Investigar y mencionar 3 sitios de datos abiertos donde se puedan obtener *datasets* relevantes para diferentes análisis a realizar. Describa el tipo de datos que se puede encontrar en cada uno.
2. ¿Cuál es la diferencia entre datos públicos y datos abiertos? Proporcione un ejemplo de cada tipo.
3. Mencione 3 tipos de licencias que pueden tener los datos abiertos, describiendo diferencias entre ellas.
4. Suponga que tiene acceso a un dataset abierto sobre los niveles de contaminación del aire en diferentes ciudades del país. ¿Qué tipos de análisis podría realizar para obtener información útil?
5. ¿Cuáles son las condiciones de la licencia creative commons?
6. ¿Qué significa tener datos IA-ready?
7. ¿Cuáles son los principios FAIR-R y sus requisitos?

### PARTE II. Visualización de Datos

Responda las siguientes preguntas sobre visualización de datos:

1. Explique que es una medida y una dimensión
2. Explique la diferencia entre un dato discreto y un dato continuo
3. ¿Por qué es importante preparar los datos?



Dadas las situaciones que se presentan a continuación, decidir qué tipo de gráfico utilizaría para visualizar la información de manera clara y efectiva. Justifique su elección indicando por qué ese tipo de gráfico es el más adecuado para cada caso.

*a. Comparación de suscripciones anuales por región geográfica*

Se cuenta con un conjunto de datos de las suscripciones anuales de una empresa de telefonía celular en distintas regiones (Norte, Sur, Este, Oeste) durante los últimos 5 años. ¿Qué tipo de gráfico utilizaría para comparar las ventas entre las regiones durante los 5 años?

*b. Análisis de la distribución de las edades de clientes*

Se tiene un dataset con datos de los clientes de una tienda virtual, entre ellos, la edad. El objetivo es entender cómo se distribuyen las edades de los clientes. ¿Qué tipo de gráfico utilizará para representar la distribución de las edades de los clientes?

*c. Relación entre el precio y la puntuación otorgada por el cliente*

Se tiene información sobre el precio de diferentes servicios ofrecidos y la calificación otorgada por los clientes. ¿Qué tipo de gráfico usaría para analizar si existe una relación entre el precio del producto y la puntuación marcada por el cliente?

*d. Análisis de los préstamos de libros por género*

Se cuenta con el registro de los préstamos de una biblioteca escolar. Entre los datos de cada uno de estos se tiene el género del libro (narrativa, poesía, cuento, novela y biografía). ¿Qué gráfico es adecuado para visualizar la proporción de préstamos de cada género?

e. Se han almacenado datos de registro de la temperatura promedio del mar, a partir de mediciones diarias frente a la costa de Las Toninas, un kilómetro mar adentro. Esta medición se viene realizando durante los últimos 7 años para estudios del comportamiento de la fauna del lugar.

Dado el siguiente esquema de la base de datos:

**TipoGradoTemp** (#tipoGradoTemp, descripcionTipoGrado)

**TemperaturaRegistrada** (#registroTemp, fecha, hora, valorTemp, #tipoGradoTemp)

**TemperaturasPromedio** (#registroProm, fecha, valorProm)

- a) ¿Qué tipo de gráfico utilizará para mostrar los cambios de la temperatura promedio a lo largo del tiempo?
- b) ¿Cuáles tablas son relevantes para presentar el análisis?

Además, el dataset muestra la cantidad de especies que se identificaron en el mar en las 5 distintas categorías (Moluscos, Artrópodos, Cnidarios, Peces y Mamíferos marinos).

f. Dado el siguiente esquema de la base de datos:



**CategoriaEspecie** (#cat\_especie, cat\_nombre )

**EspecieIdentificada** (#especie, #cat\_especie, nombre\_especie, cantidad)

- c) ¿Qué gráfico utilizaría si se quiere visualizar la proporción de especies de cada categoría?
- d) De todas las tablas propuestas, indicar cuál o cuáles son relevantes para presentar el análisis

*g. Visualización de las estadísticas de una cadena de supermercados*

Una cadena de supermercados quiere contar con estadísticas de las ventas por sucursal y por tipo de producto a lo largo del último año (mes a mes). Se quiere identificar las sucursales con mayores ventas y los tipos de productos que generan más ingresos.

Esta cadena de supermercados quiere visualizar:

- La cantidad de productos vendidos de cada categoría para todas las sucursales, para conocer qué tipo de producto es el que más se vende.
- El total ingresos de la sucursal número 10, mes a mes, durante los últimos 12 meses, para determinar si hubo o no un incremento de los ingresos.

Para ello dispone de una base de datos con el siguiente esquema:

**Venta** (id\_venta, fecha\_venta, id\_sucursal, monto\_total)

**Item\_Venta** (id\_venta, id\_producto, cant)

**Sucursal** (id\_sucursal, ubicación, cant\_empleados)

**Producto** (id\_producto, nombre\_producto, desc\_producto ,precio\_unit, categoria)

**Cliente** (id\_cliente, nombre\_cliente, apellido\_cliente, tipo\_cliente)

Determine qué tipo de gráfico podría utilizar y justifique su elección.

Con los esquemas proporcionados, elegir cuáles -con sus atributos- son relevantes para presentar el análisis visual propuesto anteriormente

### PARTE III: Graficando con Tableau

En esta sección, utilizarás [Tableau Public](#) para explorar y visualizar un datasets sobre el uso de dispositivos móviles y comportamiento del usuario. El archivo con dicho dataset estará adjunto a esta práctica, y tu objetivo será importar este en Tableau y generar gráficos que representan patrones y relaciones claves en los datos.



**Importante:** Agrega leyendas, etiquetas, títulos y otros elementos visuales a cada gráfico, según crea necesario, asegurándose de esta manera que los gráficos sean comprensibles y que proporcionen contexto suficiente.

### 1. Mobile Device Usage and User Behavior Dataset

Este dataset contiene una muestra de 700 usuarios y detalla patrones de uso de dispositivos móviles. Incluye métricas como el tiempo de uso de aplicaciones, consumo de batería y de datos móviles. Cada usuario está clasificado en una de cinco categorías de comportamiento, desde uso ligero hasta extremo. A continuación se describen las columnas más relevantes del dataset:

- User ID: Identificador único del usuario.
- Device Model: Modelo del dispositivo utilizado.
- Operating System: Sistema operativo del dispositivo (iOS o Android).
- App Usage Time: Tiempo diario en minutos dedicado a aplicaciones.
- Screen On Time: Promedio diario de tiempo de pantalla activa.
- Battery Drain: Consumo diario de batería en mAh.
- Number of Apps Installed: Total de aplicaciones instaladas en el dispositivo.
- Data Usage: Consumo diario de datos en MB.
- Age: Edad del usuario.
- Gender: Género del usuario (Masculino o Femenino).
- User Behavior Class: Clasificación de comportamiento en una escala de 1 a 5, según los patrones de uso.

#### Ejercicios:

1. Crea un gráfico de barras horizontales que muestre la cantidad de usuarios para cada modelo de dispositivo. Segmenta las barras por el color del sistema operativo (iOS y Android) para identificar las preferencias de uso por sistema.
2. Muestra una línea que refleje el promedio de aplicaciones instaladas y otra línea para el promedio de datos consumidos, ambos valores en función de la edad del usuario.
3. Crea un histograma que muestre la distribución del número total de aplicaciones instaladas en dispositivos con sistema operativo iOS (filtra los datos para que solo se incluyan dispositivos iOS)
4. Representa en un gráfico de líneas la relación entre la batería gastada y los datos consumidos. Este gráfico permitirá observar cómo el consumo de datos puede influir en el uso de batería.