ANALISIS DE DATOS DE CASOS DE DENGUE EN ARGENTINA



Autor: Karina Tapia

**Introducción:**

El presente documento muestra los casos de dengue en algunas regiones de Argentina.

El dataset se tomó de la siguiente página web:

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

Se analizo el siguiente archivo:

1. <http://datos.salud.gob.ar/dataset/ceaa8e87-297e-4348-84b8-5c643e172500/resource/8275ff70-94e5-4ee0-9cbb-630587daa618/download/1648473440-6ff92f5b32926ebfb17725a0ab8e991652470725-informacion-publica-dengue-zika-nacional-ha.xlsx>

Corresponde al periodo de **diciembre 2021**.

La tabla original de **diciembre 2021** contiene las siguientes columnas:

departamento\_id, departamento\_nombre, provincia\_id, provincia\_nombre, año, semanas\_epidemiologicas, evento\_nombre, grupo\_edad\_id, grupo\_edad\_desc, cantidad\_casos.

Se procede a realizar limpieza de los datos:

* Se observa que la columna **grupo\_edad\_id** contiene una celda vacía. Se agrega la palabra NULL.
* Las columnas **semanas\_epidemiologicas** y **cantidad\_casos** se trabajaron como tipo numero entero. Mientras que el resto de las columnas se las trabajo como tipo texto.
* Se observa que la columna **provincia\_id** contiene los nombres de las provincias, y la columna **provincia\_nombre** contiene un numero identificador. Por lo tanto, se procede a intercambiar los nombres de dichas columnas.
* La columna **departamento\_nombre** contiene el nombre “Capital” el cual hace referencia a varias provincias, por lo tanto, se procedió a combinar las columnas **departamento\_nombre y provincia\_nombre** con el fin de poder identificar a cuál Capital se hace referencia en la tabla.
* Para obtener una tabla más fácil de interpretar, se eliminan las columnas **departamento\_nombre, provincia\_id, grupo\_edad\_id**
* Se crea la columna **mes**, para indicar que los datos corresponden al mes de diciembre
* Se observa que en la columna **grupo\_edad\_desc** hay un rango de 45 a 64 y luego otro rango de 45 a 65, por lo que se unifico al rango de 45 a 65, según la lógica aplicada en dicha columna.
* En la columna **provincia\_nombre** se cambió SAnta Fe y SAnta Cruz por Santa Fe y Santa Cruz.