

MODELADO DE DATOS

	NAME	SOURCE LINK	DESCRIPTION	YEARS AVAILABLE
1	UNGRD Emergencies	https://portal.gestiondelriesgo.gov.co/Paginas/Consolidado-Atencion-de-Emergencias.aspx https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Emergencias-UNGRD-/wwkg-r6te/about_data	UNGRD: Consolidado anual de emergencias	1998-2024
2	IDEAM Precipitation	https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Precipitacion/s54a-sgyg	IDEAM: Precipitation database	2003-2026
3	IDEAM Temperature	https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Datos-Hidrometeorologicos-Crudos-Red-de-Estaciones/sbwg-7ju4	IDEAM: Temperature database	2001-2026
6	IDEAM Atmospheric Pressure	https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Presion-Atmosferica/62tk-nxj5	IDEAM: Atmospheric Pressure database	2001-2026
7	GeoJSON Col (By municipios)	https://www.colombiaenmapas.gov.co/	Colombia GeoJSON by municipio	2026
8	DIVIPOLA	https://www.datos.gov.co/Mapas-Nacionales/DIVIPOLA-C-digitos-municipios/gdxc-w37w/about_data	Standardized IDs of all municipalities in Colombia	2026

Table 1. Datasets overview.

EMERGENCIES DATASET

El dataset de emergencias es publicado por la Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD) en el sitio web de Datos Abiertos del Gobierno Colombiano¹ y contiene el registro de eventos naturales o antrópicos no intencionales ocurridos durante el periodo comprendido entre el 1 de enero de 2019 al 31 de diciembre de 2022.

Para este periodo de tiempo de 4 años se hallaron un total de 25,9K filas y 71 columnas.

¹ https://www.datos.gov.co/Ambiente-y-Desarrollo-Sostenible/Emergencias-UNGRD-/wwkg-r6te/about_data

Se indagó más a profundidad y se halló que la UNGRD se encuentra reportando en su portal web esta misma información desde el año 1998 tal como se observa a continuación:

The screenshot shows the official website of the Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD) of Colombia. The header features the GOV.CO logo and the UNGRD logo with its full name in Spanish: 'Unidad Nacional para la Gestión del Riesgo de Desastres'. A navigation bar includes links for Inicio, Transparencia, Atención y Servicios a la Ciudadanía, Participa, La UNGRD (which is highlighted in yellow), Normativa, and Prensa. Below the navigation is a search bar. The main content area is titled 'Consolidado anual de emergencias' and lists emergency reports from 1998 to 2024. At the bottom, there's a section for 'Contacto' with various communication channels and links.

Source: <https://portal.gestiondelriesgo.gov.co/Paginas/Consolidado-Atencion-de-Emergencias.aspx>

Al revisar a detalle todos los archivos anuales publicados se estableció que para años antes de 2019 el formato es variado, no tienen el mismo número de columnas y en general la calidad de los datos no es la adecuada, motivo por el cual se realizaron varias actividades técnicas de unificación que tuvieron como objetivo de obtener la unión adecuada de la información.

Como trabajo adicional se unificaron los datasets publicados por la UNGRD pero para realizar esta tarea se detectaron varios inconvenientes:

- Los reportes de emergencias para años anteriores a 2019 se encuentran publicados en archivos con extensiones: XLS y XLSX con macros internas que se ejecutan en la apertura y revisten un riesgo de seguridad y además muchos de ellos con protección por contraseña que impiden el exportado a formato CSV, pero si permiten copiado de datos.
- Los reportes de emergencias para años anteriores a 2019 no tienen el mismo número de columnas (unos años tienen más y otros menos).
- El código de municipio es un dato muy importante que es conocido como DIVIPOLA², lastimosamente es común que este registro esté ausente en los años 2018 hacia atrás.
- Los campos de texto para departamento y municipio no están estandarizados por lo que existen municipios con varios nombres diferentes debido a: falta de tildes, falta de pronomombres, adición de codificación de siglas de distritos especiales, adición de nombres etc. Ejemplo: Bogotá, Bogota, Bogota D.C, Santa fe de Bogotá.
- Ninguno de los reportes de emergencia reportados por la UNGRD tiene asociado punto GPS, solo se hace a nivel de municipio.
- Las columnas que no tienen datos en ninguno de los años reportados requieren eliminación.

Por los anteriores motivos, se generó la necesidad de hacer las siguientes actividades a fin de obtener un adecuado dataset de emergencias:

- ⊕ Unificación de veinticuatro (27) datasets de emergencias reportadas y publicadas por la UNGRD para los años comprendidos entre 1998 hasta 2024, evitando los inconvenientes de formato no compatible, protección por contraseña y ejecución en ambiente controlado de macros no identificadas.
- ⊕ Generación de un programa que, con base en la información publicada por el DANE, estandarice los nombres de departamentos y municipios de Colombia existentes en el dataset de emergencias unificado generado en con la tarea mencionada en el párrafo anterior.

² **División Político-Administrativa de Colombia – DIVIPOLA:** Según la Constitución Nacional de 1991, para la administración del Estado y la representación política, el territorio nacional se dividirá para fines administrativos en departamentos, distritos, municipios y territorios indígenas, que es lo que se conoce como División político-administrativa. El Departamento Nacional de Estadística de Colombia DANE es la entidad encargada de codificar y publicar los códigos DIVIPOLA desde 1953, Fuente: <https://www.dane.gov.co/files/investigaciones/divipola/divipola2007.pdf>

- Con los textos de departamento y municipio estandarizados por medio de un programa hecho en Python se obtuvo el código asignado por el DANE para cada registro, dicha columna se denomina DIVIPOLA.
- Unificación de columnas similares de reportes de ayuda o atención a la emergencia, esto se realizó con base en el tipo de dato y concepto o representación del mismo.
- Luego se procedió con la ejecución de técnicas de Data Wrangling y Data Cleaning para la totalidad de datos unificados, generando una mejor calidad de datos.

Al final se obtuvo un gran total de un total de 52,1K filas y 45 columnas para emergencias.

PRECIPITATION DATASET

La información es publicada por el Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM en el sitio web de Datos Abiertos del Gobierno Colombiano, y contiene datos de precipitación crudos provenientes de los sensores de las estaciones automáticas de la red propia del IDEAM y registros producto de convenios interadministrativos con terceras entidades, la información se actualiza con un retraso de un día calendario atrás, conteniendo datos desde el mes de **enero del año 2003** con una periodicidad variable entre 5 a 10 minutos por cada sensor.

El total de registros a 20 de febrero de 2026 es de 270Millones de registros distribuidos en 12 columnas.

TEMPERATURE DATASET

La información es publicada por el Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM en el sitio web de Datos Abiertos del Gobierno Colombiano, y contiene datos de temperatura crudos provenientes de los sensores de las estaciones automáticas de la red propia del IDEAM y registros producto de convenios interadministrativos con terceras entidades, la información se actualiza con un retraso de un día calendario atrás, conteniendo datos desde el mes de **enero del año 2001** con una periodicidad variable entre 5 a 10 minutos por cada sensor.

El total de registros a 20 de febrero de 2026 es de 88,7Millones de registros distribuidos en 12 columnas.

ATMOSPHERIC PRESSURE DATASET

La información es publicada por el Instituto de Hidrología, Meteorología y Estudios Ambientales IDEAM en el sitio web de Datos Abiertos del Gobierno Colombiano, y contiene datos presión atmosférica crudos provenientes de los sensores de las estaciones automáticas de la red propia del IDEAM y registros producto de convenios interadministrativos con terceras entidades, la información se actualiza con un retraso de un día calendario atrás, conteniendo datos desde el mes de **marzo del año 2001** con una periodicidad variable entre 5 a 10 minutos por cada sensor.

El total de registros a 20 de febrero de 2026 es de 32,4Millones de registros distribuidos en 12 columnas.

GEOJSON COL POR DEPARTAMENTO Y GEOJSON COL POR MUNICIPIOS

Archivo plano en formato estándar abierto JSON (JavaScript Object Notation) con la geometría aplicada basada en los polígonos determinados por los puntos GPS que delimitan la división política colombiana por departamento y por municipios, respectivamente. Es importante mencionar que el identificador de cada polígono corresponde al utilizado por diferentes entidades públicas (DIVIPOLA), lo que permite vincular información genérica con datos georreferenciados.

DIVIPOLA

DIVIPOLA es un conjunto de datos creado por el "Departamento Administrativo Nacional de Estadística" (DANE), la entidad oficial del gobierno de Colombia encargada de medir, monitorear, procesar, analizar y difundir las métricas estadísticas oficiales del país. Este conjunto de datos contiene una nomenclatura de identificación estandarizada diseñada por el DANE para la identificación de Entidades Territoriales, denominada Divipola. El DANE define la Divipola como un estándar nacional que codifica y lista las entidades territoriales: departamentos, municipios, corregimientos departamentales, así como centros poblados, inspecciones de policía, caseríos y corregimientos municipales en zonas rurales.

Diccionario de Datos — Emergencias UNGRD (1998–2024)

Fuente: Unidad Nacional para la Gestión del Riesgo de Desastres (UNGRD)

Nombre de columna	Descripción	Tipo de dato	Rango
Fecha	Fecha y hora del evento de emergencia	datetime	1995–2024
Departamento	Subdivisión del país (departamento)	object	32 departamentos
Municipio	Municipio donde ocurrió el evento	object	—
Evento	Clasificación del evento según fenómeno natural o antrópico no intencional	object	36 tipos de evento
codificación segun divipola	Código Divipola. Código único asignado por el DANE a cada municipio	object	—
RUD	Código RUD de sistema interno de la UNGRD	object	----
Muertos	Número de víctimas mortales del evento	int	[0 – ∞)
Heridos	Número de personas heridas durante el evento	int	[0 – ∞)
Desaparecidos	Número de personas desaparecidas a causa del evento	int	[0 – ∞)
Personas	Número de personas afectadas por el evento	int	[0 – ∞)
Familias	Número de familias afectadas por el evento	int	[0 – ∞)
Viviendas destruidas	Número de viviendas destruidas durante el evento	int	[0 – ∞)
Viviendas averiadas	Número de viviendas dañadas durante el evento	int	[0 – ∞)
Vías afectadas	Número de vías dañadas por el evento	int	[0 – ∞)

Puentes vehiculares dañados	Número de puentes vehiculares dañados o afectados por el evento	int	[0 – ∞)
Puentes peatonales dañados	Número de puentes peatonales dañados o afectados por el evento	int	[0 – ∞)
Acueductos afectados	Número de acueductos dañados o afectados por el evento	int	[0 – ∞)
Alcantarillados afectados	Número de alcantarillados dañados o afectados por el evento	int	[0 – ∞)
Centros de salud afectados	Número de centros de salud dañados o afectados por el evento	int	[0 – ∞)
Centros educativos afectados	Número de centros educativos dañados o afectados por el evento	int	[0 – ∞)
Centros comunitarios afectados	Número de centros comunitarios dañados o afectados por el evento	int	[0 – ∞)
Hectáreas afectadas	Número de hectáreas destruidas/afectadas por el evento	int	[0 – ∞)
Otros	Observaciones	Object	—
(...)	Columnas no usadas	NULL	NULL
DIVIPOLA	Código DANE CORREGIDO	object	—