

Universidad de Costa Rica

Escuela de Matemática

Departamento de Matemática Pura y Ciencias Actuariales

HERRAMIENTAS DE CIENCIA DE DATOS I

Análisis de las Causas de Muertes en Costa Rica durante el 2022

Jiménez Campos Cristhian Jesús C33973

Noviembre, 2025

En este documento vamos a explicar los resultados que se han obtenido en el proyecto individual, además de explicación de las librerías. Se trabajó con una base de datos del INEC, la cual fue creada con datos de las defunciones del 2022 y que fueron inscritas en este año. La base de datos tiene columnas que para el análisis que vamos a realizar son innecesarias, por ende será eliminadas para mejor comprensión y manejo de la misma.

Se realizó una filtrado para buscar las defunciones por provincia en todo el año, esto nos dio catorce gráficas, siete de mujeres y siete de hombres, mostrando las cinco principales causa de muerte en el país, tanto para hombres como para mujeres y de manera anual. Se muestran los diferentes tipos de muerte en cada provincia, siendo de utilidad para analizar las enfermedades o precauciones que se tiene que tomar, un ejemplo conciso sería en Cartago le la gráfica de los hombres, la principal causa son los Infartos agudos de miocardio, teniendo una idea de esto se puede tomar medidas y concienciar a las personas sobre esta problemática.

Después de eso se realizó un análisis por meses en el cual se muestran la cantidad de muertes que hay en el país, mostrando en que meses mueren más personas, lo cual es importante, ya que en febrero dio como resultado el mes que más muerte se ha presentado, tanto para hombres como mujeres. Para la menor cantidad en hombres está en octubre y para mujeres la menor cantidad se da en abril.

Se realizó un análisis por promedio de edad, se obtuvo la cantidad de muertes por cada caso y este se sacó un promedio de edad, es importante, ya que este resultado muestra la edad de las personas que pueden ser propensas a tener este tipo de muertes. Generalmente, las diferentes muertes pueden tener diferentes rangos, un ejemplo es que en muertes por motociclistas en hombres en Alajuela el promedio de edad son de 34.9 y 73 casos y en Puntarenas, el promedio de edad es de 33.4 con menos casos 47, entonces tenemos diferentes factores de riesgo que provocan la muerte de los ciudadanos en Costa Rica y se pueden tomar precauciones con las personas entre ese rango de edad (se me olvido filtrar los puntos fuga, entonces esto puede variar).

Luego se hizo un análisis más profundo viendo los cantones, donde se filtran y se consigue un data.frame donde se puede ver cada muerte y en que cantón pertenece, siendo de facilidad ver el lugar donde vive la persona para cierto tipo muerte, ya que se puede ver en menos zona el comportamiento de las personas tanto cuidado físico como cuidado en la alimentación. Además, en este no es de mucha ayuda para las muertes de motocicletas. Los resultados que se puede ver en esta son la cantidad de muertes totales, tanto para hombres como mujeres en los diferentes cantones, un ejemplo el lugar donde más mujeres muerte es el San Rafael con sesenta y de hombres es en Alajuela con sesenta, esto para casos del 2022.

Las variables utilizadas son las siguientes:

- **sexo:** Indica el sexo de la persona fallecida.
- **edads:** Es la edad de la persona fallecida en años cumplidos.
- **provincia:** Es la provincia donde residía la persona fallecida.
- **pc:** Es el cantón donde residía la persona fallecida.
- **causamuer:** Corresponde a la causa básica de muerte.
- **des_causa:** Etiqueta literal de la causa básica de muerte.

Las variables utilizadas son las siguientes: Las librerías usadas en el proyecto son las siguientes

- **library(tidyverse):** Nos ayuda el manejo y análisis de datos de una manera más sencilla y coherente
- **library(gridExtra):** Nos permite combinar gráficas y tablas, además de eso también permite organizar en una misma página, lo que hace que sea más fácil la presentación conjunta de elementos.
- **library(ggplot2):** Una librería que está especializada en la creación de gráficos, que nos permite visualizar los datos de diferentes maneras.
- **library(stringr):** nos ayuda a manejar cadenas de texto, lo que facilita la tarea de buscar, reemplazar, etc.
- **library(plotly):** Se usa para la generación de gráficos interactivos, en el proyecto se combinó con una gráfica de ggplot2, facilitando la comprensión de ciertos puntos de lo creado con ggplot2.
- **library(haven):** se usa para importar y exportar datos, como la base es una base “.sav”.
- **library(dplyr):** Facilita la manipulación y transformación de data frames.
- **library(tidyr):** Ayuda a transformar datos en formatos ”tidy”.
- **library(grid):** ayuda a controlar la posición, el tamaño o diseños de objetos gráficos.
- **library(gt):** Se utilizó para generar tablas de presentación, en el caso se usó en conjunto con ggplot2 y plotly, para generar una gráfica con información que facilita la comprensión en una de las gráficas.

Referencias

Instituto Nacional de Estadística y Censos. (2022). *Diccionario de datos DEFU2022* [Conjunto de datos]. Sistema Nacional de Información Censal y Estadística. https://sistemas.inec.cr/nad_a5.4/index.php/catalog/361/data-dictionary/F3?file_name=DEFU2022

Perplexity AI. (2025). *Perplexity AI* [Herramienta de inteligencia artificial]. <https://www.perplexity.ai/>