

## Data Dictionary

**Project Title:** Media Coverage and Humanitarian Crises Analysis

**Database:** humanitarian.db

**Description:** This dataset examines the relationship between media coverage and humanitarian crisis severity indicators. It integrates crisis-level, temporal, outlet-level, framing, sentiment, and attribution data.

---

### 1. Crises Table

**Table Name:** crises

**Unit of Analysis:** Crisis-level (one row per crisis)

Variable Name	Data Type	Description	Measurement / Source
crisis_id	INTEGER (Primary Key)	Unique identifier for each crisis	Generated programmatically
crisis_name	TEXT	Name of the humanitarian crisis	Source dataset
start_date	DATE / TEXT	Official start date of the crisis	Source dataset
fund_required	FLOAT	Total funding required (USD)	Humanitarian funding data
people_affected	INTEGER	Total number of people affected	Humanitarian reports
crisis_days	INTEGER	Duration of crisis (in days)	Calculated from dates
raw_coverage	INTEGER	Total number of news articles covering the crisis	Media dataset
coverage_per_day	FLOAT	Articles per crisis day	raw_coverage / crisis_days
coverage_per_funding	FLOAT	Coverage relative to funding required	raw_coverage / fund_required
coverage_per_people	FLOAT	Coverage relative to affected population	raw_coverage / people_affected

---

## 2. Monthly Coverage Table

**Table Name:** monthly\_coverage

**Unit of Analysis:** Crisis-month level

Variable Name	Data Type	Description	Measurement / Source
monthly_id	INTEGER (Primary Key)	Unique monthly observation ID	Generated programmatically
crisis_id	INTEGER (Foreign Key)	Links to crises table	Derived mapping
crisis_name	TEXT	Crisis name	Source dataset
year_month	TEXT	Year and month (YYYY-MM)	Source dataset
coverage_count	INTEGER	Number of articles in that month	Media dataset

---

## 3. Outlet Coverage Table

**Table Name:** outlet\_coverage

**Unit of Analysis:** Crisis-outlet level

Variable Name	Data Type	Description	Measurement / Source
outlet_id	INTEGER (Primary Key)	Unique outlet observation ID	Generated programmatically
crisis_id	INTEGER (Foreign Key)	Links to crises table	Derived mapping
crisis_name	TEXT	Crisis name	Source dataset
outlet_name	TEXT	Media outlet name	Media dataset
coverage_count	INTEGER	Number of articles by outlet	Media dataset

---

## 4. Framing Table

**Table Name:** framing

**Unit of Analysis:** Article-level

Variable Name	Data Type	Description	Measurement / Source
article_id	INTEGER (Primary Key)	Unique article identifier	Generated
crisis_id	INTEGER (Foreign Key)	Links to crises table	Derived mapping
crisis_name	TEXT	Crisis name	Source dataset
framing_category	TEXT	Dominant narrative frame (e.g., humanitarian, political, conflict)	Content coding

---

## 5. Sentiment Table

**Table Name:** sentiment

**Unit of Analysis:** Article-level

Variable Name	Data Type	Description	Measurement / Source
article_id	INTEGER (Primary Key)	Unique article identifier	Generated
crisis_id	INTEGER (Foreign Key)	Links to crises table	Derived mapping
crisis_name	TEXT	Crisis name	Source dataset
sentiment_score	FLOAT	Numerical sentiment polarity score	Sentiment analysis tool
sentiment_label	TEXT	Categorized sentiment (Positive, Neutral, Negative)	Derived from score

---

## 6. Victim–Causor Attribution Table

**Table Name:** victim\_causor

**Unit of Analysis:** Article-level

<b>Variable Name</b>	<b>Data Type</b>	<b>Description</b>	<b>Measurement / Source</b>
article_id	INTEGER (Primary Key)	Unique article identifier	Generated
crisis_id	INTEGER (Foreign Key)	Links to crises table	Derived mapping
crisis_name	TEXT	Crisis name	Source dataset
victim_actor	TEXT	Identified victim group	Content coding
causor_actor	TEXT	Identified responsible actor	Content coding

---

### Notes on Data Processing

- Crisis IDs are generated programmatically to ensure relational integrity across tables.
- Derived variables (e.g., coverage\_per\_day, coverage\_per\_funding) are calculated during preprocessing in Python.
- All tables are stored in a SQLite database (humanitarian.db).
- CSV files in the /data/raw/ directory serve as the original data sources.
- Data loading is handled through pandas and inserted into SQLite using to\_sql().