University of Nairobi

School of Mathematics

ISSN: 2410-1397

Master Project in Social Statistics

Modelling Fatalities Associated with Road Traffic Accidents and other Factors Using Poisson Models.

Research Report in Mathematics, Number 35, 2019

Kimani Kevin Macharia

June 2019

Submitted to the Sool of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Master Project in Social Statistics **University of Nairobi**
June 2019

Modelling Fatalities Associated with Road Traffic Accidents and other Factors Using Poisson Models.

Research Report in Mathematics, Number 35, 2019

Kimani Kevin Macharia

School of Mathematics
College of Biological and Physical sciences
Chiromo, off Riverside Drive
30197-00100 Nairobi, Kenya

Master Thesis

Submitted to the Sool of Mathematics in partial fulfilment for a degree in Master of Science in Social Statistics

Submitted to: The Graduate School, University of Nairobi, Kenya

Abstract

According to the World Health Organization (WHO, 2018), 1.35 million people died in 2016 on the world's roads, with road traffic injuries being the leading cause of death for children and young adults aged $5-29$ years.

Road traffic injuries is ranked 8th in the leading causes of death for all ages in 2016.It surpasses those dying from HIV/AIDS, tuberculosis and diarrhoeal diseases although the political commitment and financial investments in road safety is only a fraction of that made to combat these diseases (WHO, 2018).

The objectives of the study were to determine factors associated with fatal Road Traffic Accidents (RTAs) in Kenya and to determine which of the poisson model fit the RTA fatalities count data better.

Master Thesis in Mathematics at the University of Nairobi, Kenya.
ISSN 2410-1397: Research Report in Mathematics

DISTRIBUTOR: School of Mathematics, University of Nairobi, Kenya

Declaration and Approval

I the undersigned declare that this dissertation is my original work and to the best of my knowledge, it has not been submitted in support of an award of a degree in any other university or institution of learning.

Signature Date

Kimani Kevin Macharia
Reg No. I56/87243/2016

In my capacity as a supervisor of the candidate's dissertation, I certify that this dissertation has my approval for submission.

Signature Date

Dr Timothy Kamanu
School of Mathematics,
University of Nairobi,
Box 30197, 00100 Nairobi, Kenya.
E-mail: tkamanu@uonbi.ac.ke

Dedication

This project is dedicated to my parents Mr. and Mrs. Kimani, my siblings Sammy, Patricia and Melanie, my friends Mary, Carlos, Sylvester, Collins and Kevin, my extended family and to all who inspired me. Most importantly I dedicate this project to the Almighty God for always guiding and strengthening me.

Contents

Abstract .... **ii**

Declaration and Approval .... **iv**

Dedication .... **vii**

Figures and Tables .... **x**

List of Abbreviations .... **xi**

Acknowledgments .... **xii**

1 Introduction **1**

1.1 Background of the Study 1
1.2 Problem Statement 1
1.3 Objectives 2
1.4 Significance of Study 3

2 Literature **4**

2.1 Introduction 4
2.2 Road Users 5
2.2.1 Drivers 5
2.2.2 Pedestrians, Cyclists and MotorCyclist 5
2.3 Literature Relevant to this Thesis 6

3 Methodology **8**

3.1 Introduction 8
3.2 Count Data 8
3.3 Poisson Distribution 8
3.4 Zero Truncated Poisson Distribution 11
3.5 Negative Binomial Distribution 12
3.6 Model Specification 14
3.7 Parameter Estimation 14
3.8 Goodness of Fit 14
3.8.1 Akaike information criterion 14
3.8.2 Bayesian information criterion 15

4 Analysis **16**

4.1 Data Collection 16
4.2 Data Cleaning 16
4.3 Statistical Methods 17
4.4 Data Exploration 17
4.5 Case Study: Nairobi Mombasa Highway 22

4.5.1 Road Dynamics 23
4.5.2 Descriptive Statistics 23
4.6 Model Exploration along Nairobi Mombasa Highway 28
4.6.1 Poisson Model Fit 28
4.6.2 Zero Truncated Poisson Model Fit 29
4.6.3 Negative Binomial Model Fit 30
4.7 Model Insights 31
4.7.1 Day of fatal RTA 32
4.7.2 Victims of fatal RTA 32
4.7.3 Vehicle in fatal RTA 32
4.7.4 County fatal RTA occurred 32

5 Conclusion 33

5.1 Future Research 33
5.2 Limitations of the Study 33
.1 Appendix 34

Appendix: Bibliography 36

Figures and Tables

Figures

Figure 1. Rates of RTA fatalities per 100,000 by WHO: 2013, 2016 2
Figure 2. Fatal RTA's distributed by Day 18
Figure 3. Fatal RTA's distributed by Yearly Quarter 19
Figure 4. Fatalities distributed by Day 20
Figure 5. Fatalities distributed by Quarter 21
Figure 6. Fatalities distributed by Roads 22
Figure 7. Source: Google Maps Nairobi Mombasa Highway 23
Figure 8. Fatalities distributed by County along Nairobi Mombasa Highway 24
Figure 9. Fatalities distributed by Victim along Nairobi Mombasa Highway 25
Figure 10. Fatalities distributed by Vehicle along Nairobi Mombasa Highway 26
Figure 11. Fatalities distributed by Day along Nairobi Mombasa Highway 27
Figure 12. Fatal RTA's distributed by Police Stations/Bases 34
Figure 13. Fatalities distributed by Police Stations/Bases 35

Tables

Table 1. Road fatalities Distributed by Years 17
Table 2. Fatal RTA's Distributed by Years 18
Table 3. Nairobi Mombasa highway fatalities Distributed by Years 24
Table 4. Poisson Model fit for Nairobi Mombasa highway fatalities 29
Table 5. Zero Truncated Poisson Model fit for Nairobi Mombasa highway fatalities 30
Table 6. Negative Binomial Model fit for Nairobi Mombasa highway fatalities 31

List of Abbreviations

- *AIC* - Akaike Information Criterion
- *AMB* - Ambulance
- *BIC* - Bayesian Information Criterion
- *COM* - Commercial Vehicle
- *CPM* - Crash Prediction Models
- *DRV* - Driver
- *GOV* - Government Vehicle
- *HIV/AIDS* - Human Immunodeficiency Virus, Acquired Immunodeficiency Syndrome
- *MCYC* - MotorCyclist
- *MRP* - Minor Roads Programmes
- *NB* - Negative Binomial Model
- *NTSA* - National Transport and Safety Authority of Kenya
- *PAS* - Passenger
- *PCY* - Pedal Cyclist
- *PED* - Pedestrian
- *PIL* - Pillow Passenger
- *PRV* - Private Vehicle
- *PSV* - Passenger Service Vehicle
- *RAR* - Rural Access Roads
- *RTA* - Road Traffic Accident
- *ZTNB* - Zero Truncated Negative Binomial
- *ZTP* - Zero Truncated Poisson Model
- *WHO* - World Health Organisation

Acknowledgments

I would first like to express my thankfulness to Almighty God for bringing me to a successful completion of my masters studies. I will also like to express my deep appreciation for my supervisor Dr. Timothy Kamanu. No volume of words is enough to express my gratitude for his time, commitment, mentorship, and guidance throughout all stages of this thesis. My deepest gratitude goes to my family and friends, for their love, support and patience as I spent time while pursuing my studies.

Kimani Kevin Macharia

Nairobi, 2019.

1 Introduction

1.1 Background of the Study

Organisation for Economic Co-operation and Development(OECD, 2018) defines a Road Traffic Accident as an accident which occurred or originated on a way or street open to public traffic; resulted in one or more persons being killed or injured, and at least one moving vehicle was involved. These accidents therefore include collisions between vehicles, between vehicles and pedestrians and between vehicles and animals or fixed obstacles. Single vehicle accidents in which one vehicle alone (and no other road user) was involved are included. Multi-vehicle collisions are counted only as one accident provided that the successive collisions happened at very short intervals. Any person who was not killed but sustained one or more serious or slight injuries as a result of the accident is defined as injured. Serious injuries include fractures, concussions, internal lesions, crushing, severe cuts and laceration, severe general shock requiring medical treatment and any other serious lesions entailing detention in hospital is defined as sever. Slight injuries include secondary injuries such as sprains or bruises. Persons complaining of shock, but who have not sustained other injuries, should not be considered in the statistics as having been injured unless they show very clear symptoms of shock and have received medical treatment or appeared to require medical attention (WHO, 2018).

Road traffic accidents now represent the eighth leading cause of death globally for all age groups surpassing HIV/AIDS, tuberculosis and diarrhoeal diseases. They claim more than 1.35 million lives each year and cause up to 50 million injuries. The burden of road traffic injuries and deaths is disproportionately borne by vulnerable road users and those living in low- and middle-income countries, where the growing number of deaths is fuelled by transport that is increasingly motorized. Between 2013 and 2016, no reductions in the number of road traffic deaths were observed in any low-income country, while some reductions were observed in 48 middle- and high-income countries. Overall, the number of deaths from RTAs in low-income countries is 3 times higher than that in high-income countries (WHO, 2018).

1.2 Problem Statement

There are several factors that contribute to the occurrence of a fatal RTA which are but not limited to behavioral and non-behavioral factors. Non-behavioural factors such as traffic flow characteristics, road geometry and environmental conditions.(Elani, 2000) classified behavioral factors as

- Reduced capability on a long-term basis (alcoholism,disease and disability, inexperience, drug abuse, aging, alcoholism),
- Reduced capability on a short-term basis (acute psychological stress, drowsiness, acute alcohol intoxication, binge eating, short term drug effects, temporary distraction, fatigue)
- Risk taking behavior with long-term impact (macho attitude, habitual speeding, indecent driving behavior, inappropriate sitting while driving, overestimation of capabilities, accident proneness, habitual disregard of traffic regulations, non-use of seat belt or helmet)
- Risk taking behavior with short-term impact (compulsive acts, psychotropic drugs, suicidal behavior, motor vehicle crime, moderate ethanol intake)

In 2016 Kenya had a death-rate of 27.8 per 100,000 from fatal RTAs (WHO, 2018). This was an improvement from a death-rate of 29.1 per 100,000 in 2013 (WHO, 2015).However, this is higher than the average african death-rate of 26.6 deaths per 100,000 in 2016 and 26.1 deaths per 100, 000 population in 2013 (WHO, 2018). This is already higher than all other regions with South-East Asia folllowing with a death-rate of 20.7 per 100,000 in 2016 as shown.

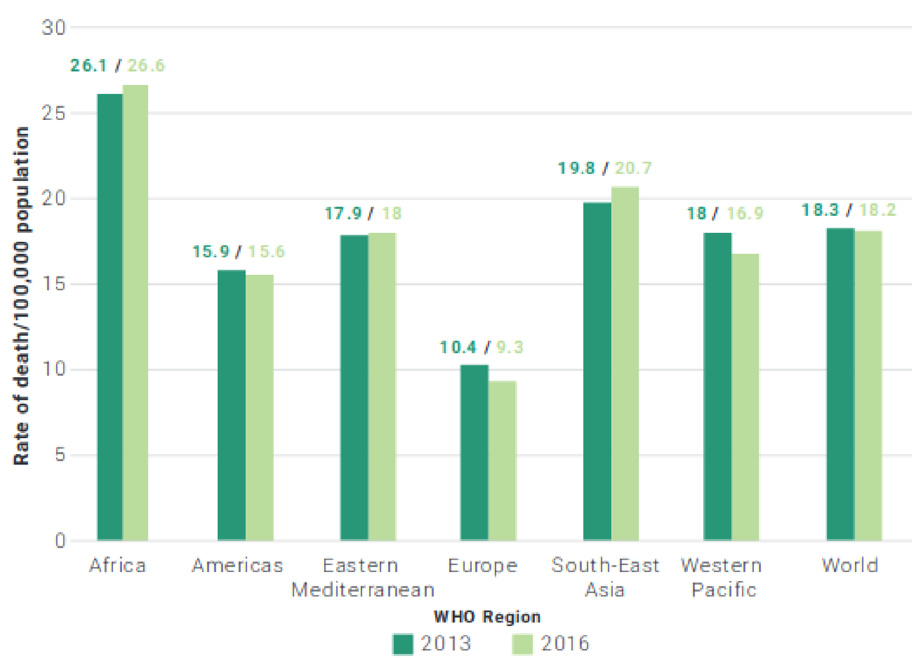

Figure 1. Rates of RTA fatalities per 100,000 by WHO: 2013, 2016

1.3 Objectives

The overall objective:

- To apply Poisson models to identify factors associated with death from RTAs.

The following are the specific objectives:

- To identify characteristics associated with road traffic safety, defined by the occurrence of fatal RTAs.
- To determine which poisson models fits better on fatal RTAs.

1.4 Significance of Study

Despite the high RTA death-rate in Kenya and a growing body of literature on factors (behavioral and non-behavioral) that are associated with RTAs on highways in other countries, to our knowledge, there have been no appropriate modelling techniques employed to estimate the incidence rate of death from RTAs and identification of factors associated with fatal RTAs in Kenya.

2 Literature

2.1 Introduction

From Kenya's independence from Britain in 1963 a lot of emphasis and resources have been used to to both upgrade and expand the road networks in order to expand the coutry's economic developments and social mobility. Road development have taken the form of reconstruction, recarpetting and road expansion. This is evidenced in the emphasis of a more connected road network which makes mobility of people, goods and services faster and more efficient which directly translates to economic progress. According to the Roads 2000 Strategic Plan (2013-2017) by the Ministry of Roads in Kenya the Rural Access Roads(RAR) and Minor Roads programmes(MRP) constructed and maintained 12,000 km of rural roads in the 70's,80's and early 90's. In the late 80's and 90's emphasis changed from construction to maintainace as it became evident that Kenya was unable to maintain her roads adequately. The roads 2000 strategy now seeks to use an optimum mix of local labour and equipment to both construct and maintain Kenyan roads.

With the transformation of roads and government policy in Kenya there is a rapid increase in vehicle population(2-wheeler and 4-wheeler) and human population. Government policy has been skewed in order to enable imports of vehicles and more specifically 2-wheeler motor-cycles with the aim of increasing employment anmong the youth. Kenya in 2007 exempted motorcycles below 250cc from a 16% value added tax to spur job creation.This led to a spike in the number of motorcycle imports and the government imposed excise duty in December 2015 so as to capitalise on the number of imports. However, this was lifted in September 2016 following lobbying from dealers and manufacturers.

Due to the increase of vehicles and better roads in Kenya inevitably came the increased road traffic accidents and resulting fatalities, injuries and property damage.Agoki (1988) noted that RTA's remained one of the biggest unsolved problems in Kenya specifically and also in other countries. It is a significand engineering and public health issue as is represents serious national losses and also loss to individual Kenyans in terms of loss of life, man-hours, injuries and the consequent effect on overall life efficiency.

The characteristics, causes and interplay of factors that lead to RTA's are least understood in many circles. Researchers have constantly sought ways to improve traffic safety and gain better understanding to better predict crash likelihood under different crash contributing factors.

A number of accident prediction models have been developed in the last two decades to estimate the expected accident frequencies on roads as well as to identify various factors associated with the occurrence of accidents.

2.2 Road Users

2.2.1 Drivers

Norman(1972) noted that Drivers, like other road users are recepients and causers of road traffic accidents and fatalities. Agoki(1988) noted that the driver's part in Road traffic accidents are a function of the adequacy of his response to changes in his road environment. There are a myriad of driver characteristics that have been studied and noticed that they can lead to fatal RTA's.

Smart (2002) noted that driver's attitude including but not limited to road courtesy and behaviour, male sex, drivers driving under the influence of alcohol, use of seat belts, driver age are among the recognised driver factors that lead to fatal RTA's.

2.2.2 Pedestrians, Cyclists and MotorCyclist

According to WHO(2015) Almost hald of all deaths that occur because of fatal RTA's are among those with the least protection - pedestrians, cyclists and motorcyclists. WHO also notes that the likelihood of dying on the road as a pedestrian, cyclist or motorcyclist varies by region and that the African Region has the highest proportion of pedestrians and cyclist deaths at 43% of all road traffic fatatilities.

According to Agoki(1988) pedestrians carry the responsibility of their own safety in RTA's.

Motorcycle fatalities can be broadyly classified into two. This involves the motorcycle driver and the pillow passanger. In developing countries like Kenya motorcycle fatalities have been on the rise due to lack of government regulations around motorcycle fatalities and there not bing a limit on pillow passangers.

The number of new motorcycle imports has increased by 150% from 125, 058 in 2013 to 186, 434 in 2017. For all new Motor Vehicle imports into Kenya in 2013 motorcycles represented 56% of all imports in 2013 compared to 66% of all imports in 2017 (NTSA, 2017). This shows that the number of motorcycles are increasing in Kenyan roads.

One of the many reasons why motorcycles constitute a huge percentage of motor vehicle imports is because they are used as used as taxis (Boda-bodas). In rural areas motorcycles are growing in popularity because of poor road networks and they are cheap to run and maintain. In urban areas motorcycles offer a easy way to escape the traffic jams experienced in these areas.

At Naivasha hospital in Kenya 36% of patients who presented to the emergency department because of a RTA's were motorcyclists.Compared to car occupants, motorcycle riders and their passengers are relatively unprotected (WHO etal, 2011). The likelihood of serious injury or death faced by motorcyclists is therefore higher than for other groups of users of motorized transport. Head injuries are a major cause of death, injury and disability among motorcyclists.

The substantial growth in the use of motorized two-wheelers, particularly in low-income and middle-income countries, is being accompanied by an increase in the number of fatalities, head and traumatic brain injuries (WHO etal, 2011).

2.3 Literature Relevant to this Thesis

Mathematical modeling is a vital tool in analyzing accidents and associated fatalities. Crash prediction models (CPMs) are very useful tools in traffic safety, with their capabilities to determine the relationship between frequency and/or severity of crashes and crash contributing factors (Hana, 2018). They are also integral in providing insight into safety levels of the roads as it helps detect unsafe road characteristics by relating crash fatalities and/or counts with many different independent variables ise road geometric characteristics, traffic volumes, segemet lengths, weather conditions and so on (Md Saidul, 2014).

Regression models have been most commonly used to relate accident frequency with explanatory variables. The result of model strongly relies on the choice of regression technique. The earlier traffic accident studies used ordinary or normal linear regression models, which follow the assumption of a normal distribution for the dependent variable, aconstant variance for the residuals, and the linear relationship existing between dependent and independent variables (Pan and Prakash, 2013).

Persaud and Dzbik (1993) modelled the relationship between freeway crash data and road traffic volumes. Using the hourly traffic, the model indicated that higher accident risk is associated with congestion and afternoon rush hour. However, they also noted that it is not possible for regression models to account for each and every factor that affects accident occurrence.

Conventional linear regression method should be used with caution because of the problems associated with non-negative and error terms (Jovanis and Chang, 1986; Abdel-Aty and Radwan, 2000).

Jovanis and Chang (1986) recommended generalized linear model using Poisson distribution error structure as a mean to describe the random, discrete and non-negative accidents. Poisson regression assumes an exponential relationship between response and explanatory variables (Eenink et al.,2007).

Abdel-Aty etal (2000) attempted to use the Poisson regression methodology and then rejected it because of different mean and variance value of the dependent variable, which indicated over dispersion in the accident data. Consequently, Poisson-gamma (or called negative binomial) model was adopted as a superior alternative to accommodate the over dispersion. The negative binomial model has been widely employed in vehicle accident analysis for rural highways, arterial roadways, urban motorways, and rural motorways (Lord, 2005; Montella, 2008).

When considerable zeros and extremely low mean value are observed in accident numbers, negative binomial model is significantly unreliable to fit the data, and the dispersion parameter can be mis-estimated.

To overcome the difficulties arising with zero accident samples, some researchers used extended Poisson and negative binomial models which can account for excessive zeros, for example, zero-inflated negative binomial (Shankar et al., 1997; Lord et al., 2005) and zero-truncated negative binomial (Chowdhury etal., 2016).

Zero-inflated probability models deal with the dual-state system: the zero-accident scenario in which no accidents is ever be observed, and the non-zero accident scenario where accident frequencies observe some known distributions ie the Poisson or negative binomial distribution.

Thus, fitting zero-inflated models to account for excess zeroes normally arises from wanting to find better fitting models which is justified; unfortunately, however, the zero-inflated model comes with "excess theoretical baggage" that doesn't have any theoretical appeal (Lord et al., 2005). To overcome this zero truncated negative binomial is used for count data. Zero truncated means the response variable doesn't have a value of 0.

3 Methodology

3.1 Introduction

This chapter is mainly concerned with formulating the models that will be used to model the accident count data in Kenya. Understanding the model formulation, parameters and validation.

3.2 Count Data

Count Data analysis comes into play when the observed outcomes are count and the desire is to estimate the covariate effects on outcomes.

It is expected that the observed outcomes on the same subject are be correlated. This type of data arises in many fields, for example, traffic accidents, health sciences, economics, social sciences, environmental studies among others. A typical example of such dependence arises in the number of traffic accidents and the number of injuries or fatalities during a specified period. However, in some situations outcomes may be truncated as zero values of counts may not be observed or may be missing for one or both of the outcomes.For example, in a sample drawn from hospital admission records, frequencies of zero accidents and length of stay are not available.

Another example is the case where the data on number of traffic accidents and related injuries or fatalities and related risk factors are collected from records and, naturally, zero counts are not available. Only those accidents that involve personal injury reported to the police using the accident reporting form are recorded. Damage-only accidents, with no human casualties or accidents on private roads or car parks, are not included generating zero-truncated count data.

3.3 Poisson Distribution

The poisson distribution is a special case of the binomial distribution where n-(number of trials) is large and the probability is small.

The binomial distribution has a fixed number of trials (n), each with a probality of success of p and failure q.

As $n \to \infty$ and $p \to 0$ the rate of success $\lambda = np$

From the binomial distribution:

$$B(n,p) = P(X=k) = \binom{n}{k} p^k (1-p)^{n-k} \tag{1}$$

Since we defined $\lambda = np$ it can be deducted that $p = \frac{\lambda}{n}$

Take the binomila llimit as n approaches inifinity

$$lim_{n\to\infty} P(X=k) = lim_{n\to\infty} \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \tag{2}$$

Exploring the Equation(2) above

$$\left(\frac{\lambda^k}{k!}\right) lim_{n\to\infty} \frac{n!}{(n-k)!} \left(\frac{1}{n^k}\right) \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k} \tag{3}$$

Further expounding the equation on $lim_{n\to\infty} \frac{n!}{(n-k)!} \left(\frac{1}{n^k}\right)$

$$lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-k)(n-k-1)...(1)}{(n-k)(n-k-1)...(1)} \left(\frac{1}{n^k}\right) \tag{4}$$

Canceling out the common denominators in both the numerator and the denominator the equation remains as follows

$$lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-k+1)}{n^k} \tag{5}$$

The n-k terms were canceled out in the numberator leaving k terms in both the numerator and the denominator.

The Equation (5) above can be expanded as follows

$$lim_{n\to\infty} \left(\frac{n}{n}\right) \left(\frac{n-1}{n}\right) \left(\frac{n-2}{n}\right) \cdots \left(\frac{n-k+1}{n}\right) \tag{6}$$

Since this has k terms and each one approaches 1 as k approaches infinity this portion of the equation simplifies to one.

Expanding on the middle term of the Equation(3) $lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n}$

By definition e=2.718...and can be expressed as follows

$$e = lim_{x\to\infty}\left(1+\frac{1}{x}\right)^{x} \tag{7}$$

Defining x as $x = -\frac{n}{\lambda}$ and replacing in Equation (7)

$$lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{n} = lim_{x\to\infty}\left(1+\frac{1}{x}\right)^{x(-\lambda)} = e^{-\lambda} \tag{8}$$

Expanding on the last term of the Equation(3) $lim_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k}$

As $n \to \infty$ this becomes 1^{-k} which is equal to one.

Putting the three distinct terms together to form the Equation(3)

$$\left(\frac{\lambda^k}{k!}\right) lim_{n\to\infty}\frac{n!}{(n-k)!}\left(\frac{1}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{n}\left(1-\frac{\lambda}{n}\right)^{-k} = \left(\frac{\lambda^k}{k!}\right)(1)(e^{-\lambda})(1) \tag{9}$$

This simplifies to the following

$$P(\lambda,k) = \left(\frac{\lambda^k e^{-\lambda}}{k!}\right) \tag{10}$$

Using the notations used by (Chowdhury et al., 2015). Let Y_1 be the count of accidents at a specific location in a given interval that has a Poisson distribution with the following poisson mass function .

$$g_1(y_1) = P(Y_1 = y_1) = \frac{e^{-\lambda_1}\lambda_1^{y_1}}{y_1!}, y_1 = 0,1,\cdots \tag{11}$$

Let Y_{2i} be a r.v with number of deaths resultiing from i^{th} RTA, and it has a poisson distribution with parameter, λ_2

$$g_2(y_{2k}) = \frac{e^{-\lambda_2}\lambda_2^{y_{2k}}}{y_{2k}!}, y_{2k} = 0, 1, \cdots \tag{12}$$

and the link function is

$$ln\,\lambda_1 = x^{'}\beta_1, where\, x^{'} = (1, x_1, \cdots, x_p), \beta_1^{'} = (\beta_{10}, \beta_{11}, \cdots, \beta_{1p}) \tag{13}$$

If Y_{2i}'s are presumed to be mutually independent, the conditional distribution of $Y_2 = Y_{2i} + \cdots + Y_{2y_1}$, the total number of deaths among Y_1 accidents occurring in the jth time interval is a Poisson distribution with parameter $\lambda_2 y_1$.

$$g_2(y_2\,|\,y_1) = \frac{e^{-\lambda_2 y_1(\lambda_2 y_1)^{y_2}}}{y_2!}, y_2 = 0, 1, \cdots \tag{14}$$

The joint distribution of number of RTAs and number of fatalities can be shown as

$$g(y_1, y_2) = g_2(y_2|y_1)\cdot g_1(y_1) = e^{-\lambda_1}\lambda_1^{y_1} e^{-\lambda_2 y_2}(\lambda_2 y_1)^{y_2} / \,(y_1!y_2) \tag{15}$$

3.4 Zero Truncated Poisson Distribution

The probability of $Y_1 = 0$ is $e^{-\lambda_1}$ using Equation (11). Therefore Y_1 is observed conditional on $Y_1 > 0$. Hence, the conditional probability mass function

$$P(Y_1 = y1\,|\,Y_1 > 0) = \frac{P(Y_1 = y_1)}{P(Y_1 > 0)} = \frac{P(Y_1 = y_1)}{1 - P(Y_1 = 0)}. \tag{16}$$

Using Equation (11) the zero truncated Poisson probability mass function for $Y_1|Y_1 > 0$

$$g_1^{*}(y_1) = P(Y_1 = y_1\,|\,Y_1 > 0) = \frac{e^{-\lambda_1}\lambda_1^{y_1}}{y_1!} \times \frac{1}{(1 - e^{-\lambda_1})} = \frac{\lambda_1^{y_1}}{y_1!(e^{\lambda_1} - 1)}. \tag{17}$$

The exponential form of the mass function is given as

$$g_1^{*}(y_1) = exp\left[y_1 ln\lambda_1 - ln(y_1!) - ln(e^{\lambda_1} - 1)\right] \tag{18}$$

The mean and variance is

$$\mu_{Y_1} = E[Y_1|Y_1 > 0] = \frac{\lambda_1 e^{\lambda_1}}{e^{\lambda_1} - 1}\; and\; \sigma_{Y_1}^2 = Var[Y_1\,|\,Y_1 > 0] = \frac{\lambda_1 e^{\lambda_1}}{e^{\lambda_1} - 1}[1 - \frac{\lambda_1}{e^{\lambda_1} - 1}]. \tag{19}$$

The zero truncated conditional distribution of $Y_2 \mid y_1, Y_2 > 0$ is

$$P(Y_2 = y_2 \mid y_1, Y_2 > 0) = \frac{P(Y_2 - y_2 \mid y_1)}{P(Y_2 > 0 / y_1)} = \frac{P(Y_2 = y_2 \mid y_1)}{1 - P(Y_2 = 0 / y_1)}. \quad \textbf{(20)}$$

The zero truncated conditional Poisson distribution is

$$g_2^*(Y_2 = y_2 \mid y_1, Y_2 > 0) = \frac{e^{-\lambda_2 y_1}(\lambda_2 y_1)^{y_2}}{y_2!} \times \frac{1}{(1 - e^{-\lambda_2 y_1})} = \frac{(\lambda_2 y_1)^{y_2}}{y_2!(e^{\lambda_2 y_1} - 1)}. \quad \textbf{(21)}$$

The exponential form of Equation (21) can be shown as

$$g_2^*(Y_2 = y_2 \mid y_1, Y_2 > 0) = exp[y_2 \, ln \, \lambda_2 + y_2 \, ln \, (y_1) - ln \, (y_2!) - ln \, (e^{\lambda_2 y_1} - 1)]. \quad \textbf{(22)}$$

The mean and variance are

$$\mu_{Y_2|Y_1} = E[Y_2|Y_1, Y_2 > 0] = \frac{\lambda_2 y_1 e^{\lambda_2 y_1}}{e^{\lambda_2 y_1} - 1} \; and \quad \textbf{(23)}$$

$$\sigma^2_{Y_2|Y_1} = Var[Y_2|Y_1, Y_2 > 0] = \frac{\lambda_2 y_1 e^{\lambda_2 y_1}}{e^{\lambda_2 y_1} - 1}[1 - \frac{\lambda_2 y_1}{e^{\lambda_2 y_1} - 1}] \quad \textbf{(24)}$$

3.5 Negative Binomial Distribution

Let Y_1 be the count of accidents at a specific location in a given interval that has a Poisson distribution with the following poisson mass function . In a Poisson model this would follow the poisson distribution as shown in Equation(11).

The Poisson regression model commonly assumes the log-linear relationship between Poisson parameter and explanatory variables as shown in Equation(13).

The major advantage of the Poisson distribution is the simplicity in calculation due to its property of the mean equalling to the variance. The relationship is termed equidispersion, which is also known as its restriction. If $E(y_i) > Var(y_i)$, it is said that the data are under-dispersed, oppositely over-dispersed if $E(y_i) < Var(y_i)$. The Poisson regression model is inappropriate to use when the variance of accident data is significantly different from the mean. In this case negative binomial regression model can be applied as an alternative to overcome the problem. The negative binomial technique relaxes the assumption of equality of the mean and variance, by adding a gamma-distributed error term.

Equation(13) is rewritten as

$$\lambda_i = E(y_i) = e^{\beta X_i + \varepsilon_i} \tag{25}$$

where ε_i is an error term, and e^{ε_i} is gamma-distributed error term with mean 1 and variance σ^2.

The addition of ε_i makes the variance to be different from the mean as follows:

$$VAR(y_i) = E(y_i)[1 + \sigma E(y_1)] = E(y_i) + \sigma E(y_i)^2 \tag{26}$$

where σ is also called the dispersion parameter, which plays an important role in the determination of choosing the Poisson regression or the negative binomial regression model.When σ is significantly different from zero, the distribution is under-dispersion or over-dispersion and the negative binomial model is appropriate. When σ approaches zero, the variation is almost equal to the mean, and the distribution can be simply modelled by the Poisson regression technique.

The negative binomial probablity distribution is as follows.

$$P(y_i) = \frac{e^{-\lambda_i e^{\varepsilon_i}}}{y_i!} \tag{27}$$

Intergrating ε_i

$$P(y_i) = \frac{\Gamma((1/\sigma))}{\Gamma(1/\sigma) y_i!} \left(\frac{1/\sigma}{(1/\sigma) + \lambda_i} \right)^{1/\sigma} \left(\frac{\lambda_i}{(1/\sigma) + \lambda_i} \right)_i^y \tag{28}$$

where $\Gamma(.)$ is a gamma function.

The negative binomial model is also estiamted by the standard maximum likelihood method. The corresponding likelihood function is as follows:-

$$L(\lambda_i) = \prod_i \frac{\Gamma((1/\sigma))}{\Gamma(1/\sigma) y_i!} \left(\frac{1/\sigma}{(1/\sigma) + \lambda_i} \right)^{1/\sigma} \left(\frac{\lambda_i}{(1/\sigma) + \lambda_i} \right)_i^y \tag{29}$$

The mean and variance for negative binomial distribution are as follows:-

$$\mu = \frac{r(1-p)}{p} \tag{30}$$

$$\sigma^2 = \frac{r(1-p)}{p^2} = \mu + \frac{1}{r}\mu^2 \tag{31}$$

where p is the probability of success and r is when the succes occurs.

3.6 Model Specification

The models have the number of fatalities from fatal RTAs within a period as function of the categorical variables; Day, Victim and County.

Each of the models parameterizes as:

$$\theta_i = exp(\beta_0 + \beta_1 Day + \beta_2 Victim + \beta_3 County) \tag{32}$$

3.7 Parameter Estimation

The parameters of the model were estimated using Maximum Likelihood Estimation. It was important to examine the signifance of the variables in the models. The estimated coefficients had to be statistically significant.

3.8 Goodness of Fit

For linear regression models the coefficient of determination R 2 test is a conventional goodness of fit measure. For non-linear model (e.g. Poisson, negative binomial regression models) to determine the goodness of fit the study will use Akaike Information Criterion(AIC) and Bayesian Information Criterion(BIC).

3.8.1 Akaike information criterion

AIC was first developed by Akaike (1973) as an estimator of the relative quality of statistical models for a given set of data. The selection of the best model is determined by an AIC score.

$$AIC = 2K - 2log(\hat{L}) \tag{33}$$

where k is the number of estimated parameters in the model, $\hat{L}$ is the maximum value of the likeliohood function in the model.

AIC deals with both the risk of overfitting and the risk of underfitting the model. The model with the least AIC value is considered to be the better model fit.

3.8.2 Bayesian information criterion

The BIC was developed by Gideon E. Schwarz (1978) as a criterion for model selection among a finite set of models; the model with the lowest BIC is preferred. It is defined as follows:

$$BIC = ln(n)k - 2ln(\hat{L}) \tag{34}$$

where:

- $\hat{L}$ is the maximum value of the likeliohood function in the model.
- x is the observed data;
- the number of data points in x, the number of observations, or equivalently, the sample size;
- k is the number of estimated parameters in the model

4 Analysis

4.1 Data Collection

The primary crash data needed for the development of the CPM were obtained from the National Transport and Safety Authority of Kenya. The data contained all accident fatalities in Kenya between January, 2015 and December, 2017. The data was collected from police stations/bases all over the country and consolidated into one dataset. The data was structured on a casualty basis ie if there was an accident involving a pedestrian and a motorcycle driver that resulted in the deaths of both the victims these are two entries in the data set. One for the pedestrian and another for the motorcycle driver. However, if there were an accident that involved two motorcycle drivers these results in one entry in the data set and the count incremented by two.

The data set included the date and time of the accident where available, the road and county of the accident, number of fatalities, the category of the accident victim,age of the victim and count of victims.

4.2 Data Cleaning

The data obtained from NTSA was the raw accident data and hence required a lot of data cleaning. The data used was collected for different purposes and not specifically to answer our research questions and hence there were some inconsistent/missing data entries and a lot of data cleaning and munging had to be done to make the data ready for analysis. The first step involved formatting the dates so as to standardize them. The data had dates had formats like dd/mm/yy, mm/dd/yy etc.

After standardizing the dates for all years in the format mm/dd/yy, two newy fields were included in the data set. Each entry was associated with the day in the week and yearly quarter that correlated to the date.

The data had a couple of fields which includes County RTA occurred, Road that the RTA occurred, Age where applicable of the RTA fatality, Gender where applicable of the RTA fatality, Police Station/Base where the RTA was reported, Details of the RTA's, Place where the RTA occurred and the victim of the RTA.

The details of raod accidents were parsed to create categorical values that could be used during analysis and create a new column labeled Vehicle in the data set.These includes but is not limited to AMB-MCYC which signified an accident between an ambulance and a Motorcycle, COM which signfied an RTA involving a commercial vehicle and COM-PRV-PSV-MCYC whcih signified a fatal RTA between a commercial vehicle, private vehicle, Passenger Service Vehicle and a Motorcycle.

Due to the nature of the dataset the Victim field was also cleaned up to be consistent through and through.This involved Unifying the categorical variables to show the casualties of the RTA's. The Place,Road and Police Stations/Bases were also updated using rugular expressions to make values that referenced similar places to be the same. This was due to the presence of spelling mistakes in the data, shothand/shortform values ie MSA to denote Mombasa etc.

4.3 Statistical Methods

Descriptive statistical analysis was done to estimate the counts, giving frequencies and-percentages. Many of the methods for count data have been advanced and these include the Poisson Model, the Zero-Truncated Poisson (ZTP), the Negative Binomial (NB) and the Zero-Truncated Negative Binomial (ZTNB). The four models, the Poisson, ZTP, NB and ZTNB were compared to see which one fits the data well. All p-values reported were two-tailed, and values below 0.05 were considered statistically significant.

The study explored all these models and the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) statistics were used in selecting the best fit model.The data was analysed using R software, version 3.4.4 (The R Foundation for Statistical Computing).

4.4 Data Exploration

A total of 7,822 RTA's were observed in Kenya between January 2015 and December 2017 that resulted in 16,678 fatalities in the period. The number of accidents distributed across the three days as follows:-

Year	Number of Fatalities	**Totals**
2015	5753	5753
2016	5562	11315
2017	5363	16678

Table 1. Road fatalities Distributed by Years

The number of accidents over the period was as shown below.

Year	Number of Accidents	**Totals**
2015	2708	2708
2016	2625	5333
2017	2489	7822

Table 2. Fatal RTA's Distributed by Years

As shown in the distribution of fatal RTA's and fatalities there is a direct correlation between the number of accidents and fatalities experienced between the years 2015 and 2017. Seasonality of fatal road accidents in the period was also examined by studying the distribution of fatal RTA's over days of the week and yearly quarters. The figure below shows the distribution of RTA's distributed by day.

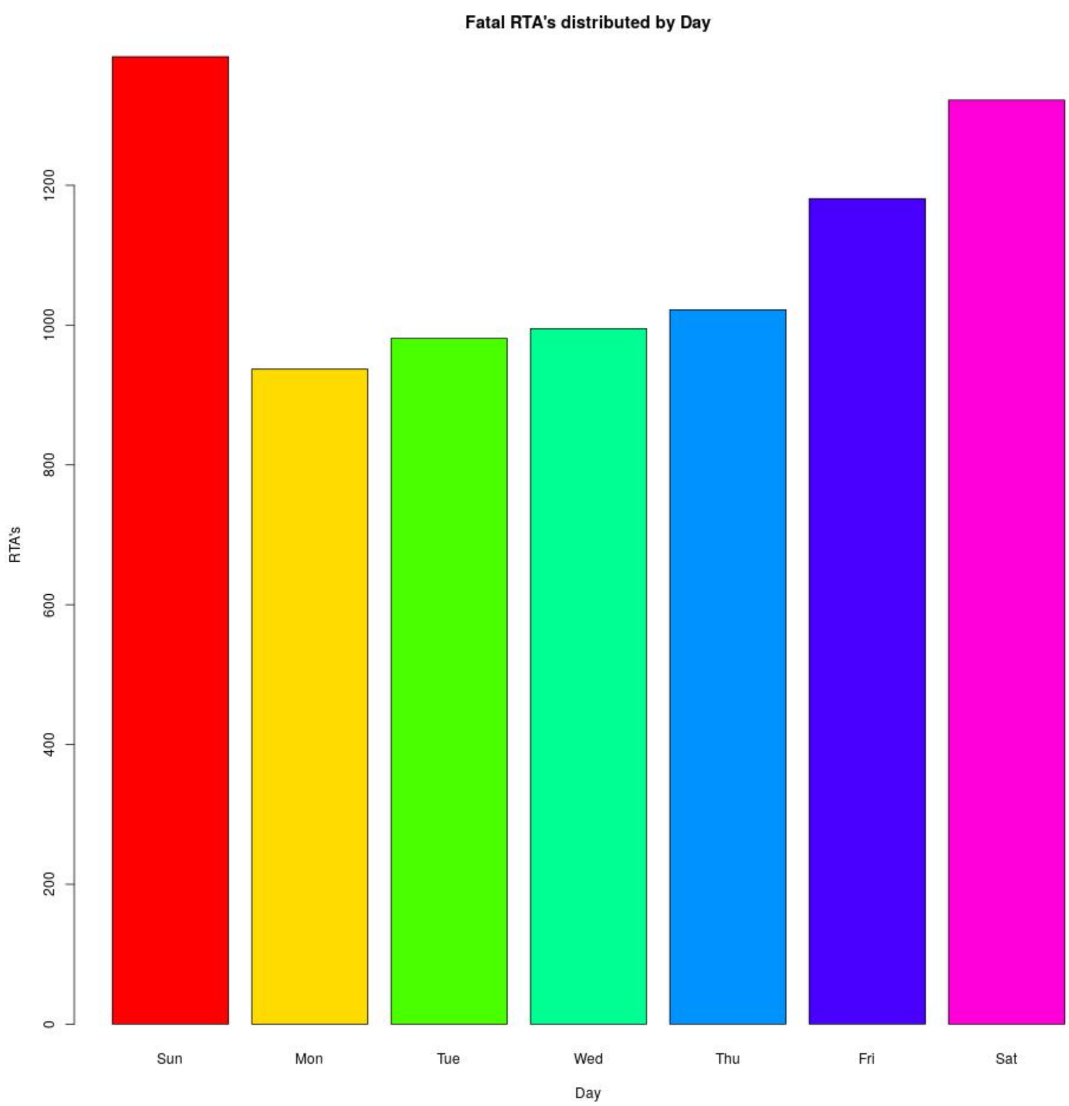

Figure 2. Fatal RTA's distributed by Day

More accidents were experienced on the weekends with Friday, Saturday and Sunday having more fatal RTA's compared to weekdays with Monday having the least number of fatal RTA's.

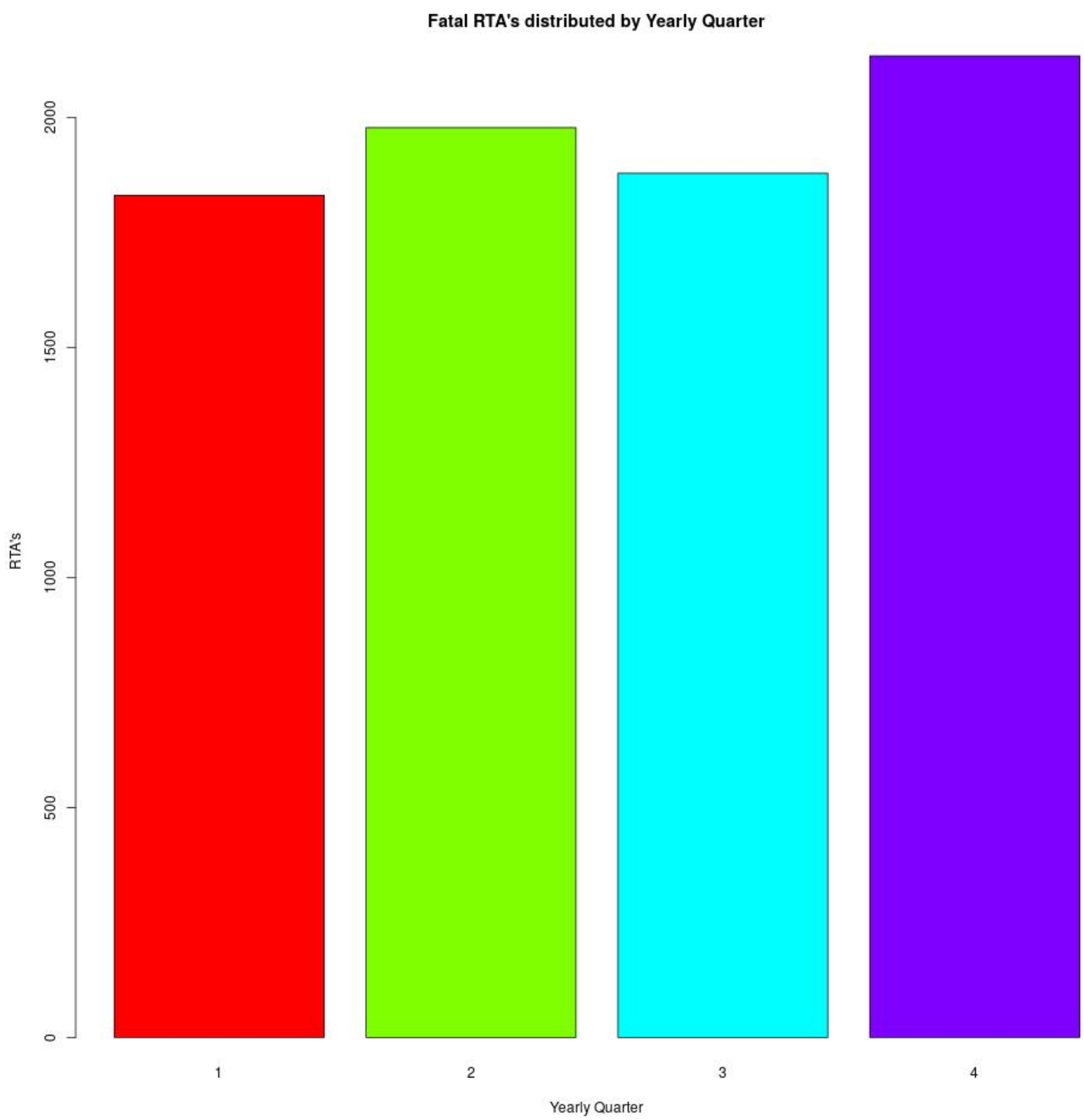

Figure 3. Fatal RTA's distributed by Yearly Quarter

The yearly quarter were defined as from January-March(1), April-June(2), July-August(3) and September-December(4). As shown in figure 3 above more RTA's were experienced in the 2nd and 4th quarters of the year. THe study also analysed the seasonality of fatalities to see whether there was any relationship between fatalities count and RTA's

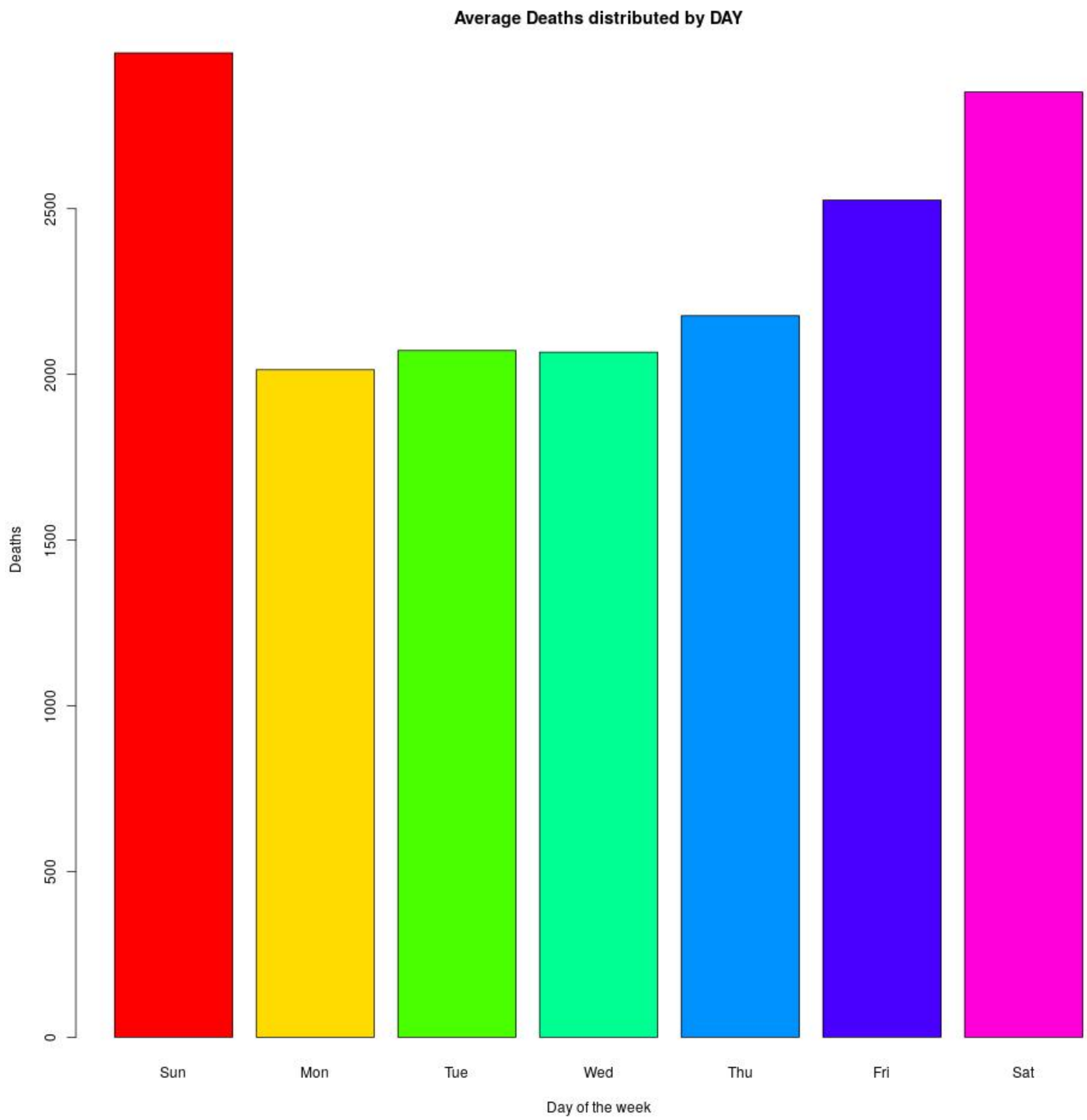

Figure 4. Fatalities distributed by Day

The number of deaths from 2015 through to 2017 is consistent with the number of fatal RTA's wiht the number of fatalites being very high over the weekend and relatively lower in the weekdays. Mondays and Wednesdays experienced relatively lower fatalities compaired to the other weekdays.

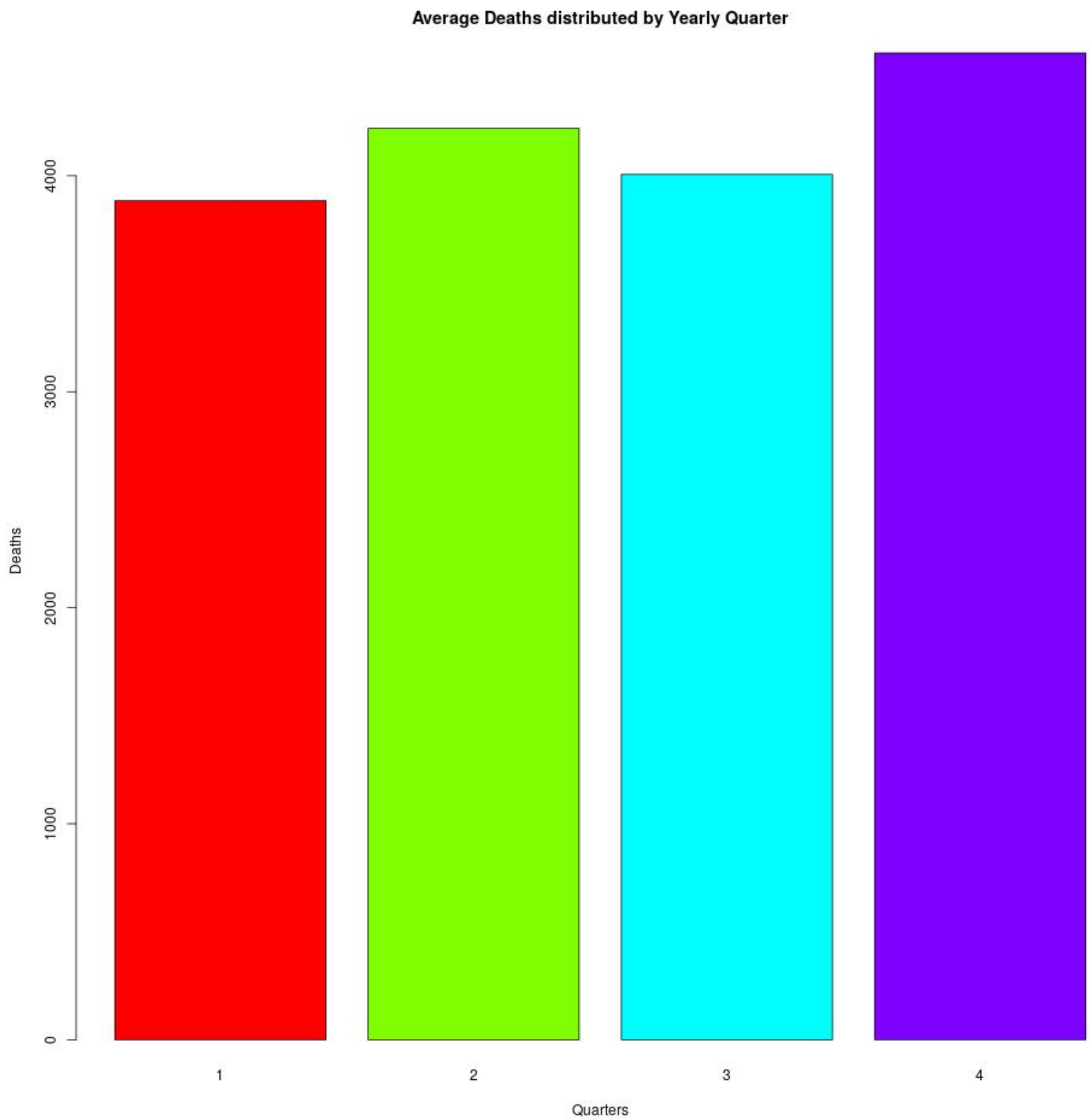

Figure 5. Fatalities distributed by Quarter

The number of quartely deaths were also higher in the second and fourth quarters of the year compared to the rest of the quarters. This shows that the numer of RTA's and fatalities were consistently higher on weekends and in the second and fourth quarters of the year(s).

The figure below shows the number of fatalities distributed over the Kenyan road network.

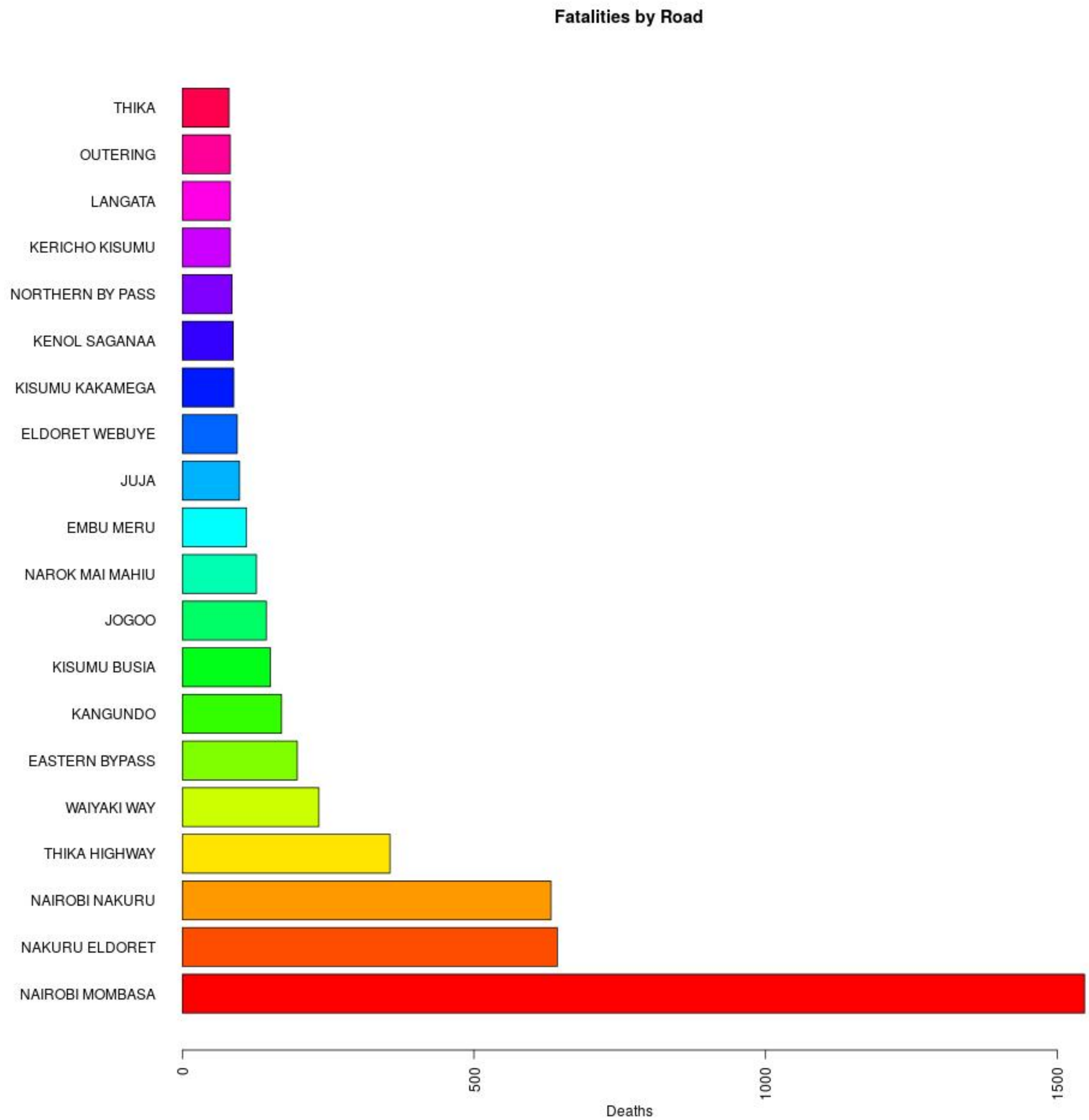

Figure 6. Fatalities distributed by Roads

The highest number of fatalities were along the Nairobi Mombasa highway and this was consistent with the high number of fatal RTA's in the road. The study also seeked to know which Plice station/bases reported the highest number of fatalities in the country.

Since the most fatal RTA's occurred along the Nairobi Mombasa highway the study used this road as a case study in order to have a granular understanding of the causes of fatal RTA's in Kenya

4.5 Case Study: Nairobi Mombasa Highway

4.5.1 Road Dynamics

The Nairobi Mombasa highway is the main road that connects the capital city of Kenya(Nairobi) to the coastal city of Mombasa which is also the larget port city in the country. The total distance from Nairobi to Mombasa is approximately 488 kilometres. According to Njoroge (2016) the highway is part of the Great North Road that moves more than 50 per cent of all goods traded in East Africa.

Gumbihi(2015) noted that the Nairobi-Malaba highway which is divided into two which is Nairobi Mombasa and Nairobi Malaba actually takes more lives than alshabaab. Tha newspaper article cited the study 'Improvement of Road Safety and Health through Road Side Stations along the Northern Corridor' in which the study concluded that in 2013 Kenya experienced 3,179 fatal RTA's along the highway.

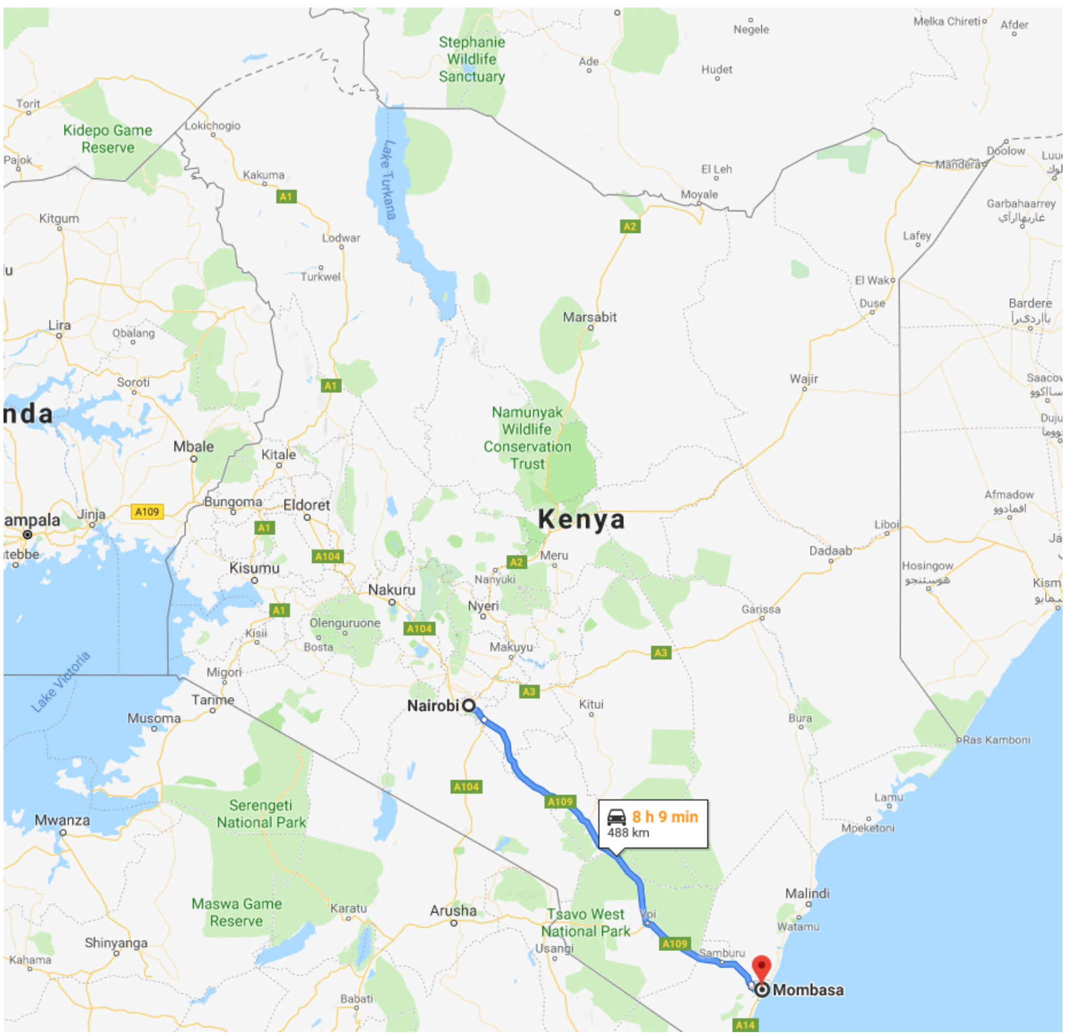

Figure 7. Source: Google Maps Nairobi Mombasa Highway

4.5.2 Descriptive Statistics

A total of 716 fatal RTA's on the Nairobi Mombasa Highway that resulted in 799 fatalities between January 2015 and December 2017. The deaths were distributed as follows.

Year	Number of Fatalities	**Totals**
2015	279	279
2016	261	540
2017	259	799

Table 3. Nairobi Mombasa highway fatalities Distributed by Years

The number of fatalities in the road have been reducing over the 3 year peiod but the reduction is not noticeably high. The Nairobi Mombasa highway passes through 7 counties in Kenya. These counties are Nairobi County, Machakos County, Kajiado County, Makueni County, Taita-Taveta County, Kwale County and Mombasa County. The figure below shows the distribution of fatalities over tehe counties along the Mombasa Nairobi Highway.

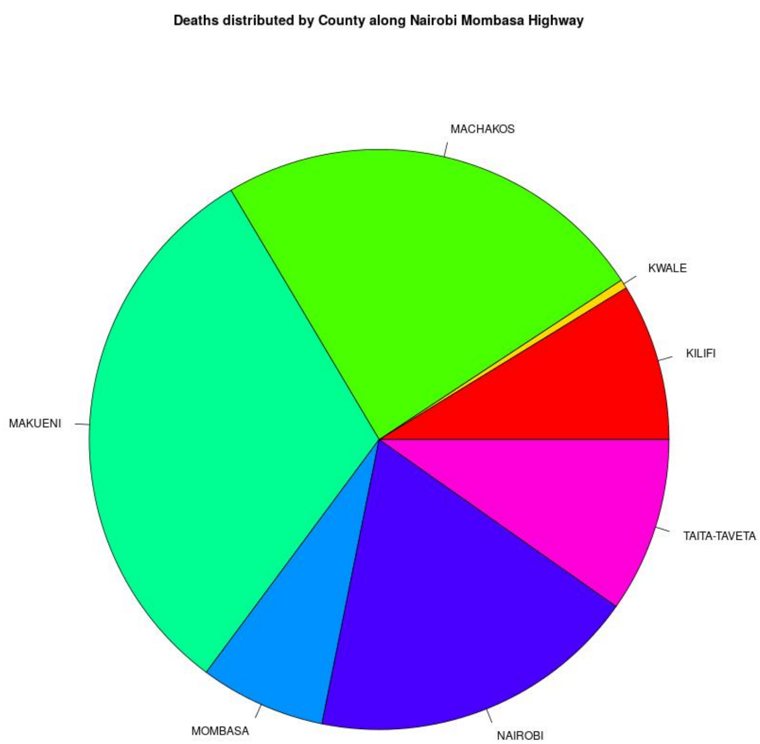

Figure 8. Fatalities distributed by County along Nairobi Mombasa Highway

Along the Nairobi Mombasa highway most of the fatalities were experienced in Makueni, Machakos and Nairobi Counties whereas the least number of fatalities were in Kwale County. The study also classified the victims of fatal RTA's along the highway and they were as follows

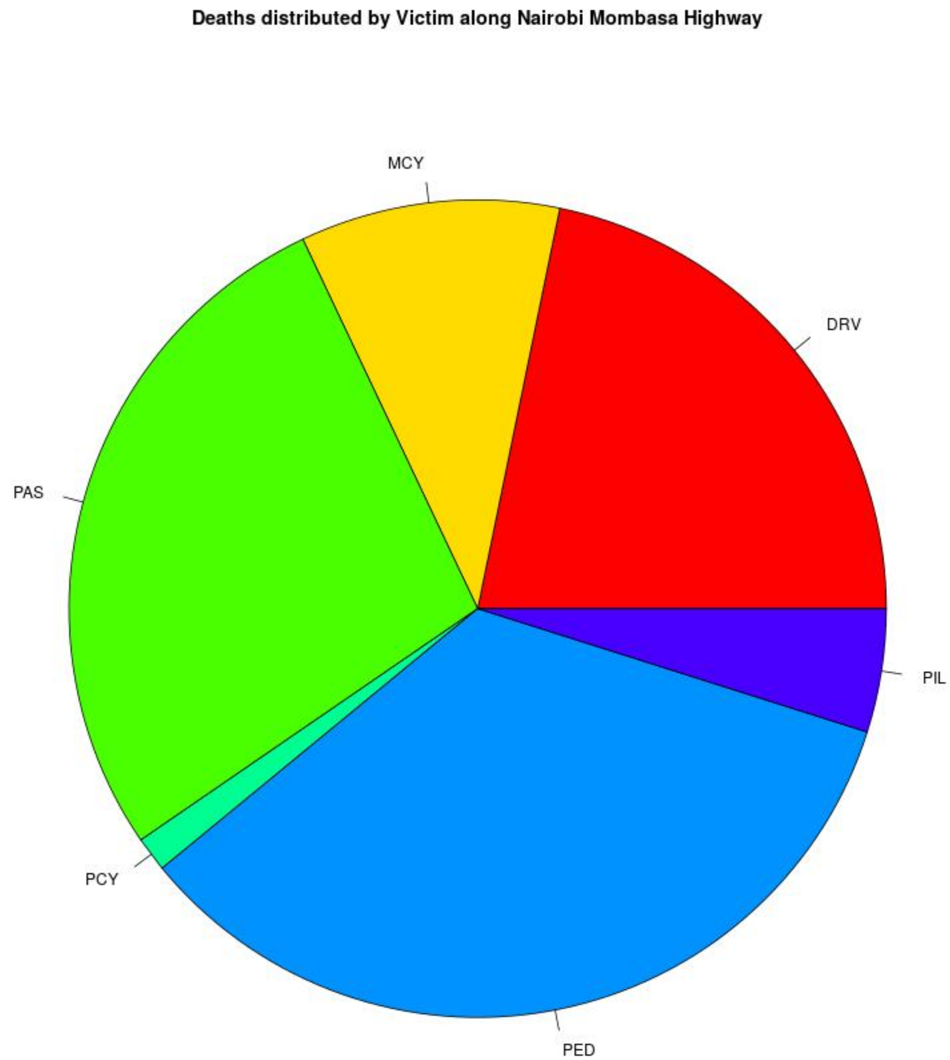

Figure 9. Fatalities distributed by Victim along Nairobi Mombasa Highway

Among the victims of fatal RTAs the highest frequency was among drivers(174) and passegers(220) with pedestrians also representing a high fatality count of 273. Motorcyclc drivers and their pillow pasengers had fatality counts of 82 and 39 respecyively with Pedal cyclists having a count of 11. This shows that the highest number of casualties along the road are vehicle users and pedestrians. The study also classified the vehicles that were involved in the fatal RTA's and the figure below shows the most frequent fatality vehicles.

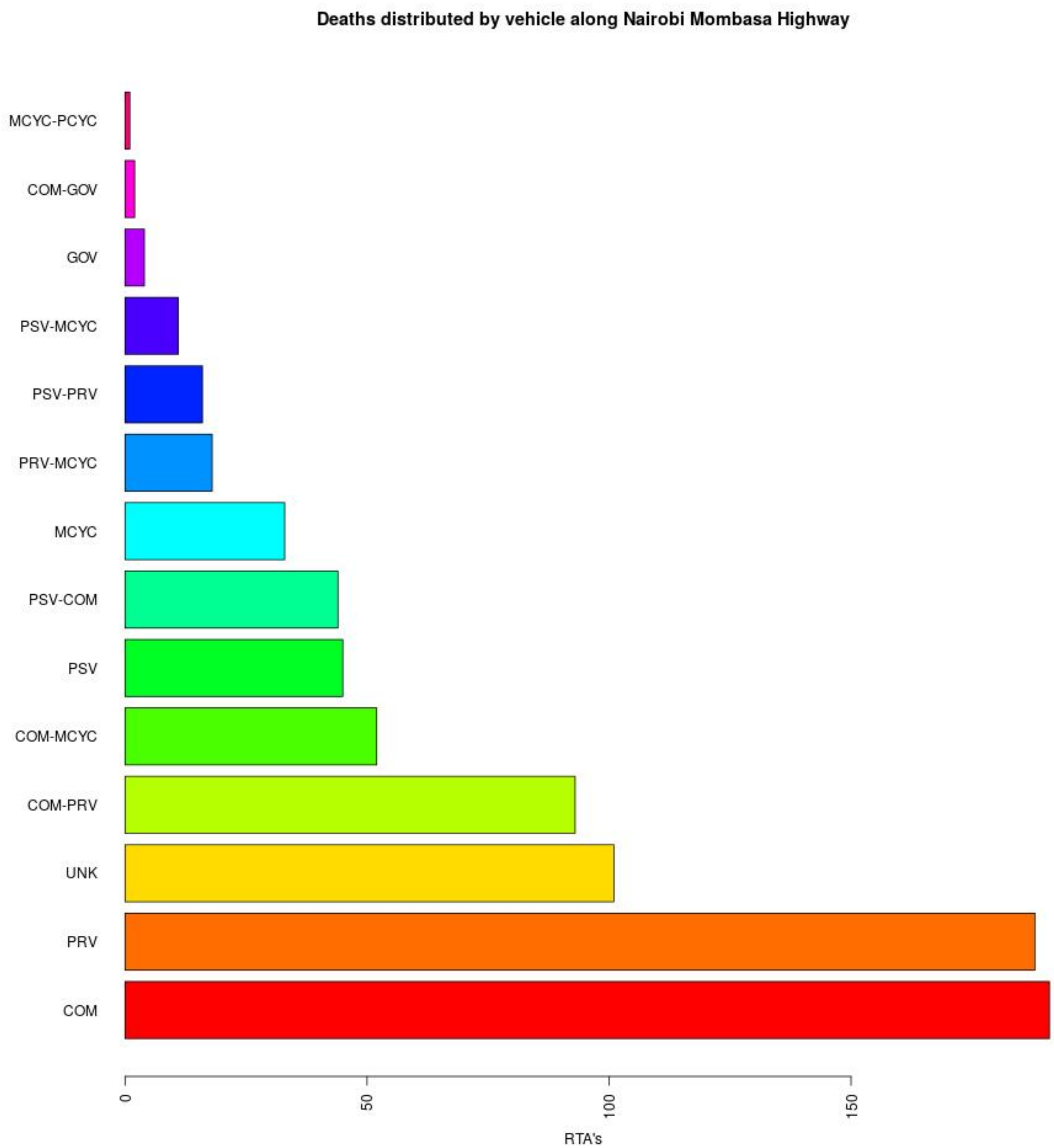

Figure 10. Fatalities distributed by Vehicle along Nairobi Mombasa Highway

Since the Nairobi-Mombasa highway forms part of the Great North Road and is mostly used to move goods into land-locked countries in East Africa it shows that the vehicles that result in most fatalities in the highway are actually Commercual vehicles followed by Private Vehicles. For fatalities that were reported but were not associated with any vehicle ie hit and run they were the third highest category of vehicular fatalities. The table also shows interactions between different RTA's vehicles along the road with accidents between MCYC -PCYC having the least count with only one fatality.

The study also studied the seasonality of accidents along the highway distributed along days of the week and the figure below show shows the distribution.

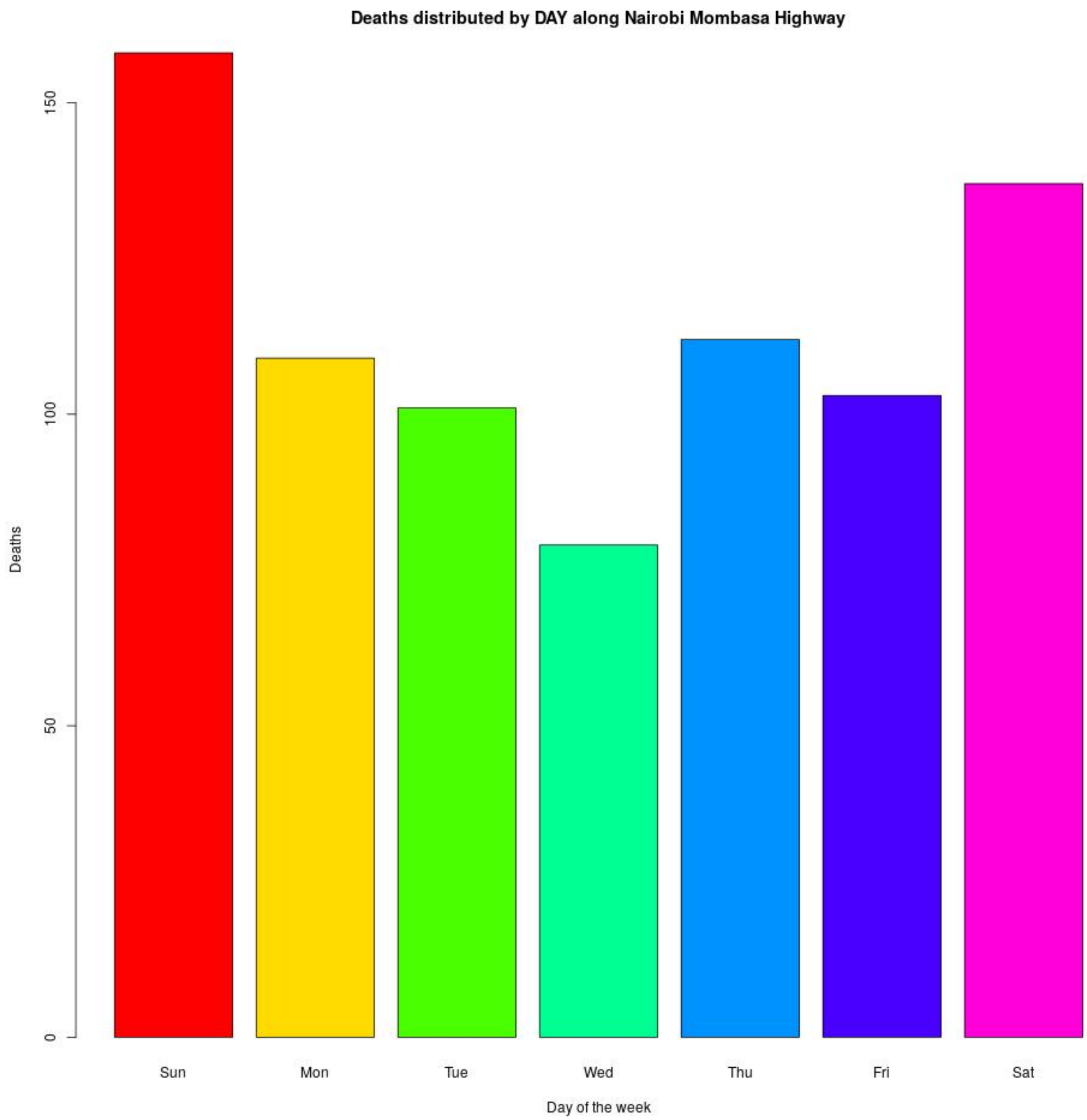

Figure 11. Fatalities distributed by Day along Nairobi Mombasa Highway

The figure above shows that the seasonality of fatal RTA's was consistent with the rest of the country since most fatalities occurred along the weekend with sunday having the highest count and wednesday having the least.

4.6 Model Exploration along Nairobi Mombasa Highway

In order to model these traffic deaths there is need for a careful selection of one or more models that may provide a good description of the traffic type, estimation of parameters such as mean and variance for the selected models and statistical testing for selection of one of the considered models and analysis of its suitability to describe the traffic type under analysis. In our data the numbers of zeros were very minimal as most of the accidents had at least one person dying underdispersion was presence since the mean was greater than the variance. When fitting the models the factors considered were the day the accident happened, the victims of the fatal RTA's, vehicles involved in the RTA and the county in which the fatal RTA occured.

4.6.1 Poisson Model Fit

The poisson model was fitted and the model had an AIC value of 1,628.447 and a BIC value of 1,710.773 .The table below shows the estimates and P-values of the model

Variable	Estimate	**P-Value**	(95% CI Unadj)
DAY Sunday	-0.035066	0.783582	(-0.131406856, 0.06127435)
Saturday	0.003796	0.976929	(-0.115526904, 0.12311890)
Monday	0.130297	0.336419	(-0.112572471, 0.37316566)
Tuesday	-0.083596	0.552061	(-0.180703889, 0.01351102)
Wednesday	-0.114480	0.445331	(-0.208793834, -0.02016547)
Thursday	0.065184	0.633148	(-0.075283192, 0.20565098)
Friday	0.234750	0.325651	(-0.054292237, 1.543227902)
Victim PAS	0.352291	0.000127	(0.235207222, 0.46937552)
PED	0.019001	0.834523	(-0.039219871, 0.07722188)
PRV	0.027386	0.785253	(-0.035668392, 0.09044003)
COM-PRV	0.163304	0.205649	(-0.001657347, 0.32826549)
PSV-COM	0.466818	0.003766	(-0.046195112, 0.97983087)
COM	-0.035123	0.727049	(-0.099433386, 0.02918760)
PSV	0.034801	0.833453	(-0.056738567, 0.12634029)
County MAKUENI	0.087633	0.374080	(-0.011071980, 0.18633837)
KILIFI	0.073572	0.608399	(-0.084769322, 0.23191311)
MACHAKOS	-0.006431	0.951853	(-0.094449367, 0.08158696)
TAITA-TAVETA	0.015326	0.912564	(-0.081857190, 0.11250967)
MOMBASA	0.567812	0.543147	(-0.081857190, 0.11250967)
NAIROBI	0.417654	0.124851	(-0.93217675, 2.23541892)

Table 4. Poisson Model fit for Nairobi Mombasa highway fatalities

From the table above the poisson model fit was not the best fit for the data since it had high values of AIC and BIC. The model also showed that only the Passenger victim and vehicle accidents between PSV vehicles and Commercial Vehicles were significant factors in the model.

4.6.2 Zero Truncated Poisson Model Fit

The zero truncated poisson model was fitted and the model had an AIC value of 429.9233 and a BIC value of 512.2496 .The table below shows the estimates and P-values of the model

Variable	Estimate	**P-Value**	(95% CI Unadj)
DAY Sunday	-0.48326	0.22975	(-1.4295540, 0.46303282)
Saturday	0.03705	0.92050	(-0.8416664, 0.91575802)
Monday	0.32978	0.28978	(-0.5132087, 1.17277555)
Tuesday	-1.14550	0.03775	(-2.2597335, -0.03126944)
Wednesday	-1.27701	0.03968	(-2.3182695, -0.23575025)
Thurday	0.55233	0.11460	(-0.2945631, 1.39922103)
Friday	0.34512	0.097562	(-0.521098, 1.3287654)
Victim PAS	2.16305	< 0.0001	(1.3808356, 2.94526204)
PED	-1.81111	0.08482	(-3.3654610, -0.25676268)
PRV	0.54505	0.24201	(-14.5401050, 15.63021199)
COM-PRV	1.04990	0.02388	(-13.8550234, 15.95482310)
PSV-COM	1.71783	0.00024	(-12.8881191, 16.32377244)
COM	0.01057	0.98277	(-14.9651891, 14.98631950)
PSV	0.59172	0.34778	(-14.0857135, 15.26914894)
County MAKUENI	1.40272	0.00791	(-14.5977094, 17.40315446)
KILIFI	1.32835	0.02318	(-14.7865647, 17.44326566)
MACHAKOS	0.86940	0.10736	(-15.4688052, 17.20759710)
TAITA-TAVETA	1.06326	0.07168	(-15.2855830, 17.41210028)
MOMBASA	0.34518	0.17852	(-10.2344561, 12.753625)
NAIROBI	0.93211	0.14389	(-15.0233554, 17.5542112)

Table 5. Zero Truncated Poisson Model fit for Nairobi Mombasa highway fatalities

From the table above the zero truncated model was the best model fit since it had the least values of AIC and BIC. The model showed significant variables in the study were a mix of the days, vehicles and county.

4.6.3 Negative Binomial Model Fit

The negative binomial model was fitted and the model had an AIC value of 1,630.451 and a BIC value of 1,717.351 .The table below shows the estimates and P-values of the model.

Variable	Estimate	**P-Value**	(95% CI Unadj)
DAY Sun	-0.035066	0.783582	(-0.28348848, 0.2176243)
Saturday	0.003795	0.976932	(-0.25220812, 0.2630112)
Monday	0.130295	0.336428	(-0.13487583, 0.3972076)
Tuesday	-0.083597	0.552063	(-0.35967227, 0.1922313)
Wednesay	-0.114480	0.445334	(-0.41087921, 0.1781467)
Thursday	0.065183	0.633153	(-0.20228189, 0.3338387)
Friday	0.082367	0.712344	(-0.39126891, 0.453212121)
Victim PAS	0.352291	0.000127	(0.17151022, 0.5320586)
PED	0.019001	0.834527	(-0.15949792, 0.1971965)
PRV	0.027386	0.785253	(-0.17041048, 0.2238374)
COM-PRV	0.163303	0.205653	(-0.09321093, 0.4131006)
PSV-COM	0.466816	0.003766	(0.14274565, 0.7754523)
COM	-0.035123	0.727052	(-0.23316618, 0.1615326)
PSV	0.034801	0.833454	(-0.30152142, 0.3487709)
County MAKUENI	0.087633	0.374086	(-0.10513285, 0.2815533)
KILIFI	0.073572	0.608401	(-0.21362489, 0.3500385)
MACHAKOS	-0.006431	0.951857	(-0.21560942, 0.2021683)
TAITA-TAVETA	0.015326	0.912564	(-0.26326358, 0.2845823)
MOMBASA	0.062541	0.341891	(-0.3189103, 0.31765182)
NAIROBI	0.091268	0.542121	(-0.4128781, 5.12345121)

Table 6. Negative Binomial Model fit for Nairobi Mombasa highway fatalities

The table above showed that the negative binomial model was not the best fit for the data since it had the largest values of AIC and BIC. Due to the presence of under-dispersion in the data there was no need to fit a zero truncated negative binomial model since it accouns for overdispersion.

4.7 Model Insights

The study applied the Zero-Truncated Poisson, Negative Binomial and the Poisson Model. The results from these models were compared to select the best fit model for this data using Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The ZTP had AIC=429.9233, BIC=512.2496, the Negative Binomial had AIC=1630.451, BIC=1,717.351

whereas the Poisson had AIC=1,628.447 and BIC=1,710.773. This indicated that the ZTP with the lower AIC and BIC was the best fit model for the data and hence the tabular values of this model were used to get the model insights.

4.7.1 Day of fatal RTA

The study found evidence that some days experienced more fatalities compared to others. More fatalities were experienced on Tuesday and Wednesday in the Nairobi-Mombasa highway. This was however in contrast to the number of RTA since more RTAs are observed over the weekend days Saturday and Sunday but the more fatal accidents occur in the middle of the week ie 2 accidents in the middle of the night that result in 30 fatalities compared to 10 RTA's that result on the same 30 fatalities. Both Tuesday and Wednesday had a significant p-value of 0.03775 and 0.03968 respectively.

4.7.2 Victims of fatal RTA

The study found that PSV passangers were more likely to die from RTA's compared to any other group of road users including pedestrians, motorcycle riders and pedal cyclists. This was backed with the number of high fatal count associated with an accident in Nairobi-Mombasa Highway since most of the PSV vehicles that plight the route are buses that have high capacity of between 50 and 70 passengers. The p-value for passangers was found to be significant with a value of less that 0.0001.

4.7.3 Vehicle in fatal RTA

The study found that most fatalities in the Nairobi-Mombasa highway as a result of Commercial and PSV Cars RTA's and PRV and Commercial Vehicles. Commercial vehicles plight this route a lot since they move goods from the port town of Mombasa to the inner parts of Kenya like Kenya and sometimes to countries like Uganda. This study shows that commercial vehicles cause more accident fatalities especially if it is involved in an accident with a PSV. Commercial and PRV RTA's had a statistical significant p-value of 0.02388 compared to commercial and PSV with a statistical significant p-value of 0.00024.

4.7.4 County fatal RTA occurred

The study found that more fatal accidents occured in Makueni and Kilifi Counties than in any other county along the Nairobi-Mombasa highway. Makueni and kilifi had statistically significant p-values of 0.00791 and 0.02318. This shows that road users were more likely to be involved in a fatal crash while they were in these two counties along the Nairobi-Mombasa highway than any where else in the road.

5 Conclusion

The study showed that there was an increased risk of death in the middle of the week and passengers were the most at risk of fatal RTA's as compared to other road users. There is also an increased risk of death if one is driving in Makueni and kilifi counties along the Nairobi-Mombasa highway. The study further observed that Commercial vehicles were involved in fatal RTA's than any other vehicle and they were even more deadly if the RTA was with a PSV. The study further revealed that the ZTP is the best fit model for data in which there are few zeros as is the case with fatal RTAs adn the data is under-dispersed.

5.1 Future Research

The study highlighted that Kilifi and Makueni counties were the most accident prone counties along the Nairobi-Mombasa highway and more research needs to be done in the stretches along these counties to establish the cause of this and how this can be addressed at a policy level. The study also observed that contrary to popular belief that accidents mostly happen over the weekend the more fatal accidents actually happen in the middle of the week and more concerted effort and research needs to be put in place in order to ensure that this is reduced. The study also noticed that commercial vehicles were involved in fatal RTA's along the Narobi-Mombasa highway and a lot of study and effort needs to be put into ensuring that commercial vehicles are driven safely along the road and in the event of a breakdown they are parked in a safe distance away from the road.

A couple of studies Huayun(2012) and Tom(2004) have used RTA data to identify black spots on roads by mapping accident place to GIS maerkers and modelling the data to correctly identify black-spots. This can be a logic research area on the Nairobi Mombasa highway so as to correctly identify black spots in the road and provide policy and road user sensitisation so as to make the highway safer.

5.2 Limitations of the Study

The data obtained from the National Transport and Safety Authority did not contain any behavioral traits that led to an accident.No proper post-accident records were obtained in the course of the study and this limited the scope of the study. Another limitation was that the data used was collected for different purposes and not specifically to answer our research questions and hence there were some inconsistent/missing data entries and a lot of data cleaning and munging had to be done to make the data ready for analysis.

.1 Appendix

The figure below show the police station/bases that reported the most RTA's in Kenyan between the periods of January 2015 and December 2017.

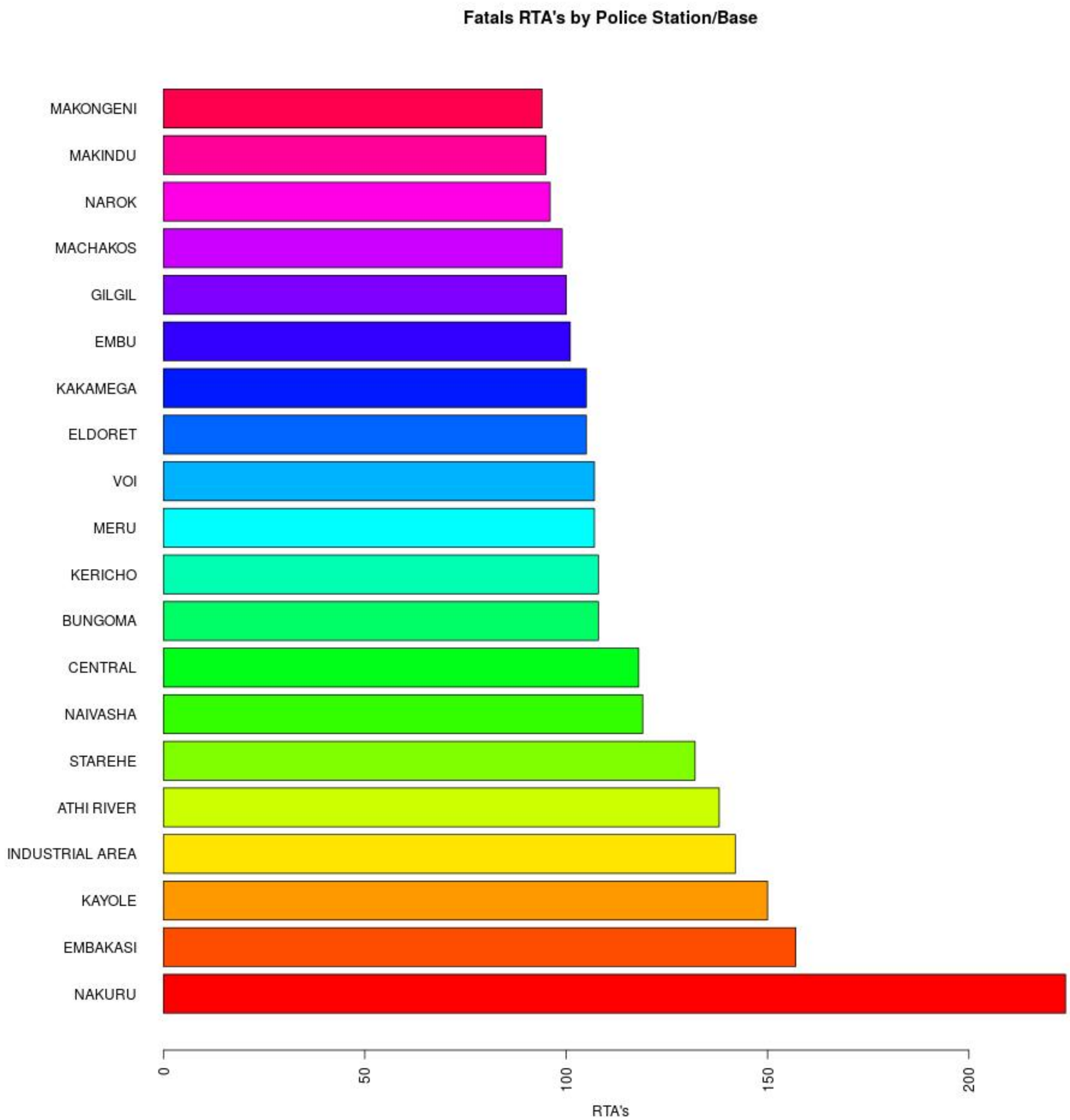

Figure 12. Fatal RTA's distributed by Police Stations/Bases

Nakuru Police station reported the highest number of fatal RTA's closely followed by both Embakasi and Kayole Police stations respectively. There were 5 police stations that were along Nairobi Mombasa highway which are indo area, athi river, voi, machakos and makindu police stations that were part of the top 20 police stations. The figure below shows the number of fatalities distributed along the police station/base that reported them.

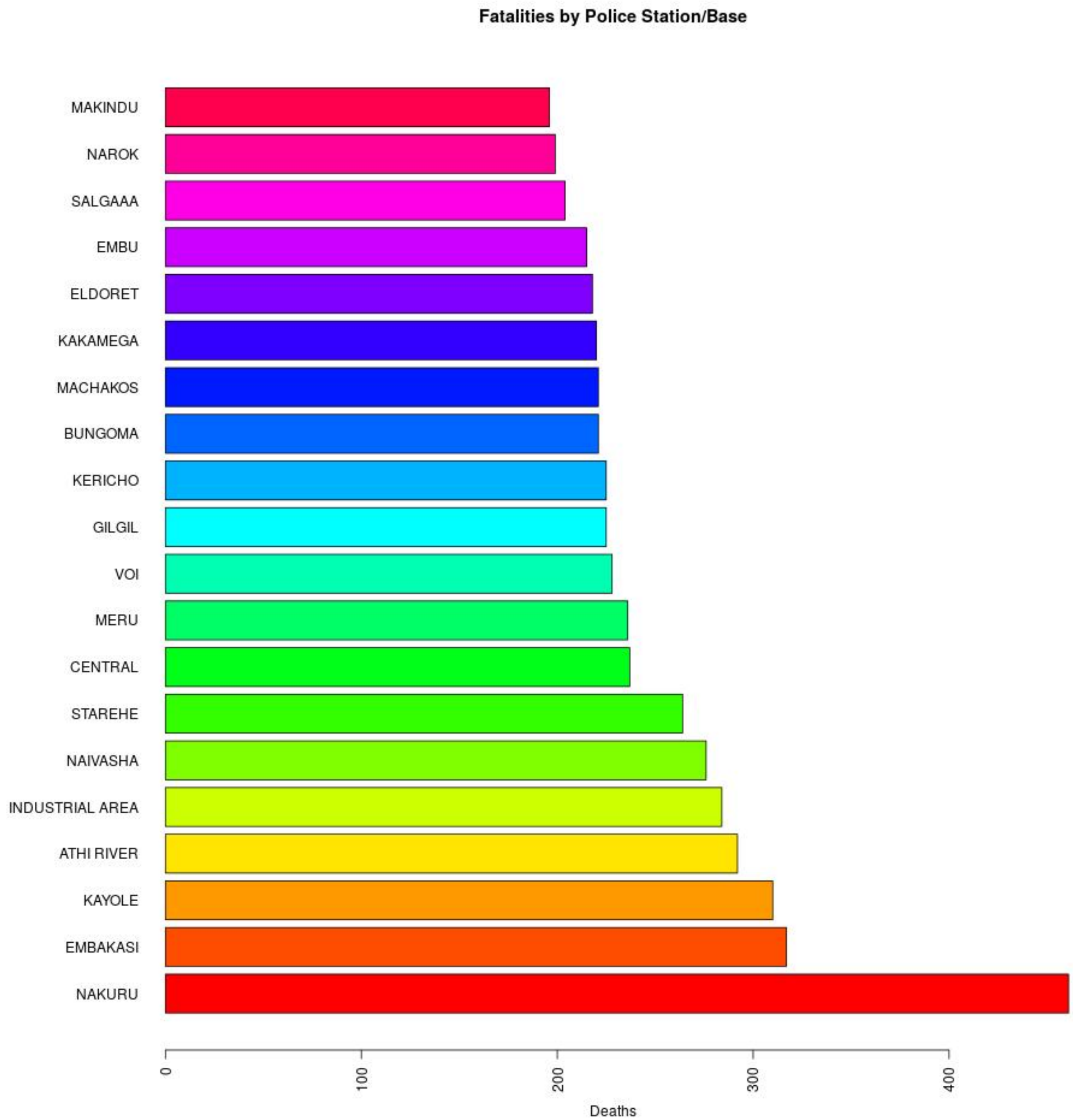

Figure 13. Fatalities distributed by Police Stations/Bases

This was consistent with the fatal RTA's with nakuru,embakasi and kayole reportiing the most fatalities. the 5 police stations along the Mombasa Nairobi highway were also consistent in the top 20 police stations that included Athi River, Indo Area, Voi, Machakos and Makindu police stations.

Appendix: Bibliography

[Abd00] Abdel-Aty, M.A., Radwan, A.E. *Modelling traffic accident occurrence and involvement. Accident Analysis and Prevention*, Vol. 32, pp. 633-642, 2000.

[Adj16] A. Y. Omari-Sasu, Adjei Mensah Isaac * , R. K. Boadi *Statistical Models for Count Data with Applications to Road Accidents in Ghana.*, International Journal of Statistics and Applications, pp. 123-127, 2016.

[Agoki88] Agoki, George Samueli. *Characteristics of Road Traffic Accidents in Kenya*, PhD dissertation, University of Nairobi., 1988.

[Akai73] Akaike, H. *Information theory as an extension of the maximum likelihood principle.*, Petrov, B.N., Csaki, F. (Eds.), Second International Symposium on Information Theory. Akademiai Kiado, Budapest, pp. 267–281, 1973.

[Chow16] Chowdhury, R.I. and Islam, M.A. *Zero Truncated Bivariate Poisson Model: Marginal- Conditional Modeling Approach with an Application to Traffic Accident Data*, Applied Mathematics, Vol. 7, pp. 1589-1598, 2016.

[Chow15] Chowdhury, R.I. and Islam, M.A. *A Bivariate Poisson Models with Covariate Dependence*, Bulletin of Calcutta Mathematical Society, Vol. 107, pp. 11-20, 2015.

[Een07] Eenink, R., Reurings, M., Elvik, R., Cardoso, J., Wichert, S., Stefan, C. *Accident prediction models and road safety impact assessment: recommendations for using these tools.*, RIPCORD-iSEREST Deliverable D2. BASt, 2007.

[Eleni2000] Eleni Petridou, Maria Moustaki *Human factors in the causation of road traffic crashes.* Eur J Epidemiol, Vol. 16, pp. 819-826, 2000. https://doi.org/10.1023/A:1007649804201

[Erso2018] European Road Safety Observatory *Traffic Safety Basic Facts 2018 - Pedestrians.* 2018

[Huay12] Huayun Chen *Black Spot Determination of Traffic Accident Locations and Its Spatial Association Characteristic Analysis Based on GIS*, Journal of Geographic Information System, Vol. 4, pp. 608-617, 2012.

[gumbihi2015] Hudson Gumbihi, Solomon Koko *Highway of death: The Mombasa-Malaba highway taking more lives than Al-Shabaab*, The Nairobian. 2015.

[Jov86] JOVANIS, P.P., CHANG, H. *Modelling the relationship of accidents to miles travelled*, Transportation Research Record, Vol. 1068, pp. 42-51, 1986.

[Lord05] LORD, D., MANAR, A., VIZIOLI, A. *Modelling crash-flow-density and crash-flow-V/C ratio relationships for rural and urban freeway segments*, Accident Analysis and Prevention, Vol. 37, pp. 185-199, 2005.

[Hana18] HANA N. *Negative Binomial Regression Model for Road Crash Severity Prediction*, Modern Applied Science, Vol. 12, pp 38-46, 2018.

[Ken07] GOVERNMENT OF THE REPUBLIC OF KENYA *Kenya Vision 2030 The Popular Version* , 2007.

[Min2013] MINISTRY OF ROADS, KENYA *Roads 2000 Strategic Plan*, 2013.

[Mont08] MONTELLA, A. *Crash prediction models for rural motorways*, Transportation Research Record, Vol. 2083, pp. 180-189., 2008.

[Md2014] MD SAIDUL ISLAM *Crash prediction models for rural motorways*, Explaining Pedestrian and Vehicular Crashes in Conjunction with Exposure Measures, 20014.

[Njoro2016] NJOROGE, KIARIE *State invites bids for expansion of Nairobi-Mombasa highway*, Daily Nation, Nairobi. 2016.

[Norm1962] NORMAN, L.G. *Road Traffic Accidents; Epidemiology, Control and Prevention*, Public Health paers, World Health Organisation. Geneva, Vol. 12, pp. 1-10., 2018.

[Ntsa2017] NATIONAL TRANSPOART AND SAFETY AUTHORITY, KENYA *Kenya Import Data*, 2017.

[Pan13] PAN C., PRAKASH R. *Modelling Motorway Accidents using Negative Binomial Regression*, Journal of the Eastern Asia Society for Transportation Studies, Vol.10, pp. 1946-1963, 2013.

[Per93] PERSAUD, B., DZBIK, L. *Accident prediction models for freeways*, Transportation Research Record, Vol. 1401, pp. 55-60., 1993.

[Ron2019] RONALD FISA, CHOLA NAKAZWE, CHARLES MICHELO, PATRICK MUSONDA. *Modelling Deaths Associated with Road Traffic Accidents and other Factors on Great North Road in Zambia between the Years 2010 and 2016 Using Poisson Models* , The Open Public Health Journal, Vol. 12, pp. 68-77., 2019.

[Shan97] SHANKAR, V., MILTON, J., MANNERING, F. *Modelling accident frequencies as zero-altered probability processes: an empirical inquiry*, Accident Analysis and Prevention, Vol. 29, pp. 829-837., 1997.

[Smart02] Smart G, Mann E. *Death and Injuries from Road Rage: Cases in Canadian Newspapers.*, CMAJ; 167(4): 761-2., 2002

[Snip2014] Snipes, M., Taylor, D.C. *Model selection and Akaike Information Criteria: An example from wine ratings and prices. Wine Economics and Policy*, 2014 http://dx.doi.org/10.1016/j.wep.2014.03.001

[Tom04] Geurts Karolien. , Wets, Geert. , Brijs Tom., Vanhoof Koen. *Identification and Ranking of Black Spots*, Transportation Research Record., Vol. 1897, 2004.

[UclNeg19] UCLA: Statistical Consulting Group. *Negative Binomial Regression | R Data Analysis Examples*, https://stats.idre.ucla.edu/r/dae/negative-binomial-regression/ (accessed July 3, 2019).

[UclPois19] UCLA: Statistical Consulting Group. *Poisson Regression | R Data Analysis Examples*, https://stats.idre.ucla.edu/r/dae/poisson-regression/ (accessed July 3, 2019).

[UclZtPois19] UCLA: Statistical Consulting Group. *Zero-Truncated Poisson | R Data Analysis Examples*, https://stats.idre.ucla.edu/r/dae/zero-truncated-poisson/ (accessed July 3, 2019).

[Who2011] WHO, The Ministry of Public Health and Sanitation*Motorcycle-related road traffic crashes in Kenya Facts & figures : The Road Safety in 10 Countries (RS10) project*; 2011.

[Who2015] *Global status report on road safety. Geneva: World Health Organisation*; 2015.

[Who2018] *Global status report on road safety 2018. Geneva: World Health Organization*; 2018. Licence: CC BY- NC-SA 3.0 IGO.