

Universidad Internacional de La Rioja

Facultad de Ingeniería y Tecnología

# Máster Universitario en Análisis y Visualización de Datos Masivos / Visual Analitics & Big Data

Visualización de evolución en el tiempo

|  |  |
| --- | --- |
| Actividad de estudio presentado por: | Juan David Escobar Escobar.  Andrés Felipe Leal Mora.  Juan Manuel Bautista Correa. |
| Tipo de trabajo: | Actividad 2 |
| Modalidad: | Individual / Grupal |
| Profesor/a: | Yusef Hassan Moreno |
| Fecha: | Febrero 2022 |

# Estructuración y análisis de datos

La siguiente imagen muestra una muestra de 5 datos del conjunto de datos original, utilizado como fuente de datos para el desarrollo de la actividad:

Table

Description automatically generated with low confidence

Se ignoraron columnas (Location, Age, Sex, Lower value, Upper value) que no generaban valor alguno al momento de representar la información mediante las herramientas de visualización. El siguiente paso fue el análisis de los datos almacenados, mediante la descripción de estos:

Table

Description automatically generated

El conjunto de datos se segmento en dos agrupaciones “Cause Death Or Injury” y “Percent total deaths”, el total de registros fue de 338, cada uno de los conjuntos mencionados quedo con 169 luego de la agrupación.

A continuación, se realizo una operación de agregación, etiquetando de una nueva categoría de las medidas, llamada “Otros”, para aquellos registros que poseen un valor por debajo del 1%.

En el siguiente paso de tratamiento de datos se realizo una operación de pivot para los valores de las categorías [“Cause Death Or Injury” y “Percent total deaths” y “Otros”] almacenadas en la variable “Measure”.

La siguiente imagen ilustra muestra de algunos registros la estructura final del Dataset consolidado y agrupado:

A screenshot of a computer

Description automatically generated with medium confidence

Por ultimo se crea una columna calculada llamada “Orden” para generar un orden principal en el Dataset agrupado por el año 2017, organizando los valores de manera ascendente los valores asociados a los % de muertes del 2017, posteriormente se genera la misma columna para el Dataset que almacena la variación porcentual de muertes entre el año 2005 y 2010, conservando el mismo orden asociado a cada enfermedad estipulado en el primer Dataset del 2017, a continuación se presenta una muestra de ambos Datasets.

A picture containing text, building

Description automatically generated

# Propuesta grafica

# Graphical user interface, table Description automatically generated

# Herramientas y recursos

* Tableau Desktop

<https://github.com/juadaves91/UnirVisualizacion/blob/main/Actividad%202/Book2.twb>

* Excel 2021 (Dataset estructurado):

<https://github.com/juadaves91/UnirVisualizacion/blob/main/Actividad%202/estructurados_act2.xlsx>

* Jupyter Notebook: A continuación, se hace referencia al script generado para el análisis descriptivo de los datos, y la estructuración de estos (<https://github.com/juadaves91/UnirVisualizacion/blob/main/Actividad%202/EstructuracionDatosAct2.ipynb>).