

# Trend analysis and fatality causes in Kenyan roads: A review of road traffic accident data between 2015 and 2020

Joseph Kamau Muguro, Minoru Sasaki, Kojiro Matsushita & Waweru Njeri |

**To cite this article:** Joseph Kamau Muguro, Minoru Sasaki, Kojiro Matsushita & Waweru Njeri | (2020) Trend analysis and fatality causes in Kenyan roads: A review of road traffic accident data between 2015 and 2020, Cogent Engineering, 7:1, 1797981, DOI: [10.1080/23311916.2020.1797981](https://doi.org/10.1080/23311916.2020.1797981)

**To link to this article:** <https://doi.org/10.1080/23311916.2020.1797981>



© 2020 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.



Published online: 05 Aug 2020.



Submit your article to this journal



Article views: 27077



View related articles



View Crossmark data



Citing articles: 9 View citing articles



Received: 22 April 2020  
Accepted: 05 July 2020

\*Corresponding author: Minoru Sasaki, Mechanical Engineering, Gifu University, Gifu, Japan  
E-mail: [sasaki@gifu-u.ac.jp](mailto:sasaki@gifu-u.ac.jp)

Reviewing editor:  
Ahmed Haj Darwish, Faculty of Informatics Engineering, University of Aleppo, Aleppo, Syrian Arab Republic

Additional information is available at the end of the article

## CIVIL & ENVIRONMENTAL ENGINEERING | RESEARCH ARTICLE

# Trend analysis and fatality causes in Kenyan roads: A review of road traffic accident data between 2015 and 2020

Joseph Kamau Muguro<sup>1,2</sup>, Minoru Sasaki<sup>1\*</sup>, Kojiro Matsushita<sup>1</sup> and Waweru Njeri<sup>2</sup>

**Abstract:** With increasing population and motorization, Kenya as well as other African countries are faced with a tragic road traffic accidents (RTA). This paper looks at 5-year (2015–2020) data downloaded from National Transport and Safety Authority (NTSA) website, to identify trends and review progress of the traffic accidents in the country. The objective is to assess the prevalence of accidents within affected groups and location to identify trends and generalized causative agency from the reported data. From literature review, research activity focused on RTA in the country is minimal compared to the social significance accidents poses. The data were extracted and classified using Latent Dirichlet Allocation, a machine learning algorithm modelled in Matlab to group reported accident briefs into general categories/topic which are closely related. Four categories were identified as leading causes of fatality in the country: Knocking down victims, hit-and-run, losing control and head on collision. The identified causes point to preventable driver's errors which agrees with other

### ABOUT THE AUTHORS

Joseph Kamau Muguro is a Ph.D. Fellow in Faculty of Engineering, Gifu University, and a Lecturer at DeKUT, Kenya. His research interest includes Machine Learning, Driver Behavior and Accident Analysis, among others.

Waweru Njeri is currently a Post-Doctoral Fellow at Gifu University, and a Lecturer at DeKUT. Current research interests include Intelligent Control of Electro-mechanical System, Robotics, and others.

Minoru Sasaki is a Senior Professor in Gifu University, Faculty of Engineering, Intelligent Mechanical Engineering Course. He holds a Bachelor's degree from Yamagata University, Master and Doctorate degrees from Tohoku University, in Mechanical Engineering. He has collaborated with research institutions including JAXA, University of California, Los Angeles, DeKUT among others. His current research interests include Intelligent Control, Mechatronics, and Robotics.

Kojiro Matshushita is an Associate Professor, Faculty of Engineering, Gifu University in Intelligent Mechanical Engineering Course. His current research interest includes Welfare Robots, Human Machine Interface, Mechatronics, and others.

### PUBLIC INTEREST STATEMENT

Currently, Kenya is witnessing an unprecedented challenge of tragic road accidents. This paper looks at data from National Transport and Safety Authority (NTSA) website for the period between the year 2015 and 2020. The objective is to assess the prevalence of accidents within affected groups and the general causes of deaths in the reported data. From machine learning, we identified four major leading causes of deaths in the country as: knocking down victims, hit-and-run, losing control and head on collision. All these points to preventable errors which can be remedied using proper policies. In addition, accident related deaths have increased by 26% while injuries have increased by 46.5% between 2015 January and 2020 January. Injuries to pedestrians, motorcycle (boda-boda) passengers and motorcyclist (vulnerable road users), have increased by over 250% for the same period. From the paper, urgent action is needed by all stakeholders to deal with the menace as well as curb the overly decried driver behavior.

researchers. From trend analysis, fatalities and injuries have increased by 26% and 46.5%, respectively since January 2015 to January 2020. This paper found that injuries in vulnerable road users: pedestrians, pillion passengers and motorcyclist, has seen a foldfold increment compared to 2015 data. From the discussion, urgent fine-tuning of policing to protect vulnerable road user as well as curb the overly decried driver behavior is needed. The paper recommends fine-tuning of data collection, capturing details of accident that will be useful in modeling and data analysis for future planning.

**Subjects:** Environmental Impact Assessment; Machine Learning; Transportation Engineering; Road Transport Industries

**Keywords:** Road traffic accidents (RTA); machine learning; LDA; Kenya

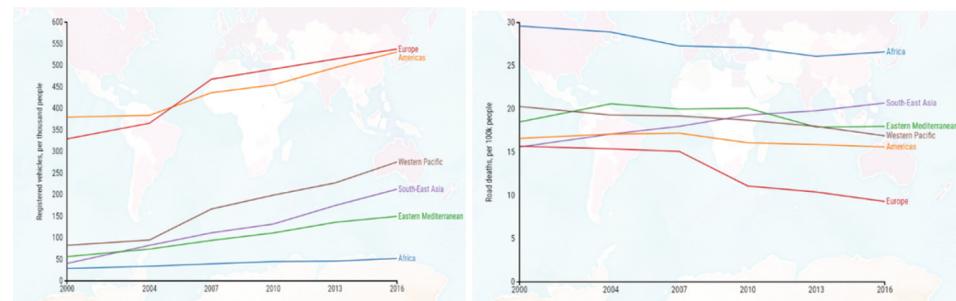
## 1. Introduction

According to global status report on road safety in 2018 by World Health Organization (WHO), fatalities emanating from road traffic accidents (RTAs) have increased to 1.3 million per year (2018). More than 93% of these accidents, occur in low-income and middle-income countries. The report also identified road traffic deaths (RTDs) as the leading cause of death for people groups between 5 and 29 years of age. Furthermore, research on RTA estimates that more than 50% of injuries/deaths emanating from road accident occur to age groups 15–49 years which is considered the economically productive period (Sapkota et al., 2016), (Macharia et al., 2009, June). From previous and current reports and various research undertakings, it has been shown that RTDs are severe on vulnerable road users, pedestrians, cyclists and riders of motorbikes and their passengers accounting for staggering 46% of global traffic deaths (World Health Organization, 2018), (World Health Organization, 2015), (World Health Organization, 2009).

A major imbalance that was shown in the 2018 WHO report is on the number of registered vehicles per 1,000 people and the rate of deaths per 100,000 people in African region compared to the world. This is shown in Figure 1. As of 2016, Africa as a region had the highest death rate (more than 26 per 100,000) with the lowest motorization (less than 50 per 1000) (2018). The financial implications that RTAs places on these developing economies is far worse than acceptable. Studies estimating the financial implications of road traffic injuries of a country ranges between 1.3% and 3.0% of gross domestic product (GDP) (Manyara, 2016). This means that concerted efforts are needed to combat the menace through public sensitization, proper policy formulation and/or enforcement as well as road infrastructural development.

Kenya as a country, being ranked as a lower middle-income, has experienced an increase in RTAs over the last decade. This has to do with urbanization and increasing motorization in the country. Being a low income-economy, road infrastructural development is still lagging as well as policy

**Figure 1. RTA trends per capita according to WHO global report 2018.**



a) Registered vehicles per 1000 people

b) Road deaths per 100,000 people

challenges in following through the international safety standards. According to National Transport and Safety Authority (NTSA), the body in charge of transport in Kenya, the country recorded 3572 fatalities, 6938 serious injuries and 5186 slight injuries as at December 2019 (NTSA, 2020). The recorded numbers have been questioned by various parties (Kelly, 2018). The general explanation on this has to do on the collection of data and the categorization of accident data. NTSA seemingly adopts a dead-on-the-spot for fatality report; no follow up is done with hospitals to determine which injuries led to death. This is contrary to international recommended standard of reporting which considers RTDs within a 30-day window (OECD, 1980) (World Health Organization, 2015).

### **1.1. Traffic situation in Kenya**

Previous research has identified various transportation and traffic situation as identifiers for accident prevalence. Such identifiers include but not limited to, traffic congestion, human factors and behaviors, road types and sections, vehicle conditions and motorization, policing, weather, etc. (Zheng et al., 2020; Hordofa et al., 2018; Bucsuhaázy et al., 2020; Sun et al., 2019; Mohammed et al., 2019). In this section, we explore some of the prevailing situations thought to fuel RTAs in the country and some remedial actions and policies taken by the relevant stakeholders.

#### *1.1.1. Public service vehicles and motorbikes*

The country relies majorly on public service vehicles (PSVs) to meet the commuter needs of diverse groups of people. It has been reported that mode of public transport (Matatu industry) in Kenya is at fault when it comes to issues on safety (Manyara, 2016) (Mogambi & Nyakeri, 2015). PSVs in Kenya are legalized to carry between 10 and 50 passengers depending on the category of the vehicle (R. of K. (GoK), 2018). The vehicles often ferry passengers in excess, in disregard to the laid out recommendation for each category (B. Team, 2019). In the occurrence of a traffic accident, this increases the number of casualties and impacted people. Additionally, a PSV driver is paid on a per-trip basis, driving on unmarked and unsigned roads, with dense traffic jams with another competing motorist plying the same route. This causes these drivers to have a general aggressive tendency while driving on any road.

Besides the Matatu industry, motorcycle transport (Boda-boda) has been on the rise in the country (Murumba, 2017). From literature, growth in usage of motorized two/three-wheeled vehicles has been reported to increase injuries and fatalities among users (Diaz Olvera et al., 2019) (D. Wang et al., 2019) (Soehodho, 2017). The situation is made worse by the rise in registered vehicles in the country. According to (Islam & Al Hadhrami, 2012), higher rate of motorization increases RTAs.

#### *1.1.2. Disaster preparedness*

Availability of first aid and emergency centers/hospitals has been identified as a challenge in many different developing countries Kenya included. In the country, it has been reported that only 16% of road accidents casualties received first aid, 76.5% of the injured persons were transported to hospitals by well-wishers (other motorists) while Police and ambulance vehicles transported 6.1% and 1.4%, respectively (Macharia et al., 2009, June).

#### *1.1.3. Policing*

In several papers, it is reported that 85% of the accidents are caused by driver error in Kenya ranging from over speeding, intoxication and/or plain recklessness among others (Manyara, 2016), (Mogambi & Nyakeri, 2015). The Kenya police in conjunction with NTSA has been enforcing adherence of traffic rules in the country. Touching on drunk-driving, NTSA introduced a breathalyzer to detect the alcohol content in drivers, infamously called alcoblow in 2012. This was in a bid to reduce drug and substance abuse as defined by traffic act (R. of K. (GoK), 2018). Alcoblown was however removed from the traffic monitoring personnel in November 2019 citing that the devices were not effective in deterring drunk-driving (Tendu, 2019). There is no published information of its reintroduction since then, but the problem of drunk driving still persists.

The Kenya police actively monitor reckless driving, over speeding and other malpractices that impact road accidents. Traffic police performs random “Crack-down” on vehicles that does not conform to the road safety standards laid out by the laws (Kipkemoi, 2020). By use of speed-traps and checkpoints, the traffic behavior is altered but the effects are mild in areas with no proper police supervision. In the eyes of the public, crackdowns have been quite disruptive to commuters who have tight travel schedules (Kipkemoi, 2020). It is worth noting that the strategy used in the country is mostly checkpoint system with few cases of patrol. The definitiveness of checkpoints may have drawbacks when it comes to enforcement of reckless driving behavior outside of the monitored zone.

### **1.2. Research and public awareness in RTAs**

We found few research related to RTAs in the period 2015–2020 reported in literature in relation to the prevalence of the menace (Manyara, 2016), (Fraser et al., 2020) (Bachani et al., 2017), (Myers et al., 2017), (Walcott-Bryant et al., 2016). This is also true for durations not covered in the paper. In the reviewed literature, authors focused on different aspects of RTAs in the country ranging from road and infrastructure, impacts of RTAs, access to hospital and emergency centers, trauma in accidents among others. There is still a big research need for inquiries to understand the intricacies of RTA in the country. There are no definitive answers as to what really causes the increases in accidents in the country. This is heightened by over-reliance on media stations for accident reporting. Authors in (Mogambi & Nyakeri, 2015) pointed out the tendency of media stations to prime (overreport crash events) and over rely on eyewitnesses. Under normal circumstances, news reports and police report of accidents is usually based on eyewitness accounts which is highly subjective. This is a major shortcoming in the publicly availed data source (NTSA, 2020).

#### **1.2.1. Zusha campaign**

One of the successful awareness campaigns in the country is Zusha campaign, a nation-wide awareness effort to combat RTAs. Zusha! -the Kiswahili word for “protest” or “speak up,” is meant to encourage passengers in PSV to speak up against reckless driving (Zusha, 2018), (GiveWell, 2018). The campaign distributed stickers on PSV with graphics and messages encouraging passengers to speak directly to the drivers against dangerous driving. From the campaign, cost-effectiveness ratio of the intervention is estimated to lie between US\$6.50 and \$11.70 per disability-adjusted life year (DALY) (Zusha, 2018). The campaign focuses and places the mandate of supervision on the passengers. Challenges facing the approach include complacency on the side of the passengers when they allow recklessness to gain a minute or two in the middle of a traffic jams, hostility from other passengers who are late or partisans in the transport industry to mention just but a few. Additionally, there has been violent exchanges between passengers and PSV handlers (drivers and conductors) as noted in (Wa Mungai & Samper, 2006).

### **1.3. Present study**

As mentioned in section 1.2, there has been few research activities in the country compared to the number of incidences. In this study, we seek to explore a 5-year (2015–2020) publicly accessible data to uncover trends and traffic situation of roads in Kenya. The main target is to assess the prevalence of accidents by affected groups, causative agent, and location within the country. We will utilize machine learning approaches to explore the data. Machine learning has been applied in various aspects of traffic analysis ranging from infrastructure design to prediction models (Gichaga, 2017; Pakgohar & Kazemi, 2015; Hadji Hosseini et al., 2018; Karimzadeh and Shoghli, 2020; Syssoev et al., 2020; Heyns et al., 2019). To our knowledge, there has not been publications on the said dataset. The target of the paper is to analyze eyewitness type of data (police reported) on the cause of the accident and come up with general categorization.

One of the contributions of the paper is on text mining of public accident records (eyewitness brief descriptions of the accident) to extract meaningful categorization of accidents. This is done using unsupervised machine learning model. The necessity of this inquiry is made clear by the fact that classification of fatalities plays a major role in establishing working solution. Table 1 shows agency's cause of crash for period between January and October 2019. From the table, the

**Table 1. NTSA report on leading causes of traffic crashes by October 2019**

No.	Cause	Incidence count
1	Hit and run	925
2	Losing control	451
3	Failing to keep left	235
4	Overtaking improperly	225
5	Misjudging clearance	140
6	Error of judgement	129
7	Excessive speeding	120
8	Stepping/walking running off vehicle	63
9	Crossing road in nondesignated area	44
10	Walking/standing in road	40
11	Vehicle turning without due care	31
12	Swerving	27
13	Brake failure	22
14	Drunk driving	19
15	Stationary vehicle dangerous parking	19
16	Falling from moving vehicle	17
17	Boarding or alighting from PSV	16
18	Failure of wheel-tires/wheels	14
19	Slipping or falling	14
20	Lane changing	14

categorization employed does not avail much information and is redundant in some. No other research has focused on the inquiry; most of the research in the country rely on police-recorded causes or a localized region as is the case in (Osoro et al., 2015). Understanding the causes at a national level will avail information regarding to what actions pose the greatest threats particularly to the vulnerable road users.

Additionally, the paper extract 5-year trends from public accident records to make inferences on trends. The agency approach on reporting has been on an annual-based comparative analysis; that is, the previous year is compared to the current year. With a unified 5-year data distributed monthly, we hope to identify trends that will avail more insights as well as enable projections of state of RTAs for the near future. The paper also discusses possible ways RTAs can be improved in the country and recommends best-practices touching on data collections and reporting. This conclusions and discussion will prove useful to stakeholders in planning and in the formulation of traffic policies.

The rest of the document is broken down as follows. Section 2 explains research demography, data collection and processing methods used. Section 3 gives the results of machine learning classification of the fatal accident causes and trend analysis of 5-year data. Section 4 discusses the results and the significance of the results and lastly a conclusion is drawn from the paper.

## 2. Methodology

### 2.1. Data collection and coverage

The data was downloaded from NTSA online database, the body in charge of safety and transportation oversight in Kenya. NTSA has two main types of data availed to the public as a transparency and sensitization of safety: *Daily reports* and *fatal accident report* covering the entire country. Data

collection involved downloading all entries availed in the website entry by entry. The data are published as excel documents as shown in samples below (Tables 2 and 3).

*Daily reports* show the fatalities and injuries categorized as: pedestrians, passengers, drivers, pillion passengers (motorcycle passengers), pedal cyclist and motor-bike cyclist as shown in Table 2. In fatal reports, the agency avails fatality incidences citing the place (County), time, details of the victim and category as shown in Table 2.

*Fatal accident report* shows time, Base/Subbase, County, Road, Place, motor vehicle (MV) Involved, Brief Accident Details, (Name of Victim), gender, age, cause code, victim, and no. of fields as entry columns. Of this, we are concerned with time, County, brief accident details and victim. The data covers the entire country subdivided into 47 regions (county). Demography of the country is shown in Table 4 highlighting the populations and road size. Population data is from 2018 census by Kenya bureau of statistics (KNBS, 2018). Paved road in this case is asphalt as described in Kenya Roads Board report for year 2018 (KRB, 2018).

## 2.2. Data processing

Two distinct data entries are considered in this paper. The general distinctions are as shown in Figure 2. We performed data processing for the two data types, herein referred as *fatal report* and *daily report*.

### 2.2.1. Daily report data processing

As shown in Table 2, *daily reports* contain statistics in a cumulative form; that is, from 1<sup>st</sup> January to 30 December 2019. From the table, every report has two entry dates current and previous year for comparison purposes. We applied data extraction for all categories using Matlab and separated the data into years. In the end, data for year 2015 up to 2020 was acquired.

To get a distribution and individual daily entries, we did some data processing and interpolation for missing date times. We were interested in monthly distribution of entries. As such, interpolating between existing data points to get a month-by-month datum was performed in Matlab®. Further data cleaning was done to omit entries with missing and or irrelevant fields. There was uncertainty in differentiating between *slightly injured* and *seriously injured* since no clear distinction was availed in the online data. In our analysis, we combined the two fields to one (Injured) for all the categories.

### 2.2.2. Fatal reports data processing

There is a “CAUSE CODE” column appended for each entry of *fatal accident report* incidences. However, we could not ascertain its significance anywhere in the database. *Brief accident details* column provides a free form of input text that describes the incidence. This is a form of eyewitness brief of accident. Sentences like “THE VEHICLE KNOCKED DOWN THE VICTIM,” “HEAD ON COLLISION,” “UNKNOWN M/V HIT UNKNOWN M/A PEDESTRIAN WHO DIED ON SPOT,” “THE CANTER RAMMED INTO THE TRAILER,” “THE TRACTOR LOST CONTROL AND HIT THE VICTIM,” and many others are common in the reports. We sought to cluster accidents into similar causes (description) using machine learning and establish the relationship/trends that exist in the data by developing a machine learning model that will “mine” through the description given and extract a general cause of the accident.

In machine learning and natural language processing (NLP), topic models are generative models, which provide a probabilistic framework (Blei et al., 2003). Topic modelling methods are used for automatic searching, understanding, and summarizing large electronic archives (Tong & Zhang, 2016, May) (Blei et al., 2003) (De Smet & Moens, 2009, November). The topics refer to the hidden variable relation that link words in a vocabulary and their context in documents. A document is viewed as a text containing a mixture of topics (Tong & Zhang, 2016, May). Topic models searches for the hidden themes in the dataset and annotate the documents according to found themes.

**Table 2. Sample Daily report from NTSAs (report as at 30 December 2019 data snippet)**

NATIONAL TRANSPORT AND SAFETY AUTHORITY

COMPARATIVE STATISTICS TRENDS FOR 2018 AND 2019 AS AT 30TH DECEMBER 2019										PILLION PASSENGER	PEDAL CYCLIST	MOTOR CYCLISTS	Total
Victim	PEDESTRIANS			DRIVERS			PASSENGERS						
	Category	Dead	Seriously Injured	Slightly Injured	Dead	Seriously Injured	Slightly Injured	Dead	Seriously Injured	Dead	Seriously Injured	Slightly Injured	Dead
1 JAN - 30 DECEMBER 2018	1202	851	256	306	495	305	745	2082	4073	246	498	205	63
1 JAN - 30 DECEMBER 2019	1382	1275	282	344	648	464	704	2770	3691	345	832	386	73

**Table 3. Sample Fatal report from NTSAs (report data as at 14 February 2016 data snippet)**

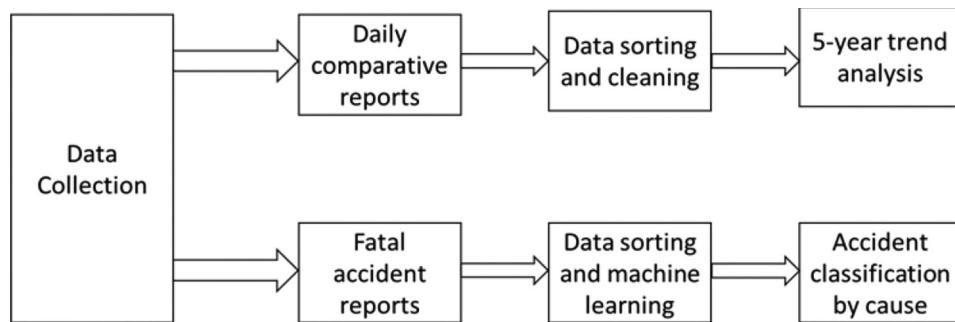
FATAL REPORT AS AT 14TH FEBRUARY 2016							
NATIONAL TRANSPORT AND SAFETY AUTHORITY							
S/N	TIME 24HR	BASE	COUNTY	ROAD	PLACE	MV INVOLVED	BRIEF ACCIDENT DETAILS
1	545	MTWAPA	KILIFI	MOMBASA KILIFI	MSAMBARAUNI	KBZ 884F ISUZU LORRY, KTWA 833W TUUK TUUK	HEAD ON COLLISION
2	300	INDO AREA	NAIROBI	MOMBASA	CAPITAL CENTRE	KYQ 080 VAN	THE VEHICLE KNOCKED DOWN THE VICTIM WHILE CROSSING THE ROAD
3	300	KAREN	NAIROBI	NGONG ROAD	MIOTONI	KBS 001C M/BENZ, KMDD 597 YAMAHA T/BUS	THE VEHICLE HIT THE MOTOR CYCLIST
4	620	DJANI	KWALE	BEACH ROAD	ASILIA COMPLEX	KBT 998R T/BUS	THE VEHICLE KNOCKED DOWN THE VICTIM WHILE CROSSING THE ROAD
5	UNKNOWN	CHANGAMWE	MOMBASA	MIRITINI	MAGANDA	KBK 897Q	THE VEHICLE KNOCKED DOWN THE VICTIM WHILE CROSSING THE ROAD
NATIONAL TRANSPORT AND SAFETY AUTHORITY							FATAL REPORT AS AT 14TH FEBRUARY 2016
S/N	NAME OF VICTIM	GENDER	AGE	CAUSE CODE	VICTIM		No
1	UNKNOWN	M	32	26	DRIVER		1
2	UNKNOWN	M	43	63	PEDESTRIAN		1
3	G** WALUMBO	M	46	7	M/CYCLIST		1
4	M** ATIENO	F	26	63	PEDESTRIAN		1
5	UNKNOWN	F	6	68	PEDESTRIAN		1

**Table 4. Kenya demographics showing population according to 2018 census and road inventory report as at 2018**

County	Population (at 2018)	Area (km <sup>2</sup> )	Total Roads (Km)	Paved (km)	County	Population (at 2018)	Area (km <sup>2</sup> )	Total roads (km)	Paved (km)
Nairobi	4,397,073	704	4,558.25	1780*	Nandi	885,711	2,856	3,340,415	290
Kiambu	2,417,735	2,539	7927.54	1158	Bomet	875,689	2,531	3,863,448	207
Nakuru	2,162,202	7,462	15,902.87	1063	Mandera	867,457	25,940	3649,754	87
Kakamega	1,867,579	3,020	5292.237	281	Kwale	866,820	8,267	5,455.46	193*
Bungoma	1,670,570	3,024	5218.076	251	Garissa	841,353	44,736	4,646.40	38*
Meru	1,545,714	7,006	8570.895	635	Wajir	781,263	56,773	7435.635	25
Kilifi	1,453,787	12,540	5,888.49	356*	Nyeri	759,164	3,325	5,356,244	777
Machakos	1,421,932	6,043	9,181.26	509*	Baringo	666,763	10,976	3829.591	500
Kisii	1,266,860	1,323	3127.237	342	Nyandarua	638,289	3,286	4,855,894	312
Mombasa	1,208,333	220	1,493.85	424*	West Pokot	621,241	9,123	2551.789	182
Uasin Gishu	1,163,186	3,392	5761.615	470	Kirinyaga	610,411	1,478	3,146,063	250
Narok	1,157,873	17,950	7,145.76	242*	Embu	608,599	2,821	3989.511	297
Kisumu	1,155,574	2,085	3786.595	434	Nyamira	605,576	897	1,698,531	12
Kitui	1,136,187	30,430	13,313.44	263*	Vihiga	590,013	564	1042.503	153
Homa Bay	1,131,950	3,153	4241.071	245	Laikipia	518,580	9,532	4,635,294	326
Kajiado	1,117,840	21,871	9,076.21	585*	Marsabit	459,785	70,944	5654,896	389
Migori	1,116,436	2,614	3967.712	216	Elgeyo-Marakwet	454,480	3,032	2195.314	204
Murang'a	1,056,640	2,524	5480.144	763	Tharaka-Nithi	393,177	2,564	2706.617	108
Siaya	993,183	2,530	4271.545	367	Taita Taveta	340,671	17,152	6,126.31	336*
Trans Nzoia	990,341	2,495	2778.846	197	Tana River	315,943	37,951	4,422.06	365*
Makueni	987,653	8,170	11,999.16	387*	Samburu	310,327	21,065	2836.786	111
Turkana	926,976	68,233	6190.443	483	Isiolo	268,002	25,350	3033.414	34
Kericho	901,777	2,436	4299.328	388	Lamu	143,920	6,253	1,306.69	93*
Busia	893,681	1,696	3233.455	167					

Notes: \* represent only the classified roads according to KRB

**Figure 2. Flow diagram of methodology.**



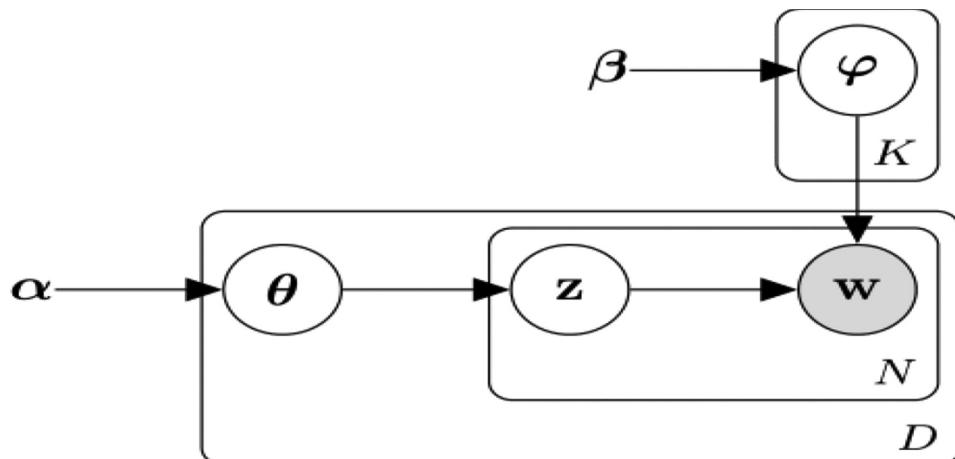
Probability distribution of topics is generated from the document which provides a way to explore the data on the perspective of topics.

In this paper, we applied Latent Dirichlet allocation (LDA) model for text mining. LDA was proposed by J. K. Pritchard, M. Stephens and P. Donnelly in 2000 as a clustering model (Tong & Zhang, 2016, May). LDA has been applied in NLP as an unsupervised machine learning algorithm to uncover categories in texts. In literature, it has been used to classify documents, formulate topic models from tweets, cluster web-data into relevant topics amongst other applications (B. Wang et al., 2017) (Kolini & Janczewski, 2017) (Lau et al., 2013, June). Clustering similar texts into meaningful categories (topics) is often more beneficial before an advanced analysis (Kong et al., 2016).

We used Matlab Text Analytics Toolbox™ to perform LDA modeling in our data (Mathworks, 2020). First, the document was passed through a pre-processing phase, followed by a bag-of-words (term-frequency counter) model phase before applying the clustering algorithm. Each entry of “accident brief” text is treated as our document that represents a combination of different topics. Matlab representation of the model is shown in Figure 3. LDA takes in a collection of  $D$  documents with a topic mixtures  $\theta_1, \dots, \theta_D$ , contained in  $K$  topics. Each topic is characterized by word probabilities  $\varphi_1, \dots, \varphi_K$ . The assumption made is that the topic mixtures and the words in the topics follow a Dirichlet distribution with concentration parameters  $\alpha$  and  $\beta$  respectively (Mathworks, 2020).

The generative process  $p(\theta, z, w|\alpha, \beta)$  of a document with words  $w_1, \dots, w_N$ , topic mixture  $\theta$ , and with topic indices  $z_1, \dots, z_N$  is given by

**Figure 3. LDA implementation:**  
source (Mathworks, 2020).



$$p(\theta, z, w|\alpha, \varphi) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta)p(w_n|z_n, \varphi) \quad (1)$$

Equation 1 above is further integrated to give the probability of marginal distribution  $p(w|\alpha, \varphi)$  of document  $w$  as shown in equation (2).

$$p(w|\alpha, \varphi) = \int_{\theta} p(\theta|\alpha) \prod_{n=1}^N \sum_{z_n} p(z_n|\theta)p(w_n|z_n, \varphi)p(w_n|z_n, \varphi)d\theta \quad (2)$$

### 3. Results

#### 3.1. Overview of available data

The available daily reports have been reducing every year as described in the bar chart of Figure 4. In 2016, over 200 entries were made compared to 2019 where less than 20 entries were made for fatal incidences and daily reports. This is one of the reasons as to why interpolation to get monthly estimate was necessitated.

Data in *fatal accident report* is as shown in Figure 5 based on county and time of day. From the figure, accidents are prevalent in rush hours (from 1600 hrs to 2100 hrs). A peak is also present at 0000hrs which is attributed to any entry with “unknown” time. In terms of county, Nairobi county has the highest incidences followed by Nakuru and Kiambu county.

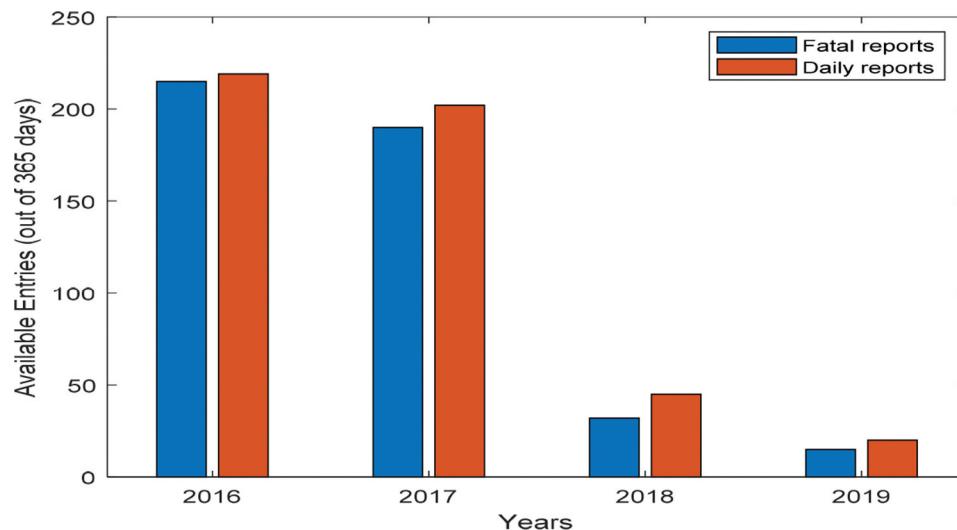
#### 3.2. Text mining results

The results of LDA with four selected topics is as shown in Wordcount clouds below. Wordcount shows the frequent words in bold colored letters and less frequent are faded out. From Figure 6, four topics are apparent; hit/run, head/collision, lost/control and victim/knocked/down. Model trained with four topics captured the categories accurately. A three-topic model joined hit/run with head/collision while six-topic model identified hit/run, and head/collision while the rest were spread through remaining categories.

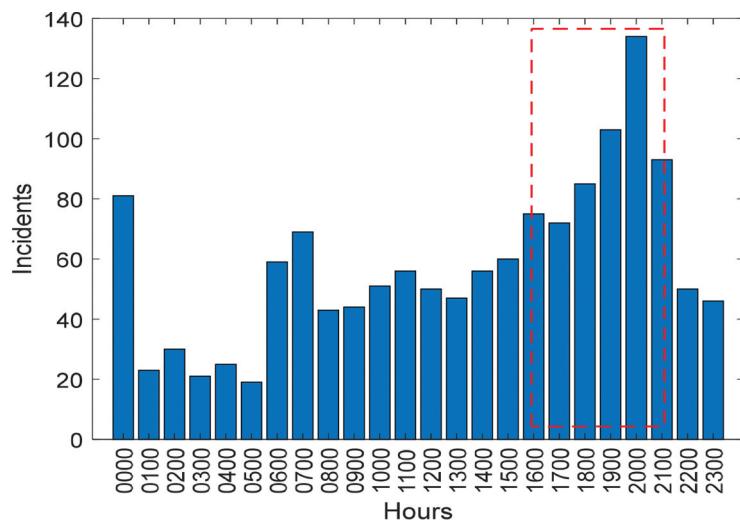
#### 3.3. Corpus topic probability

This is the probabilities of observing each of the topics in the entire data set used to fit the LDA model as shown in Figure 7. The displayed probabilities correspond to the probability of observing topic  $k$  in the data. Higher number of topics (five- and six-topic models) resulted in seemingly equal

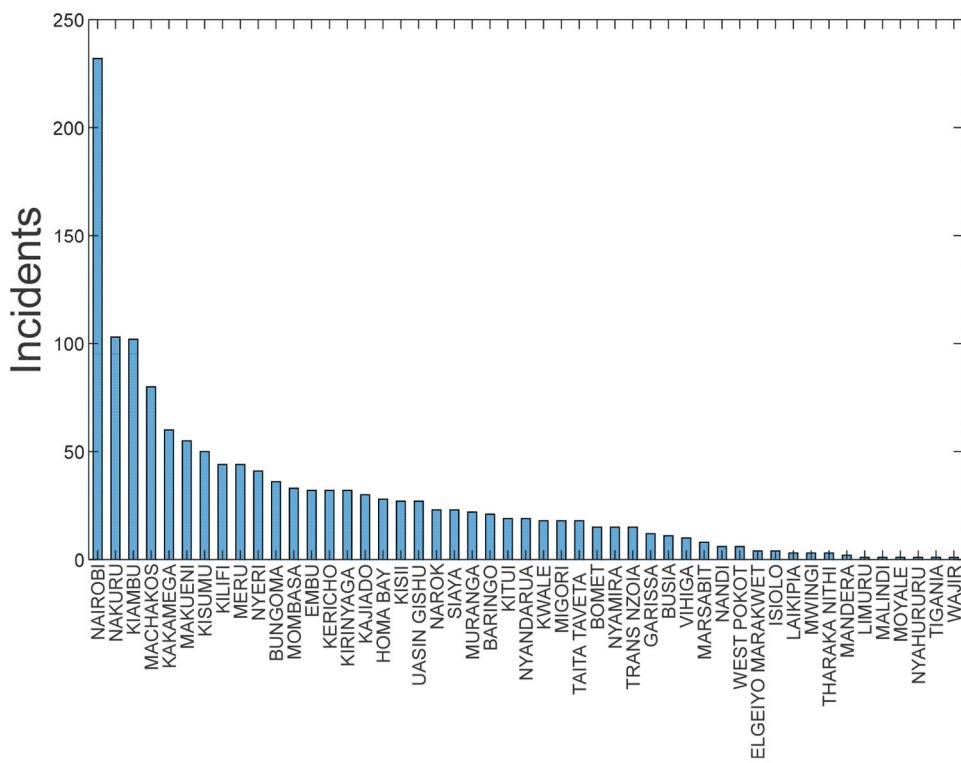
**Figure 4. Total available reports (data entries) per year.**



**Figure 5. Hourly and County distribution of fatal incidences.**



a) Hourly distribution of accidents with peak hours highlighted



b) County distribution of accidents ranked with prevalence

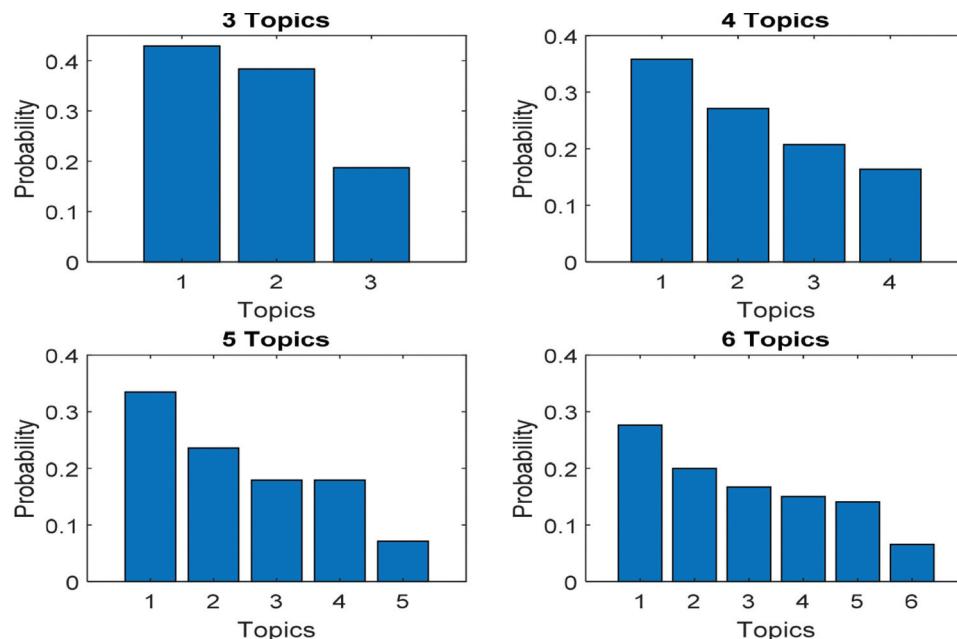
intertopic probabilities as well as cases of very low probabilities (0.05) in at least one of the topics. From this, we resulted in using four-topic model.

From the four-topic model, we performed an hourly distribution of fatal accidents for visualization as shown in Figure 8. The topics have been identified as knock/down, hit/run, lost/control, and

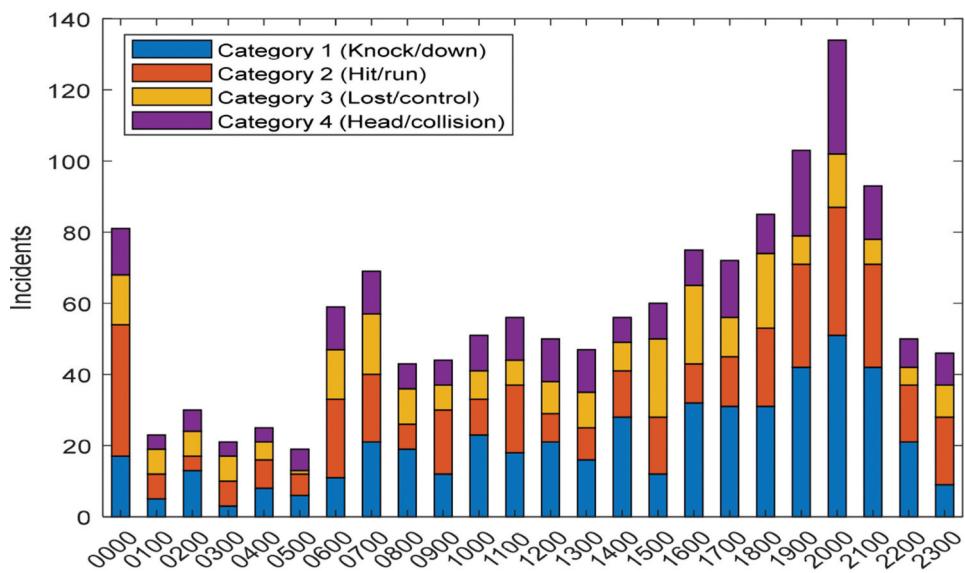
**Figure 6. Word cloud of topic models.**



**Figure 7. Corpus topic probability for different topic index.**



**Figure 8. Distribution of accident types with hour of day.**



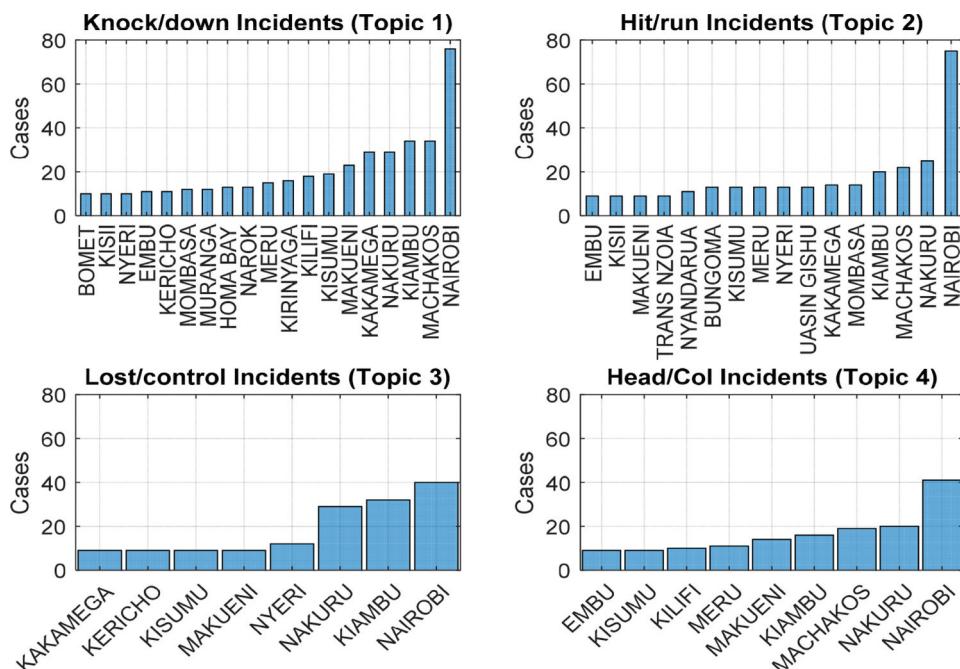
head/collision. Similar with Figure 5, incidences increase near/about rush hours; 1600 hrs up to 2100 hrs.

Similarly, we performed a topic-based ranking of counties to identify which places have prevalence on an accident type. We specified incidence number to greater than 10 to remove counties with lower reported incidences as shown in Figure 9. In all the categories, Nairobi, Nakuru, Kiambu, Machakos, and Kakamega have highest count with Nairobi being the severest of the five counties.

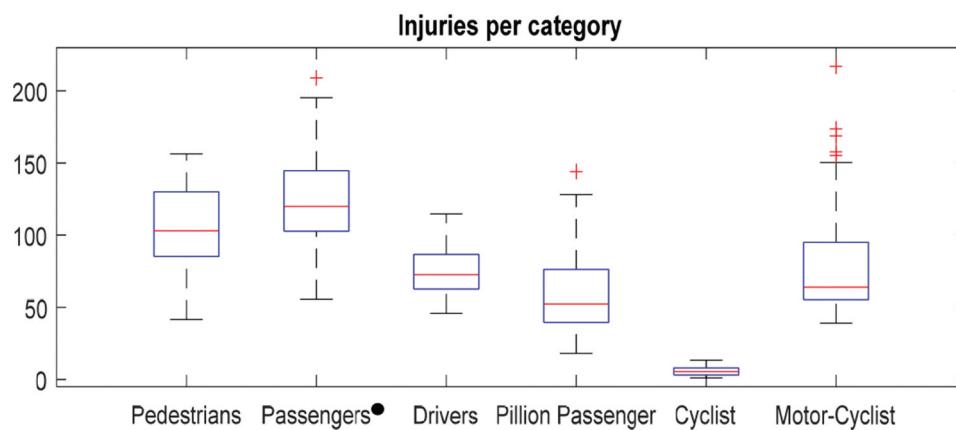
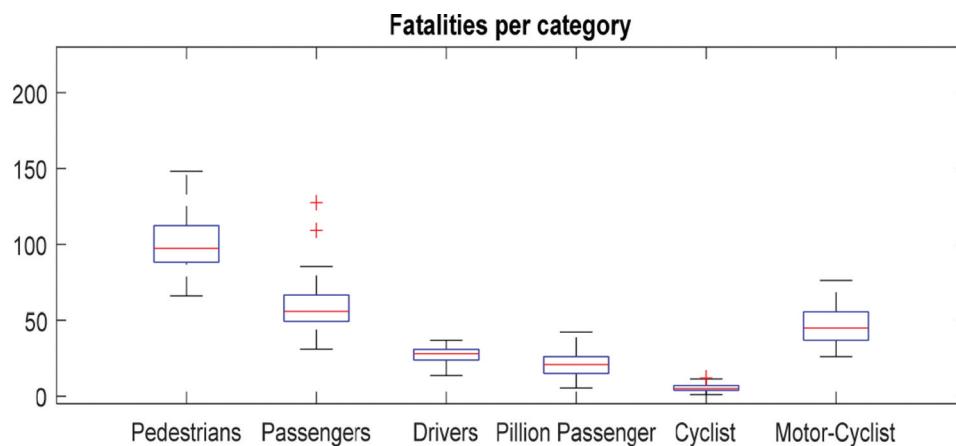
### 3.4. Daily reports trends analysis

We performed data analysis for the 5-year broken down into monthly entries. Figure 10 shows boxplot of six categories (Pedestrian, Passenger, Drivers, Pillion Passengers (Motorcycle passenger),

**Figure 9. County ranking per prevailing cause and accident type.**



**Figure 10. Boxplot of categories in daily reports.**



Cyclists, and Motorcyclist) for fatalities and injuries of the available data. In the figure, Injuries recorded in passenger category (shown with asterisk \*) has been factored by 0.25 to fit within the scope of other categories. The mean and standard deviation (std) of all categories is shown in Table 5.

**Table 5. Monthly statistics (mean and standard deviation)**

Category	Type	Average	std	Min	Max
Pedestrian	Dead	100.15	16.60	66	148.2
	Injured	103.29	26.34	41.56	156.3
Passenger	Dead	57.30	12.47	31.06	127.5
	Injured	490.27	129.83	222.3	836
Driver	Dead	27.36	4.84	13.59	36.8
	Injured	72.46	14.61	45.89	114.8
Pillion Passenger	Dead	20.23	6.84	5.33	42.3
	Injured	56.74	25.22	18	144.1
Pedal Cyclist	Dead	5.56	2.55	1	12
	Injured	5.77	3.17	1	13.3
Motor Cyclist	Dead	45.13	10.90	26	76.2
	Injured	77.10	33.10	39	217

### 3.5. Regression

With the data available, we sought to uncover any patterns or trends that would be meaningful as inference to forth coming years using linear regression estimates. Figure 11 shows the total incidences of fatality and injuries reported in the country. Fitted regression (polynomial of order 3) is shown in yellow line and a predicted value for 2020 December. Root mean square error is indicated for each plot.

Additional, vulnerable groups have been singled out to project the progress of fatal and injury incidences in the group. In all the three categories, incidences are increasing at an exponential rate as shown in Figure 12(a)–(f).

Finally, we sought to understand the relationship that exists between deaths and injuries from the reported data. Intuitively, an increase in injuries will lead to increase in overall fatalities. The plot in Figure 13 is derived from plotting monthly fatalities per injuries. The target is to visualize the correlation of injuries to deaths over the years. From the data, average fatality per death is 3.16, with a standard deviation (STD) of 0.6.

## 4. Discussion

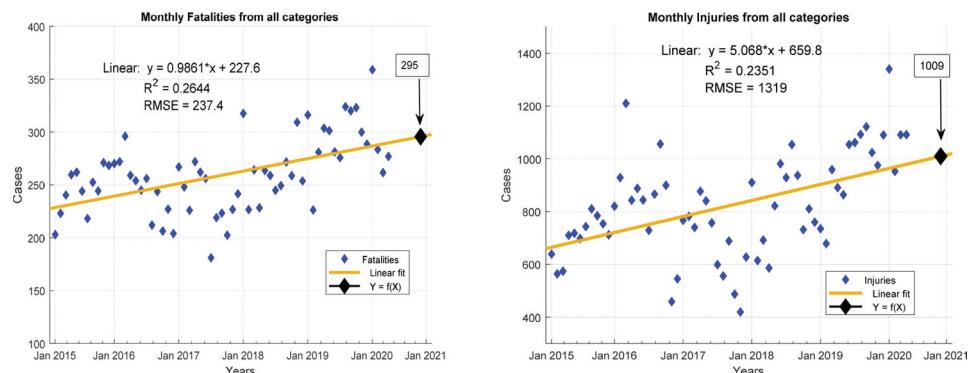
This paper utilized publicly available data from NTSA for the period between January 2015 and April 2020 to analyze traffic situation in the country. Broadly, the data is divided into fatal report and daily report. The target on the fatal report data was to analyze eyewitness type of reports (police reported) to come up with general cause of fatality using machine learning algorithm. Daily reports are analyzed for trends and other parameters that are informative.

### 4.1. Machine learning categorization

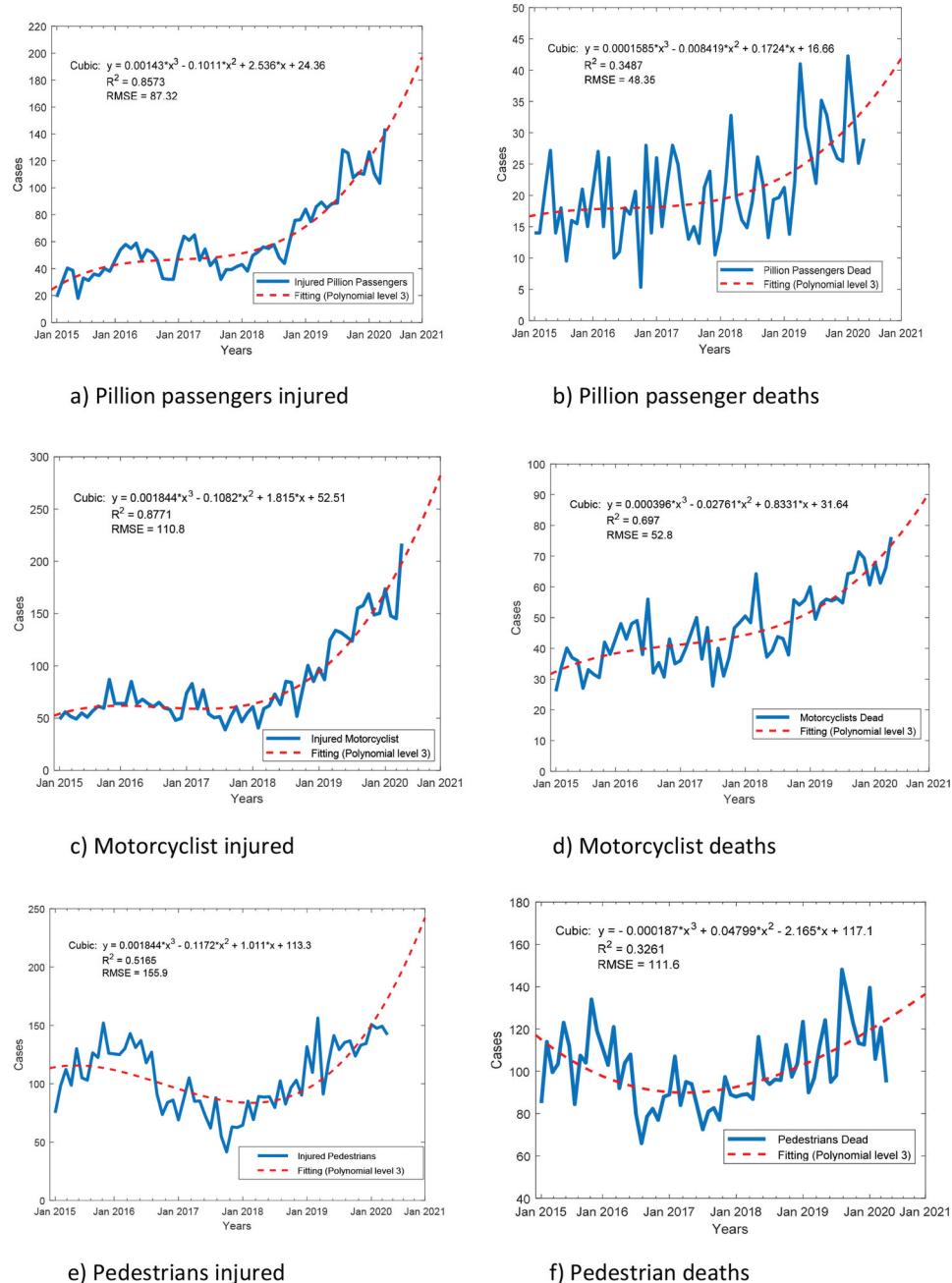
The paper identified four leading categorization of accident causes in the country as, knocking down victim (run over victim), hit-and-run, vehicle losing control and head on collision. Compared to Table 1, this categorization agrees with only two of NTSA's leading causes of crashes, i.e. hit and run and lost control with difference in the order of prevalence. From Figure 6(b) (working with the 4-topic model), the overall probabilities are given as 0.3534, 0.2773, 0.1803, and 0.1889 from topic 1 to 4, respectively. This translates to a 35.34% prevalence in running over victim followed by 27.73% for hit and run, 18.03% for lost control and 18.89% for head-on collision. From the categories, it is clear how the vulnerable groups (pedestrians motorcycle users) are most affected. In knocked down victims and hit and run categories, the target victims are the vulnerable road users. This is in agreement reported literature focusing on susceptible road-users (Kelly, 2018) (Mohammed et al., 2019).

The four identified categories are shedding more light on the general cause of fatalities. Arguably, the driver errors and/or negligence is significant in all the four categories. Particularly, the leading cause of fatality, knocked down victim, points to the modality of driving and road

**Figure 11. Trend of fatalities and injuries in 5 years showing projected estimate for December 2020.**



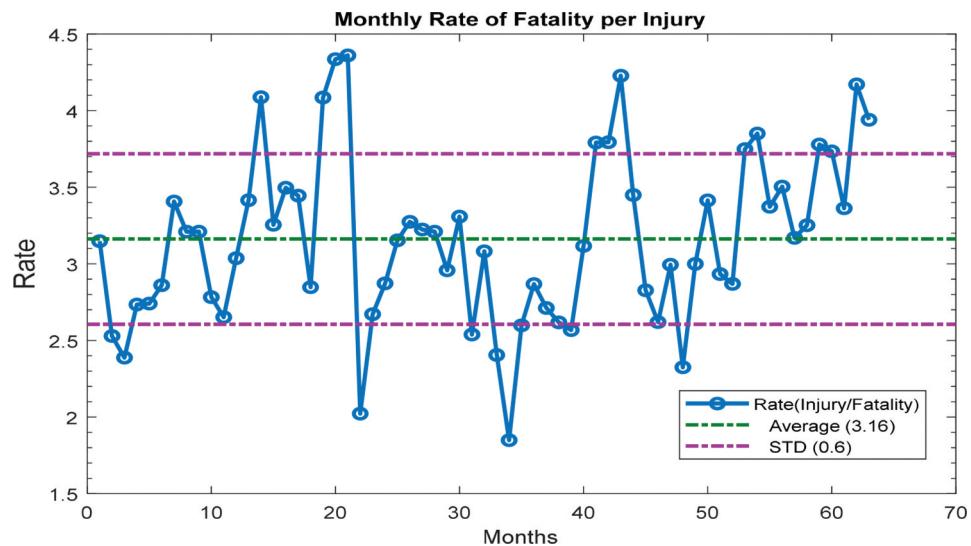
**Figure 12. Trend analysis of vulnerable road users (million passengers, motorcyclist, and pedestrians).**



safety standards observed by the country. Speeding, careless driving, drunk driving, and other detrimental driver behaviors can be linked with each of this category. The presence of police patrols has been proven to alter drivers behavior (Nakano et al., 2019). In order to combat the mitigate dangerous driving behavior, policies should explore ways of approaching policing as a supplement to current system of fixed checkpoints. Additional research is also needed to ascertain the effectiveness of existing policing activities.

A discrepancy was noted in the taries of *fatal report* and total reported deaths in *daily report*. As at 25 November 2016, total fatalities were at 2643 from daily report. In our analysis, we had a total of

**Figure 13.** Rate of fatality per recorded injury.



1392 instances of fatalities for the entire 5-year period. This led us to believe that the report does not capture all incidences but rather, a representative of the incidences that take place in the country.

Table 5 gives a descriptive statistic of available data showing monthly breakdown of the fatalities and injuries per category. Pedestrian fatality is the highest with an average of 100.15 persons per month with a standard deviation of 16.6. Passengers and motor cyclists follow closely in terms of fatality rate with 57.30 and 45.13, respectively. In terms of injury, passengers top the list with an average of 490.27 people.

County based ranking found Nairobi, Nakuru, Kiambu, Machakos and Kakamega to represent the top five accident-prone places. Compared to the demographics of these regions, population plays a major role in the statistics. Road network development has been pointed out by (Gichaga, 2017) and (Walcott-Bryant et al., 2016) to impact RTA in the country. From Table 4, only three out of the top five counties have a paved road greater than 1000 km<sup>2</sup>. Further studies should be performed to ascertain what other factors are in play in the face of accidents.

#### 4.2. Trend analysis

With the daily reported data, we sought to uncover any patterns or trends to make inference for coming years using linear regression. Figure 11 shows the monthly incidences of fatality and injuries reported in the country for the period under review. Both trends are positive indicating a worsening situation of traffic accidents. Between January 2015 and January 2020, the country reported an increase of 26.31% and 46.5% in fatalities and injured persons, respectively.

The exponential growth of injuries and fatalities in the vulnerable-people group shown in Figure 13 is another significant finding. The increase of motorization mentioned in the introduction section particularly in motorcycle (boda-boda) industry will continue pushing the numbers higher unless an intervention method is put in place. It is notable that injured motorcyclist and motorcycle passengers have seen a threefold increment over the past years.

Average and maximum monthly entry for entire period places the annual fatality between 3000 and 4500 deaths per year. Compared to the total population, this translates to fatality rate between 5.8 and 8.7 per 100,000 people. We have all reasons to question these numbers given that, the global fatality rate as at 2016 was 18.2 per 100,000 with African countries most severe. WHO and other

researchers have placed the estimates for most African countries at 26.6 and above per 100,000 population quotas. Authors in (Osoro et al., 2015), placed Kenyan fatality at 34 per 10,000 in 2015. Considering the increase of RTAs noted in the presented paper, the calculated fatality of below 8.7 per capita points to underreporting or incomplete data collection. Proper adjustment ought to be made to capture accurately the present situation of the country. This also affirms the conclusion made on the data being incomprehensive but a representative sample of the traffic accidents situation in the country. As an alternative to express the fatality per capita in the presented data, we sought to understand the correlation between reported fatalities in relation to incidences of injured people. From Fig. 16, the monthly data suggests that one person dies for every 3.2 injury on the road.

The cost implications for the accidents and financial burden it imposes on the country should be considered a threat to the country and addressed by stakeholders and policy makers. The country is said to be losing close to 170 million USD in traffic jams (Atieno, 2019). The cost of accidents and the damages on infrastructure has not been estimated yet but should be equal or greater than traffic jams.

### 5. Challenges and recommendations

This research faced a challenge with the available data and its scarcity as shown in Figure 6. The publicly available reports have been reducing per year. As highlighted in previous section, the available data are heavily underreported and does not capture the true position of road accidents. Records with reliable data cannot be overemphasized as it is essential in assessing the impacts in health, financial and social burden of traffic accidents. For proper risk assessment, development and evaluation of interventions, and general public safety sensitization, accurate data recording is indispensable. We recommend planned dissemination of data, e.g., per monthly or so, to enable researchers, policy-makers, stakeholders, etc. access the progress of the situation as well as develop prevention models.

The information availed on the data is another challenge that ought to be addressed. Currently, the reports record motor vehicle number, victim details and accident location and affected victim category. The challenge associated with the data is that, it is hard to make inferences on prevailing conditions or accident analysis.

We recommend inclusion of accident-relevant details to enable further accident analysis. Details like weather conditions, road intersection and type, drivers age, drunk/speeding behavior, vehicle category (type), etc. should be included for further analysis. This recommendation is informed by parameters used in analysis by researchers to make statistical inferences and modeling (Yousefzadeh-Chabok et al., 2016), (D. Wang et al., 2019), (Aldred et al., 2020). Further, the data reporting should be standardized to comply with international standards, e.g., following the recommended quantifying RTD as deaths within 30 days.

Additionally, fatal reports contain personally identifiable information (name, age, and gender) of the victim as well as details like motor vehicle registration number. This raises concerns about data protection policies employed. Readily available data for 2016 and 2017 has *Name of Victim* column as shown in Figure 3. The data starting from 2018 has no names but the report maintained other identifying details. Considering what the data represents (fatalities) socially, we recommend the removal of identifying information since the presence of such does not have statistical significance.

### 6. Conclusion

In conclusion, this paper has presented a 5-year look into present state of RTAs in Kenya from available data. The objective being to uncover trends and severity of RTA and steps being taken in the country to combat the menace. From literature review, we identified that for the last half decade, not much research activity has focused on RTA despite its proven importance and continued prevalence. In this paper, we employed machine learning algorithms for text mining to uncover general cause of fatality from accident description provided. From the results, four categories were identified as leading causes of fatality in the country: Knocking down victims, hit-and-run, vehicle losing control and head on collision. The overall probabilities for the causes are

given as 0.3534, 0.2773, 0.1803, and 0.1889, respectively. This translates to a 35.34% prevalence in running over victim followed by 27.73% for hit and run, 18.03% for lost control, and 18.89% for head-on collision. All this causes point to driver errors and are preventable with proper measures.

Monthly fatalities in the country have increased by 26% since January 2015 to January 2020 while injuries have increased by 46.5% over the same period. The trend is projected to increase with a positive gradient unless a remedial action is taken. The paper also identified three vulnerable road users: pedestrians, pillion passengers and motorcyclist (boda-boda). From the data, the injuries emanating from this group has seen a threefold increment compared to 2015 data.

From the results, RTA seemingly is prevalent in regions with high population and at part, by urbanization and increased motorization in the region. Nairobi, Nakuru, Kiambu, Machakos and Kakamega Counties were leading with highest reported fatal accidents. Further studies should be performed to ascertain factors in play in the face of accidents in specific counties.

A limitation identified by the findings is that the data is grossly underreported which may lead to misinformation when it comes to fatality rate for the said period. As an alternative to fatality per capita, we resulted in computing fatality rate by comparing the reported injuries with the deaths. From the reported data, a death occurred for every 3.2 injuries present in the roads.

The paper faced a challenge with scarcity of accident-related information from the available data. Availed data seemed to be victim-oriented, reporting personal information (name, gender, age, and car registration details) of the victim. The challenge associated with the data is that, it is hard to make inferences on prevailing conditions or accident analysis. Details like weather conditions, road intersection and type, drivers age, drunk/speeding behavior, vehicle category (type), etc. should be included for further modeling and analysis. We recommend accident-oriented information that might prove useful for further analysis.

#### Funding

The authors received no direct funding for this research.

passengers in the country ranging from buses to taxis

#### Author details

Joseph Kamau Muguro<sup>1,2</sup>

RTA/D:

Road Traffic Accident/Death

E-mail: [joseph.muguro@dkut.ac.ke](mailto:joseph.muguro@dkut.ac.ke)

WHO:

World Health Organization

Minoru Sasaki<sup>1</sup>

Zusha:

An awareness campaign that encourages passengers to speak up against reckless driving

E-mail: [sasaki@gifu-u.ac.jp](mailto:sasaki@gifu-u.ac.jp)

Kojo Matsushita<sup>1</sup>

E-mail: [waweru.njeri@dkut.ac.ke](mailto:waweru.njeri@dkut.ac.ke)

Waweru Njeri<sup>2</sup>

<sup>1</sup> Department of Mechanical Engineering, Gifu University, Yanagido 1-1, Gifu, Japan.

<sup>2</sup> School of Engineering, Dedan Kimathi University of Technology, Nyeri 657-10100, Kenya.

#### Abbreviation

Boda-boda: Local name for motorcycle (taxis) in Kenya

#### Declaration of interest statement

The authors declare no conflict of interest on the subject matter or materials discussed herein.

LDA: Latent Dirichlet allocation

#### Citation information

Matatu: Local name for Public service vehicle in Kenya

Cite this article as: Trend analysis and fatality causes in Kenyan roads: A review of road traffic accident data between 2015 and 2020, Joseph Kamau Muguro, Minoru Sasaki, Kojo Matsushita & Waweru Njeri, Cogent Engineering (2020), 7: 1797981.

NTSA: National Transport and Safety Agency, the body in charge of transport regulation in Kenya

#### References

PSV: Public service vehicles, all vehicles allowed to carry

Aldred, R., Garcia-Herrero, S., Anaya, E., Herrera, S., & Mariscal, M. Á. (2020). Cyclist injury severity in spain: A bayesian analysis of police road injury data focusing on involved vehicles and route environment. *International Journal of Environmental Research and*

- Public Health, 17(1). <https://doi.org/10.3390/ijerph17010096>
- Atieno, Barbara. (2019). Traffic Jam: Kenya Loses 18.25 billion annually. In Science Africa, Retrieved March 30, 2020, from. [Online]. <https://scienceafrica.co.ke/traffic-jam-kenya-loses-18-25-billion-annually/>
- B. Team. What laws? Full-blown matatu chaos returns. Business Daily, Nairobi, Retrieved October 09, 2019, from.
- Bachani, A. M., Hung, Y. W., Mogere, S., Akunga, D., Nyamari, J., & Hyder, A. A. (2017, March). Helmet wearing in Kenya: Prevalence, knowledge, attitude, practice and implications. *Public Health*, 144, S23-S31. <https://doi.org/10.1016/j.puhe.2016.12.005>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet allocation. *J. Mach. Learn. Res.*, 3(4-5), 993-1022. <https://doi.org/10.1016/b978-0-12-411519-4.00006-9>
- Bucsu házy, K., Matuchová, E., Zúvala, R., Moravcová, P., Kostíková, M., & Mikulec, R. (2020). Human factors contributing to the road traffic accident occurrence. *Transportation Research Procedia*, 45, 555-561. doi: [doi.10.1016/j.trpro.2020.03.057](https://doi.org/10.1016/j.trpro.2020.03.057)
- De Smet, W., & Moens, M-F.. (2009, November). Cross-language linking of news stories on the web using interlingual topic modelling. In Proceedings of the 2nd ACM workshop on Social web search and mining (pp. 57-64). doi:[doi.10.1145/1651437](https://doi.org/10.1145/1651437)
- Diaz Olvera, L., Plat, D., & Pochet, P. (2019, July). Looking for the obvious: Motorcycle taxi services in Sub-Saharan African cities. *Journal of Transport Geography*, 102476. <https://doi.org/10.1016/j.jtrangeo.2019.102476>
- ECMT (1980), Resolutions of the Council of Ministers of Transport and Reports Approved in 1979. OECD Publishing, Paris. doi:[https://doi.org/10.1787/ecmt\\_resolutions-1979-en](https://doi.org/10.1787/ecmt_resolutions-1979-en).
- Fraser, M. S., Wachira, B. W., Flaxman, A. D., Lee, A. Y., & Duber, H. C. (2020, March). "Impact of traffic, poverty and facility ownership on travel time to emergency care. K. Nairobi African Journal of Emergency Medicine, 10(1), 40-45. <https://doi.org/10.1016/j.afjem.2019.12.003>
- Gichaga, F. J. The impact of road improvements on road safety and related characteristics. *IATSS Research*, 40 (2), Elsevier B.V., 72-75. Retrieved March 01, 2017, from. <https://doi.org/10.1016/j.iatssr.2016.05.002>
- GiveWell. (2018). Georgetown University Initiative on Innovation, Development, and Evaluation — Zusha! Road Safety Campaign. [Online]. Retrieved March 30, 2020, from. <https://www.givewell.org/charities/zusha-road-safety-campaign>
- Hadj Hosseiniou, M., Mahdavi, A., & Alinezhad, M. (2018). Optimization of the allowable speed on Iran's freeways to reduce violations and accidents, using zero-truncated Poisson regression model. *Civil Engineering Journal*, 4(2), 443. <https://doi.org/10.28991/cej-0309104>
- Heyns, E., Uniyal, S., Dugundji, E., Tillema, F., & Huijboom, C. (2019). Predicting traffic phases from car sensor data using machine learning. *Procedia Computer Science*, 151(2018), 92-99. <https://doi.org/10.1016/j.procs.2019.04.016>
- Hordofa, G. G., Assegid, S., Girma, A., & Weldemariam, T. D. (2018, September). Prevalence of fatality and associated factors of road traffic accidents among victims reported to Burayu town police stations, between 2010 and 2015, Ethiopia. *Journal of Transport & Health*, 10, 186-193. <https://doi.org/10.1016/j.jth.2018.06.007>
- Islam, M. M., & Al Hadhrami, A. Y. S. (2012). Increased motorization and road traffic accidents in Oman. *Journal of Emerging Trends in Economics and Management*, 3(6), 907-914. <https://hdl.handle.net/10520/EJC130254>
- Karimzadeh, A., & Shoghli, O. (2020). Predictive analytics for roadway maintenance: A review of current models, challenges, and opportunities. *Civil Engineering Journal*, 6(3), 602-625. <https://doi.org/10.28991/cej-2020-03091495>
- Kelly, K. J. WHO: Kenya road deaths four times higher than NTSA reported. *Daily Nation*, Retrieved December 13, 2018, from. [Online]. <https://www.nation.co.ke/news/Kenya-road-deaths-grossly-underreported-WHO/1056-4893792-ve7d07z/index.html>
- Kipkemoi, W. Over 200 vehicles seized, 100 drivers arrested in Mombasa police crackdown," Standard Media, Retrieved January 07, 2020, from. [Online]. <https://www.standardmedia.co.ke/article/2001355618/over-200-vehicles-seized-100-drivers-arrested-in-police-crackdown>
- KNBS. (2018). Kenya National Bureau of Statistics (KNBS): Census report 2018. Retrieved from <https://www.knbs.or.ke/>
- Kolini, F. & Janczewski, L. (2017). Clustering and Topic Modelling: A New Approach for Analysis of National Cyber security Strategies. Pacific Asia Conference on Information Systems (PACIS- 2017) Proceedings. 126. <http://aisel.aisnet.org/pacis2017/126/>
- Kong, J., Scott, A., & Goerg, G. M. (2016). Improving topic clustering on search queries with word co-occurrence and bipartite graph co-clustering. Google, Inc. <http://static.googleusercontent.com/media/research.google.com/en/pubs/archive/45569.pdf>
- KRB. (2018). Kenya roads board (KRB): Road inventory and conditions report. Retrieved from <https://www.krb.go.ke/road-networks/>
- Lau, J. H., Cook, P., & Baldwin, T. (2013, June). unimelb: Topic modelling-based word sense induction for web snippet clustering. In Second Joint Conference on Lexical and Computational Semantics (\* SEM), Volume 2: Proceedings of the Seventh International Workshop on Semantic Evaluation (SemEval 2013) (pp. 217-221). <https://www.aclweb.org/anthology/S13-2039.pdf>
- Macharia, W. M., Njeru, E. K., Muli-Musiime, F., & Nantulya, V. (2009, June). Severe road traffic injuries in Kenya, quality of care and access. *African Health Sciences*, 9 (2), 118-124. <https://www.ajol.info/index.php/ahs/article/view/43772>
- Manyara, C. G. (2016). Combating road traffic accidents in Kenya: A challenge for an emerging economy. *Kenya After*, 50, 1, 101-122. doi:[https://doi.org/10.1057/9781137574633\\_7](https://doi.org/10.1057/9781137574633_7)
- Mathworks. (2020). A text analytics toolbox™ user's guide (R2020a). [Online]. Retrieved March 01, 2020, from. [https://au.mathworks.com/help/pdf\\_doc/textanalytics/textanalytics\\_ug.pdf](https://au.mathworks.com/help/pdf_doc/textanalytics/textanalytics_ug.pdf)
- Mogambi, H., & Nyakeri, F. (2015). Media priming of road traffic accidents in Kenya: Praxis, patterns, and issues. *SAGE Open*, 5(4). <https://doi.org/10.1177/2158244015606491>
- Mohammed, A. A., Ambak, K., Mosa, A. M., & Syamsunur, D. (2019). A review of the traffic accidents and related practices worldwide. *The Open Transportation Journal*, 13(1), 65-83. <https://doi.org/10.2174/18744478013010065>
- Murumba, S. *Motorcycle registrations spike on popularity of 'boda boda' transport*. Business Daily, December 26, 2017. [Online]. <https://www.businessdailyafrica.com/corporate/shipping/Motorcycle-registrations-spike-on-popularity/4003122-4241930-as340oz/index.html>

- Myers, J. G., Hunold, K. M., Ekernas, K., Wangara, A., Maingi, A., Mutiso, V., Dunlop, S., & Martin, Ian B K. (2017) Characteristics of the accident and emergency department of Kenyatta National Hospital, Nairobi, Kenya: A cross-sectional, prospective analysis. *BMJ Open*, 7 (10),e014974. BMJ Publishing Group. Retrieved October 01, 2017 doi: [doi:10.1136/bmjopen-2016-014974](https://doi.org/10.1136/bmjopen-2016-014974) 10 7 e014974
- Nakano, Y., Okamura, K., Kosuge, R., Kihira, M., & Fujita, G. (2019). Effect of visible presence of policing activities on drivers' vigilance and intention to refrain from non-driving activities: A scenario-based survey of general Japanese drivers. *Accident Analysis & Prevention*, 133, 105293. <https://doi.org/10.1016/j.aap.2019.105293>
- NTSA. (2020). *National Transport and Safety Authority (NTSA)*. [Online]. Retrieved March 01, 2020, from. [http://www.ntsa.go.ke/index.php?option=com\\_content&view=article&id=237](http://www.ntsa.go.ke/index.php?option=com_content&view=article&id=237)
- Osoro, A. A., Ng, Z., & Yitambe, A. (2015). An analysis of the incidence and causes of road traffic accident in Kisii, Central district, Kenya. *IOSR Journal of Pharmacy*, 5(9),2250–3013. <http://iosrphr.org/papers/v5i9/F0590041049.pdf>
- Pakgohar, A., & Kazemi, M. (2015). An examination of accident severity differences between male and female drivers, using logistic regression model. *Civil Engineering Journal*, 1(1), 31–36. <https://doi.org/10.28991/cej-2015-00000003>
- R. of K. (GoK). (2018). *Traffic Act (Act No. CAP. 403)*. *KenyaLaw.org*. [Online]. [http://kenyalaw.org:8181/exist/kenyalex/acitview.xql?actid=CAP.403#part\\_IX](http://kenyalaw.org:8181/exist/kenyalex/acitview.xql?actid=CAP.403#part_IX)
- Sapkota, D., Bista, B., & Adhikari, S. R. (2016). Economic costs associated with motorbike accidents in Kathmandu, Nepal. *Frontiers in Public Health*, 4(DEC), 1–8. <https://doi.org/10.3389/fpubh.2016.00273>
- Soehodho, S. (2017, January). Public transportation development and traffic accident prevention in Indonesia. *IATSS Research*, 40(2), 76–80. <https://doi.org/10.1016/j.iatssr.2016.05.001>
- Sun, L. L., Liu, D., Chen, T., & He, M. T. (2019, October). Road traffic safety: An analysis of the cross-effects of economic, road and population factors. *Chinese Journal of Traumatology*, 22(5), 290–295. <https://doi.org/10.1016/j.cjtee.2019.07.004>
- Sysoev, A., Khabibullina, E., Kadasev, D., & Voronin, N. (2020). Heterogeneous data aggregation schemes to determine traffic flow parameters in regional intelligent transportation systems. *Transportation Research Procedia*, 45 (01),507–513. doi:[doi:10.1016/j.trpro.2020.03.063](https://doi.org/10.1016/j.trpro.2020.03.063)
- Tendu, B. (2019). Alcoblown blown away from Kenyan roads. *Standard media*, Nairobi, 22 November.
- Tong, Z., & Zhang, H. (2016, May). A text mining research based on LDA topic modelling. In International Conference on Computer Science, Engineering and Information Technology (pp. 201–210). doi:[10.5121/csit.2016.60600](https://doi.org/10.5121/csit.2016.60600)
- Wa Mungai, M., & Samper, D. A. (2006). "No Mercy, No Remorse": Personal experience narratives about public passenger transportation in Nairobi, Kenya. *Africa Today*, 52(3), 51–81. <https://doi.org/10.1353/at.2006.0033>
- Walcott-Bryant, A., Tatsubori, M., Bryant, R. E., Oduor, E., Omondi, S., Osebe, S., Wamburu, J., & Bent, O. (2016). Harsh brakes at potholes in Nairobi: Context-based driver behavior in developing cities. *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, Rio de Janeiro, 2016. <https://doi.org/10.1109/ITSC.2016.7795626>
- Wang, B., Liakata, M., Zubiaq, A., Procter, R. (2017) A Hierarchical Topic Modelling Approach for Tweet Clustering. In: Ciampaglia, G., Mashhadi, A., Yasseri T. (eds) Social Informatics 2017, Lecture Notes in Computer Science, vol 10540. Springer, Cham. doi: [https://doi.org/10.1007/978-3-319-67256-4\\_30](https://doi.org/10.1007/978-3-319-67256-4_30)
- Wang, D., Liu, Q., Ma, L., Zhang, Y., & Cong, H. (2019, September). Road traffic accident severity analysis: A census-based study in China. *Journal of Safety Research*, 70, 135–147. <https://doi.org/10.1016/J.JSR.2019.06.002>
- World Health Organization. (2009). Road safety in ten countries: Kenya. [Online] [https://www.who.int/violence\\_injury\\_prevention/road\\_traffic/countrywork/rs10\\_kenya\\_en.pdf](https://www.who.int/violence_injury_prevention/road_traffic/countrywork/rs10_kenya_en.pdf)
- World Health Organization. (2015). Global status report on road safety 2015. [Online] [https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2015/en/](https://www.who.int/violence_injury_prevention/road_safety_status/2015/en/)
- World Health Organization. (2018). Global status report on road safety 2018. [Online]. [https://www.who.int/violence\\_injury\\_prevention/road\\_safety\\_status/2018/en/](https://www.who.int/violence_injury_prevention/road_safety_status/2018/en/)
- Yousefzadeh-Chabok, S., Ranjbar-Taklimie, F., Malekpour, R., & Razzaghi, A. (2016). A time series model for assessing the trend and forecasting the road traffic accident mortality. *Archives of Trauma Research*, 5 (3). <https://doi.org/10.5812/atr.36570>
- Zheng, Z., Wang, Z., Zhu, L., & Jiang, H. (2020, March). Determinants of the congestion caused by a traffic accident in urban road networks. *Accident Analysis & Prevention*, 136, 105327. <https://doi.org/10.1016/j.aap.2019.105327>
- Zusha. (2018). *The life you can save*. [Online]. Retrieved March 31, 2020, from. <https://www.thelifeyoucansave.org/best-charities/zusha/>



© 2020 The Author(s). This open access article is distributed under a Creative Commons Attribution (CC-BY) 4.0 license.

You are free to:

Share — copy and redistribute the material in any medium or format.

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

Under the following terms:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made.

You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions

You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.



**Cogent Engineering (ISSN: ) is published by Cogent OA, part of Taylor & Francis Group.**

**Publishing with Cogent OA ensures:**

- Immediate, universal access to your article on publication
- High visibility and discoverability via the Cogent OA website as well as Taylor & Francis Online
- Download and citation statistics for your article
- Rapid online publication
- Input from, and dialog with, expert editors and editorial boards
- Retention of full copyright of your article
- Guaranteed legacy preservation of your article
- Discounts and waivers for authors in developing regions

**Submit your manuscript to a Cogent OA journal at [www.CogentOA.com](http://www.CogentOA.com)**

