

# Kenya - Road Traffic Crashes 2012-2023

**Sveta Milusheva**

Report generated on: June 6, 2024

Visit our data catalog at: <https://microdata.worldbank.org/index.php>

## Identification

### SURVEY ID NUMBER

KEN\_2012-2023\_RTC\_v01\_M

### TITLE

Road Traffic Crashes 2012-2023

### SUBTITLE

Derived from Crowdsourced Reports from Ma3Route

### ABBREVIATION OR ACRONYM

RTC 2012-23

### COUNTRY/ECONOMY

Name	Country code
Kenya	KEN

### ABSTRACT

This project geolocated the location of road traffic crashes based on crowdsourced reports of crashes from Ma3Route, a mobile/web/SMS platform that crowdsources transport data

### KIND OF DATA

Observation data/ratings [obs]

### UNIT OF ANALYSIS

Road traffic crashes

## Version

### VERSION DESCRIPTION

The datasets contain the time and location of road traffic crashes in Kenya (primarily Nairobi); crash information is derived from crowdsourced reports from @Ma3Route. Ma3Route is a mobile/web/SMS platform that crowdsources transport data and provides users with information on traffic, road traffic crash (RTC), matatu directions and driving reports. Users post RTC or traffic information to Ma3Route, where Ma3Route then publishes the post on Twitter. Tweets from @Ma3Route were queried using the Twitter API (tweets were no longer queried once Twitter rebranded to X).

### VERSION DATE

2024-05-20

## Coverage

### GEOGRAPHIC COVERAGE

Primarily Nairobi, Kenya

## Producers and sponsors

### PRIMARY INVESTIGATORS

Name	Affiliation
Sveta Milusheva	World Bank

### FUNDING AGENCY/SPONSOR

**Name**

World Bank

## Sampling

**SAMPLING PROCEDURE**

All tweets from @Ma3Route from August 2012 to July 2023

## Data collection

**DATES OF DATA COLLECTION**

Start	End	Cycle
2012-08-01	2023-07-12	1

**DATA COLLECTION MODE**

Internet [int]

## Access policy

**CONTACTS**

Name	Affiliation	Email
Sveta Milusheva	WBG	smilusheva@worldbank.org
Guadalupe Bedoya	WBG	gbedoya@worldbank.org
Robert Marty	WBG	rmarty@worldbank.org
Amy Dolinger	WBG	adolinger@worldbank.org
Arianna Legovini	WBG	alegovini@worldbank.org

**CITATION REQUIREMENTS**

Milusheva S, Marty R, Bedoya G, Williams S, Resor E, et al. (2021) "Applying machine learning and geolocation techniques to social media data (Twitter) to develop a resource for urban planning." PLOS ONE 16(2): e0244317.

<https://doi.org/10.1371/journal.pone.0244317>

**ACCESS AUTHORITY**

Name	Affiliation	Email
Sveta Milusheva	WBG	smilusheva@worldbank.org
Guadalupe Bedoya	WBG	gbedoya@worldbank.org
Robert Marty	WBG	rmarty@worldbank.org
Amy Dolinger	WBG	adolinger@worldbank.org
Arianna Legovini	WBG	alegovini@worldbank.org

## Disclaimer and copyrights

**DISCLAIMER**

The user of the data acknowledges that the original collector of the data, the authorized distributor of the data, and the relevant funding agency bear no responsibility for use of the data or for interpretations or inferences based upon such uses.

# Metadata production

---

DDI DOCUMENT ID  
DDI\_KEN\_2012-2023\_RTC\_v01\_M

PRODUCERS

Name	Abbreviation	Affiliation	Role
Development Data Group	DECDG	World Bank	Documentation of the DDI

DATE OF METADATA PRODUCTION  
2024-05-23

**Data Dictionary**

<b>Data file</b>	<b>Cases</b>	<b>Variables</b>
<b>ma3route_crashes_algorithmcode.dta</b> Includes crashes from August 2012 through July 2023; whether a tweet reports a crash and the geolocation of the crash are determined by the described in Milusheva et al. (2021)	31064	10
<b>ma3route_crashes_manualcode.dta</b> Includes crashes from July 2017 through July 2018; whether a tweet reports a crash and the geolocation of the crash are manually coded.	2595	10



**Data file: ma3route\_crashes\_algorithmcode.dta**

Includes crashes from August 2012 through July 2023; whether a tweet reports a crash and the geolocation of the crash are determined by the described in Milusheva et al. (2021)

Cases: 31064

Variables: 10

**Variables**

ID	Name	Label	Question
V2	crash_id	Unique crash ID	
V3	crash_datetime	Date/time of the crash (using date/time of the first tweet that reported the crash)	
V4	crash_date	Date of the crash (using date of the first tweet that reported the crash)	
V5	latitude	Latitude	
V6	longitude	Longitude	
V7	n_crash_reports	Number of tweets that reported crash	
V8	contains_fatality_words	Whether the tweet contains one of the words: 'dead', 'died', 'body', 'killed', or 'fatal'	
V9	contains_pedestrian_words	Whether the tweet contains the word: 'pedestrian'	
V10	contains_matatu_words	Whether the tweet contains the word: 'matatu'	
V11	contains_motorcycle_words	Whether the tweet contains one of the words: 'boda', 'motorycle', or 'motor cycle'	

Total: 10

**Data file: ma3route\_crashes\_manualcode.dta**

Includes crashes from July 2017 through July 2018; whether a tweet reports a crash and the geolocation of the crash are manually coded.

Cases: 2595

Variables: 10

**Variables**

ID	Name	Label	Question
V12	crash_id	Unique crash ID	
V13	crash_datetime	Date/time of the crash (using date/time of the first tweet that reported the crash)	
V14	crash_date	Date of the crash (using date of the first tweet that reported the crash)	
V15	latitude	Latitude	
V16	longitude	Longitude	
V17	n_crash_reports	Number of tweets that reported crash	
V18	contains_fatality_words	Whether the tweet contains one of the words: 'dead', 'died', 'body', 'killed', or 'fatal'	
V19	contains_pedestrian_words	Whether the tweet contains the word: 'pedestrian'	
V20	contains_matatu_words	Whether the tweet contains the word: 'matatu'	
V21	contains_motorcycle_words	Whether the tweet contains one of the words: 'boda', 'motorcycle', or 'motor cycle'	

Total: 10





**CONTAINS\_MATATU\_WORDS: Whether the tweet contains the word: 'matatu'****Data file:** ma3route\_crashes\_algorithmcode.dta**Overview**

Valid: 31064    Invalid:    Minimum: 0    Maximum: 1  
 Type: Discrete    Decimal: 0    Width: 1    Range: 0 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
0	No	28523	91.8%
1	Yes	2541	8.2%

**CONTAINS\_MOTORCYCLE\_WORDS: Whether the tweet contains one of the words: 'boda', 'motorycle', or 'motor cycle'****Data file:** ma3route\_crashes\_algorithmcode.dta**Overview**

Valid: 31064    Invalid:    Minimum: 0    Maximum: 1  
 Type: Discrete    Decimal: 0    Width: 1    Range: 0 - 1    Format: Numeric

**Questions and instructions**

## CATEGORIES

Value	Category	Cases	
0	No	29922	96.3%
1	Yes	1142	3.7%

**CRASH\_ID: Unique crash ID****Data file:** ma3route\_crashes\_algorithmcode.dta**Overview**

Valid: 31064    Invalid:    Minimum: 1    Maximum: 31064  
 Type: Continuous    Decimal: 0    Width: 5    Range: 1 - 31064    Format: Numeric

**CRASH\_DATETIME: Date/time of the crash (using date/time of the first tweet that reported the crash)****Data file:** ma3route\_crashes\_algorithmcode.dta**Overview**

Valid: 31064    Invalid:

Type: Continuous    Width: 20    Range: -    Format: character

### CRASH\_DATE: Date of the crash (using date of the first tweet that reported the crash)

Data file: ma3route\_crashes\_algorithmcode.dta

#### Overview

Valid: 31064    Invalid:

Type: Continuous    Width: 10    Range: -    Format: Date

### LATITUDE: Latitude

Data file: ma3route\_crashes\_algorithmcode.dta

#### Overview

Valid: 31064    Invalid:    Minimum: -3.10000000200807    Maximum: -0.565402313796744

Type: Continuous    Decimal: 0    Width: 18    Range: -3.10000000200807 - -0.565402313796744    Format: Numeric

### LONGITUDE: Longitude

Data file: ma3route\_crashes\_algorithmcode.dta

#### Overview

Valid: 31064    Invalid:    Minimum: 36.2839500025994    Maximum: 37.8794900005407

Type: Continuous    Decimal: 0    Width: 16    Range: 36.2839500025994 - 37.8794900005407    Format: Numeric

### N\_CRASH\_REPORTS: Number of tweets that reported crash

Data file: ma3route\_crashes\_algorithmcode.dta

#### Overview

Valid: 31064    Invalid:    Minimum: 1    Maximum: 66

Type: Discrete    Decimal: 0    Width: 2    Range: 1 - 66    Format: Numeric

### Questions and instructions

#### CATEGORIES

Value	Category	Cases	
1		25252	81.3%
2		3551	11.4%
3		1047	3.4%
4		445	1.4%
5		241	0.8%
6		163	0.5%
7		89	0.3%

8		86	0.3%
9		40	0.1%
10		23	0.1%
11		24	0.1%
12		20	0.1%
13		17	0.1%
14		10	0%
15		4	0%
16		13	0%
17		5	0%
18		3	0%
19		5	0%
20		3	0%
21		3	0%
22		1	0%
23		2	0%
25		2	0%
26		2	0%
27		1	0%
28		1	0%
30		1	0%
31		1	0%
32		1	0%
34		1	0%
35		1	0%
38		1	0%
40		1	0%
41		1	0%
46		1	0%
51		1	0%
66		1	0%

**CONTAINS\_FATALITY\_WORDS:** Whether the tweet contains one of the words: 'dead', 'died', 'body', 'killed', or 'fatal'

Data file: ma3route\_crashes\_algorithmcode.dta

### Overview

Valid: 31064    Invalid:    Minimum: 0    Maximum: 1  
Type: Discrete    Decimal: 0    Width: 1    Range: 0 - 1    Format: Numeric

## Questions and instructions

---

### CATEGORIES

Value	Category	Cases	
0	No	28780	92.6%
1	Yes	2284	7.4%

---

### **CONTAINS\_PEDESTRIAN\_WORDS: Whether the tweet contains the word: 'pedestrian'**

Data file: ma3route\_crashes\_algorithmcode.dta

#### Overview

Valid: 31064   Invalid:   Minimum: 0   Maximum: 1  
 Type: Discrete   Decimal: 0   Width: 1   Range: 0 - 1   Format: Numeric

## Questions and instructions

---

### CATEGORIES

Value	Category	Cases	
0	No	30120	97%
1	Yes	944	3%

---

**CRASH\_ID: Unique crash ID****Data file:** ma3route\_crashes\_manualcode.dta**Overview**

Valid: 2595   Invalid:   Minimum: 1   Maximum: 3115  
 Type: Continuous   Decimal: 0   Width: 4   Range: 1 - 3115   Format: Numeric

---

**CRASH\_DATETIME: Date/time of the crash (using date/time of the first tweet that reported the crash)****Data file:** ma3route\_crashes\_manualcode.dta**Overview**

Valid: 2595   Invalid:  
 Type: Continuous   Width: 20   Range: -   Format: character

---

**CRASH\_DATE: Date of the crash (using date of the first tweet that reported the crash)****Data file:** ma3route\_crashes\_manualcode.dta**Overview**

Valid: 2595   Invalid:  
 Type: Continuous   Width: 10   Range: -   Format: Date

---

**LATITUDE: Latitude****Data file:** ma3route\_crashes\_manualcode.dta**Overview**

Valid: 2595   Invalid:   Minimum: -4.059868   Maximum: 1.257331  
 Type: Continuous   Decimal: 0   Width: 18   Range: -4.059868 - 1.257331   Format: Numeric

---

**LONGITUDE: Longitude****Data file:** ma3route\_crashes\_manualcode.dta**Overview**

Valid: 2595   Invalid:   Minimum: 34.145808   Maximum: 40.171389  
 Type: Continuous   Decimal: 0   Width: 16   Range: 34.145808 - 40.171389   Format: Numeric

---

**N\_CRASH\_REPORTS: Number of tweets that reported crash****Data file:** ma3route\_crashes\_manualcode.dta**Overview**

Valid: 2595   Invalid:   Minimum: 1   Maximum: 67  
 Type: Discrete   Decimal: 0   Width: 2   Range: 1 - 67   Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
1		2014	77.6%
2		321	12.4%
3		112	4.3%
4		54	2.1%
5		29	1.1%
6		16	0.6%
7		8	0.3%
8		8	0.3%
9		9	0.3%
10		5	0.2%
11		2	0.1%
12		2	0.1%
13		1	0%
14		4	0.2%
15		2	0.1%
16		1	0%
18		2	0.1%
20		1	0%
35		1	0%
42		1	0%
49		1	0%
67		1	0%

**CONTAINS\_FATALITY\_WORDS:** Whether the tweet contains one of the words: 'dead', 'died', 'body', 'killed', or 'fatal'

Data file: ma3route\_crashes\_manualcode.dta

### Overview

Valid: 2595   Invalid:   Minimum: 0   Maximum: 1  
 Type: Discrete   Decimal: 0   Width: 1   Range: 0 - 1   Format: Numeric

## Questions and instructions

### CATEGORIES

Value	Category	Cases	
0	No	2429	93.6%

1	Yes	166	6.4%
---	-----	-----	------

### CONTAINS\_PEDESTRIAN\_WORDS: Whether the tweet contains the word: 'pedestrian'

Data file: ma3route\_crashes\_manualcode.dta

#### Overview

Valid: 2595 Invalid: Minimum: 0 Maximum: 1  
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

#### Questions and instructions

##### CATEGORIES

Value	Category	Cases	
0	No	2534	97.6%
1	Yes	61	2.4%

### CONTAINS\_MATATU\_WORDS: Whether the tweet contains the word: 'matatu'

Data file: ma3route\_crashes\_manualcode.dta

#### Overview

Valid: 2595 Invalid: Minimum: 0 Maximum: 1  
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

#### Questions and instructions

##### CATEGORIES

Value	Category	Cases	
0	No	2386	91.9%
1	Yes	209	8.1%

### CONTAINS\_MOTORCYCLE\_WORDS: Whether the tweet contains one of the words: 'boda', 'motorycle', or 'motor cycle'

Data file: ma3route\_crashes\_manualcode.dta

#### Overview

Valid: 2595 Invalid: Minimum: 0 Maximum: 1  
Type: Discrete Decimal: 0 Width: 1 Range: 0 - 1 Format: Numeric

#### Questions and instructions

##### CATEGORIES



Value	Category	Cases	
0	No	2500	96.3%
1	Yes	95	3.7%

# Documentation

## Technical documents

### Road Traffic Crashes Derived from Crowdsourced Reports from Ma3Route - Dataset Description

---

Title	Road Traffic Crashes Derived from Crowdsourced Reports from Ma3Route - Dataset Description
Country	Kenya
Language	English
Filename	documentation.pdf

---