



ANÁLISIS DE DATOS

# DESASTRES NATURALES

IMPACTO HUMANO Y ECONÓMICO

**Soluciones aplicadas de Análisis de Datos y Aprendizaje Automático para entender y prever la magnitud destructiva de fenómenos naturales, con el fin de mitigar su impacto.**

**Equipo 2**  
Alejandra Eng  
Ángel Bardon  
Elena Blanco  
Rafael De Marco



# OBJETIVOS

## Académicos:

Adquirir conocimientos y habilidades en un contexto real, mediante metodología CRISP-DM.

## Aplicación Práctica:

Conocer la magnitud y el impacto socio-económico para evaluar riesgos y diseñar soluciones prácticas: EDA, modelos predictivos y dashboards.



**EDA:**  
Terremotos  
Sequías  
Tormentas  
Inundaciones



**Modelo Predictivo:**  
Nivel de Destrucción



**Modelo de Negocio**

## DESASTRES NATURALES

# Base de datos

### EM-DAT (Emergency Events Database)

Fuente: Centre for Research on the Epidemiology of Disasters (CRED), Univ. Católica de Lovaina (Bélgica).

Ocurrencia de desastres masivos desde 1900 hasta la actualidad.

<https://www.emdat.be/>

<https://www.emdat.be/guidelines>

#### Fuente



Excel

#### Número de registros

**26.004** filas

50 columnas

#### Grupos de desastres

**Naturales**

Tecnológicos  
Complejos

#### Selección

**16.636** filas

26 columnas

#### Período que abarca

**122 años**

1900 a 2023

#### Tipos de desastres

**15** ..... 4

● Inundación

● Sequía

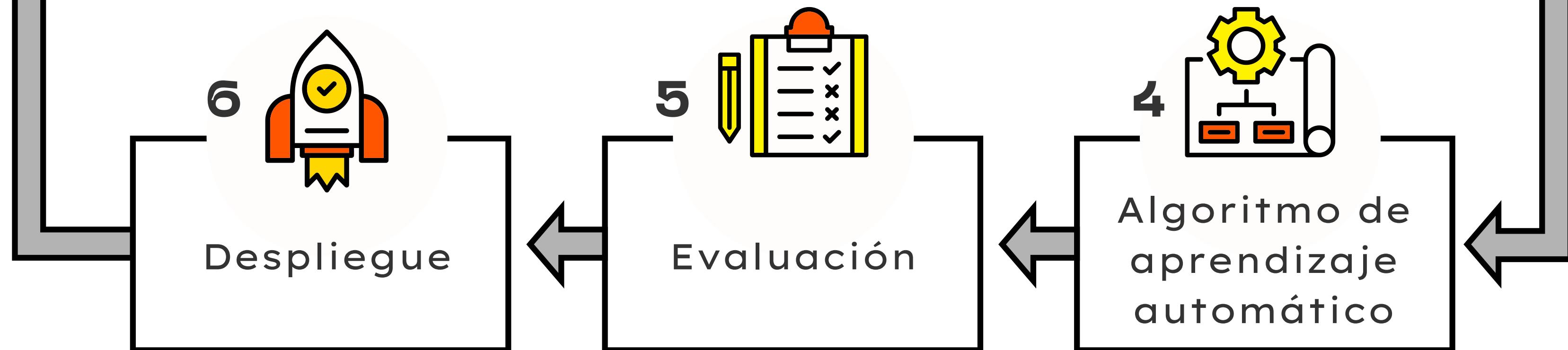
● Terremoto

● Tormenta

## DESASTRES NATURALES

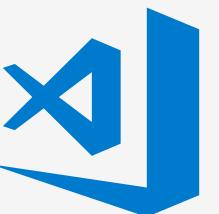


## METODOLOGÍA CRISP-DM



## DESASTRES NATURALES

## PROCESO Y HERRAMIENTAS



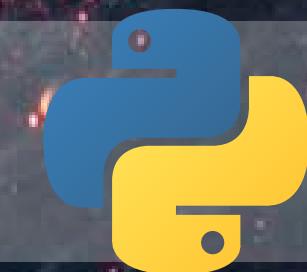
Comprensión del problema



Entendimiento de los datos



Preparación de datos



Modelado

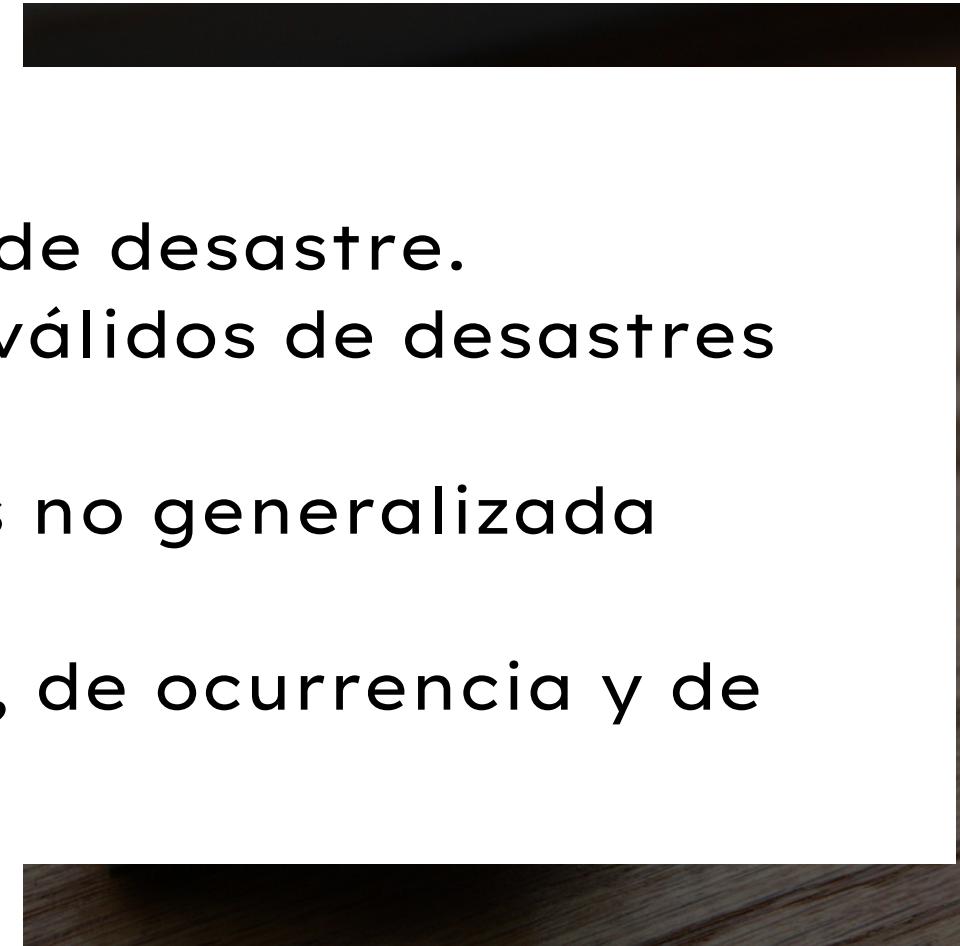


Business Intelligence



# PROCESOS GENERALES EDA

- 4 tipos de desastres.
- Limpieza y segmentación por tipo de desastre.
- Outliers no se eliminan: son datos válidos de desastres con grandes impactos.
- Limpieza y relleno de valores nulos no generalizada debido a naturaleza del dataset.
- Análisis de categorías geográficas, de ocurrencia y de impacto.



# BUSINESS INTELLIGENCE

Gestión del socorro en casos de desastre



**DISASTER RELIEF**  
BI Dashboard



ANÁLISIS DE DATOS. EQUIPO 2

Herramientas





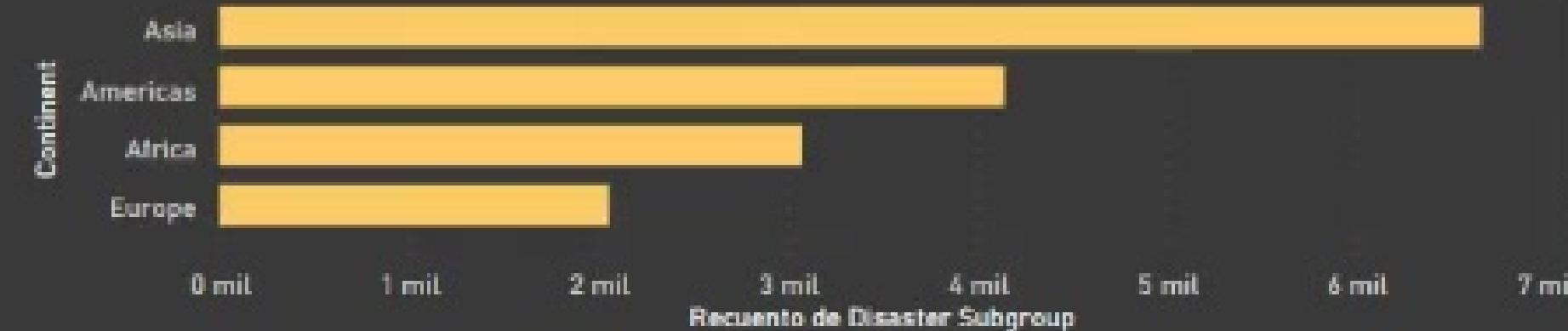
# Desastres Naturales

## Subgrupos y Datos Generales

Fuente: EM-DAT con datos desde 1900-2023.

Clasificación original de desastres naturales:  
Meteorológica, Hidrológica, Geofísica, Climatológica y Biológica.

### Por Continente



### Tipos de Desastres dentro de Subgrupos de Desastres

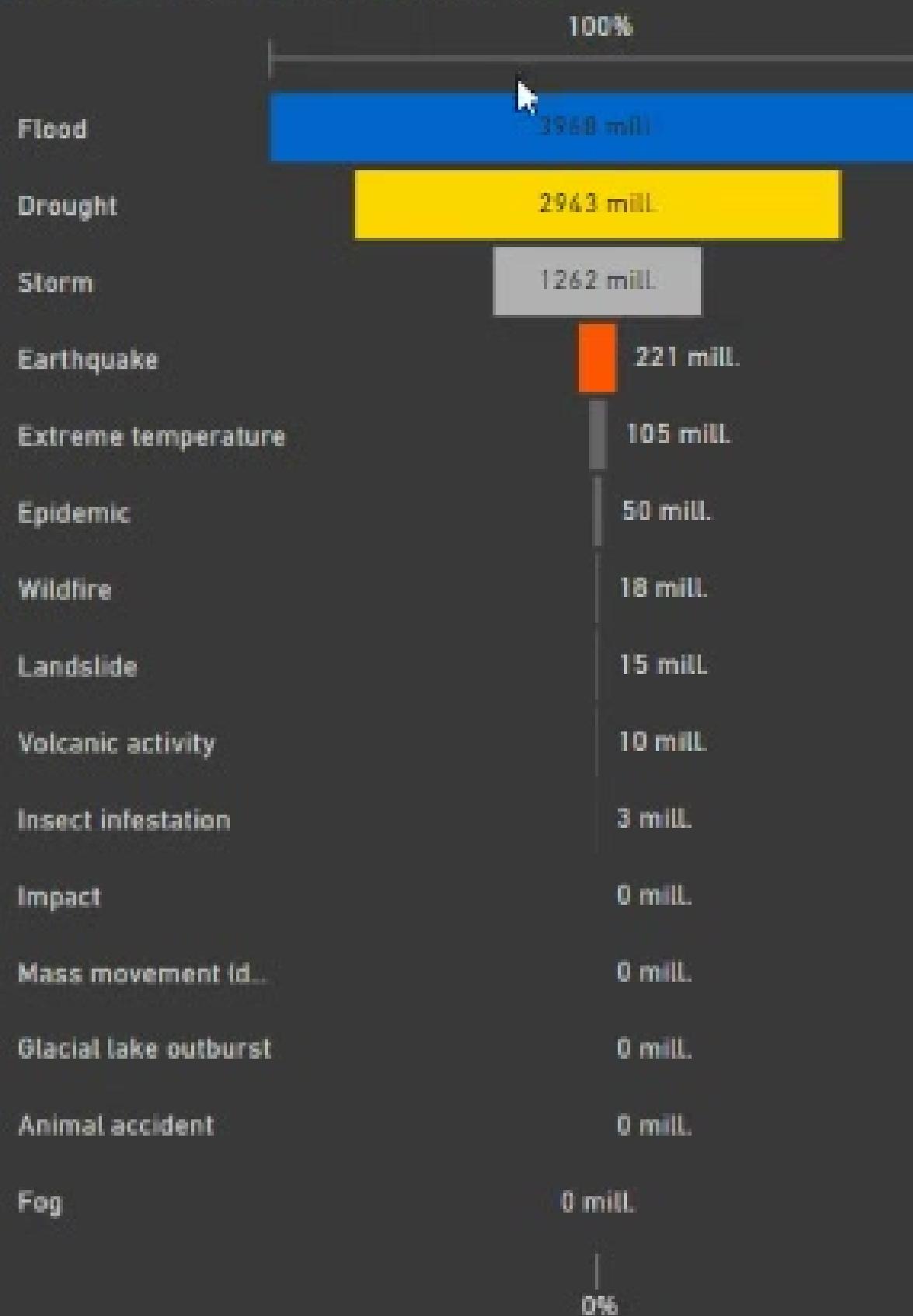




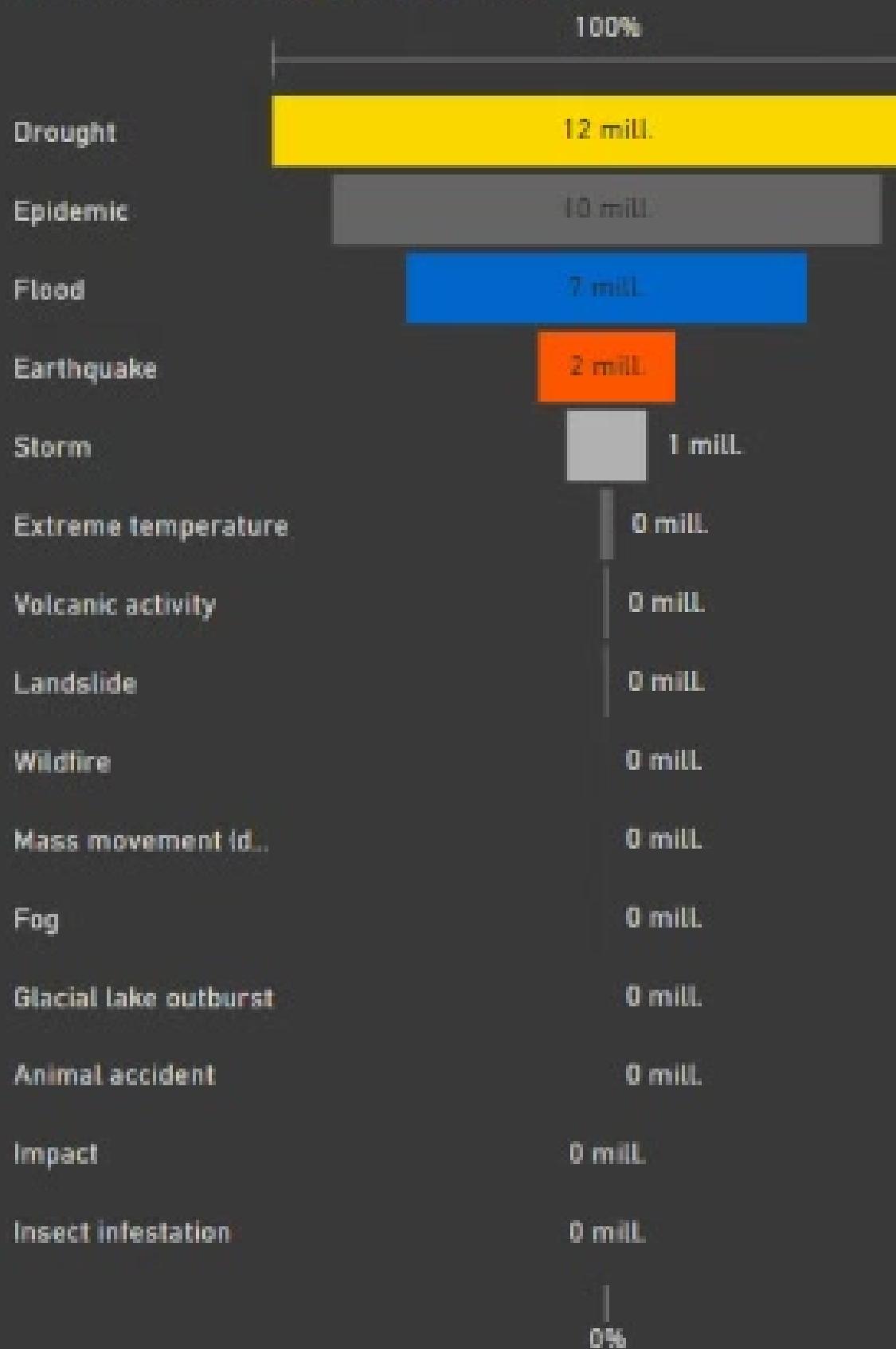
# Desastres Naturales

## Impactos por Tipo

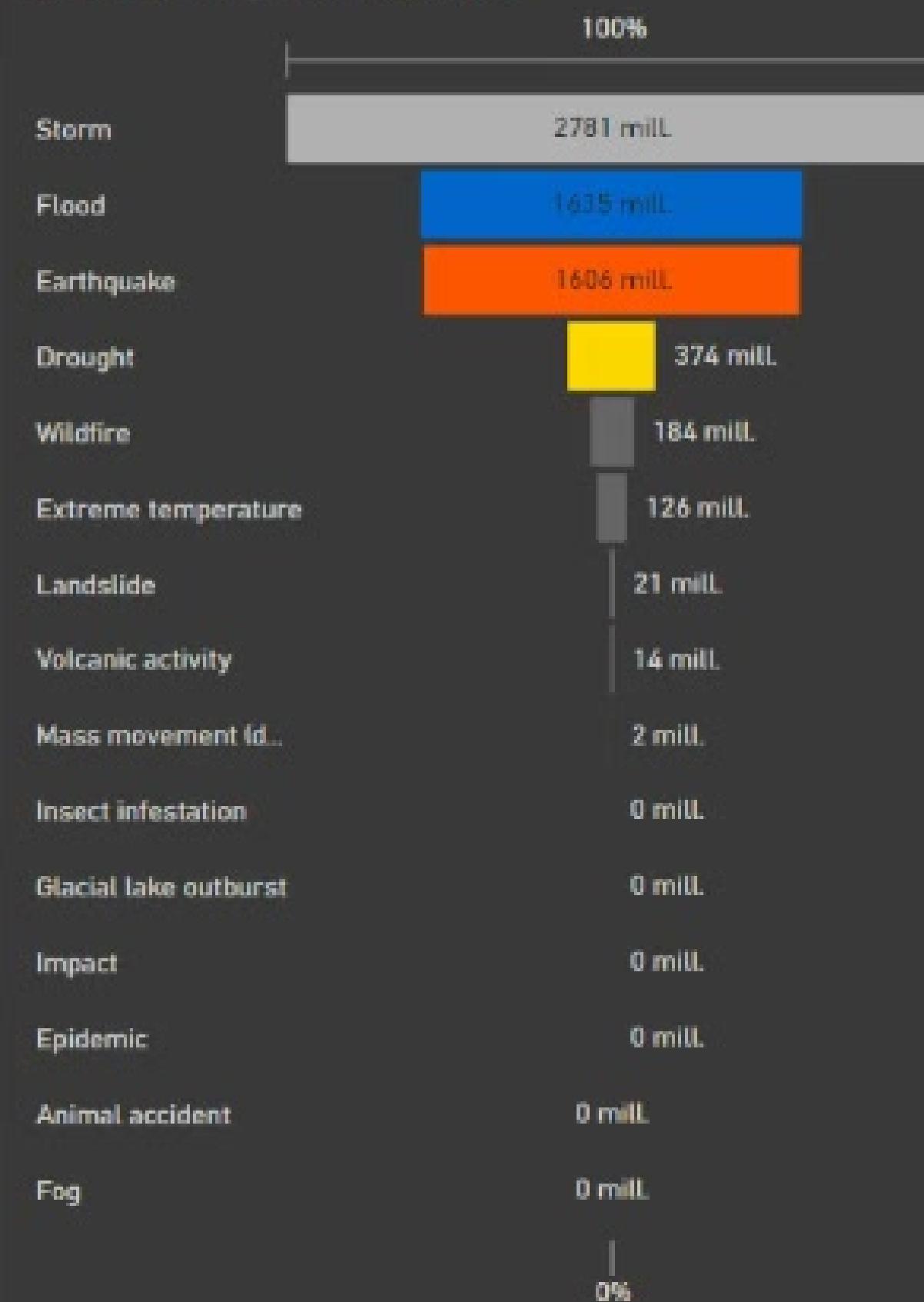
### Afectados por Tipo de Desastre



### Fallecidos por Tipo de Desastre



### Daños por Tipo de Desastre

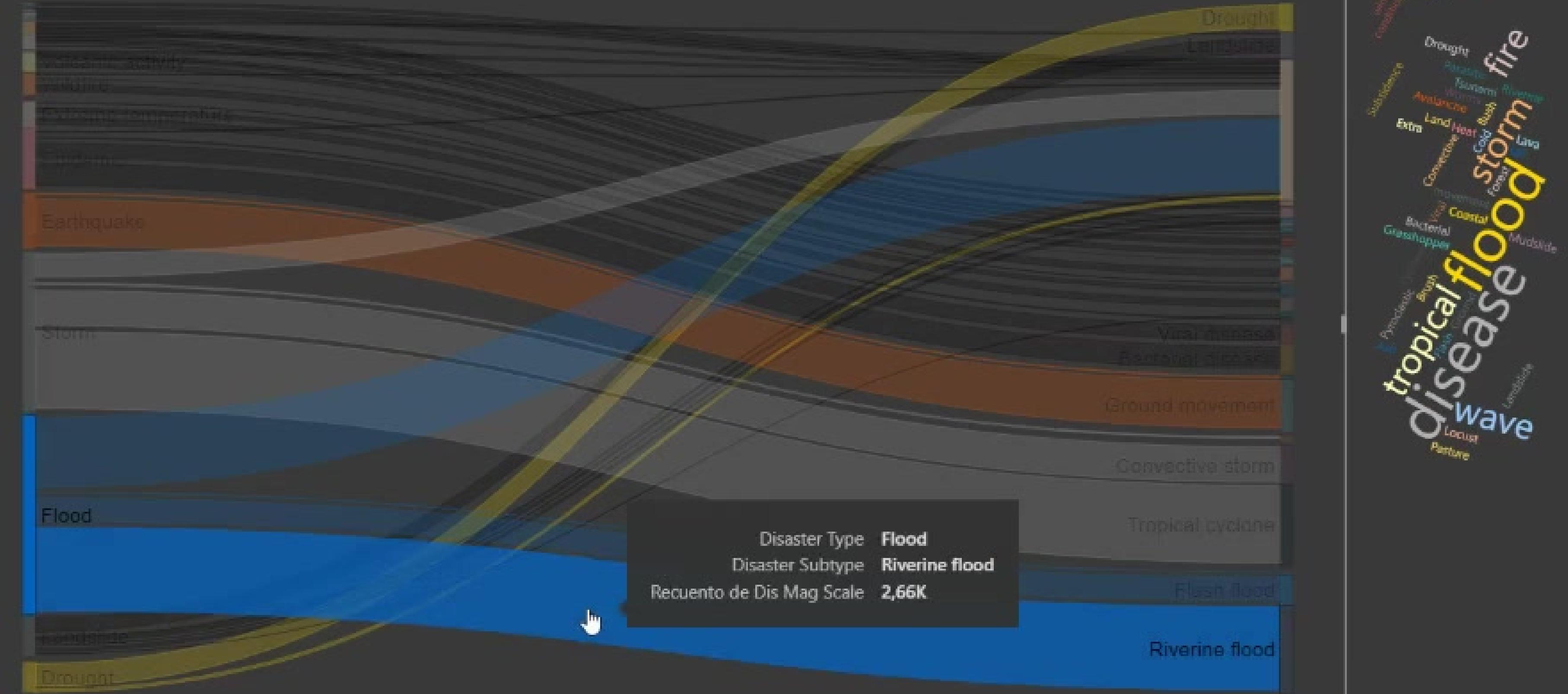


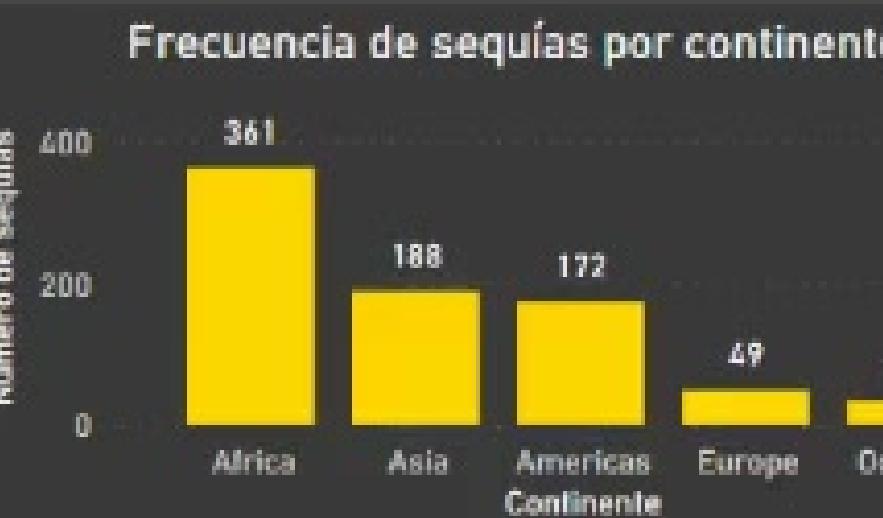
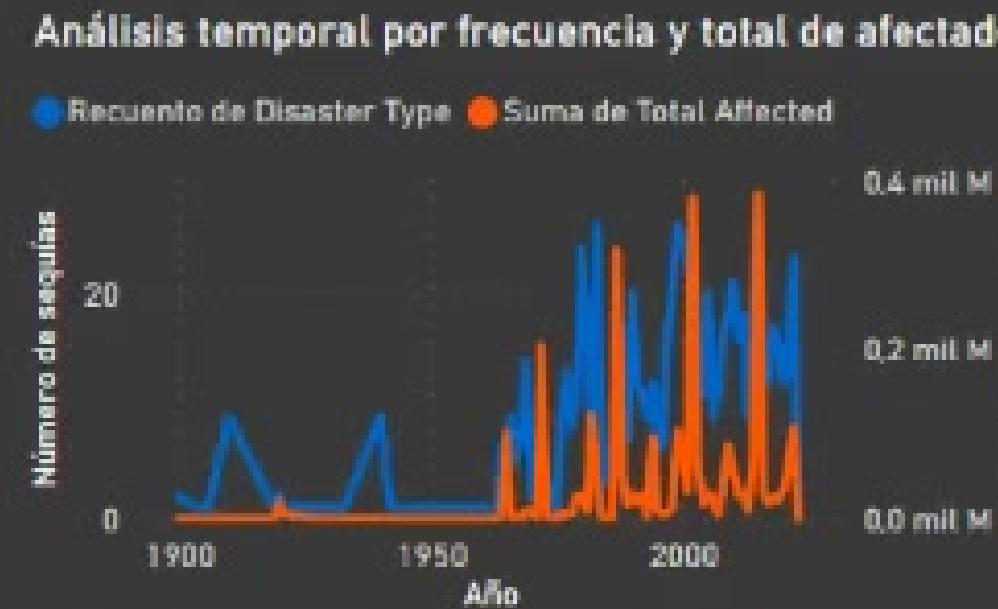


# Desastres Naturales

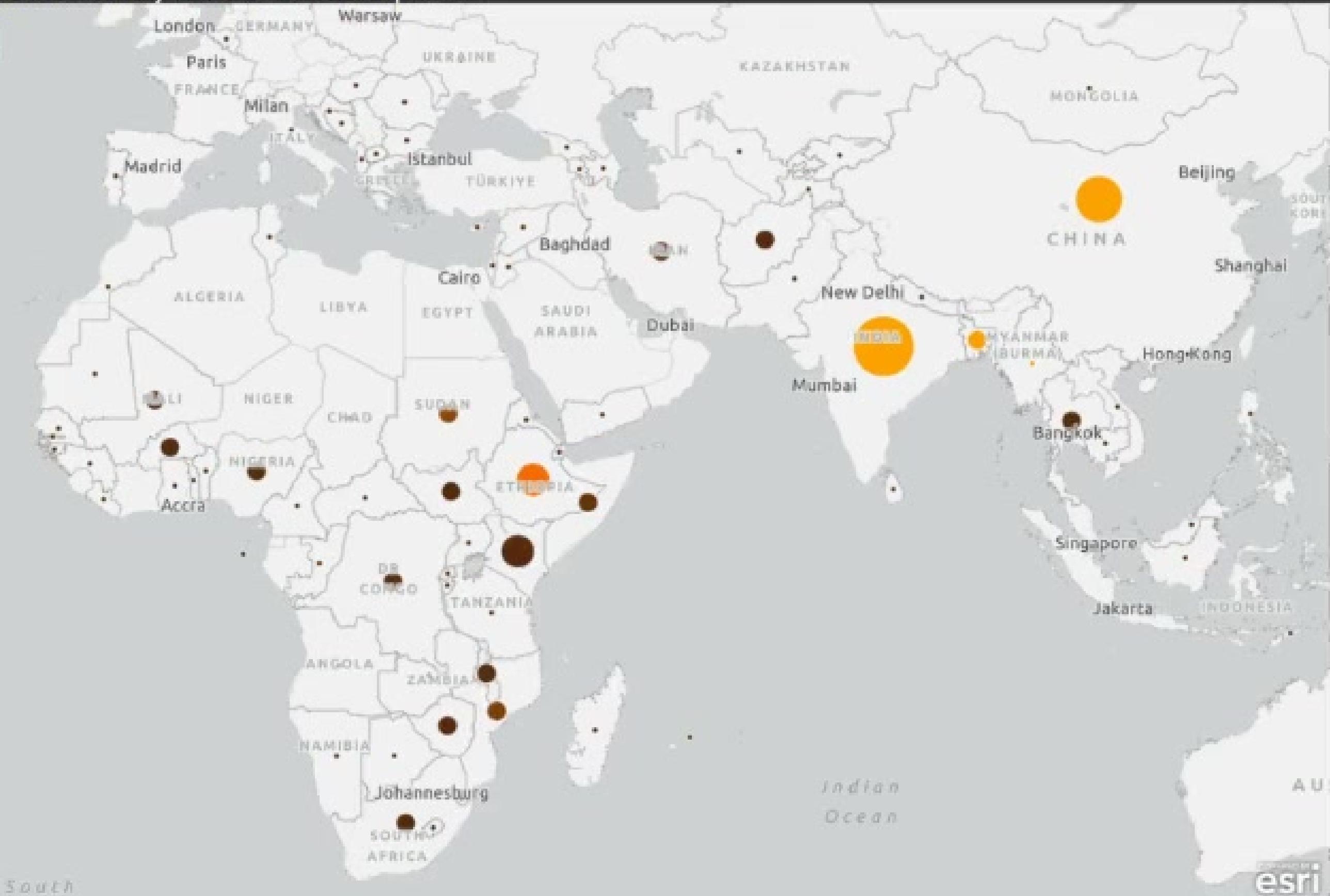
## Tipos y Subtipos

**Diagrama Sankey: representación del flujo de tipos a subtipos.**





Total de Afectados y Total de Fallecidos por País



## Frecuencia de sequías por países



Disaster Type • Drought

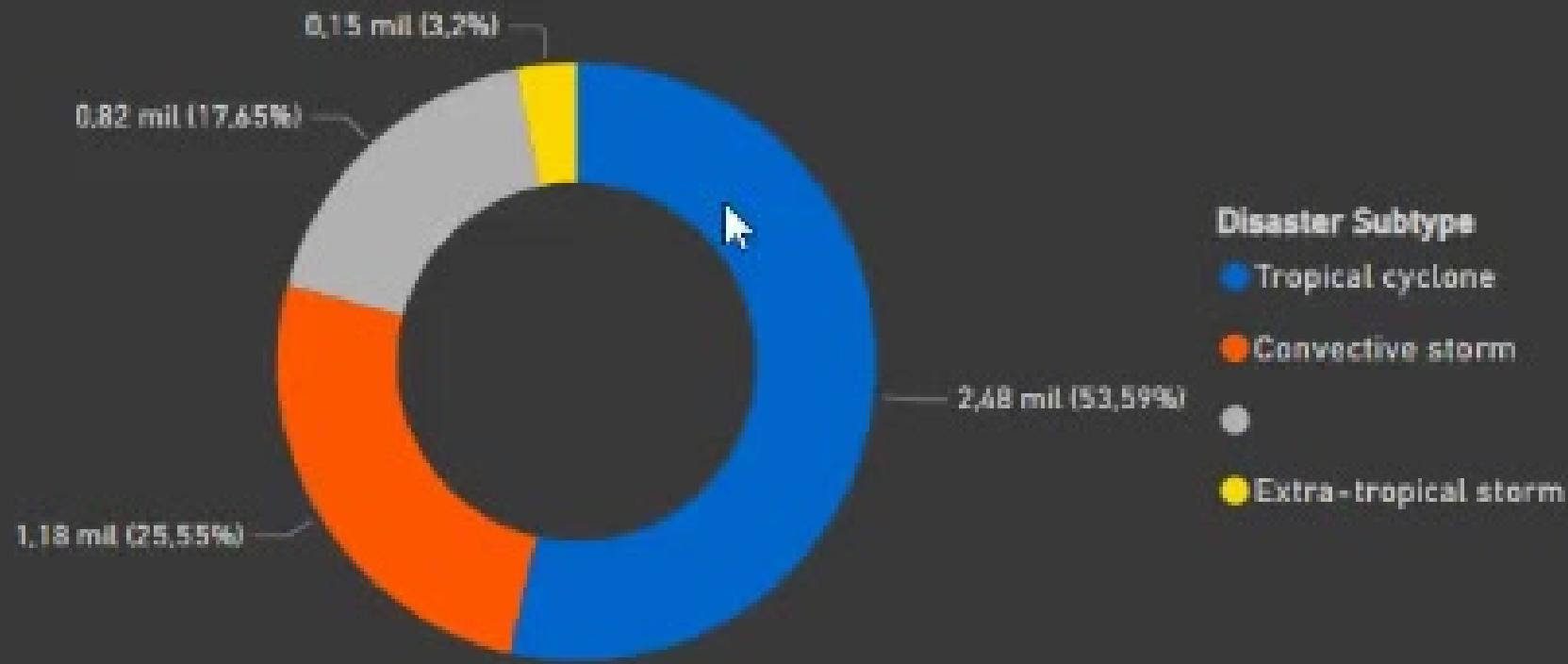




# Desastres Naturales

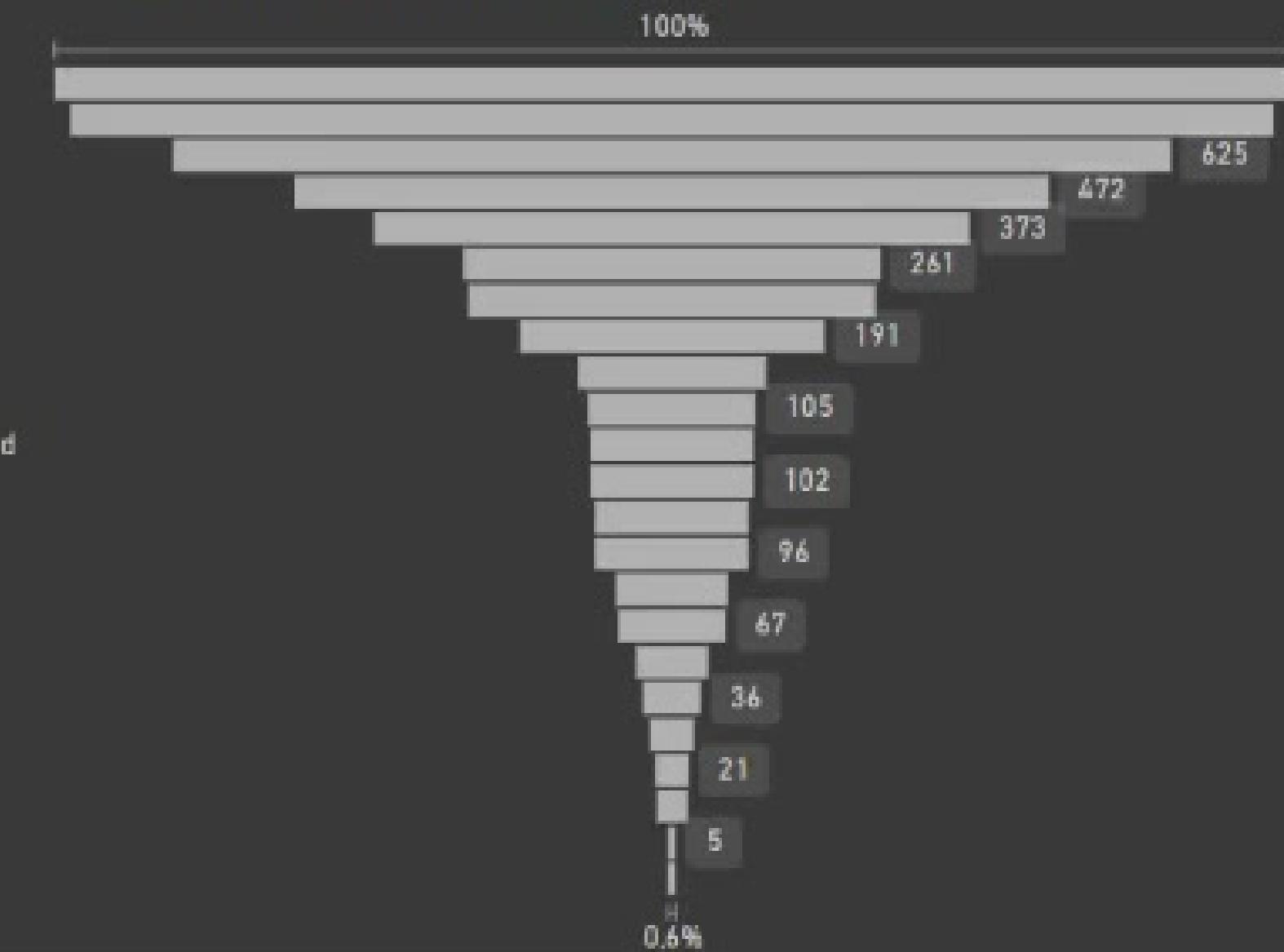
## Tormentas

### Distribución por subtipos de tormenta

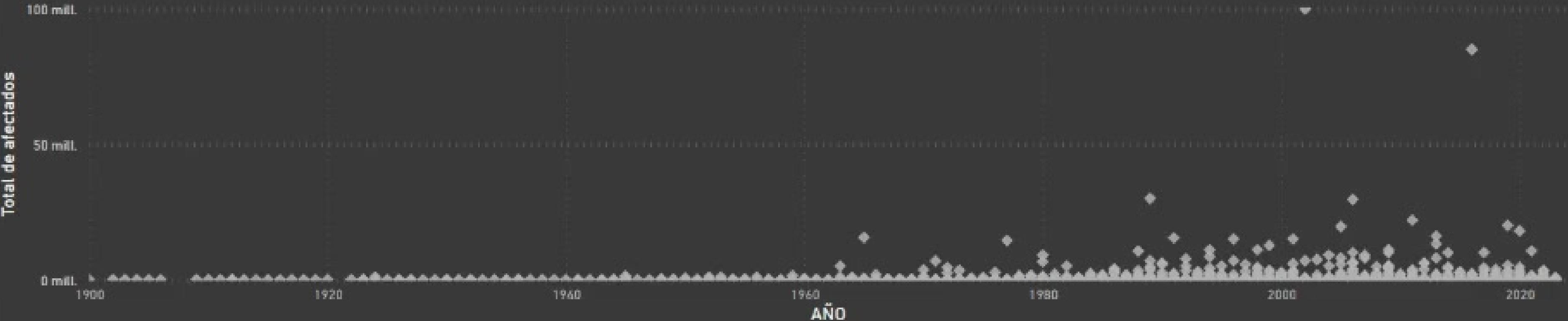


Northern America  
Eastern Asia  
South-Eastern Asia  
Southern Asia  
Caribbean  
Western Europe  
Central America  
Eastern Africa  
Melanesia  
Northern Europe  
Australia and New Zealand  
Eastern Europe  
South America  
Southern Europe  
Polynesia  
Western Asia  
Southern Africa  
Western Africa  
Micronesia  
Middle Africa  
Northern Africa  
Central Asia  
Russian Federation

### Frecuencia de tormentas por región



### Evolución temporal del número de afectados





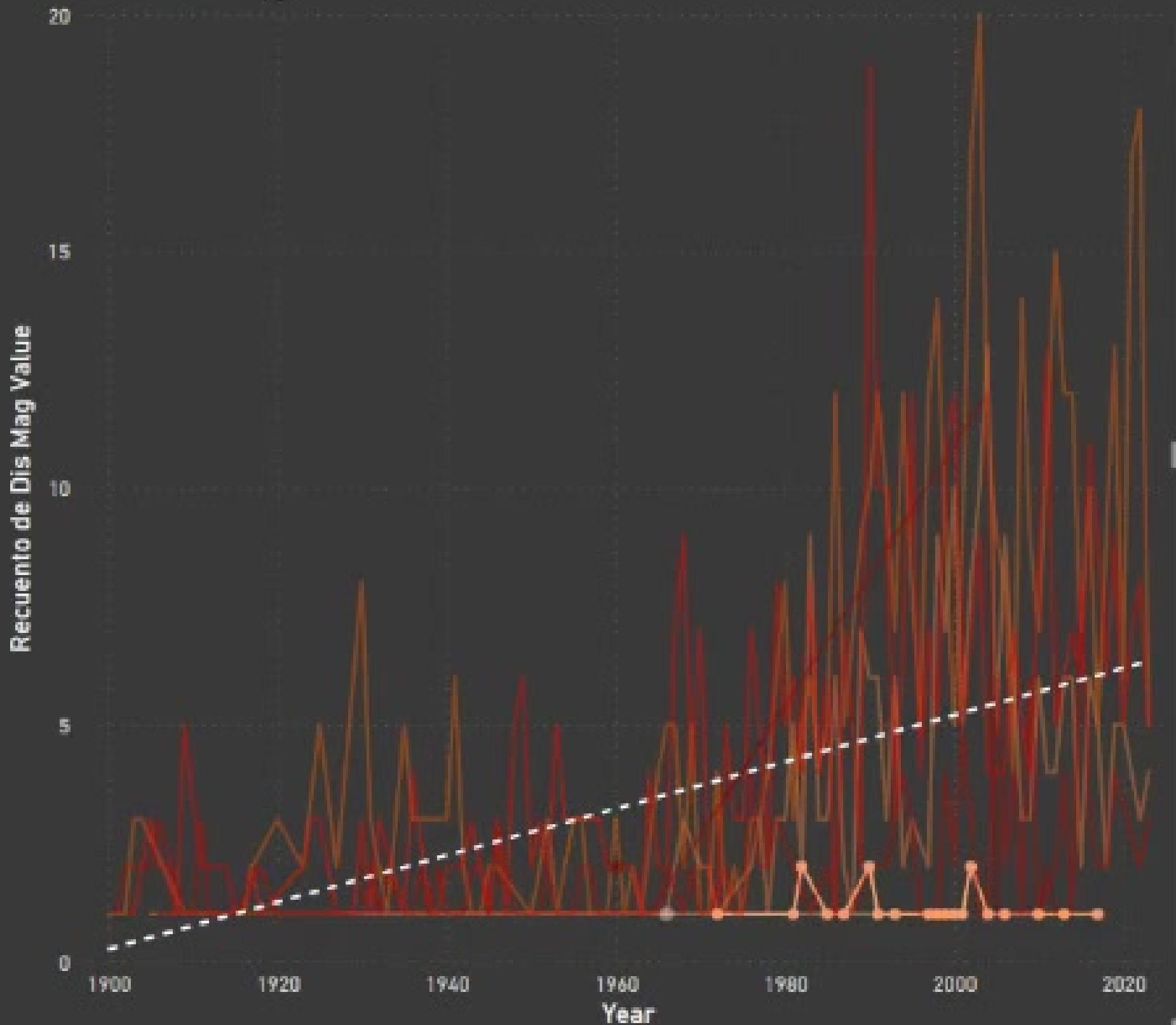
# Desastres Naturales

## Terremotos

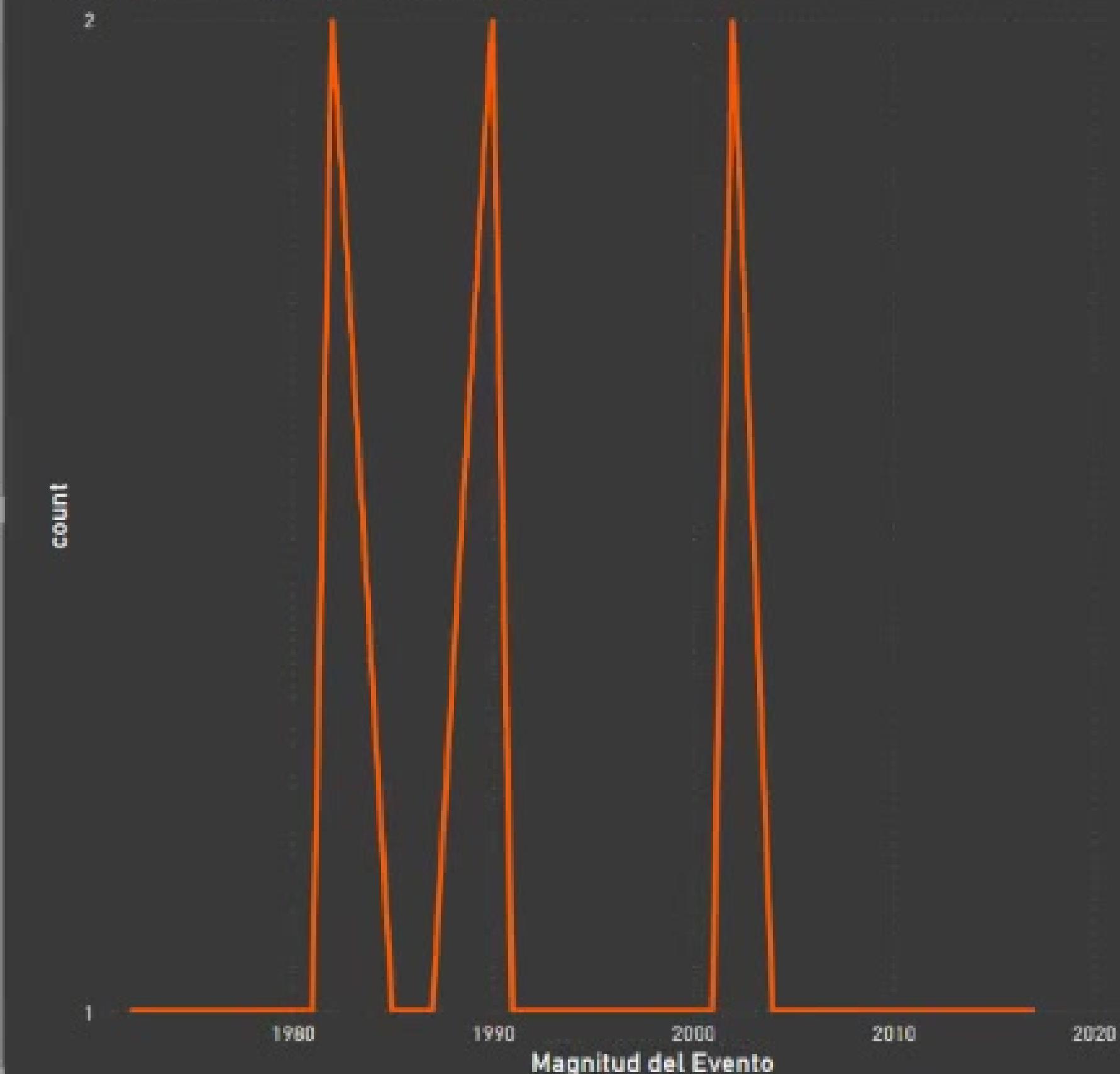


Magnitud por Año

Dis Mag Value ● 1 ● 3 ● 4 ● 5 ● 6 ● 7 ● 8 ● 9 ● 10



Terremotos a través de los años





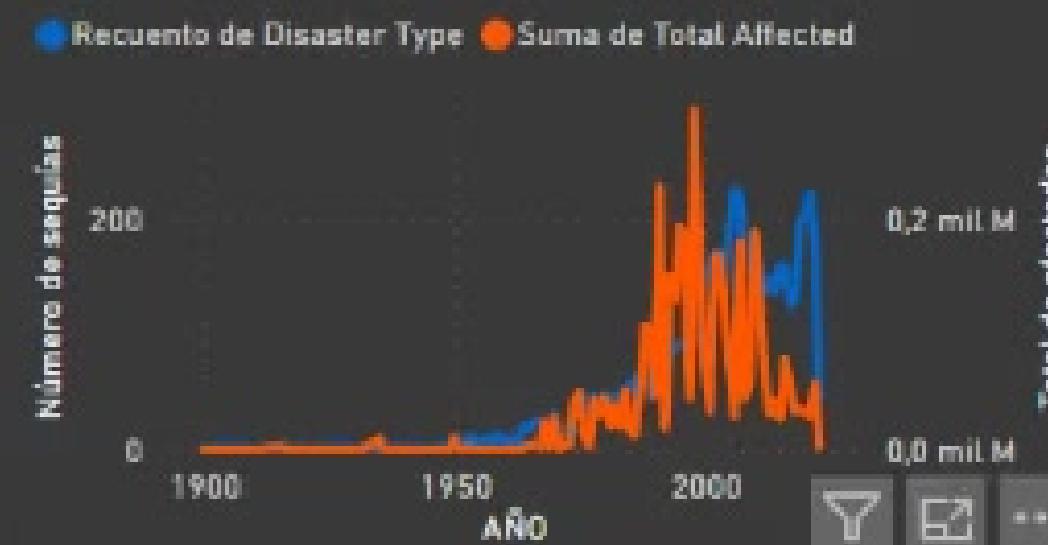
# Desastres Naturales

## Mapa de Geolocalización de Terremotos

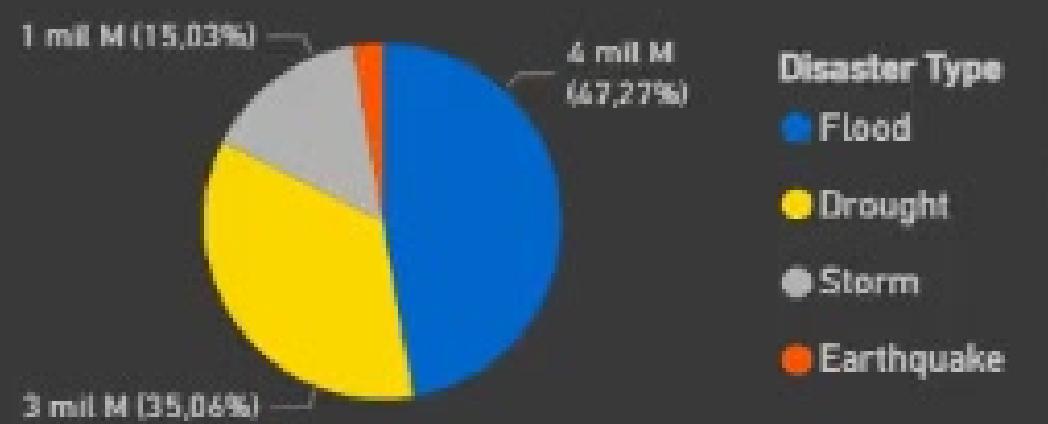




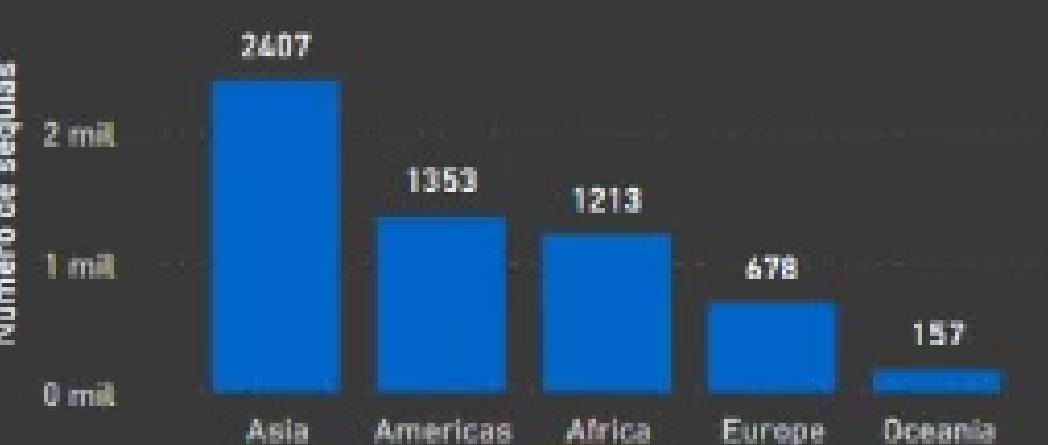
### Análisis temporal por frecuencia y total de afectados



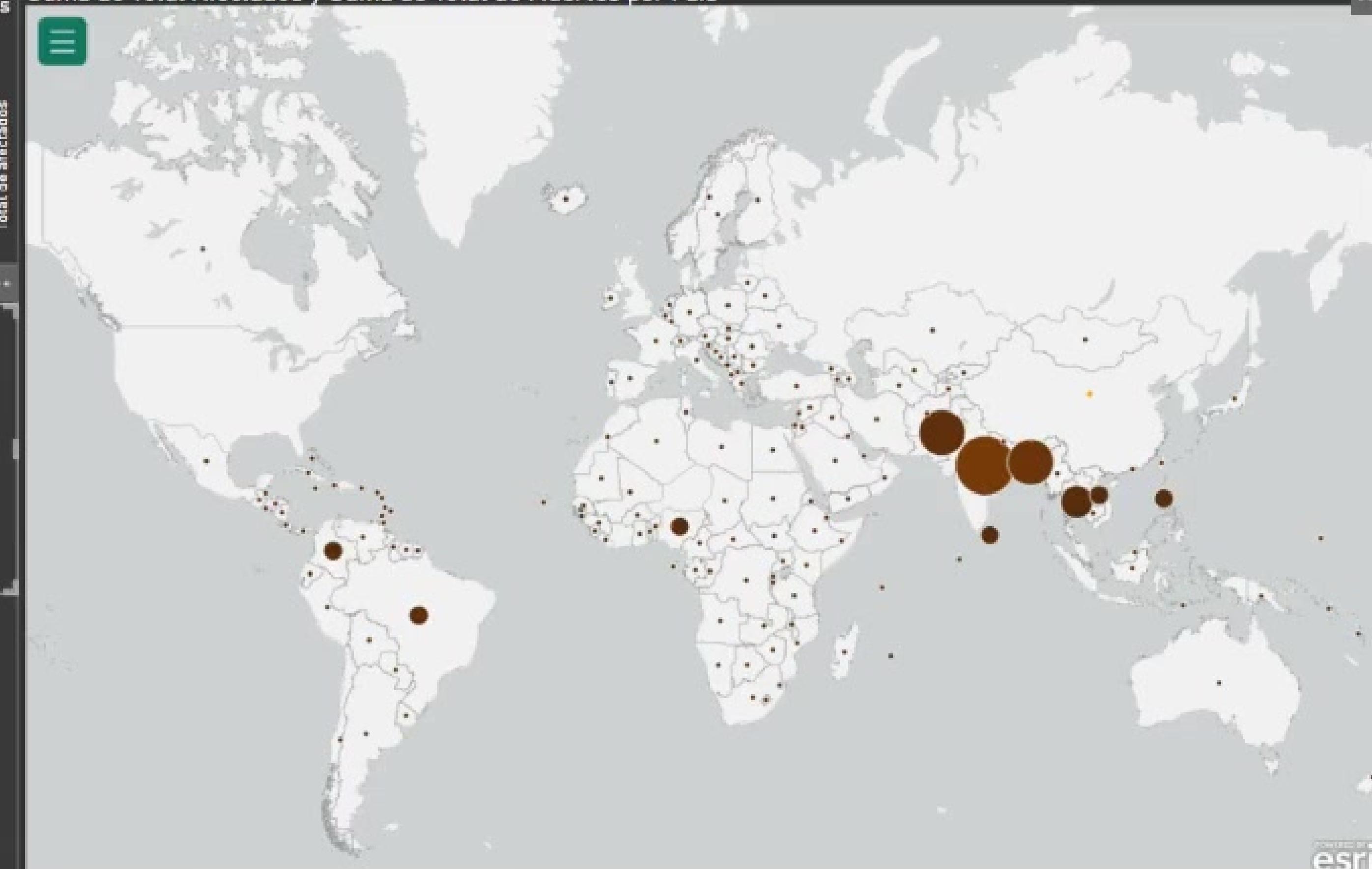
### Fallecidos por tipo de desastre



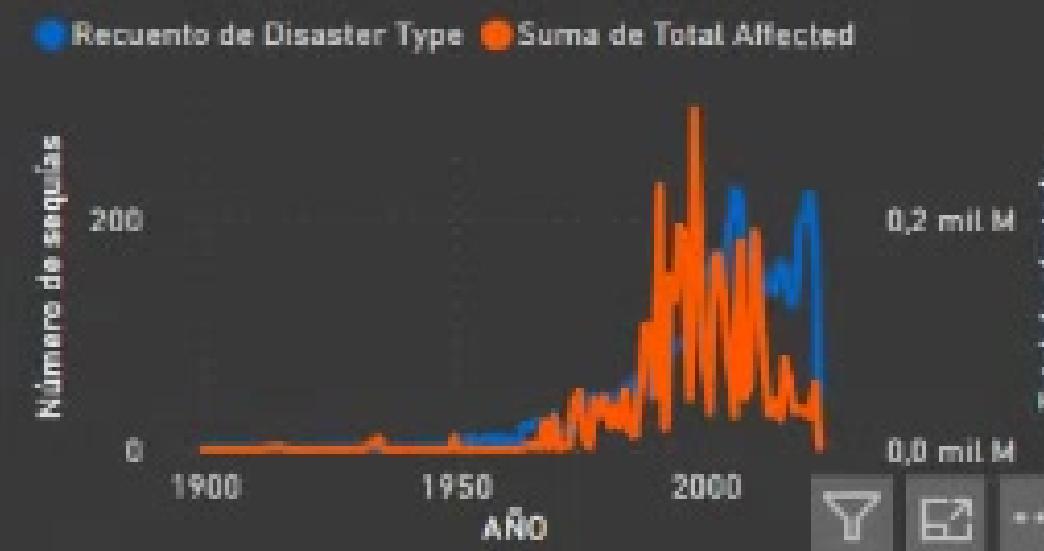
### Frecuencia de inundaciones por continente



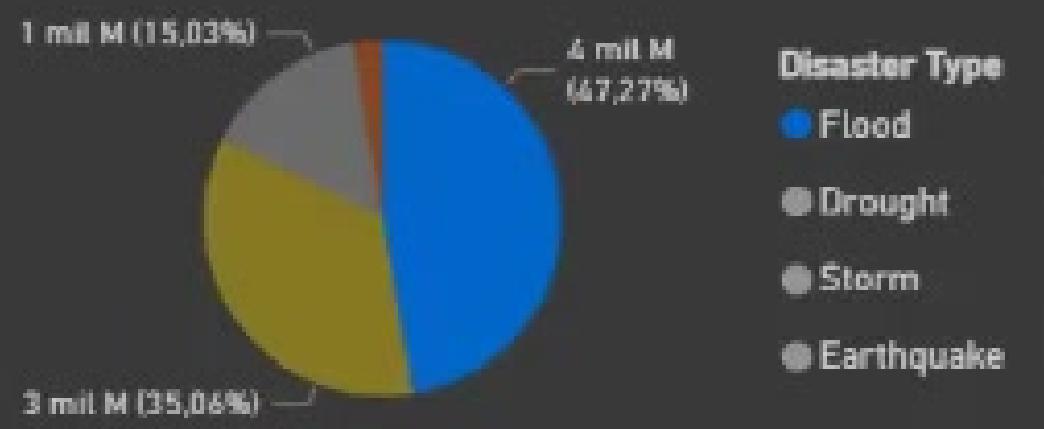
### Suma de Total Afectados y Suma de Total de Muertes por País



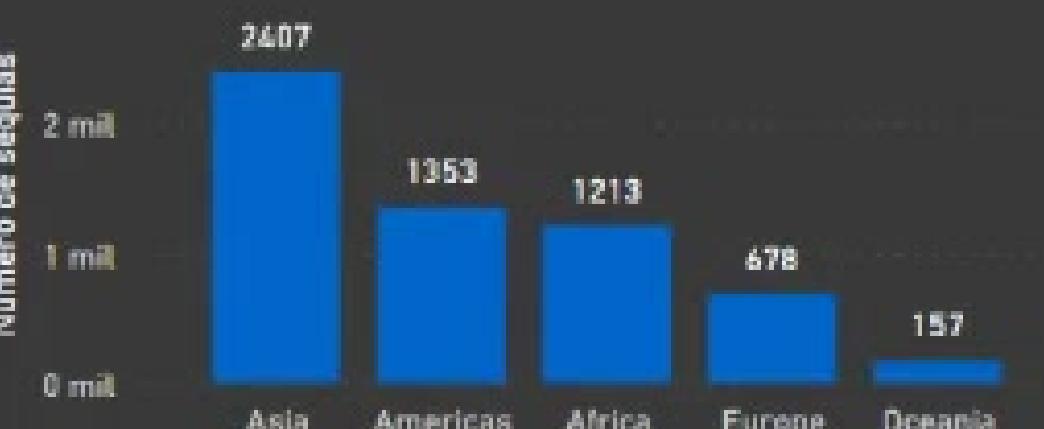
### Análisis temporal por frecuencia y total de afectados



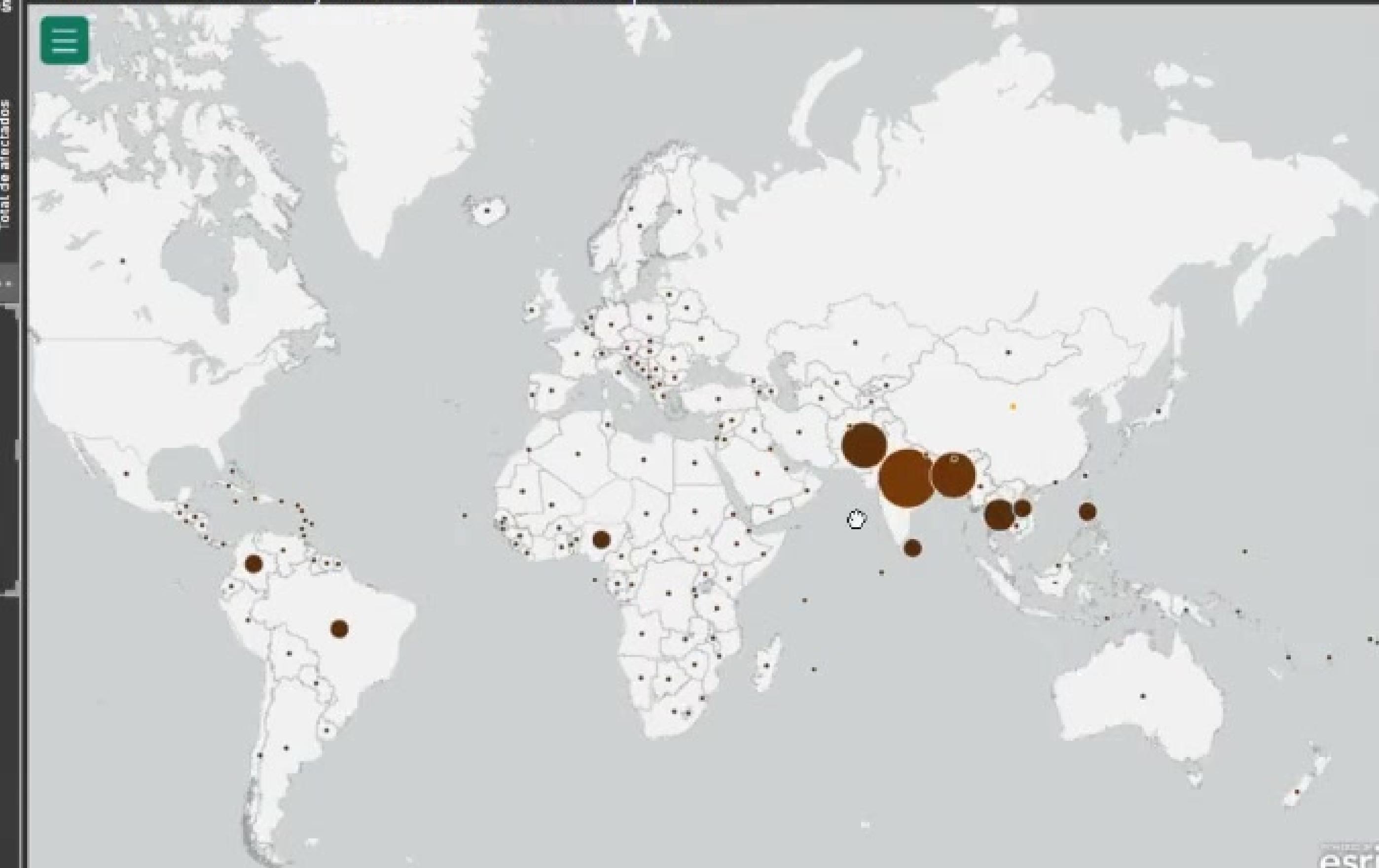
### Fallecidos por tipo de desastre



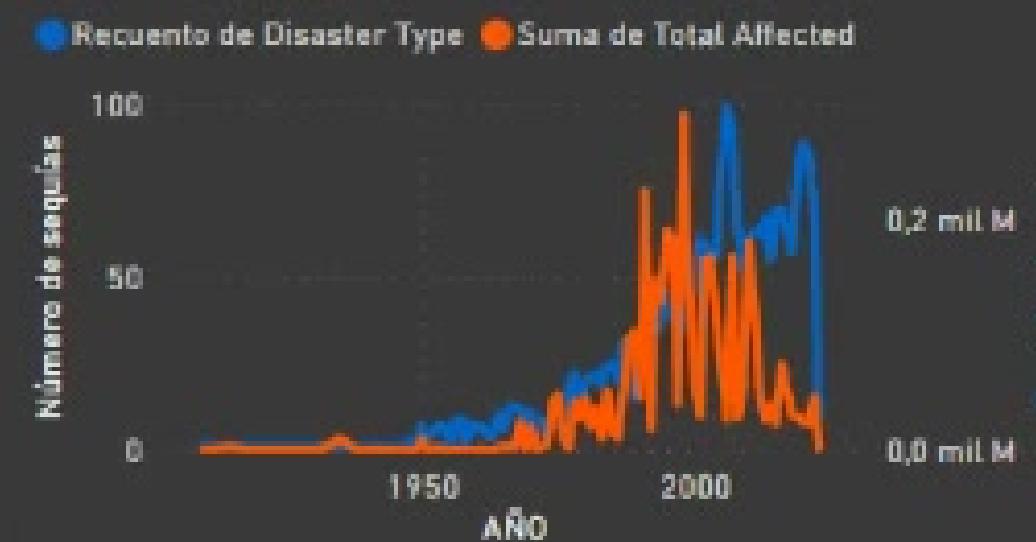
### Frecuencia de inundaciones por continente



### Suma de Total Afectados y Suma de Total de Muertes por País



### Análisis temporal por frecuencia y total de afectados



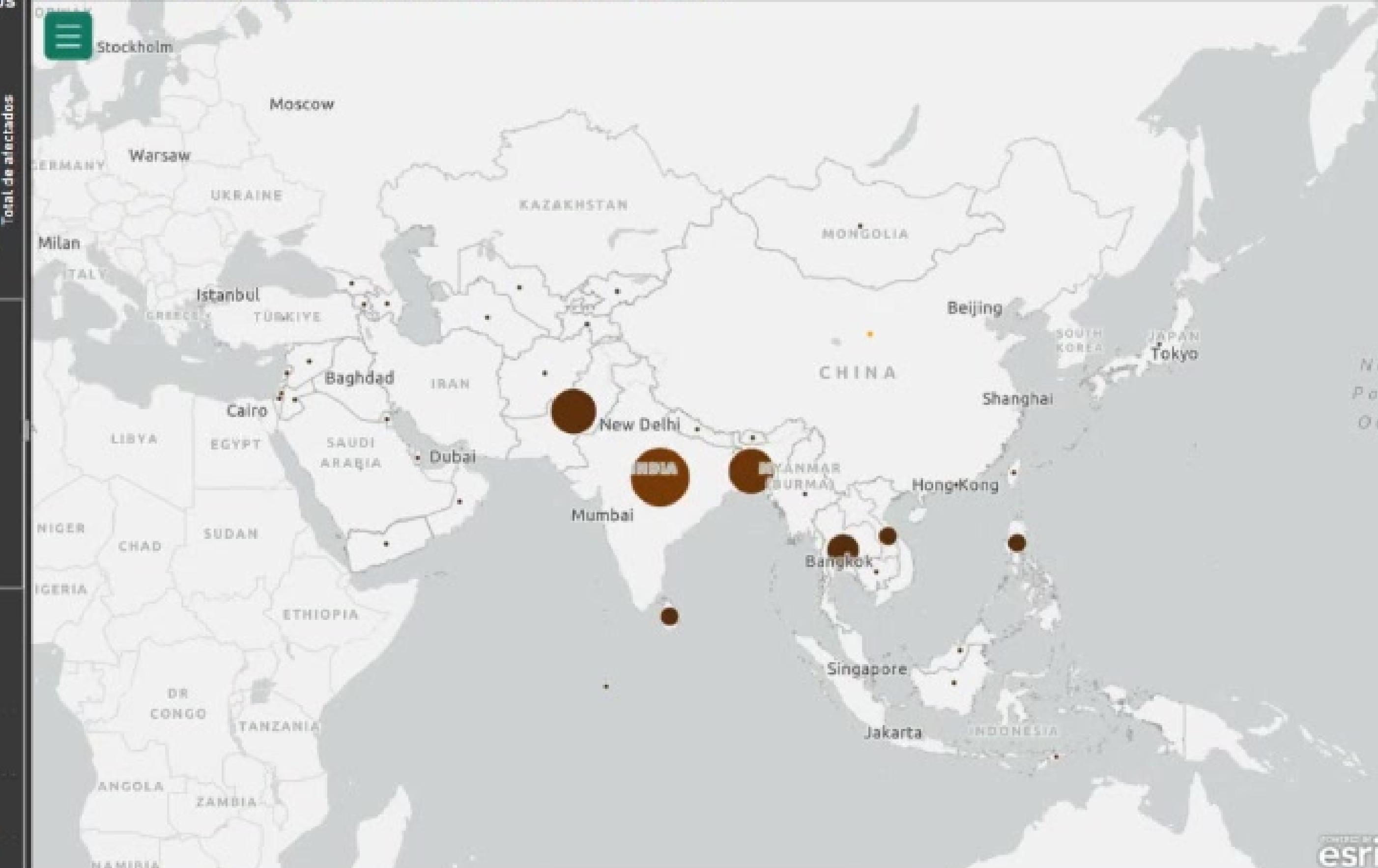
### Fallecidos por tipo de desastre



### Frecuencia de inundaciones por continente



### Suma de Total Afectados y Suma de Total de Muertes por País



## DESASTRES NATURALES

## CONCLUSIONES

## GENERALES

- Los DDNN se están incrementando con el tiempo.
- Reducción paulatina en pérdidas de vidas humanas, probablemente debido a medidas de protección poblacional.
- A partir de los 60 se evidencia tendencia a la alza en daños, que coincide con el aumento en la frecuencia de desastres.

## SÍNTESIS DE IMPACTO:

MUERTES  
DESCENSO

AFFECTADOS  
AUMENTO

DAÑOS  
AUMENTO



## DESASTRES NATURALES

### ESPECÍFICAS POR TIPO DE DESASTRE

#### Sequías:

- Las más letales.
- Incremento en sucesos y reducción del período entre los mismos.

#### Terremotos:

- Frecuencia irregular y a la alza hasta 2004.
- A partir de 70s, incremento de daños cuando magnitud supera 8.

#### Tormentas:

- Top 10 años con máximo record: 9 años son del SXXI.
- Picos de afectados cada 5 ó 6 años.

#### Inundaciones:

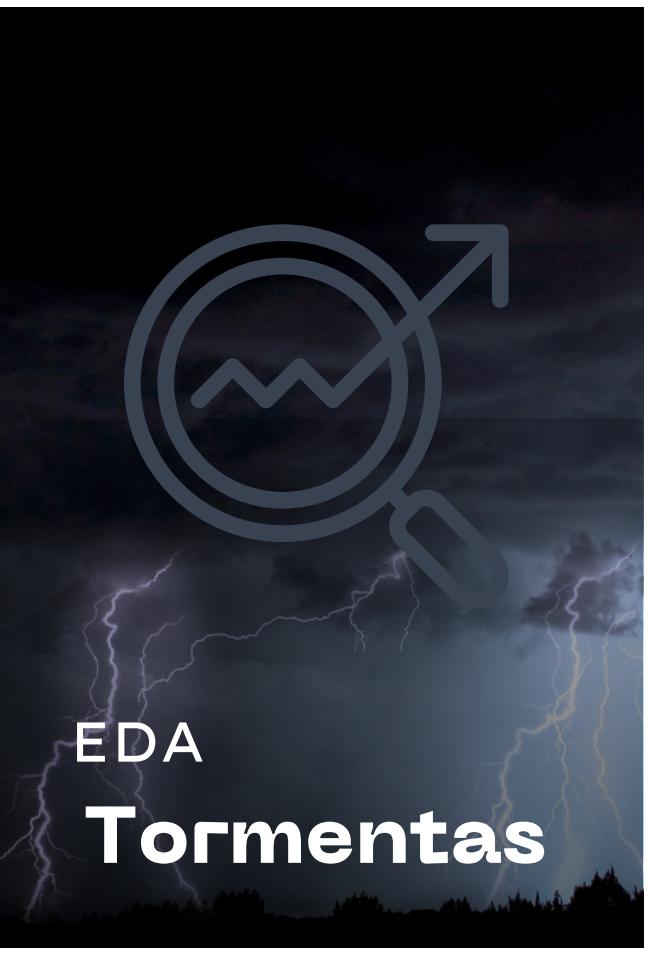
- Las más habituales.
- Más afectados.



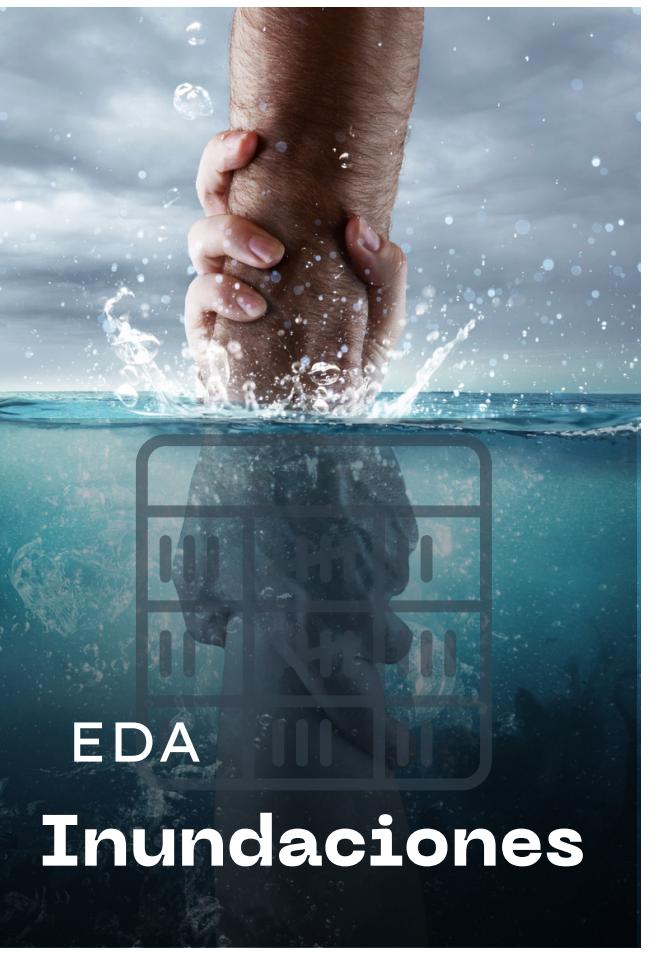
EDA  
**Sequías**



EDA  
**Terremotos**



EDA  
**Tormentas**



EDA  
**Inundaciones**

# MACHINE LEARNING

Modelo predictivo de Nivel de Destrucción



Herramientas



# MODELO DE APRENDIZAJE SUPERVISADO CLASIFICACIÓN DE NIVELES DE DESTRUCCIÓN

- Basado en Valor de **Magnitud** de inundaciones (km<sup>2</sup>)
- Definición de clasificaciones del '**'Low'** 1, '**'Moderate'** 2, '**'High'** 3 y **'Very High'** por umbrales basados en cuartiles.
- Función para asignar a cada clase de '**Nivel de Destrucción**' una combinación de **umbral** y **máximos de cada tipo de impacto**: fallecidos, afectados y daños.

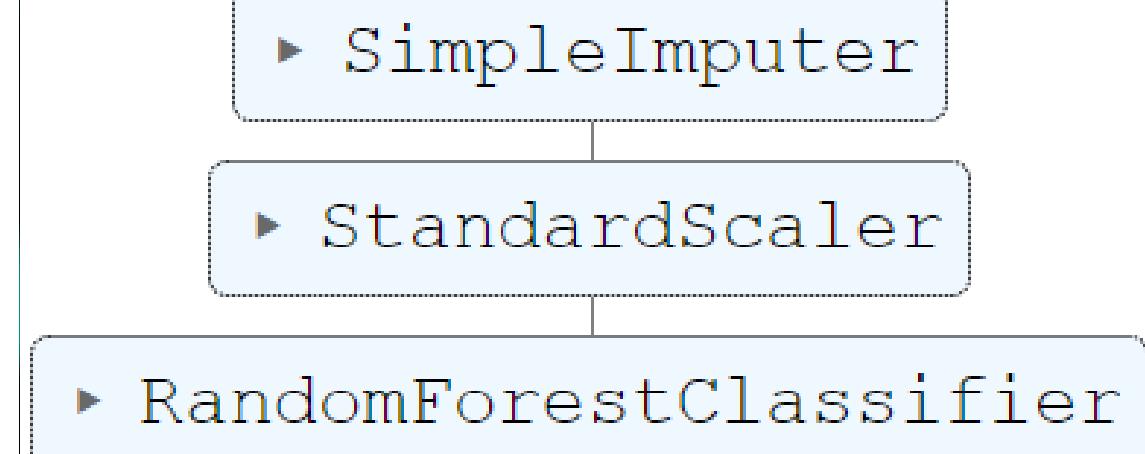
## Balanceado de clases

```
original: (1421, 10) (1421,)  
resampled: (1876, 10) (1876,)
```

4.0	469
3.0	469
1.0	469
2.0	469

Name: Destruction Level, dtype: int64

## Pipeline



## DESASTRES NATURALES

### MODELO PREDICTIVO

```
# Filtro principal para Flood
df_flood = df[df['Disaster Type'] == 'Flood']
filter = df_flood['Dis Mag Value'] != 0
df_flood_filtered = df_flood[filter]
df_flood_filtered.head()
```

Filtros

```
type1_threshold = np.percentile(df_flood_filtered['Dis Mag Value'], 25)
type2_threshold = np.percentile(df_flood_filtered['Dis Mag Value'], 50)
type3_threshold = np.percentile(df_flood_filtered['Dis Mag Value'], 75)
print(f'Umbral 1: {type1_threshold}')
print(f'Umbral 2: {type2_threshold}')
print(f'Umbral 3: {type3_threshold}')
```

Umbrales por cuartiles

```
# Función previa que no considera valores nulos
def assign_flood_type(row):
    if row['Dis Mag Value'] < type1_threshold and row['Total Deaths'] < 100 and row['Total Affected'] < 10000 and row['Total Damages Adj'] < 1000000:
        return 'Low' # 1
    elif row['Dis Mag Value'] < type2_threshold and row['Total Deaths'] < 500 and row['Total Affected'] < 100000 and row['Total Damages Adj'] < 10000000:
        return 'Moderate' # 2
    elif row['Dis Mag Value'] < type3_threshold and row['Total Deaths'] < 1000 and row['Total Affected'] < 50000 and row['Total Damages Adj'] < 100000000:
        return 'High' # 3
    else:
        return 'Very High' # 4
```

Función asignar clases Low, Moderate, High y Very High

```
floodtype_to_encode = df_flood_filtered['Flood Type'].values.reshape(-1, 1)
ordered_categories = ['Low', 'Moderate', 'High', 'Very High']
encoder = OrdinalEncoder(categories=[ordered_categories])
df_flood_filtered['Destruction Level'] = encoder.fit_transform(floodtype_to_encode) + 1
df_flood_encoded = df_flood_filtered.copy()
df_flood_encoded.head(10)
```

OrdinalEncoder clases numéricas 1–4 Destruction Level

# PREDICCIÓN DE NIVEL DE DESTRUCCIÓN

Modelo ML con Random Forest Classifier

```
# Evaluar desempeño del dataset
print("Reporte de clasificación:")
print(classification_report(y_pred, y_pred))
```

Reporte de clasificación:

	precision	recall	f1-score	support
1.0	1.00	1.00	1.00	33
2.0	1.00	1.00	1.00	114
3.0	1.00	1.00	1.00	98
4.0	1.00	1.00	1.00	111
accuracy			1.00	356
macro avg	1.00	1.00	1.00	356
weighted avg	1.00	1.00	1.00	356

## RESULTADOS DE CLASSIFICATION REPORT

Métricas colectivas de nivel de precisión y capacidad predictiva:

**MUY ALTA**

Iteración para mejora continua: tests y validación

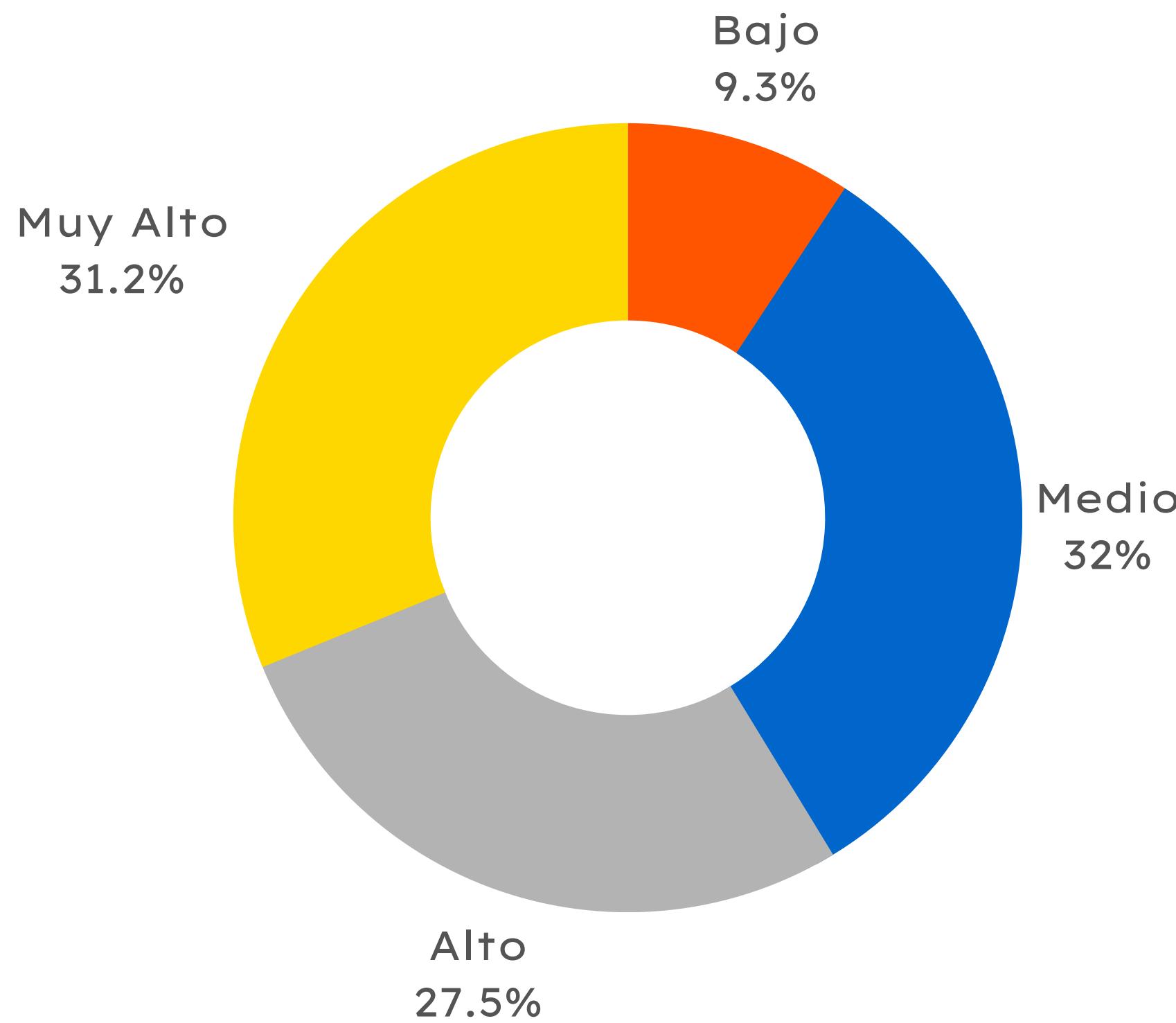
### Diagnóstico:

- 1) Posible fuga de X a y por umbrales.
- 2) Escasez de datos para niveles Bajo y Alto.
- 3) Entrenar con menos columnas.
- 4) Necesario ajustar umbrales con percentiles también en columnas de impacto.
- 5) Explorar otra configuración para umbrales no basada en estadística.

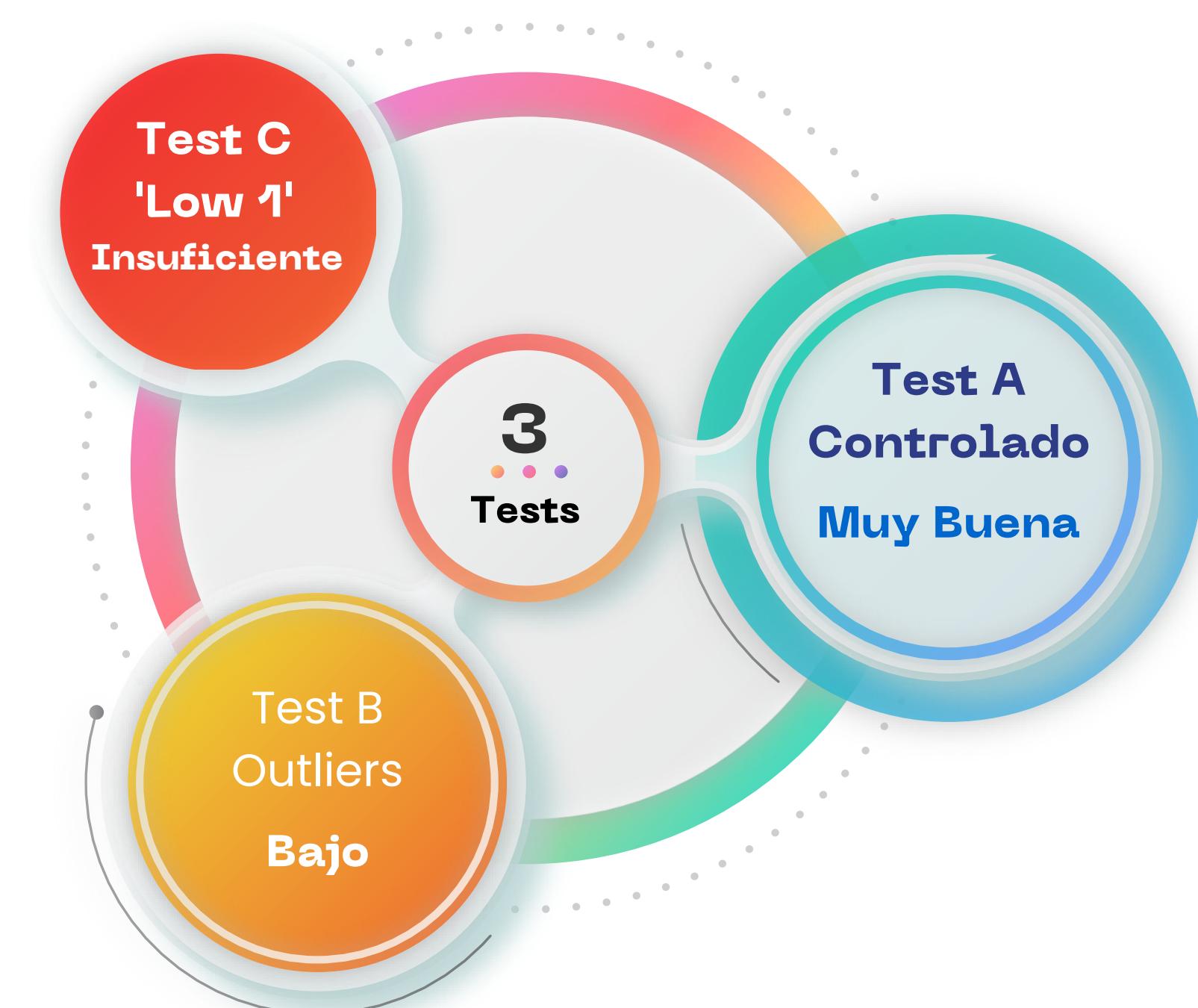
## DESASTRES NATURALES

### MODELO PREDICTIVO: PREDICCIÓN Y TESTS

#### INUNDACIONES Y\_PRED

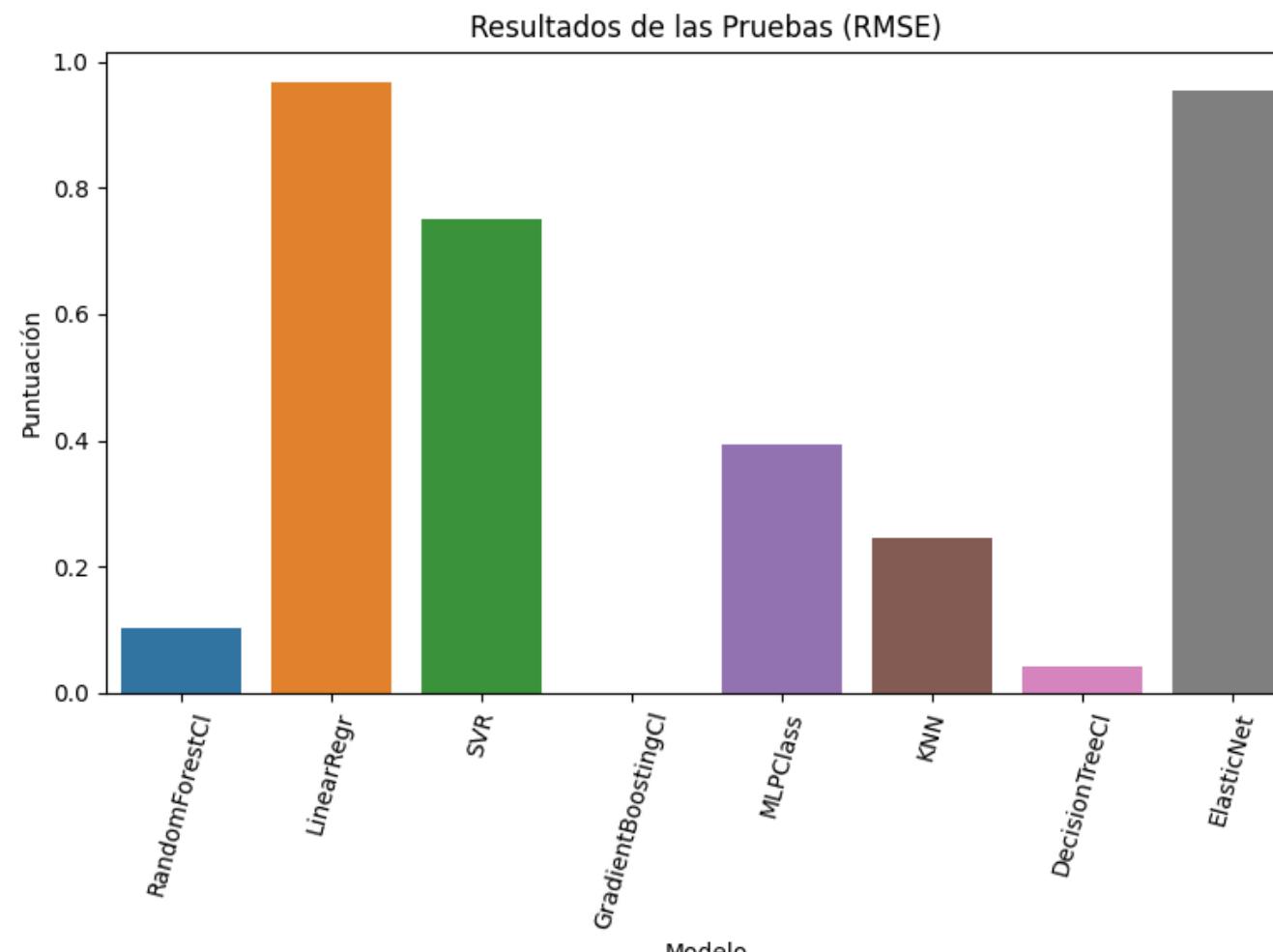
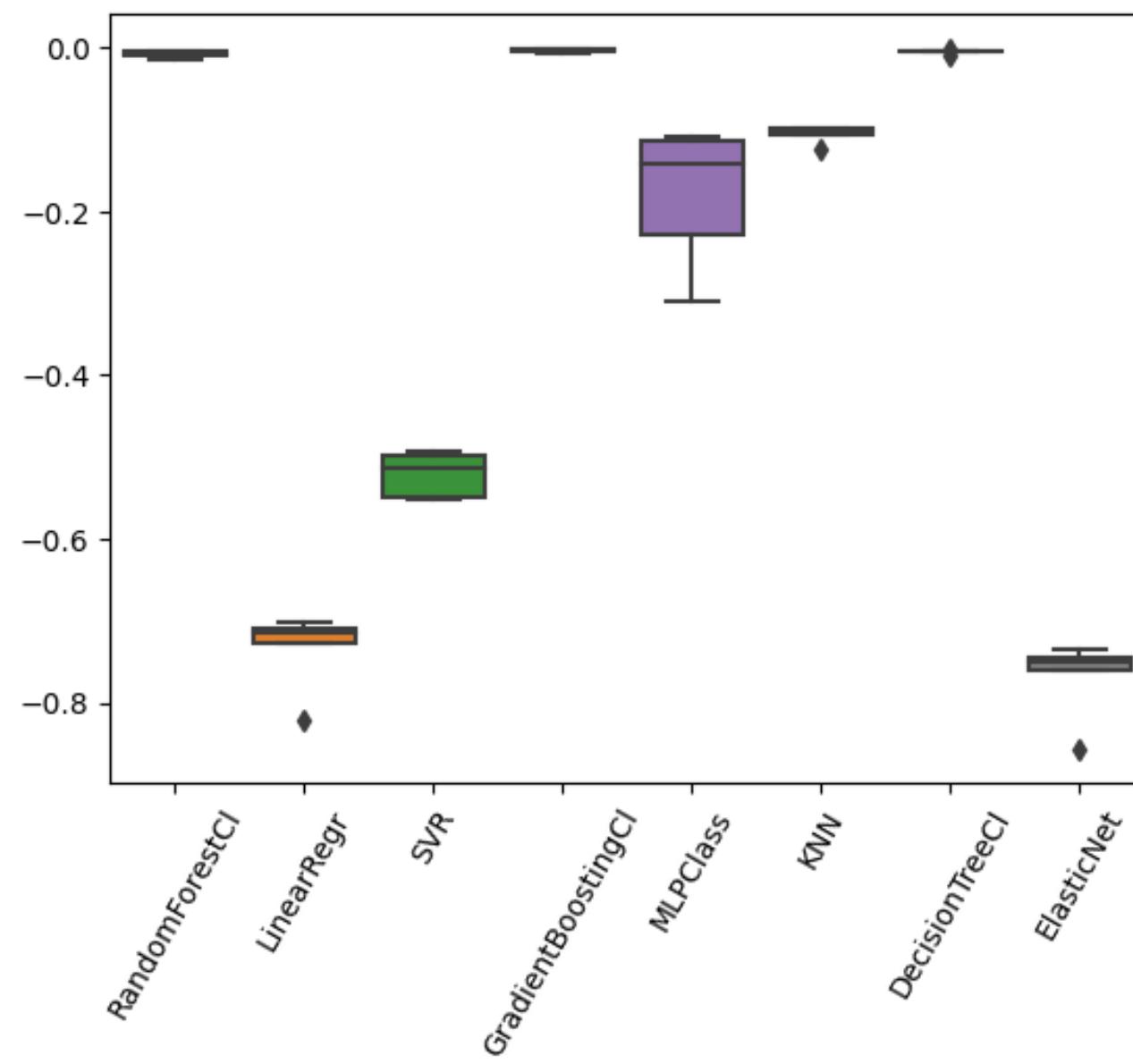


#### TESTS A, B Y C

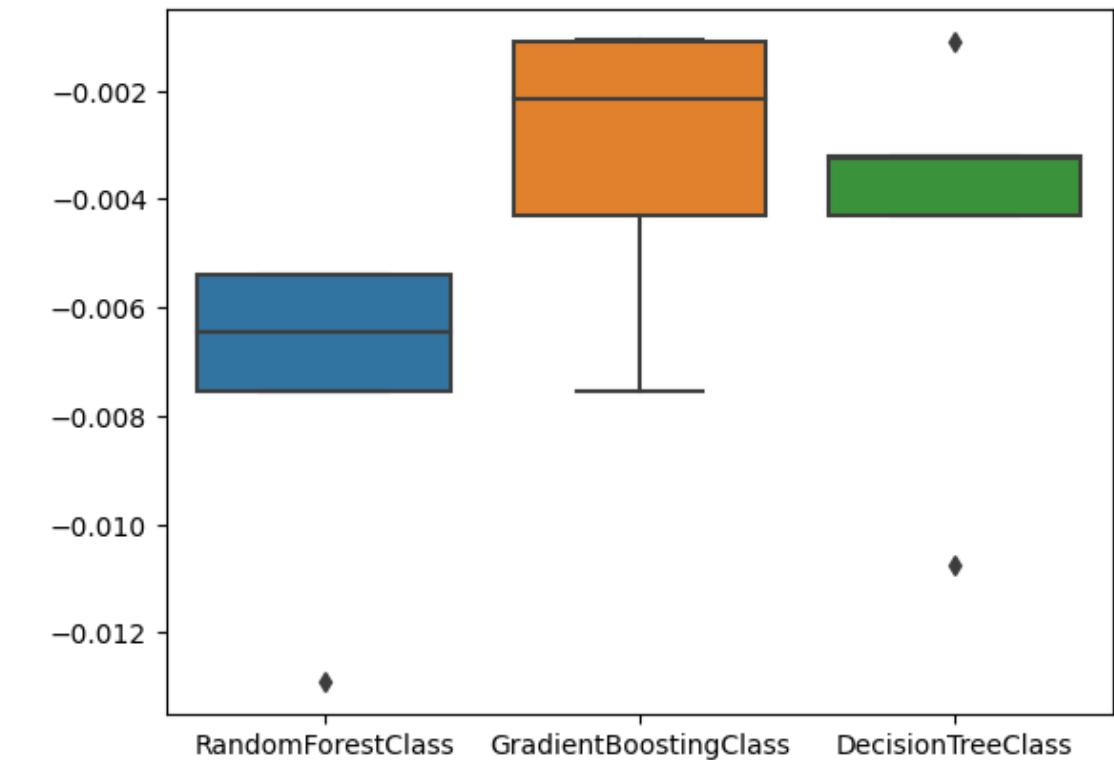


# VALIDACIÓN CON ENSAMBLE DE ALGORITMOS

*Análisis comparativo*



**Modelos finalistas:**  
 RandomForestClassifier  
 GradientBoostingClassifier  
 DecisionTrees



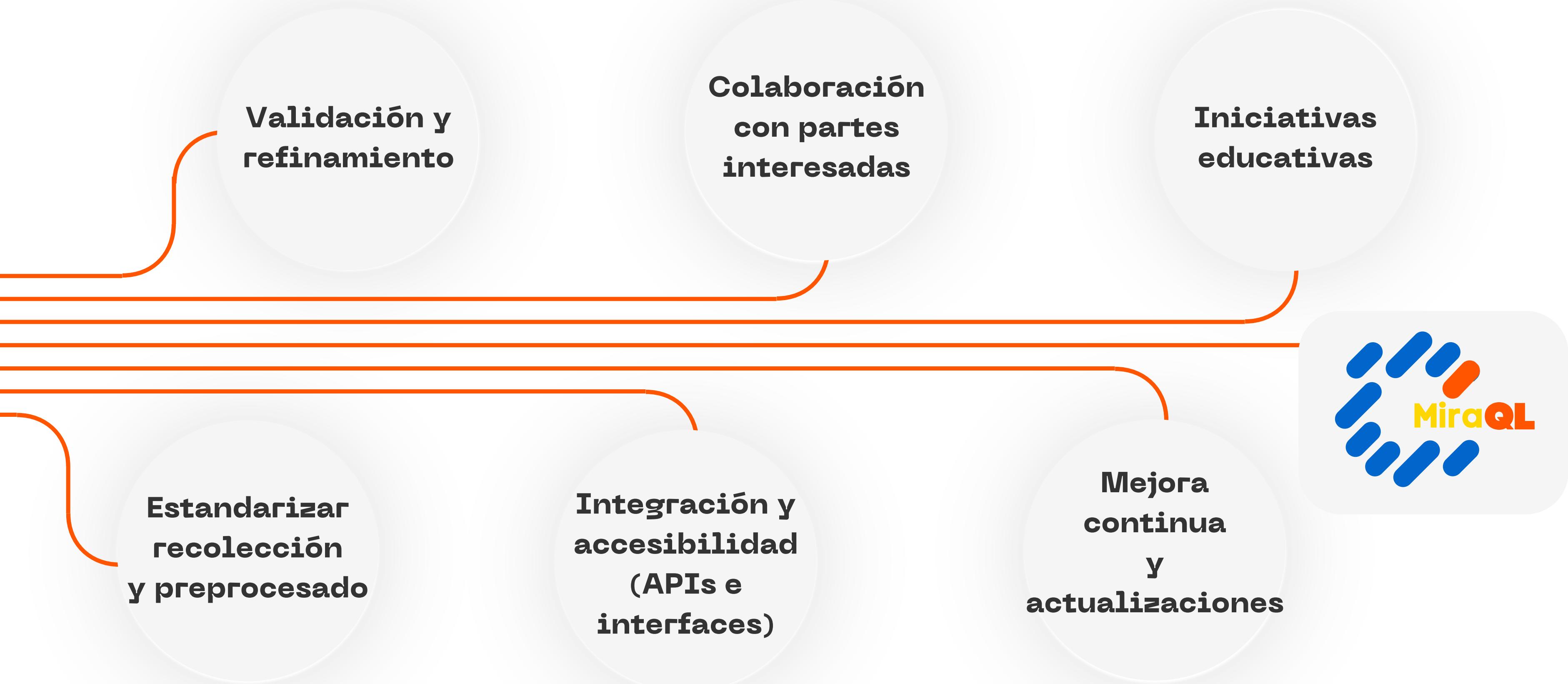
# **RECOMENDACIONES**

**Iteración para mejora de resultados de predicciones**

- Umbrales de percentiles a variables de impacto.
- Reducir complejidad: menos etiquetas para X.
- StratifiedKFold para asegurarnos de que los conjuntos de entrenamiento y prueba contengan una distribución equilibrada.
- Ajustes a hiperparámetros.
- Data Augmentation previo a columnas de impacto y reportes de clasificación de tests con datos nuevos de mayor volumen.

## DESASTRES NATURALES

### MODELO PREDICTIVO: IMPLEMENTACIÓN



**DESASTRES NATURALES**

**PROCESO DE MEJORA CONTINUA**

## **BENEFICIOS DE MEJORAS**

- Mayor Precisión
- Planificación Eficiente
- Respuesta Rápida
- Adaptabilidad a diferentes escenarios

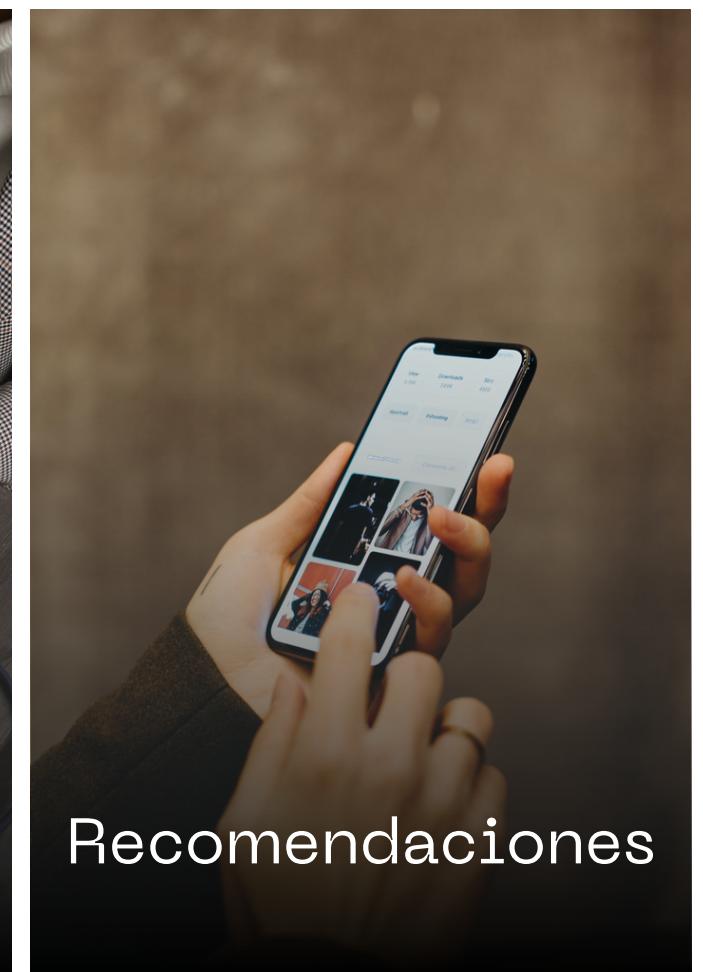
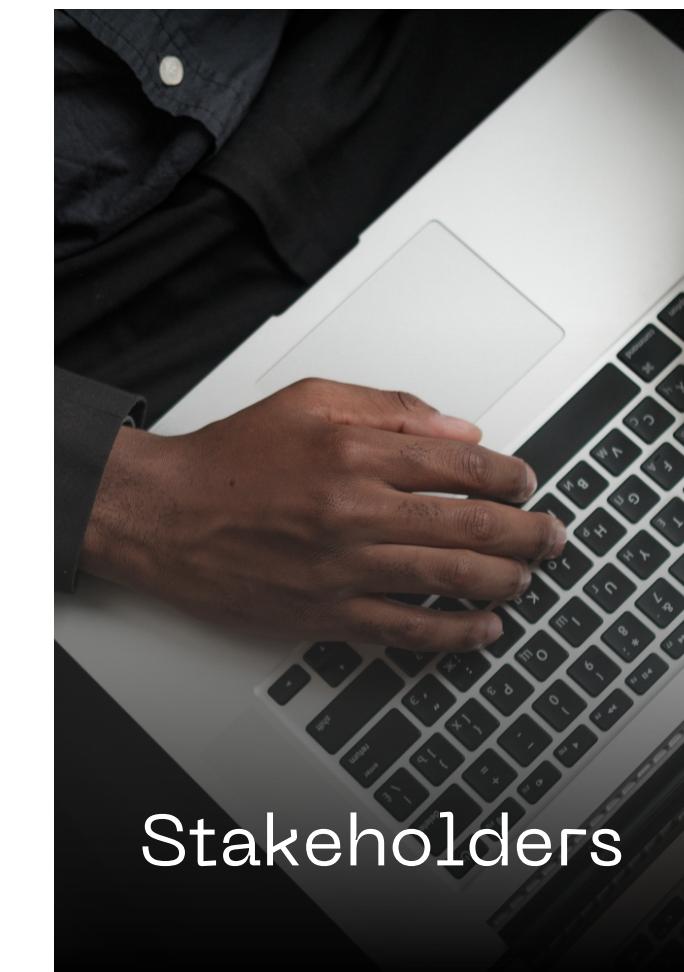


**DESASTRES NATURALES**  
**CONCLUSIONES**

**MODELO DE  
NEGOCIO,  
CONCLUSIONES Y  
RECOMENDACIONES**

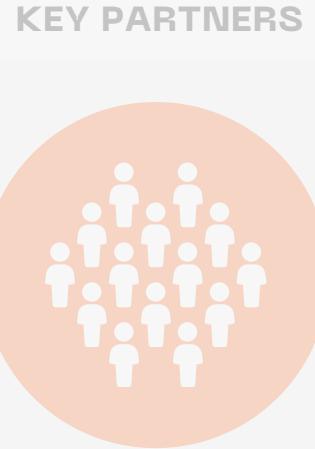


ANÁLISIS DE DATOS. EQUIPO 2



## DESASTRES NATURALES

## MODELO DE NEGOCIO EN CANVAS



Convenios de colaboración público-privada

- Administraciones públicas.
- Universidades.
- Agencias de recolección de datos sobre desastres.
- NASA y otras agencias aeroespaciales de observación metereológica satelital.
- Comunidades en riesgo.



## Predicción es igual a prevención, protección y resiliencia

- Hardware, licencias de software y actualización prof.
- Obtención de bases de datos de pago o por suscripción.
- Contratación de personal cualificado.
- Tecnología cloud para almacenamiento y gestión de grandes volúmenes de datos.
- Viáticos de desplazamiento para investigación en zonas en riesgo (en caso de que se requiera).
- Mejora continua de modelos y testing.

## KEY ACTIVITIES

- Consultoría BI.
- EDA.
- Diseño de Dashboards.
- CRISP-DM a BBDD sobre desastres naturales.
- Adaptación del modelo predictivo a diferentes tipos de desastres y regiones.
- Capacitación en uso de herramientas.



## KEY RESOURCES

- BBDD de instituciones reconocidas con una metodología de recolección confiable.
- Patentes de modelos predictivos.
- Inteligencia de negocio.

## PROPIEDAD DE VALOR

### Estrategia y planificación

### Innovación

### Gestión de ayuda y optimización del rescate

### BI y especificidad sobre riesgos

## CUSTOMER RELATIONSHIPS

- In-company por consultoría y capacitación.
- Remotas para proyectos específicos como cualquier etapa de la metodología CRISP-DM EDA, dashboards, BI y ML.



## CHANNELS

- Ferias y foros de tecnología de IA e inversión.
- In-company y remoto
- Blog, e-mail marketing y RRSS.
- Streaming para probar herramientas durante desastres naturales en tiempo real.



Los clientes están dispuestos a pagar por:

- Tiempo es vida, no solo dinero.
- Minimización de desastres.
- Planificación de la reconstrucción en zonas de menor o cero impacto.
- Gestión y mejora continua de los modelos predictivos y la transformación de los datasets.
- Soluciones globales y regionales.

## CUSTOMER SEGMENTS

- Administraciones públicas.
- Compañías de Seguros y Administradoras de Riesgos.
- Promoción Inmobiliaria.
- Cadena de Suministro y Logística.
- Desarrollo de Infraestructura y Planificación Urbana
- Servicios de Emergencia y Respuesta a Desastres.
- Comunidades y ONG.





# DIRIGIDO A:

**GOBIERNOS Y AGENCIAS  
DE GESTIÓN DE  
DESASTRES**



## Estrategia

Mejorar estrategias de prevención, mitigación y respuesta.

**EMPRESAS  
DE SEGUROS**



## Especificidad sobre riesgos

Integrar en procesos de suscripción y evaluación de riesgos para ofrecer pólizas más precisas y ajustadas a riesgos específicos de cada región.

**EMPRESAS DE  
CONSTRUCCIÓN E  
INFRAESTRUCTURA**



## Resiliencia

Identificar zonas de alto riesgo y diseñar edificaciones e infraestructuras resistentes.

**ORGANIZACIONES  
DE AYUDA  
HUMANITARIA**



## Gestión de ayuda

Anticipar y preparar recursos y dirigir la ayuda hacia las áreas más afectadas.

**COMUNIDADES  
VULNERABLES**



## Protección

Preparación ante el riesgo y toma de decisiones informadas sobre la protección y evacuación.

## DESASTRES NATURALES

### RECOMENDACIONES

#### GENERALES

- Consideraciones éticas sobre fines de uso y sesgos.
- Combinar tecnología, experiencias multidisciplinarias y colaboración público-privada para abordar desafíos y estrategias de gestión de desastres.
- Integrar análisis de datos para infraestructuras y planes de emergencia.



## DESASTRES NATURALES

### RECOMENDACIONES

#### BASE DE DATOS:

- Incorporar magnitudes de densidad demográfica, densidad de construcción y/o duración.
- Recopilar BBDD específicas, normalizadas y homogéneas para cada tipo de desastre y modelo.
- Definir medida de magnitud para las sequias y su correlación con las personas afectadas.
- Reforzar y apoyar entes como EMDAT (1988), etc.



# CONCLUSIÓN

- Concientizar sobre la dimensión del impacto de los desastres naturales
- La recurrencia de ciertos desastres naturales que afectan a determinadas poblaciones limitan seriamente las posibilidades de desarrollo de esas poblaciones.
- Impulsar el desarrollo de modelos predictivos de Desastres Naturales de grandes magnitudes.



# GRACIAS

CONTACTA CON EL EQUIPO

Si te interesa nuestro proyecto, contacta con Alejandra para solicitar acceso al repositorio GitHub:  
[https://github.com/AleDataAnalyst/grupo2\\_desastresnaturales](https://github.com/AleDataAnalyst/grupo2_desastresnaturales)

Elena Blanco

blancoelenal11@gmail.com

LinkedIn:  
[@elenablancodataanalytics](https://www.linkedin.com/in/elenablancodataanalytics)

Alejandra Eng

nomadicblink@gmail.com  
<https://nomadicblink.com>  
<https://ongdigital.es>  
[Github.com: AleDataAnalyst](https://github.com/AleDataAnalyst)  
[LinkedIn: @alejandraengbroca](https://www.linkedin.com/in/alejandraengbroca)

Rafael De Marco

rdemarcoz@outlook.com

LinkedIn:  
<https://www.linkedin.com/in/rafael-de-marco-z-53097014>

Ángel Bardon

bardon.angel@gmail.com

LinkedIn:  
<https://www.linkedin.com/in/angel-bardon-diez>

# AGRADECIMIENTOS



A todo el equipo, especialmente a Cristina, Pilar y Alan,  
por habernos introducido al mundo del Análisis de Datos

A nuestros mentores:  
Gema Piña de Minsait y José Luis Carvajal de INDRA

A las empresas que apoyan el programa de Escuela Digital  
y a la Agencia para el Empleo del Ayuntamiento de Madrid