

Asignatura: Gestión Y Almacenamiento De Datos

Profesor: Juan Manuel Núñez Velasco

Programa: Esp. Analítica de Big Data

Estudiantes: Estefany Vargas Montaña
Gustavo Adolfo Henao
Jhon David Caicedo
Brayan Camilo Valencia
Diana Patricia Segura

Fecha de entrega: marzo 16 2025

IDENTIFICAR LAS FUENTES DE DATOS – CARACTERIZACIÓN DE LOS DATOS.

Para realizar la caracterización de los datos tuvimos en cuenta estos pasos:

- **Identificar la fuente de datos:** determinar de donde provienen los datos que vamos a analizar

El archivo fuente está en formato .csv y los datos se presentan de manera estructurada; estos datos nos brindan información de lesiones fatales por causas externas obtenidas de diferentes fuentes.

El sitio web del Instituto Nacional de Medicina Legal y Ciencias Forenses publica estadísticas detalladas sobre lesiones de causa externa, incluyendo homicidios, suicidios y accidentes.

Tabla 1. Fuente de dato e identificación del tipo

Nombre de la fuente de datos	Descripción	Tipo de datos	Formato de los datos	Fecha de actualización	Ubicación	Frecuencia de muestreo
Lesiones_fatales_de_causa_externa_2015_a_2023	Lesiones fatales por causas externas.	Cadena (Estructurado)	Filas, columnas (csv)	2024	Local	Anual
Lesiones_fatales_de_causa_externa_2024	Lesiones fatales por causas externas.	Cadena (Estructurado)	Filas, columnas (csv)	2025	Local	Anual

- **Caracterizar datos:** revisar y comprender las propiedades y calidad de los datos antes de analizarlos

La siguiente tabla hace referencia a los nombres de las columnas y lo que los datos contienen.

Tabla 2. Caracterización del tipo de datos de la fuente

Nombre de la Columna	Descripción	Tipo de Dato
ID	Identificador único de cada registro	Entero
Año del hecho	Registra el año en el que ocurrió el hecho	Entero
Grupo de edad de la víctima	Clasifica a las personas por rangos de edad	Texto
Grupo Mayor Menor de Edad	Distingue entre mayores y menores de edad	Texto
Edad judicial	Contextualiza la edad conforme a categorías legales	Texto
Ciclo Vital	Indica el estado en el ciclo de vida	Texto
Sexo de la víctima	Define si es hombre, mujer, u otro	Texto
Estado Civil	Registra el estado civil de la persona	Texto
País de Nacimiento de la Víctima	Muestra el país natal	Texto
Escolaridad	Niveles de educación alcanzados por la víctima	Texto
Pertenencia Grupal	Afiliaciones específicas como adicción a drogas u otros grupos	Texto

Mes del hecho	Indican el mes y día específicos	Texto
Día del hecho		Texto
Rango de Hora del Hecho X 3 Horas		Texto
Código Dane Municipio	Código del Municipio del hecho DANE	Entero
Municipio del hecho DANE	Municipio del hecho DANE	Texto
Departamento del hecho DANE	Departamento del hecho	Texto
Código Dane Departamento	Código del departamento	Entero
Escenario del Hecho	Lugar específico	Texto
Zona del Hecho	Especifica si es cabecera municipal, rural	Texto
Actividad Durante el Hecho	Describe qué realizaba la víctima en ese momento	Texto
Circunstancia del Hecho	Proporciona contexto adicional	Texto
Manera de Muerte	Define el tipo de evento	Texto
Mecanismo Causal	Describe el instrumento usado	Texto
Diagnóstico Topográfico de la Lesión	Área del cuerpo afectada	Texto
Presunto Agresor	Especifica si hay relación con un agresor	Texto
Razón del Suicidio	Solo relevante en casos relacionados	Texto
Condición de la Víctima	Medio usado	Texto
Medio de Desplazamiento o Transporte		Texto
Servicio del Vehículo		Texto
Clase o Tipo de Accidente		Texto
Objeto de Colisión		Texto
Servicio del Objeto de Colisión		Texto
Localidad del Hecho		Texto
Ancestro Racial	Categoriza según el ancestro étnico	Texto
Pueblo Indígena	Identifica si pertenece a una comunidad indígena	Texto
Orientación Sexual	Información relacionada con la identidad de la víctima.	Texto
Identidad de Género		Texto
Transgénero		Texto



UNIVERSIDAD AUTÓNOMA DE OCCIDENTE

El informe se sube al GitHub, enlace: https://github.com/evm2024/Proyecto_G-A