Kenya - Manually Labelled Crash Reports from Ma3Route 2017-2018

World Bank

Report generated on: March 24, 2021

Overview

Identification

ID NUMBER
KEN_2017-2018_MLCRM_v01_M

Version

VERSION DESCRIPTION
- v2.1: Edited, anonymous dataset for public distribution.

Overview

ABSTRACT
The purpose of the Tweet IDs and Manually Labelled Crash Reports from Ma3Route 2017-2018 project is identify tweets from the @Ma3Route twitter handles that report road traffic crash reports. Using the Twitter API, tweets were scraped from Ma3Route, which is a mobile/web/SMS platform that crowdsources transport data and provides users with information on on road traffic crash reports as well as traffic, matatu directions, and driving reports.

KIND OF DATA
Observation data/ratings [obs]

UNITS OF ANALYSIS
Road traffic crash reports

Coverage

GEOGRAPHIC COVERAGE
Kenya (primarily Nairobi)

UNIVERSE
Tweets reporting road traffic crash reports, scraped from twitter handle @Ma3Route

Producers and Sponsors

PRIMARY INVESTIGATOR(S)

Name	Affiliation
World Bank	

OTHER PRODUCER(S)

Name	Affiliation	Role
Sveta Milusheva	Development Impact Evaluation Department, World Bank	
Robert Marty	Development Impact Evaluation Department, World Bank	
Guadalupe Bedoya	Development Impact Evaluation Department, World Bank	
Sarah Williams	School of Architecture and Planning, Massachusetts Institute of Technology	

Elizabeth Resor	School of Information, University of California, Berkeley	
Arianna Legovini	Development Impact Evaluation Department, World Bank	

Metadata Production

METADATA PRODUCED BY

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

DATE OF METADATA PRODUCTION
2021-03-24

DDI DOCUMENT VERSION
Version 01 (March 2021)

DDI DOCUMENT ID
DDI_KEN_2017-2018_MLCRM_v01_M_WB

Sampling

No content available

Questionnaires

No content available

Data Collection

Data Collection Dates

Start	End	Cycle
2017-07-01	2018-07-31	N/A

Data Collection Mode

Other [oth]

Data Collection Notes

Using the Twitter API, tweets were scraped from the twitter handle @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 were obtained in order to identify tweets that reported RTC. Tweets from May 2012 to July 2020 were scraped and a "truth dataset", of tweets manually coded to determine if they reported crash reports and the location of the reported crashes, was generated. Additional information on the data is provided in additional documents found under the 'Documentation' tab.

Data Processing

No content available

Data Appraisal

No content available

File Description

Variable List

twitter_truth

Content	Using the Twitter API, tweets were scrapped from the twitter handle @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 were obtained in order to identify tweets that reported RTC. This dataset includes a manually labelled dataset of a subset of tweets indicating which tweets report a crash and the location of crashes.
Cases	9479
Variable(s)	9
Structure	Type: Keys: ()
Version	
Producer	
Missing Data	

Variables

ID	NAME	LABEL	TYPE	FORMAT	QUESTION
V1	uid	Unique ID	contin	numeric	Unique ID
V2	tweet_id	Tweet ID	discrete	character	Tweet ID
V3	created_at	Time Date/Time (EAT)	contin	numeric	Time Date/Time (EAT)
V4	crash_report	Tweet reports crash	discrete	numeric	Tweet reports crash
V5	latitude	Latitude of crash	contin	numeric	Latitude of crash
V6	longitude	Longitude of crash	contin	numeric	Longitude of crash
V7	crash_id_c1	Crash ID (from coder 1)	contin	numeric	Crash ID (from coder 1)
V8	crash_id_c2	Crash ID (from coder 2)	contin	numeric	Crash ID (from coder 2)
V9	crash_landmark	Landmark used to geocode	discrete	character	Landmark used to geocode

Unique ID (uid)
File: twitter_truth

Overview

Type: Continuous
Format: numeric
Width: 4
Decimals: 0
Range: 1-9479

Valid cases: 9479
Invalid: 0
Minimum: 1
Maximum: 9479

Literal question

Unique ID

Tweet ID (tweet_id)
File: twitter_truth

Overview

Type: Discrete
Format: character
Width: 19

Valid cases: 7724
Invalid: 0

Literal question

Tweet ID

Time Date/Time (EAT) (created_at)
File: twitter_truth

Overview

Type: Continuous
Format: numeric
Width: 13
Decimals: 0
Range: 1814496118000-1848681624000

Valid cases: 9479
Invalid: 0
Minimum: 1814496118000
Maximum: 1848681624000

Literal question

Time Date/Time (EAT)

Tweet reports crash (crash_report)
File: twitter_truth

Overview

Type: Discrete
Format: numeric
Width: 1
Decimals: 0
Range: 0-1

Valid cases: 9479
Invalid: 0
Minimum: 0
Maximum: 1

Literal question

Tweet reports crash

Latitude of crash (latitude)
File: twitter_truth

Overview

Type: Continuous
Format: numeric
Width: 9
Decimals: 0
Range: -4.059868-1.257331

Valid cases: 4193
Invalid: 5286
Minimum: -4.1
Maximum: 1.3

Literal question

Latitude of crash

Longitude of crash (longitude)
File: twitter_truth

Overview

Type: Continuous
Format: numeric
Width: 9
Decimals: 0
Range: 34.145808-40.171389

Valid cases: 4191
Invalid: 5288
Minimum: 34.1
Maximum: 40.2

Literal question

Longitude of crash

Crash ID (from coder 1) (crash_id_c1)
File: twitter_truth

Overview

Type: Continuous
Format: numeric
Width: 5
Decimals: 0
Range: 907-41898

Valid cases: 3796
Invalid: 5683
Minimum: 907
Maximum: 41898

Literal question

Crash ID (from coder 1)

Crash ID (from coder 2) (crash_id_c2)
File: twitter_truth

Overview

Type: Continuous
Format: numeric
Width: 5
Decimals: 0
Range: 909-47227

Valid cases: 3674
Invalid: 5805
Minimum: 909
Maximum: 47227

Literal question

Crash ID (from coder 2)

Landmark used to geocode (crash_landmark)
File: twitter_truth

Overview

Type: Discrete
Format: character
Width: 71

Valid cases: 4210
Invalid: 0

Literal question

Landmark used to geocode

Documentation

Technical documents

Documentation - Tweet IDs and Manually Labelled Crash Reports from Ma3Route 2012-2020

Title	Documentation - Tweet IDs and Manually Labelled Crash Reports from Ma3Route 2012-2020
Author(s)	World Bank
Country	Kenya
Language	English
Filename	Doc/Technical/documentation.pdf

Ma3Map Social Media Classification

Title	Ma3Map Social Media Classification
Author(s)	World Bank
Country	Kenya
Language	English
Filename	Doc/Technical/ma3map_coding_instructions.pdf

Tweet Crash Clustering Protocol

Title	Tweet Crash Clustering Protocol
Author(s)	World Bank
Country	Kenya
Language	English
Filename	Doc/Technical/crash_clustering_protocol.pdf