

# Bases de Datos 1

## Práctica 5 - Análisis de datos y visualización

<b>Parte 1 - Datos abiertos.....</b>	<b>2</b>
1).....	2
2).....	2
3).....	2
4).....	3
<b>Parte 2 - Visualización de datos.....</b>	<b>4</b>
a).....	4
b).....	4
c).....	4
d).....	4
e).....	4
f).....	4
g).....	4
h).....	5
i).....	5
j).....	5
k).....	5

# Parte 1 - Datos abiertos

1)

<https://datos.gob.ar/>

<https://data.gov/>

<https://data.europa.eu/en>

En los tres se pueden encontrar todo tipos de datos pertenecientes a Argentina(1), Estados Unidos (2) y la Unión Europea(3)

2)

Un dato público es cualquier dato generado en el ámbito gubernamental, o que se encuentra bajo su guarda

Los datos abiertos son aquellos de origen público o no a los que cualquier persona puede acceder, usar y compartir libremente. Sólo deben atribuirse y compartirse con la misma licencia con la que fueron publicados

Ejemplo de datos públicos: Registro de votantes:

En muchos países, las autoridades electorales mantienen registros de votantes que son considerados datos públicos. Esto incluye información como nombres, direcciones y el estado de inscripción de los votantes. Aunque es un dato público, el acceso a esta información puede estar regulado y no siempre está disponible para el público en general sin restricciones.

Ejemplo de datos abiertos: Datos de Tráfico Abierto:

Muchas ciudades publican datos abiertos sobre el tráfico, como las tasas de congestión en tiempo real, información sobre accidentes y flujos de vehículos. Estos datos están disponibles en plataformas de datos abiertos y pueden ser utilizados, analizados y compartidos libremente por cualquier persona, siempre que se atribuya correctamente la fuente y se respete la licencia bajo la cual fueron publicados

3)

Open Data Commons Public Domain Dedication and License (PDDL): permite difundir, reutilizar o adaptar los datos sin restricción alguna

Open Data Commons Attribution License: exige la referencia a la autoría o fuente de los datos para la reutilización de la información.

Open Data Commons Open Database License (ODbL): permite la reutilización de los datos siempre que se reconozca la autoría de la información original y se mantenga la misma licencia en las obras derivadas, de las cuales se puede restringir su uso si, además, se distribuye una versión sin dichas Open Data Commons restricciones

## 4)

### Análisis Descriptivo:

- Calcular estadísticas descriptivas de los niveles de contaminación para cada ciudad.

### Análisis Comparativo:

- Comparar los niveles de contaminación entre diferentes ciudades para identificar cuáles tienen los niveles más altos y más bajos.
- Realizar un análisis de correlación para ver si hay alguna relación entre los niveles de contaminación y variables socioeconómicas (como la población, el ingreso promedio, etc.) si se dispone de esos datos.

### Tendencias Temporales:

- Analizar cómo han cambiado los niveles de contaminación a lo largo del tiempo en diferentes ciudades, identificando patrones estacionales o tendencias a largo plazo.
- Crear series temporales para visualizar las fluctuaciones diarias, mensuales o anuales en los niveles de contaminación.

### Análisis de Fuentes de Contaminación:

- Si el dataset incluye información sobre diferentes contaminantes, analizar cuáles son los más prevalentes en cada ciudad y en qué momentos del año son más altos.
- Cruzar datos con información sobre actividades industriales, tráfico o clima para identificar posibles fuentes de contaminación.

### Evaluación del Cumplimiento Normativo:

- Comparar los niveles de contaminación con los límites establecidos por las autoridades ambientales para evaluar si las ciudades cumplen con las normativas de calidad del aire.
- Identificar períodos de incumplimiento y posibles implicaciones para la salud pública.

### Impacto en la Salud Pública:

- Si se dispone de datos sobre salud pública, realizar estudios para correlacionar los niveles de contaminación con enfermedades respiratorias u otros problemas de salud en la población.

## Parte 2 - Visualización de datos

a)

Usaría un gráfico de barras. Esto es porque un gráfico de barras apiladas permite representar eficazmente los datos a lo largo del período de 5 años solicitado, donde cada barra sería un año y dentro de la misma habría 4 partes apiladas, una para cada región

b)

Usaría un gráfico de líneas ya que dicho tipo de gráficos permite mostrar e identificar fácilmente las tendencias, en este caso de decrecimiento, en los datos. Además, al ser mediciones mensuales de los últimos 10 años, hay datos suficientes para poder realizar un gráfico de líneas continuo y fluido

c)

Usaría un gráfico de barras, dado que, se podrían hacer barras para distintos grupos etarios (ej.: <18, 18-25, etc.) lo cual permite visualizar claramente cómo se distribuyen las edades en el conjunto de datos

d)

Usaría un gráfico de dispersión, ya que se quiere comparar dos valores y, para cada par de valores, ver dónde se intersectan

e)

Usaría un gráfico de torta, ya que al ser poca cantidad de categorías, se podrán visualizar de manera correcta los valores en relación al total

f)

Usaría un treemap porque me permite mostrar qué actividades utilizan un mayor presupuesto y a la vez, agruparlas por color, pudiendo ver a qué departamento pertenecen

g)

Usaría un gráfico de líneas, ya que es el tipo más adecuado para identificar tendencias

h)

Usaría una nube de palabras, ya que permiten visualizar la frecuencia de aparición de las palabras en un determinado texto, en este caso, en los comentarios del trailer de la película

i)

Usaría un gráfico de líneas, ya que es el tipo más adecuado para identificar tendencias y evolución en un período de tiempo

j)

Usaría un mapa de área, donde los diferentes tonos en las provincias indicarían la densidad poblacional

k)

Usaría un gráfico de tortas, ya que permite visualizar de manera acertada las proporciones de las categorías sobre el total de ventas y dado que no son muchas las categorías a graficar