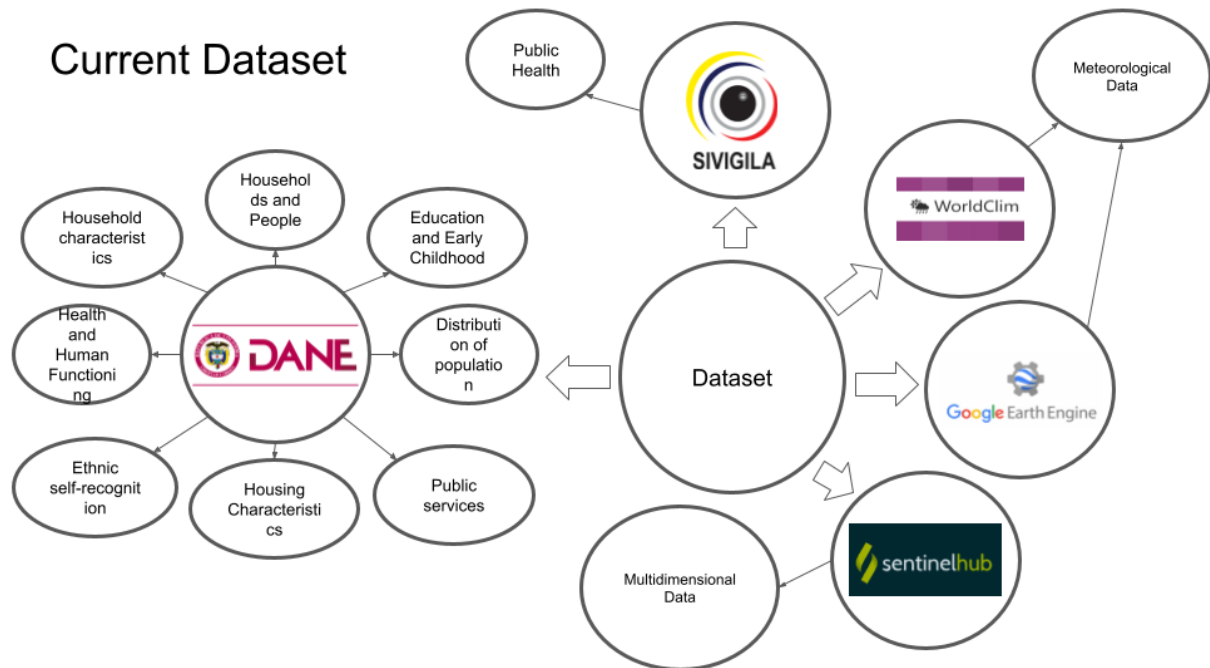


CONJUNTO DE DATOS ACTUAL



1- DATOS TABULARES (Compuesto por 2 archivos)

1. Set de datos con datos sociodemográficos y socioeconómicos (1.2), casos de dengue (1.1) y temperatura y precipitación (1.3) para todos los municipios:

- **dengue_data_all_municipalities.csv**

Municipality code	Municipality	Population/Year	...	Cases/Year	Temperature_month	Precipitation_month	Year/epiweek
5002	Abejorral	20000		23	19	12	12

2. Set de datos para el Municipio de Medellín que contiene información más refinada de Temperatura y precipitación (1.4).

- **Dengue_Dataset(Medellin).csv**

cases_m edellin	DATA	YEAR	YearWeek	LastDayWeek	MONTH	precipitation_ medellin	precipitation_m edellin_urbano	precipitation_m edellin_rural	temperature
1	2007-w1	2007	200701	16/01/2007	1	12	13	12	19

1.1. Casos de Dengue:

(Fuente de Datos -> SIVIGILA):

- Casos de dengue semanales basado en la semana epidemiológica: desde la primera semana epidemiológica del año 2007 a la última semana epidemiológica del año 2019.
- También se puede encontrar casos anuales de dengue entre 2007 y 2019.

Nombre de la columna:

- Year/epiweek (Año/Semana epidemiológica)
 - Ejemplo:
 - 2007/w01, 2007/w02, ..., 2019/w26

1.2. Datos Sociodemograficos y Socioeconómicos:

(Fuente de Datos -> DANE)

- Identificador único del Municipio.
 - Nombre de la columna:
 - Municipality code.
- Nombre del Municipio.
 - Nombre de la columna:
 - Municipality.
- Población para cada municipio entre el año 2007 y el 2019.
 - Nombres de las columnas:
 - Population2007, Population2008, ..., Population2019.
- Porcentaje de población que pertenece a una cierta edad.
 - Nombres de las columnas:
 - Age0-4(%), Age5-14(%), Age15-29(%), Age>30(%)
- Porcentaje de población afrocolombiana.
 - Nombre de la columna:
 - AfrocolombianPopulation(%)
- Porcentaje de población india.
 - Nombre de la columna:
 - IndianPopulation(%)
- Porcentaje de personas con discapacidades: esta variable describe el grupo de personas que tienen alguna limitación física, psicológica o mental.
 - Nombre de la columna:
 - PeoplewithDisabilities(%)
- Porcentaje de personas que no pueden leer o escribir
 - Nombre de la columna:
 - Peoplewhocannotreadorwrite(%)
- Porcentaje de personas que tiene nivel de educación secundario/superior
 - Nombre de la columna:
 - Secondary/HigherEducation(%)
- Porcentaje de población empleada
 - Nombre de la columna:
 - Employedpopulation(%)
- Porcentaje de población desempleada
 - Nombre de la columna:
 - Unemployedpopulation(%)
- Porcentaje de personas que realizan trabajo doméstico
 - Nombre de la columna:
 - Peopledoinghousework(%)
- Porcentaje de personas jubiladas
 - Nombre de la columna:
 - Retiredpeople(%)

- Género o población en porcentaje de hombres y mujeres.
 - Nombres de las columnas:
 - Men(%), Women(%)
- Hogares sin acceso a agua.
 - Nombre de la columna:
 - Householdswithoutwateraccess(%)
- Hogares sin acceso a internet.
 - Nombre de la columna:
 - Householdswithoutinternetaccess(%)
- Estratificación de las viviendas entre 1 y 6.
 - Nombres de las columnas:
 - Buildingstratification1(%), Buildingstratification2(%), ..., Buildingstratification6(%)
- Número de hospitales por Km2:
 - Nombre de la columna:
 - NumberofhospitalsperKm2
- Número de casas por Km2
 - Nombre de la columna:
 - NumberofhousesperKm2

1.3. Temperatura y Precipitaciones:

c. Temperatura:

- Temperatura mensual para cada municipio en Colombia.

d. Precipitaciones:

- Precipitaciones mensuales para cada municipio en Colombia.

Nombre de la columna:

- VariableName_Month_year (variable Nombre_mes_año)
 - Ejemplo:
 - PRECIPITATION_jan_07, ..., PRECIPITATION_dec_18
 - TEMPERATURE_jan_07, ..., TEMPERATURE_dec_18

1.4. Sólo para Medellín:

- **Dengue_Dataset(Medellin).csv**

Set de datos con casos de dengue en Medellín, pero con datos de temperatura y precipitación semanales basados en la semana epidemiológica.

2- IMÁGENES SATELITALES

- Imágenes bisemanales basadas en la semana epidemiológica.
- 10 m/px, 12 bandas (interpolación de vecinos más cercanos para bandas con resolución menor que 10 m/px)
- Las mejores imágenes se obtuvieron usando la menor cantidad de algoritmos de nubes.

En el datathon sólo se usarán los embeddings de Medellín

Enlace: https://github.com/MITCriticalData-Colombia/SatDengue_MakeHealth

- **Nota:** Un ejemplo de como leer y trabajar con estos archivos se puede encontrar en el siguiente enlace: <https://www.kaggle.com/code/davidrestrepo/read-embeddings/notebook>

Las características satelitales de 164 imágenes de Medellín se pueden descargar aquí en formato csv. La forma de cada archivo csv viene dada por la estructura (164, n_características + 1), donde n_características representa el número de características obtenidas para cada modelo y la columna adicional es la fecha de la imagen:

- **2.1. Extracción de características de imágenes satelitales usando modelos de deep learning**
 - [features_resnet50.csv](#) - [Download](#): Extracción de características basada en resnet50 preentrenado en Imagenet - Extraído de Sentinel 2 en Medellín entre 2015-2018
 - [features_transformer.csv](#) - [Download](#): Extracción de características basada en vision transformers preentrenado en Imagenet - Extraído de Sentinel 2 en Medellín entre 2015-2018.
- **2.2. Reducción de dimensionalidad de imágenes satelitales con Autoencoders Variacionales**
 - [embeddingsmedellin100features.csv](#) - [Download](#): Características generadas usando un autocodificador variacional con espacio latente de 100 (100 características) en formato csv - Extraído de Sentinel 2 en Medellín entre 2015-2018
 - [embeddingsmedellin200features.csv](#) - [Download](#): Características generadas usando un autocodificador variacional con espacio latente de 200 (200 características) en formato csv - Extraído de Sentinel 2 en Medellín entre 2015-2018
- **2.3. Reducción de dimensionalidad de imágenes satelitales con análisis de componentes principales (PCA)**
 - [pcamedellin100features.csv](#) - [Download](#): Embeddings generados usando los primeros 100 componentes principales en formato csv - Extraídos de Sentinel 2 en Medellín entre 2015-2018
 - [pcamedellin120features\(per_band\).csv](#) - [Download](#): Embeddings generated using the first 10 principal components in each band (120 features in total per image) in csv format - Extracted from Sentinel 2 in Medellin between 2015-2018