ABDIKARIN ISAK HASHI

GIS-BASED SPATIAL ANALYSIS OF ROAD TRAFFIC ACCIDENTS: A CASE STUDY OF MOGADISHU CITY, SOMALIA

NEU

2023

NEAR EAST UNIVERSITY

INSTITUTE OF GRADUATE STUDIES

DEPARTMENT OF CIVIL ENGINEERING

GIS-BASED SPATIAL ANALYSIS OF ROAD TRAFFIC ACCIDENTS: A CASE STUDY OF MOGADISHU CITY, SOMALIA

M.Sc. THESIS

Abdikarin Isak HASHI

Nicosia

February, 2023

NEAR EAST UNIVERSITY
INSTITUTE OF GRADUATE STUDIES
DEPARTMENT OF CIVIL ENGINEERING

GIS-BASED SPATIAL ANALYSIS OF ROAD TRAFFIC ACCIDENTS: A CASE STUDY OF MOGADISHU CITY, SOMALIA

M.Sc. THESIS

Abdikarin Isak HASHI

Supervisor

Assist. Prof. Dr. Mustafa ALAS

Nicosia
February, 2023

Approval

We certify that we have read the thesis submitted by **Abdikarin Isak HASHI** titled **"GIS-BASED SPATIAL ANALYSIS OF ROAD TRAFFIC ACCIDENTS: A CASE STUDY OF MOGADISHU CITY, SOMALIA"** and that in our combined opinion it is fully adequate, in scope and in quality, as a thesis for the degree of Master of Educational Sciences.

Examining Committee	Name-Surname	Signature
Head of the Committee:	Assoc. Prof. Dr. Shaban Ismael Albrka	………………………
Committee Member:	Dr. Allam Musbah	………………………
Supervisor:	Assist. Prof. Dr. Mustafa Alas	………………………

Approved by the Head of the Department

.20/.02/2023

Prof. Dr. Kabir Sadeghi
Head of Department

Approved by the Institute of Graduate Studies

……/……/20..

Prof. Dr. Kemal Hüsnü Can Başer
Head of the Institute

Declaration

I hereby declare that all information, documents, analysis and results in this thesis have been collected and presented according to the academic rules and ethical guidelines of the institute of graduate studies, Near East University. I also declare that as required by these rules and conduct, I have fully cited and referenced information and data that are not original to this study.

Abdikarin Isak HASHI

..06./..02../..23..

Acknowledgment

I would like to extend my sincere gratitude to my advisor ASSIST. PROF. DR. MUSTAFA ALAS for his kindness, motivation, and knowledgeable counseling throughout this thesis. It has been a privilege for me to work and learn under his helpful advice and without his support and advice, this research could not have been done. I want to express my appreciation to all of the professors and instructors at Near East University for spreading knowledge and offering sincere and valuable support during the course.

My sincere gratitude and appreciation to my parents for their encouragement and support in helping me finish my master degree both directly and indirectly. Finally, I want to thank my brothers, sisters, and friends for helping me develop emotionally and physically throughout my life.

Abdikarin Isak HASHI

Abstract

GIS-BASED SPATIAL ANALYSIS OF ROAD TRAFFIC ACCIDENTS: A CASE STUDY OF MOGADISHU CITY, SOMALIA

Abdikarin Isak HASHI and Assist. Prof. Dr. Mustafa ALAS MA,

Department of Civil Engineering, Faculty of Civil and Environmental Engineering, Near East University, Nicosia.

February, 2023, 67 Pages

Concerns have been raised by an increase in traffic accident rates in recent years. This study highlighted the significance of geographical factors in the development of traffic accident data from Mogadishu, the capital city of Somalia, using spatial statistics and geo-information systems. The understanding of change is greatly aided by the spatiotemporal analysis when combined with spatial and statistical analysis. Researchers might gain a higher comprehension of the patterns of traffic accidents in complex metropolitan networks by examining four clustering studies. To apply the Using clustering on urban roadways, the point characteristics for inner-city crashes in vehicles between (2019–2021) have been registered in geographic information systems according to their x and y locations (GIS). The aforementioned areas were subjected to mean center analysis, kernel density estimation (KDE), Moron's I, and Hotspot analysis using ArcGIS. The goal of this study is to provide policymakers for road safety with a spatial understanding about how accidents are distributed in the city of Mogadishu so they can create efficient plans for managing traffic safety and develop the best possible solutions for the causes of such accidents when necessary.

Keywords: Spatial analysis, traffic accidents, GIS, Moran’s I, hotspot.

Table of Contents

Page

List of Figures

List of Tables

List of Abbreviations

AASHTO: American Association of Highway and Transportation Officials

GPS: Global Positioning System

GIS: Geographic Information System

HSM: Highway Safety Manual

KDE: Kernel Density Estimation

RTIs: Road Traffic Injuries

RTAs: Road Traffic Accidents

SANET: Spatial Analysis Along Network

STAA: Spatial Traffic Accident Analysis

CHAPTER I

Introduction

1.0 Background.

Nearly 1.3 million of the world's road accidents result in fatalities; 90% of these accidents happen within low-income areas, like countries in Africa, while 20–50 million non-fatal accidents result in disability (Chen, Simiao, et al. 2019). According to statistical records, low- and middle-income people contribute significantly to the occurrence of traffic accidents globally, and the overall cost of crushes is estimated to be close to 3% of the world's gross domestic product. (Raul Sanchez and Patricia Yaez-Pagans, 2019). Bicyclists, motorcyclists, and pedestrians are the most vulnerable road accident perpetrators. Shares of the cost of road accidents represent 3% of the GDP. From 2000 to 2018, there was a constant rise in the number of fatal accidents, between 1.15 million to 1.35 million (Mahmoodi Khaniabadi, Shadi, et al. 2022). Around the world, road accidents are a serious problem. One of the main causes of road accidents is irresponsible driving, which is innately influenced by infrastructure and traffic conditions, among other things (Hosseinian et al., 2020).

Road networks are intricate, dynamic, and unpredictable systems that are impacted by factors in the human, technology, and environmental spheres, leading to fatalities (Athiappan, K., et al. 2022). The likelihood and severity of traffic accidents depend on a variety of variables, such as road and environmental conditions, vehicle characteristics, the mental and physical health of other drivers, the presence of passengers, traffic flow circumstances, and many others (Zimmerman K, Jinadasa D, Maegga, et. Al 2015). In order to take the most effective actions to improve road safety and prevent accidents, it is vital to understand how the four components of human, vehicle, road, and environment affect the likelihood of traffic accidents. It was discovered that there was a significant correlation between both the road crash and its geometric features, including sight distance, curvature radius, and slope. Accident rates rise with peak hour traffic, tangent length and longitudinal pitch, but fall with curve radius (Thakali L, Kwon TJ, Fu LJ. 2015). The relation between speed variations and longitudinal gradient is another efficiency-related factor that contributes to traffic congestion, decreased safety, and an increased chance of accidents. Driving range and driver behaviors, including passing other vehicles, are impacted. The risk of occurrence rises as a result of the magnified impacts of pavement and longitudinal surface friction in the falling slope. (Khanal et al., 2016)

Road accidents are more likely to occur when the following human elements are present: the driver's age, alcohol use, the usage of cell phones whilst driving, distraction from billboards, and the physical condition of the pedestrian and driver. Fog, ice rain, and rain are three environmental elements that are linked to road accidents and will be more relevant to their occurrence (Mahmoodi Khaniabadi, Shadi, et al. 2022). Road accidents in bad weather were highly significant, according to accident analysis based on road type and environmental exposure (Martinez, Sebastian, Raul, and Patricia, 2019). The study found that the texture depth of the road surface and the accompanying skid resistance rating have a bigger influence on the likelihood of an accident.

Road traffic accidents (RTAs) are destroying property and killing hundreds of people every day without regard to race or gender, yet little effort has been made to lessen their severity. However, it is one of the incidents that poses the greatest risk of loss of lives and property damage worldwide. In order to reduce the severity of the harmful consequences on human and damage to property, it will be helpful to identify the main causes of traffic accidents.

The number of traffic accidents has considerably increased in recent years. According to the World Bank, a fatality rate of 85.6 per 1,000 vehicles is acceptable; but, in Mogadishu, the speed is dangerously high. The severity of a road is not random; it has patterns that can be recognized and avoided. The conclusion is that accidents are "events which may be evaluated, analyzed, and prevented" (Gissane W. 1965). Accidents are defined as "Fatalities are not fated; accidents do not just happen; disease is not random; they are created," by the workers' health organization.

In order to better understand how many elements like route shape, driver behavior, traffic patterns, and environmental elements affect the likelihood of crashes, transportation safety studies have been conducted. Without thorough knowledge of both the incident itself and its surrounding conditions, the impact of those elements on road crashes can't be fully understood. The Highway Safety Manual (HSM) offers common scientific methods and information to assist transportation officials in making wise decisions about traffic safety (AASHTO, 2010).

There are numerous ways and procedures that can lower the accident rate and can save many lives in order to contribute to our society and minimize the number of accidents occurring in our daily lives.

1.1 Problem statement

Over the past several decades, there have been a considerable increase in traffic-related fatalities and injuries, and Somalia is one of the countries with the worst scores for road safety. For every 100,000 inhabitants, there are around 27.1 traffic-related fatalities, and the annual rate of urban expansion is 4.25 percent, according to the WHO. As a result, a lot of research has been done in recent years on the instruments and techniques used to examine accidents and road design. Geographical Information Systems (GIS) stand out among these technologies for their capacity to do in-depth geographical analysis.

1.2 Aims and objectives of the study

1.2.1 Aim of the study

The primary objective of accident hotspot analysis is to identify and generate the information needed to assist the decision makers in adopting suitable measures to prevent and to reduce the accident happenings. This study's main goal is to look at how road accidents are distributed by locating accident hotspots using GIS & spatial statistics. Through the application of spatial statistics within the context of geographical information technology, this study seeks to identify and represent the accident hotspots in Mogadishu city by modeling actual accident location information in conjunction with a number of spatial attributes.

1.2.2 Objectives of the study

- ✓ The study's specific goals were to examine road crashes using a methodology or technique and to identify and avoid them using GIS tools.
- ✓ To summarize the existing situation so that governmental organizations can focus enforcement efforts in zones with a high crash risk and take appropriate action. The government can effectively and efficiently ensure the public safety by using this optimization plan and procedures.

1.3 Research Questions

How can GIS be used to evaluate the effectiveness of existing road safety measures in a specific region?

Can GIS be used as a predictive tool for road traffic accidents in a specific region?

How can GIS-based spatial analysis be used to support decision making and policy development related to road traffic accidents in a specific region?

1.4 Scope of the study

The model developed in this work is to identify hotspots on some roads. Both spatial or non-spatial data are analyzed by GIS in order to determine the geographic distribution of traffic accidents. As a result, it is simple to use GIS to add a model for locating accident areas on roads.

1.5 Limitations

Although the scope of this research is limited to identifying hot spots in particular years, it may also be used to pinpoint cold spots. This work can be used as a starting point for future research on the impact of road conditions, driving behavior and time of day. An example would be a study that determines which particular day of the week has the highest rate of accidents and where they occur.

CHAPTER II

Literature Review

2.0 Background

Many people's lives and entire populations have improved thanks to motorization, however there are costs associated with the advantages. Even while the number of people killed in accidents in high-income nations has been trending downward in recent years, the burden of road traffic injury—in terms of societal and economic costs—is increasing dramatically for the bulk of the world's population. In poor nations, road traffic accidents account for 90% of years lost to disability-adjusted life expectancy and more than 85% of all fatalities.

When an automobile collides with another automobile, an animal, a person, road wreckages, or another stationary object like a structure, pole or tree, the result is a traffic crash, also known as a motor automobile crash, automotive accident, or car crash. Traffic accidents commonly cause damage to individuals and property as well as cost the affected parties and society money. The most dangerous situation that individuals frequently experience is on the roads, although the number of fatalities from these accidents receives less public attention than that of other, less frequent types of catastrophes.

In 2018, deaths from road accidents were the leading cause of death, according to WHO. With the current pandemic (coronavirus), it's possible that the ranking of traffic accidents as one of the leading causes of mortality will slightly shift. The pandemic has decreased the number of persons traveling as a result of the barrier measures put in place during the pandemic, even if this may not show that the risk of accidents will decline in the coming years. Approximately 3700 persons every day worldwide pass away in a car, truck, bus, bicycle, motorcycle, or pedestrian accidents. The position of road accidents between 1998 and 2020 was forecast by WHO (Table 1).

Table 1

Range of Road Accident Prediction.

NO	1998 Disease or Injury	2020 Disease or Injury
1	Lower respiratory contaminations	Ischaemic heart disease
2	HIV/AIDS	Unipolar major depression
3	Perinatal conditions	Road traffic accident
4	Diarrhoeal diseases	Cerebrovascular diseases
5	Unipolar major depression	Chronic obstructive pulmonary diseases
6	Ischaemic heart disease	Lower respiratory contaminations
7	Cerebrovascular diseases	tuberculosis
8	Malaria	War
9	Road traffic accident	Diarrhoeal diseases
10	Chronic obstructive pulmonary diseases	HIV/AIDS

2.1 Traffic Accident Causes

The very worst thing that may occur to a road user is a traffic accident, notwithstanding how often they happen. The worst thing is that we don't take anything away from our road trips. Most people who use the roads are fairly aware of the general safety procedures and regulations that are in place while using them, but accidents and wrecks only happen as a consequence of carelessness on the part of those who use the roads. Josephine and colleagues draw attention to some of the typical causes of accidents.

2.1.1 Speeding

The majority of deadly accidents are the result of excessive speed. People have a natural desire to achieve. We shall, however, always adhere to one or more of them when sharing the road with other motorists. Speed raises the risk of an accident and the seriousness of any injuries. Accidents involving faster vehicles are more likely to occur, and the severity of such accidents will also be greater.

2.1.2 Inattentional driving

Any form of distraction might make it difficult to operate a vehicle. Mobile device use while driving is becoming more dangerous. Compared to not using a phone while driving, using one raises the likelihood of an accident by roughly four times.

2.1.3 Unsafe Road construction

The design of a road may have a big influence on how safe it is. Roadways should ideally be built with everyone's safety on the road in mind. This would require ensuring that there are enough facilities for walkers, motorcycles, and bicycles. To lessen the danger of injury to these road users, strategies like sidewalks, safe crossing sites, bike lanes, and other traffic calming measures could be required.

2.1.4 Dangerous vehicles

Vehicle safety is essential if collisions are to be avoided and the likelihood of serious injuries is to be decreased. If various UN automobile safety laws were enforced in nations' manufacturing and production standards, they may save a lot of lives. In an emergency, a well-built, well-kept automobile with dependable tires, brakes, and suspension will be simpler to control and better equipped to avoid collisions.

2.1.5 Poorly enforced traffic regulations

If traffic laws governing driving while intoxicated, wearing seat belts, adhering to speed limits, wearing helmets, and utilizing child restraints are not imposed, the anticipated decline in injuries and fatalities from traffic accidents linked to particular habits will not occur. As a result, it is likely that traffic laws won't be followed and won't have much of an impact on behavior if they aren't enforced or are considered to be disregarded.

2.2 Impacts of Traffic Accident

There are incredibly wide-ranging effects of road accidents. For everyone involved, its impacts might be physical, emotional, social, and economic.

2.2.1 Physical effects of automobile collisions

Head and brain trauma, such as a traumatic brain injury, neck injuries, such as disk damage, neck strains and whiplash, and spine or back injuries, such as strains, fractures, or sprains are some of the most severe physical ailments that are usually linked to vehicle accidents. Physical harm with long-term effects includes traumatic brain injuries (TBIs), paralysis, and amputations (Health safety & Environmental Encyclopedia).

2.2.2 Emotional impact of automobile collisions

Even while certain accident-related impairments are not immediately apparent, they may nonetheless have long-term consequences. Car accident injuries that result in emotional or psychological distress are among those that commonly have long-term impacts. Following a car accident, victims may experience mental pain, emotional distress, dread, and fury.

2.2.3 General impacts of automobile collisions

Serious accidents cause fatalities. As a result of tragic traffic accidents, many people have lost their jobs. Road accidents have resulted in the loss of many people's crucial body parts. The output at the hospital would be much diminished if people were to work instead than just lounging about. To boost their effectiveness, efficiency, profitability, and production, all businesses rely on human labor. Their squad must remain tough in this situation.

The loss of victims' productivity, the cost to the legal system, the cost of their agony and suffering, and the decline in the quality of life for the sufferer and their family are some of the social effects of traffic accidents. The loss of production is responsible for the bulk of all societal costs.

A new World Bank research suggests that reducing traffic-related fatalities and injuries might greatly increase low- and middle-income countries' long-term revenues. In the research "The High Toll of Traffic Injuries: Unacceptable and Preventable," a brand-new international method for assessing the financial effects of traffic safety and studies is presented.

Although road traffic deaths and accidents are well documented, little is known about the connection between economic growth and traffic injuries.

The new study quantifies the relationship between spending money on people and spending money on road safety.

The study comes to the conclusion that governments that do not invest in road safety might potentially miss out on per capita GDP growth of anywhere between 7 and 22% over a 24-year period. Authorities must prioritize expenditures in road safety in order to remedy this. With over 1.25 million deaths occur annually as a result of delays, which also lower productivity and reduce economic potential.

By 2030, it is anticipated that road traffic injuries (RTIs) would rank as the fifth largest contributor to the global disease burden, raising serious public health issues, according to Josephine et al. (2012). But regrettably, accidents on the road frequently result in fatalities, damage to infrastructure, and issues with people's health (Aldegheishem et al, 2018). A very traumatic event like RTI can hurt a person's body, psyche, and property. Ihueze et al. (2018) found that 93% of these fatal traffic accidents occurred in middle- and low-income countries (2018). According to the study, Africa is not the place to go to for best practices in road safety (Road Accidents Statistics Worldwide (2019)). In 2018, there were 26.6 road traffic fatalities for every 100,000 people on the continent, according to a WHO estimate. Recognizing the terrible situation involving road deaths and injuries and taking the necessary steps are therefore essential. Most nations lose 4% or so of their GDP as a result of traffic accidents. Road accidents have a significant psychological impact on quality of life, but they also have a significant socioeconomic cost to the country (Wim et al, 2020). Medical expenditures, lost productivity, property damage, human costs and administrative costs are some of the costs associated with traffic deaths.

RTAs have many different contributing elements, which can be categorized into three groups: highway, vehicle, and driver factors (Olemo et al, 2016). All proximal incidents associated with the driver's actions that could cause severe injury are referred to as "driver factors." RTIs have an adverse effect on Pacific peoples, as raised by Josephine et al. They emphasize the need to address issues like drinking and driving, bad driving, not wearing seatbelts, and poorly maintained cars and roads.

Traffic accidents are typically caused by vehicles traveling at high speeds and/or motorists disobeying traffic signals (Batamag et al, 2020). Drivers frequently overload their vehicles, which leads to numerous traffic accidents on our roadways, particularly in developing nations. Understanding the many elements that lead to road accidents is crucial (Zhou et al, 2020). Through empirical study, a number of factors have been discovered, the majority of which are related to the environment, drivers, cars, and roads. These results were obtained from the research carried out by Batamag *et al*, 2020.

A thorough inquiry into a traffic accident revealed numerous elements that may or may not have caused the collision. Frequent occurrences may point to a typical accident type or trend at a particular site. Although the causes of disability and death are not well quantified, published data indicates that RTIs and the hazards associated with them are important. Road accidents are foreseeable and avoidable, but reliable data is necessary to comprehend how road safety actions may be successful. To solve this global issue, cost-effective preventive methods can be developed. In order to reduce fatalities, Aldegheishem et al, 2018) suggested that a protocol be created to foresee or prevent traffic incidents at the severe level. Reduced traffic accidents will result in a drop in the number of fatalities they cause. The largest contributing elements to accidents are the driver's behavioral errors. The last group of accident causation factors is ranked as malfunctioning vehicle systems. Vehicle systems are a significant contributor to a significant number of traffic accidents (Zhou et al, 2020). The injuries brought on by driver mistake and other factors related to the road had received the majority of attention in previous studies of auto accidents.

2.3 GIS program in analyzing road accidents

The economic, environmental, social, and sustainable growth of a city are all at risk from traffic accidents. For urban parts of a metropolis, traffic control was always a problem. Geographic information systems (GIS) enable individuals locate the necessary information on traffic, roads, locations, and directions to a destination. GIS provides an excellent solutions to the road accidents of a city.

Many traffic agencies employ GIS technology since it is a popular tool for analyzing hot spots and visualizing accident data (DeepthiJayan and Ganeshkumar, 2010). The safety experts can identify the parts having a larger frequency of crashes by comparing those sections to other similar places with an understanding of spatial and temporal crash patterns. These areas are known as hotspots (Elvik, 2008 and Mohaymany 2013).

The methodical process of locating road segments with an unacceptably high crash risk is known as "hotspot identification." In order to further diagnose specific issues, choose cost-effective solutions, and prioritize treatment sites, a limited subset of road network locations are chosen from a broad population as part of a low-cost road safety management approach. In the literature, these locations are referred to as dangerous sites, hotspot, black spots, priority investment spots, collision-prone locations, or hazardous locations (Thakali et al, 2015).

Numerous scholars have been examining GIS technology and their uses in the spatial pattern of crash investigation since 1990. These cross-reference models provide geographical spatial queries, segment & intersecting analysis, pattern analysis, accident analysis and proximity analysis. Through traffic safety research, the impacts of numerous elements on safety performance are explored. These include how geographic circumstances, environmental factors, such as weather, and geometric elements of road design, affect the likelihood of accidents (Khan, G.; Qin, X.; Noyce, 2006).

The identification of hotspots on roadways has made tremendous strides in the last few years. This is made feasible through the use of GIS and GPS applications in transportation research. The literature has employed a variety of hotspot detection techniques, including global indexes like Geary's C, GOG and Global Moran's I.

In a geographic hotspot identification of point data, KDE is used to address the number of measurements inside a unit area at a certain place (point pattern analysis). Kriging is a more advanced method of spatial analysis that is usually utilized in a variety of study domains (Thakali et al, 2015).

The limitations of planar spatial analysis for point events constrained to linear networks are solved by the SANET toolset. The above tool case is a spatial network analysis that identifies the network segments with high intensities and assesses the intensities of points on a network. It is significantly more effective than the planar spatial analysis approach for networks with error-prone Euclidean distances (Düzgün et al, 2009). The STAA technique analyzes historical accident data using a hazard-based methodology that takes socioeconomic factors, accident frequency, and severity into account (Zahran et al, 2017). Contrary to SANET-KDE, STAA stipulates that the accident spots cross the centerline of the roadway. The accident points' original coordinates are preserved as a result. It may be used to analyze both individual roads and networks of roads (El-Said Mahmoud et al, 2019).

The KDE, nearest-neighbor, K-function, hierarchical clustering (HC), dangerousness index (DI), and climbing are the approaches utilized in cluster analysis of traffic accidents. The nearest-neighbor and K-function approaches can show that there is a propensity for clustering on a road stretch, but they are unable to identify the precise location where it happened. Therefore, the localization of the clusters inside the section is not aided by these approaches. The real cluster position inside a sector or a network may be determined using the KDE and DI approaches. The statistical significance of clusters cannot be determined using the HC approach. It could only distinguish between traffic accident clusters. The DI approach, which uses "points of measurement," is a specific instance of KDE. Although the climbing approach can locate clusters, it is very vulnerable, suggesting that even a minor change in the placement of incidents outside of the cluster can have a significant impact on the cluster's significance.

According to Anderson, T.K., KDE permits the dispersal of the risk of an accident. KDE is better suited for display than hotspot identification, as Plug et al. noted. There is currently no comprehensive analysis of KDE's statistical relevance in the literature. For the analysis of incidents in a 1D linear space, network KDE is more effective (e.g., a road). Using the kernel function KDE+, an enhanced KDE technique assesses the probability density function of the event points. The plus symbol shows the importance of carefully choosing relevant clusters. The technique has a drawback, too, in that it works best for event sites along the road segments between junctions because there are more accidents at intersections than there are hazardous spots there (Andráik et al, 2015).

2.4 Geospatial assessment methods

GIS spatial approaches are frequently used in geographical studies to identify road traffic incidents. This spatial technique pinpoints the crash's exact position and assesses the map's visual appeal using distribution patterns. The following list includes the GIS tools utilized in this study:

- **Mapping clusters**
- **Spatial analysis techniques**
- **Hotspot Analysis**

2.4.1 Mapping clusters

The mapping clusters do cluster analysis to identify hotspots, coldspots, geographical outliers, and comparable features or zones. This method is particularly useful when taking action is necessary based on the location with one or more clusters. Below are detailed descriptions of Global Moran's I and Incremental Spatial Autocorrelation, two forms of mapping clusters that have been investigated:

2.4.1.1 Moran's Index Tool

The Spatial autocorrelation (Moran's I technique) operates both on feature sites & feature values simultaneously, not just on attribute values or feature locations separately. To assess the presence of clusters inside the spatial arrangement, one sort of indicator of spatial association is used. The outcome of the cluster and outlier analysis, which is based on the Z score, P-value, and cluster type for each data set, is a Local Moran's I index value. Z scores and p-values are statistical significance measures that show us, feature by feature, whether we can successfully reject the null hypothesis or not. They effectively show if there is more apparent resemblance (or dissimilarity) between a feature's values and those of its neighbors than one might anticipate from a random distribution (Amiri et al, 2021).

2.4.1.2 Incremental Spatial Autocorrelation

We will determine Moran's I index value and related Z score, which denote a statistical significance at various threshold distances, using the Incremental Spatial Autocorrelation tool. Using the Getis-Ord Gi* function, the mapping cluster will select the minimum distance with the highest Z score. The best grouping of minimum and maximum values may be found at the distance with the highest Z score. The results of the ISA tool are displayed graphically along with various distance thresholds and the corresponding Z scores. For the best clustering, the initial peak or tallest peak may be selected from among many values. The scale of your study will matter with a lot of analysis of spatial data you perform. For instance, the Hot Spot Analysis tool's fixed distance band default Conceptualization of Spatial Relationships asks you to enter a distance value. You will be requested to specify a radius value for numerous density tools. The distance you choose should be in line with the scope of the inquiry you are attempting to address or the scope of the proposed remedy (Amiri et al, 2021).

2.4.2 Methods of spatial analysis

A very simple type of point pattern analysis includes summary statistics such as the standard distance, standard deviational ellipse and mean center. Such point pattern analysis techniques were widely employed prior to the recent revolutionary advancement in the power and capability of computers. There have been two categories for more complex spatial analysis techniques: approaches based on density and approaches based on distance (Manuel, 2017). The three most well-known spatial analysis methods are LDE, PDE, and KDE, as was before indicated. KDE and the mean center were the only variables employed in this experiment.

2.4.2.1 Method of Kernel Density

A variety of spatial software is available to comprehend patterns of shifting geographic places. Kernel density estimation is one of the most suitable tools for this (Chainey and Ratcliffe, 2005). Numerous benefits of this study include the use of K-means clustering techniques and the identification of static hotspots. The primary benefit of this approach relates to the assessment of accident risk increase rate. Risk expansion is the geographic region that, according to a geographical study, has the highest accident rate around a certain cluster. A contractual spatial analysis unit can be created utilizing the density technique consistently over the whole study region in order to define criteria for grouping and comparison. According to prior research, traditional kernel density estimation (KDE) was used in a number of safety studies to identify crash danger locations (Anderson, 2009, Pulugurtha et al, 2007). Density estimate is based on the notion that a point pattern has a density at any position within the research region, not simply where an event takes place or is visible (Lloyd, 2006; O'Sullivan and Unwin, 2002). In order to estimate the kernel density, a symmetric plane must be placed on each point.

2.4.2.2 Mean-center assessment method

The mean center establishes a new category of point features, but each feature in this class serves as a representative of a mean center. In addition to the X and Y mean center values, the mean dimension field is recognized as an output feature property. The mean center identifies the hotspot centers or geographic centers for categorizing the features.

2.4.3 Hot Spot research

A hot spot is a place or a small part that falls inside a specific range and has a higher concentration of occurrences. Density, cluster mapping, and event gathering are the three primary approaches used to determine the desired hotspots for accidents. Getis and Ord first introduced the Getis-Ord (Gi*) spatial statistics, an autocorrelation technique (Getis & Ord, 2010). This technique addresses the problem that other earlier techniques had with distinguishing between hot and cold areas. The G* can detect spatial clustering and can tell where there are high and low value concentrations in a set of nearby inputs (Songchitruksa & Zeng, 2010). This method finds low or high equivalent values by comparing each feature to its surrounding characteristics. A feature must not only have a high value in and of itself but also have high values in the features that surround it in order to be labeled a hotspot.

CHAPTER III

Methodology

3.0 Overview

This chapter presents how the GIS software works while the study's main goal is to look at how road accidents are distributed by locating accident hotspots using GIS & spatial statistics. The primary goal of accident hotspot analysis is to locate and gather the data necessary to help decision-makers adopt the most effective strategies for preventing and minimizing accident occurrences. Combinations of approaches and procedures will be used in the research analysis to thoroughly investigate the existence or absence of accident clusters. If they exist, it indicates that certain areas of the research region are more likely than others to experience accidents. As a result, additional money and resources must be set aside for research and problem-solving in these areas.

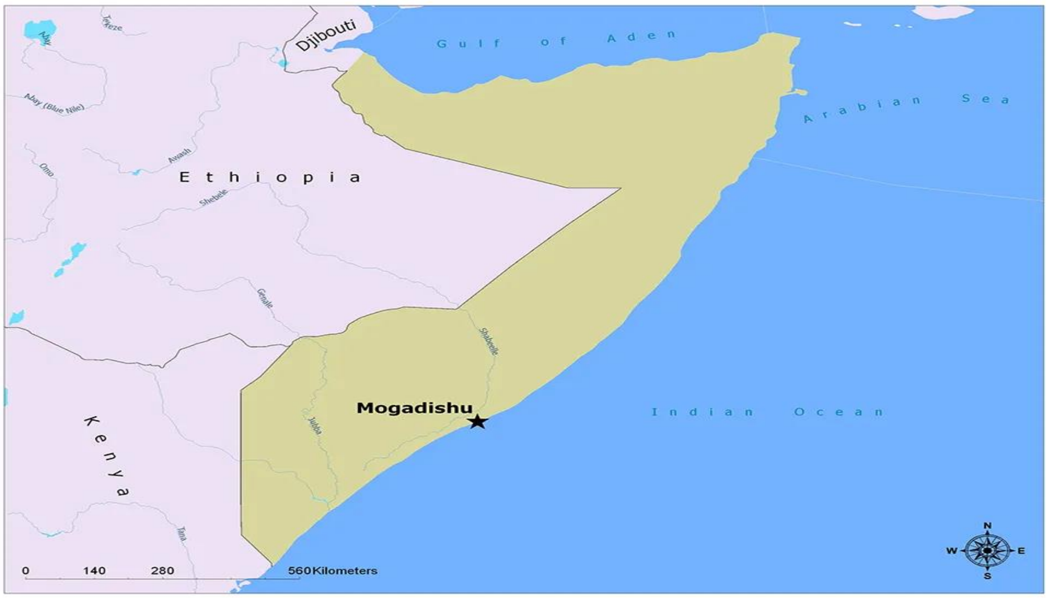
The initial phase in the research was to identify the accident locations on roads in Mogadishu City as shown in Figure 2. After data gathering, accident locations were converted from Word files into Excel sheets so they could be imported into geodatabases. The properties of accident points were organized in databases before data were prepared for transference to a GIS. By identifying and removing points with the incorrect coordination and location, the quality of the data gathered was improved.

3.1 Study Area

The capital city of Somalia is Mogadishu, which is located in eastern Africa close to the Indian Ocean. It serves as the state capital of Banadir. There are 2,497,000 people living in Mogadishu, a growth of 4.56 percent over the previous year. (Integrated Land Information for the Mogadishu Administration, 2022). There are 91 square kilometers on the surface. Mogadishu is connected to other cities in Somalia as well as to other nations through roads. A significant road network divides the capital's physical form into various grid patterns. Almost all of the main roads in Mogadishu's 1,000 km long road network are in bad shape as a result of improper maintenance and repairs brought on by the protracted civil conflict. There are four major roads in the nation, and because they were all constructed more than 30 years ago and did not receive regular maintenance, they are all full of potholes that make it difficult for large vehicles to go on them.

Figure 1

Mogadishu Map



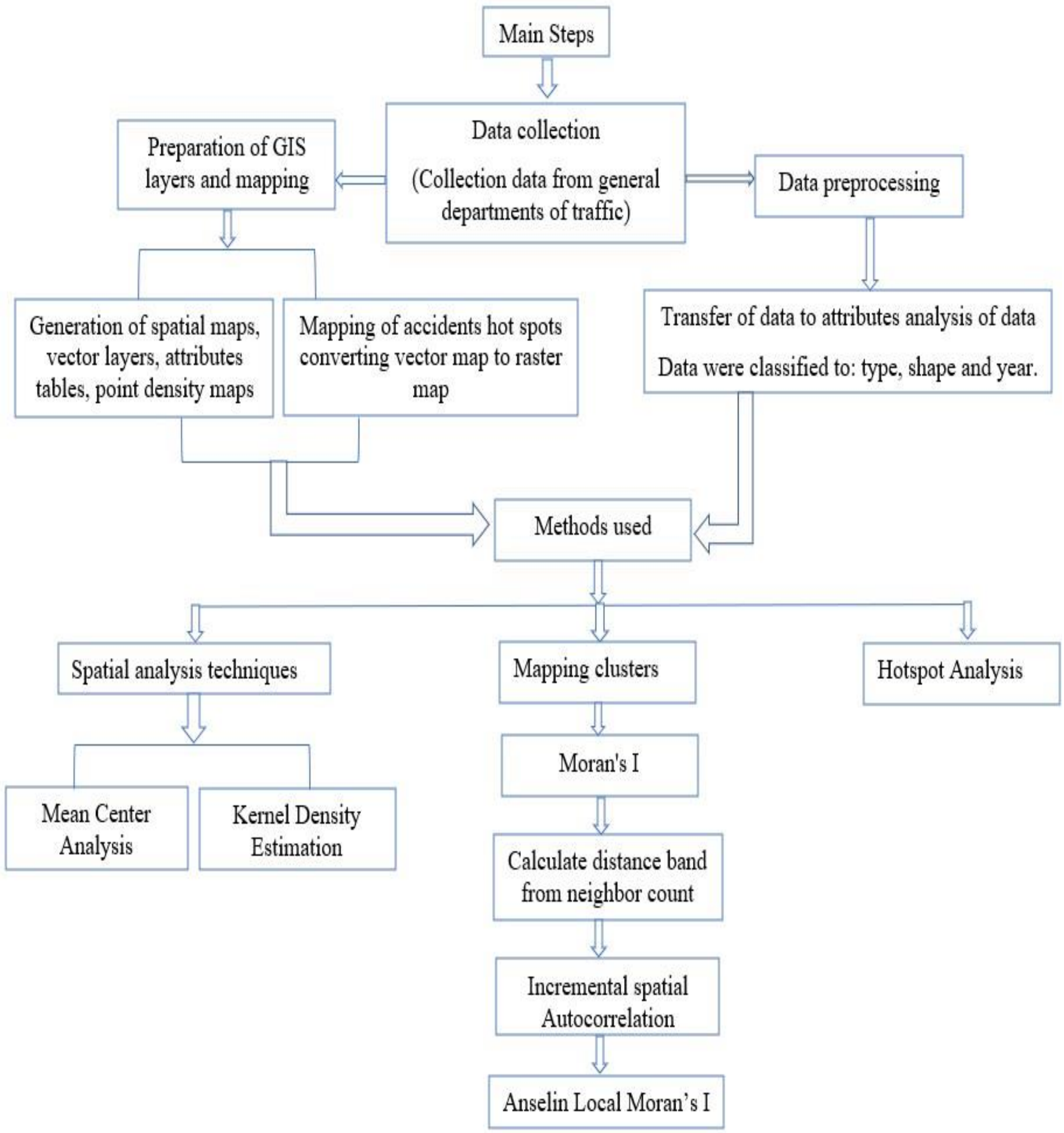
3.2 Data Collection

Road crash data, including geometric features and human and environmental elements, were gathered from the Mogadishu traffic police headquarters to execute the accident identification of hotspots through geo-spatial techniques and to undertake factor analysis. Total collision accidents during the past three years, which include all of the months from 2019 to 2021, have resulted in 7560 fatal accidents and terrible injury accidents. Three years of data were used to level out random deviations. The police agency gave facts on the accident, including the date, location, kind of incident, and other specifics like the cars involved.

To do statistical mapping and modeling with specific data, researchers must have a comprehensive knowledge of the temporal and geographical events that led to the accident. The accident sites in this study were analyzed using the Arc GIS program. High levels of accuracy were field validated for the pertinent data used. The ArcGIS program was used to carry out hotspot analysis procedures. X and Y in this study identify the places where transactions take place.

Figure 2

Outline of the Study



3.3 GIS- Based Analysis

On maps, charts, and tables, the traffic police's record of accident causes were summarized. Four distinct types of ArcGIS analyses were carried out to evaluate the traffic police accident records and locate hotspots throughout the research area: cluster analysis, hotspot analysis, KDE and mean center analysis.

KDE: The density surface shows the high-to-low spatial changes in accident frequency. The statistical significance of both the high and low accident frequency at various places within the research region is not revealed, though. These variances could arise by chance and are not always connected to a cause.

Mean center examination: it gives the optimum position for traffic control center in that particular location.

Spatial autocorrelation analysis: means the relationship between one point and the other.

Step 1: Moron' 1: it tries to check if a point has a relationship with their neighbors. If we find a relationship, we go to step two.

Step 2: Calculating Distance range from neighbor count: this method tries to compute the minimum distance from which one point can have a relationship with another. After we calculate that distance, we use it for step three.

Step 3: Incremental spatial autocorrelation: we divide that distance by two, after we do that, we form the X axis data. A graph will be generated by the tool with peaks, which means we will have a hotspot in those regions.

Step 4: Anselin Morans I: this is the results and the final step of cluster analysis which shows the physical locations of points which has relationship with each other. Grouping the points to indicate locations with high- high or low- low accident occurrence near each other, and if such locations are statistically significant.

Hotspot Analysis: shows the points with the highest/lowest frequency of accident occurring and the resources have to be put in place to ensure the accidents don't occur.

3.3.1 Mean Center Analysis

Using the frequency of the accident as a weight, using this tool, the geographic average of the traffic accident along the roadway system was computed. The geographic center values or frequencies are drawn toward accident sites with higher frequency features using the weighted mean center technique. The results of this calculation might help the analyst identify areas of the study area where crushes are more likely to occur. To illustrate any discernible movement in the mean center, the computation was specifically performed for each year and shown in a separate window.

3.3.2 Kernel Density:

Estimating the density of a cluster of parameters in the area around those parameters is the task of the kernel density tool. For both point and line characteristics, it is calculable.

It might be useful to investigate how utility lines or streets affect an environmental habitat. One point might stand in for several observations by using the population field to give some attributes more weight than others. For example, one address may represent a condo building having seven units, or the weighting of certain offences in calculating total crime rates may vary. A divided highway may have a greater influence on line characteristics than a narrow rural road.

3.3.2.1 How to calculate Kernel Density

For point features

Kernel density determines the density of point characteristics near each output raster cell.

Theoretically, a surface with a gentle curve surrounds each point. The surface value rises with increasing proximity to the point until falling to its maximum value at the search radius from the point. The only useful neighborhood is a circle. The population parameter value for the location, or 1 if NONE is specified, determines the volume below the surface. The values of all the kernel sections where they overlap the raster cell center are added to determine the density at each output raster cell. The quartic kernel function established by Silverman serves as the foundation for the kernel function (1986, p. 76, equation 4.5).

The value of each item dictates how several times the point will be tallied if a populations field setting other than NONE is selected. A point with a value of 3 would be counted as three points, for example. It is possible to use integer or floating-point numbers.

Unless otherwise provided in the Output Coordinate System environment option, a unit is automatically chosen based on the linear unit in the projection specification of the input point feature data. The estimated density for the cell is multiplied by the appropriate factor before being recorded to the output raster if an area unit is used.

For instance, the output area units will by default be square kilometers if the input units are meters. When comparing measurements in meters and kilometers, there might be a factor of 1,000,000 difference in readings (1,000 m, x 1,000 m).

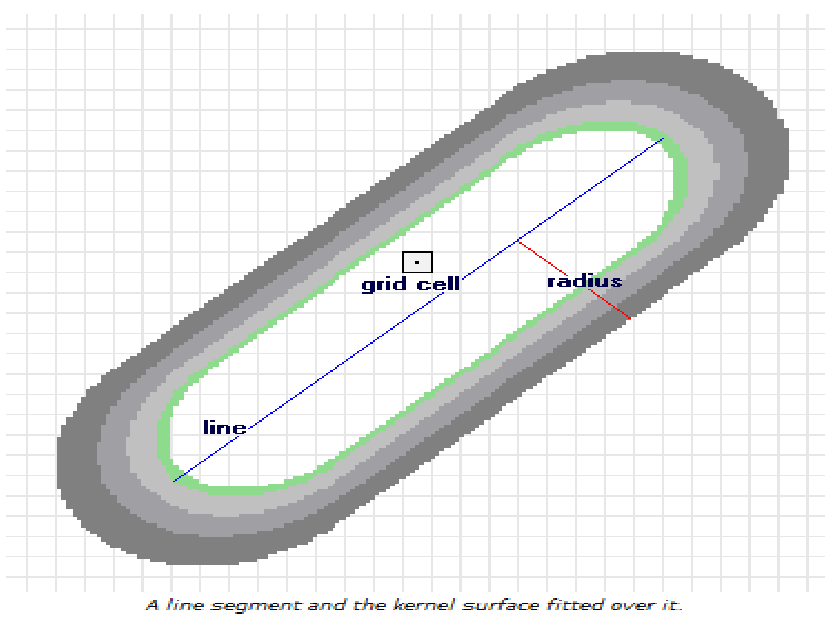
For characteristics of lines

The density of geometric properties around each output raster cell may also be determined using kernel density.

Conceptually, each line is covered by a surface with a little curvature. As you get farther from the line, its value decreases until it zeroes out at the selected Search radius. It is the highest along the route. The surface is made to have a volume underneath it equal to the product of the value of the population field and the length of the line. The frequencies of all the kernel surfaces where they overlap the raster cell center are added to determine the density at each output raster cell. The quartic parameter for point densities proposed by Silverman serves as a model for the kernel function for lines.

Figure 3

Line Segment of Kernel density



In the picture above, a line segment is fitted across the kernel surface. How much the line segment contributes to density depends on the rate of the kernel surface at the raster cell's center.

By default, the output coordinate system environmental parameter is determined by the projected feature of the input polyline feature data.

The length and area units are converted each time the output Area unit's ratio is given. The resulting area units should default to SQUARE KILOMETERS if indeed the linear unit is metered, for instance, and the resultant line density units will be altered to kilometers per square kilometer. Eventually, a factor of 1,000 will separate the density figures while measuring a neighborhood scale factor of meters to kilometers. You may manually control the density units by selecting the appropriate factor. Adjust the frequency from the standard of kilometers per square kilometer to meters per square meter by setting the area units to square meters. Set the area units to sq. miles to have your output's density stated in miles per square mile.

The length of the line is presumed to be equal to its actual length times the value of the population data for that line if a population data other than NONE is given.In the picture above, a line segment is fitted across the kernel surface. How much the line segment contributes to density depends on the rate of the kernel surface at the raster cell's center.

By default, the output coordinate system environmental parameter is determined by the projected feature of the input polyline feature data.

The length and area units are converted each time the output Area unit's ratio is given. The resulting area units should default to SQUARE KILOMETERS if indeed the linear unit is metered, for instance, and the resultant line density units will be altered to kilometers per square kilometer. Eventually, a factor of 1,000 will separate the density figures while measuring a neighborhood scale factor of meters to kilometers. You may manually control the density units by selecting the appropriate factor. Adjust the frequency from the standard of kilometers per square kilometer to meters per square meter by setting the area units to square meters. Set the area units to sq. miles to have your output's density stated in miles per square mile.

The length of the line is presumed to be equal to its actual length times the value of the population data for that line if a population data other than NONE is given.

Kernel density calculation formulas:

The procedures for choosing the main search radius and calculating the kernel density for points are described in the following formulas.

Estimating the point density

The following equation yields the estimated density at a new (x, y) location:

$$Density = \frac{1}{(radius)^2} \sum_{i=1}^{n} \left\{ \frac{3}{\pi} . pop_i (1 - \left(\frac{dist_i}{radius} \right)^2)^2 \right\}$$

$$For\ dist_i < radius$$

where the input points are:

- I = 1, 2,..., n. Only points that are inside the range of (x, y) location should be included in the total.
- Popi is an optional parameter that indicates the population value of the variable of the point I.
- disti is the separation of the point I and the coordinates (x, y).

The number of points or, if one was supplied, the sum of the population field is multiplied by the estimated density. With this adjustment, the spatial integral is no longer necessarily equal to 1, but rather equal to the number of points (or sum, or population field). This version makes use of a Quartic kernel (Silverman, 1986). Every place you wish to estimate the density will need a calculation of the formula. The computations are applied to the center of each cell in the output raster since a raster is being constructed.

Default search radius

The method for choosing the main search radius was enhanced for ArcGIS 10.8.

The main search radius, commonly referred to as bandwidth, is calculated using the following algorithm:

1. Determine input points' mean center. If a Population data was provided, the values in that field will be used to weight this computation as well as any subsequent calculations.
2. Determine the separation of each point from the (weighted) mean center.

3. Determine the median (weighted) distance between these points, Dm.
4. Determine the Average Distance (weighted).
5. Use the calculation below to get the bandwidth.

$$SearchRadius\ = 0.9 * \min\left(SD, \sqrt{\frac{1}{In(2)}} * D_m\right) * n^{-0.2}$$

where:

- Dm is the weighted median's separation from the center of the weighted mean.
- • When a population variable is given, n represents the field's total value; if not, n represents the number of points.
- The standard distance is SD.
- SD is the standard distance.

In order to generate an arbitrary heat surface representing the range of values for road accidents spanning high to low, the kernel density estimation was applied to the data. This measurement calculates the percentage of all accidents that are predicted to occur at each particular junction and location on the map.

3.4 Cluster Analysis:

3.4.1 Spatial Autocorrelation (Morans I),

This spatial Autocorrelation analysis determines Morans I and use both parameter location and parameter values concurrently. When a group of traits and an related attribute are present, It determines if the pattern displayed is random, distributed, or clustered. The program calculates the value of the Moran's I Index together with a p-value, z-score, and other metrics to determine its importance. P-value are numerical estimates of the area under the curve for a given distribution that are bound by the test statistic.

Figure 4

Calculations for Spatial Autocorrelation (Morans I)

The Moran's I statistic for spatial autocorrelation is given as:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} z_i z_j}{\sum_{i=1}^{n} z_i^2} \tag{1}$$

where z_i is the deviation of an attribute for feature i from its mean $(x_i - \bar{X})$, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features, and S_0 is the aggregate of all the spatial weights:

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{i,j} \tag{2}$$

The z_I-score for the statistic is computed as:

$$z_I = \frac{I - \mathrm{E}[I]}{\sqrt{\mathrm{V}[I]}} \tag{3}$$

where:

$$\mathrm{E}[I] = -1/(n-1) \tag{4}$$

$$\mathrm{V}[I] = \mathrm{E}[I^2] - \mathrm{E}[I]^2 \tag{5}$$

The Global Moran's I statistic is explained mathematically in the previous section. For the characteristic under consideration, the calculator finds the mean and variance. The deviation from the mean is then calculated by deducting the mean from each feature value. By multiplying the deviation values of all neighboring features, a cross-product is created. Be aware that the integrand of the spatial Autocorrelation statistic contains these summing cross-products. Consider characteristics A-B to be near to one another, with the mean result of all attributes being 10. Note the range of possible cross-product results:

Table 2

Potential Cross-product Outcomes

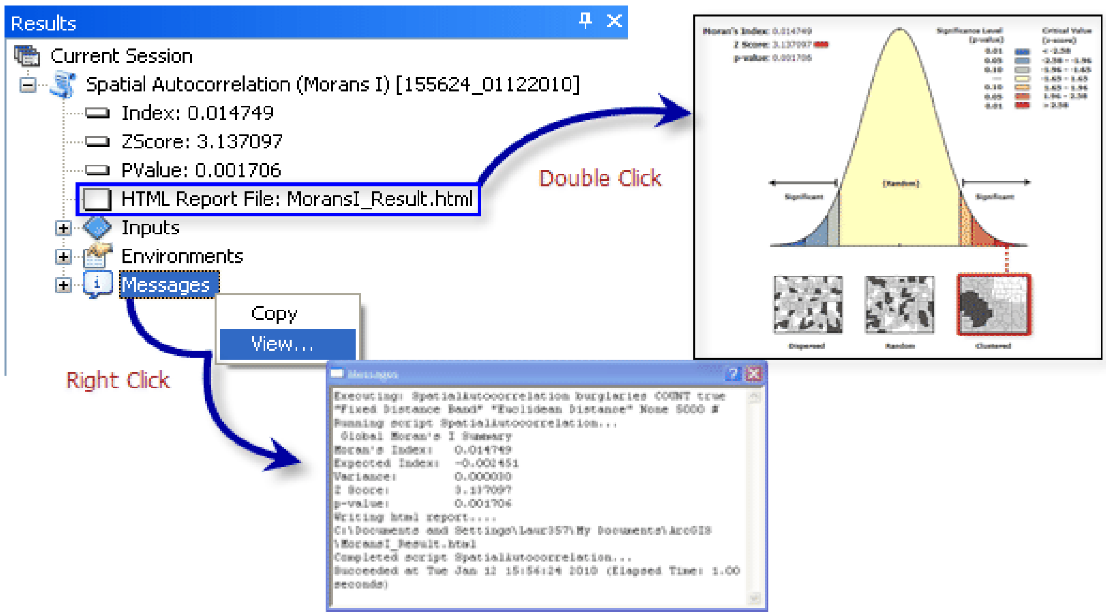
Feature values		Deviations		Cross-products
A=50	B=40	40	30	1200
A= 8	B=6	-2	-4	8
A=20	B=2	10	-8	-80

The cross-product will be positive if the values for the adjacent features are both larger than or less than the mean. The cross-product would be negative if one value is lower than the meanwhile the other is just above mean. Every instance demonstrates that the stronger the cross-product result, the larger the divergence from the mean. The Moran's Index will be positive if the geographic clustering of the dataset's values tends to be favorable (high values cluster close to other high values; low values cluster close to other low values). The index will be negative if high values gravitate toward low values while rejecting other high values. The Index will be extremely close to 0 if positive cross-product values equal negative cross-product values. In order to make the range of Index values between -1.0 and +1.0, the numerator is normalized by the variance.

Utilizing the Spatial Autocorrelation (Global Moran's I) tool, the Expected Index value is calculated following the calculation of the Index value. Then, the values of the Expected and Observed Index are contrasted. The tool offers a z-score and p-value indicating whether or not this difference is statistically significant based on the number of features in the dataset and the variance for the data values as a whole. Index values must be understood within the context of such null hypothesis in order to be fully understood.

Figure 5

Output of Moran's I



Tool output is accessible from the Results window.

The outcomes will also be shown in a Message dialog box if you right-click the Message element in the outcomes window and choose View.

3.4.1.2 Calculate the Distance Band Using the Neighbor Count

Provides, for a collection of characteristics, maximum, the minimum and average distances to the detailed Nth closest neighbor (N is an input feature). Results are expressed as signals from a tool's execution.

3.4.2 Incremental Spatial Autocorrelation

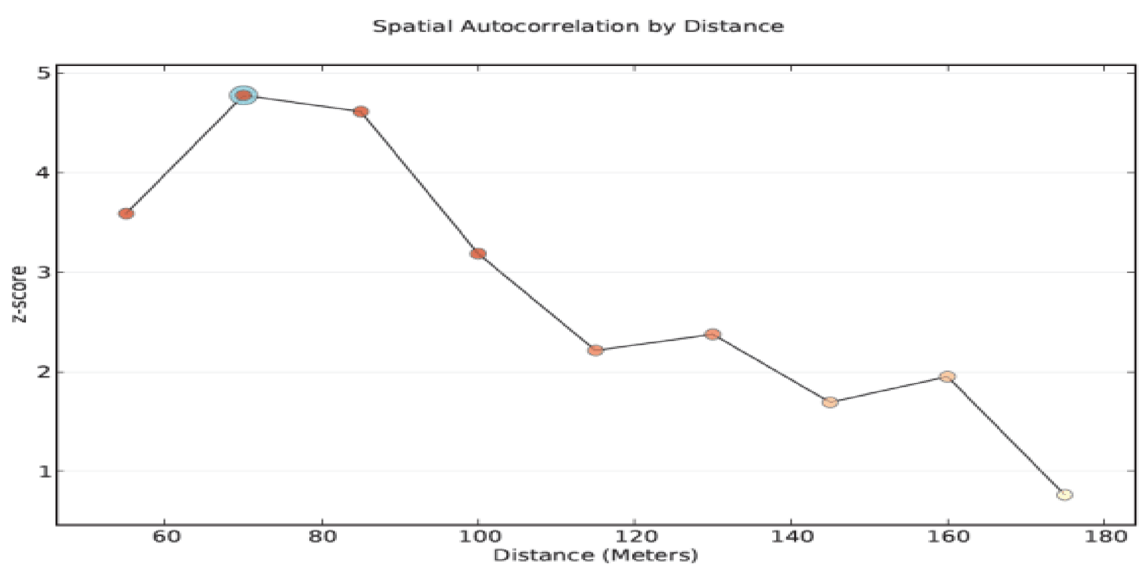
With a lot of the geographical data analysis you do, the size of your research will important. for instance, the fixed distance band of the hot spot analysis tool You must input a distance value when using Conceptualization of Spatial Relationships by default. For several density tools, you will be asked to provide a radius value. Your choice of distance should be consistent with the size of the problem you're trying to solve or the size of the suggested solution. Let's pretend you want to know why kids nowadays are so fat.

What geographic area do you analyze? For each person, is it at the household or communal level? The modest distance you use to choose your scale of study in this example will encompass the houses within a block or two of one another. Where to develop after-school physical education programs in order to perhaps reduce childhood obesity may be the solution to your question. In this case, your distance will generally be a good indicator of school zones. The selection of the appropriate scale of study while looking into commute patterns could be very straightforward. For instance, choosing a scale of analysis of 12 miles would be appropriate if you are aware that the average commute is 12 miles long. The defense of a certain analytical distance can occasionally be more difficult. At this time, the incremental spatial autocorrelation tool is useful.

When spatial clustering is observed in the landscape, it means that there are active spatial processes at work. By having some grasp of the geographic scale at which those underlying processes occur, you may select an appropriate analytical distance. This spatial Autocorrelation examination is run by the Incremental Spatial Autocorrelation tool to determine the degree of spatial clustering at each distance using a sequence of growing distances. However, the z-score frequently summits over a particular distance. On rare occasions, you could see many peaks.

Figure 6

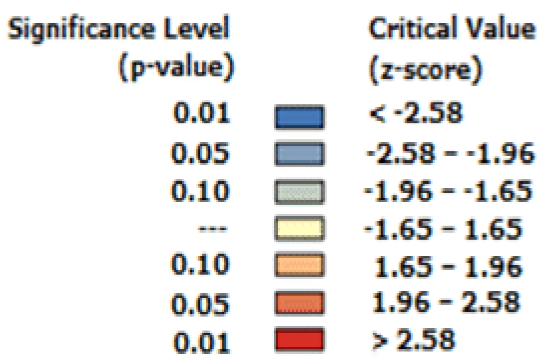
Outcome of Incremental Spatial Autocorrelation



The strongest spatial processes favoring clustering are shown as peaks at specific distances. The color of each point on the graph denotes the statistical significance of the z-score values.

Figure 7

P-value and Z-score Values.

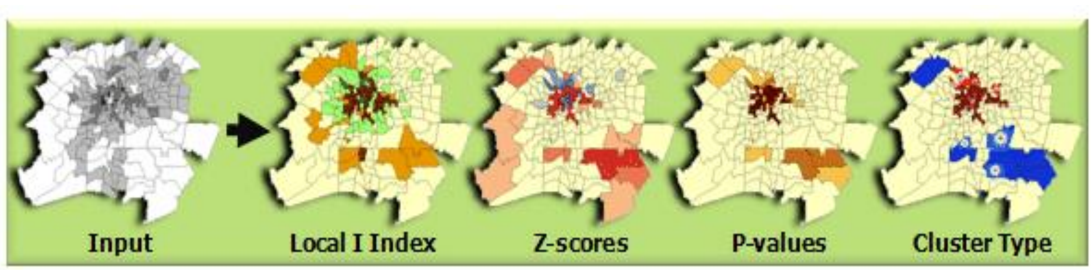


3.4.3 Cluster and Outlier Analysis

Find statistically significant hot spots, cold spots and geographic outliers given a set of weighted characteristics by using the Anselin Moran's I measurement.

Figure 8

Illustration for Anselin Local Moran's I



Usage:

- This tool creates a new Output Feature Class for each feature in the input feature class, and it has the following properties: cluster/outlier type (CO Type), pseudo p-value, z-score, and Local Moran's I index.
- A strong positive z-score indicates that the surrounding characteristics' values are comparable for that trait (either high values or low values). The CO Type field in the Output Feature Class will be HH for a statistically significant cluster of high values and LL for a statistically significant cluster of low values.
- A statistically significant geographic data outlier is indicated by a parameter with tiny negative z-score (less than -4.96, for example). If a feature has a low value and features with high values around it (HL), or if a feature has a high value and features with low values surround it (HL), it will be shown in the CO Type field in the Output Feature Class (LH).
- Chordal measures are used to calculate distances when the Given Input Class is not extended (i.e., when coordinates are supplied in degrees, mins, and seconds) or if the output coordinate system is set to a Geographic Coordinate System. Since chordal length measurements can be made rapidly and offer accurate estimates of genuine geodesic distances, they are frequently used, at least for places that are within 30 degrees of one another. Chordal distances are calculated on an oblate spheroid. The length of a line that would need to travel through the three-dimensional earth in order to connect any two points on the globe's surface is known as the chordal distance between those points. The Distance Band or Threshold Distance parameter, if provided, should be specified in meters when chordal distances are used in the analysis. Chordal distances are reported in meters.
- Feature centroids are utilized in distance calculations for features such as lines and polygons. The weighted mean center among all feature components is used to compute the centroid for multi-points, polylines, and polygons having multiple parts. One is given as the weighting for point features, one for length for lines, and one for area for polygon features.

The Moran's I method of spatial autocorrelation evaluates a characteristic's locations & values concurrently. No matter how the data are arranged—scattered, clustered, or randomly—it shows the pattern that the data reflect. An inferential statistical tool called Moran's I is used to interpret the study's findings in the context of the null hypothesis. The null hypothesis takes full spatial randomization as a given. In other words, there is erratic value distribution across features, which points to erratic production of the spatial mechanism. Statistically significant difference of clustering with either high or low data on traffic accidents was evaluated using this method. Since the data linked with the accident locations are not fairly evenly distributed (as established by KDE) over the research region, Moran's I is known as the high or low cluster (Getis Ord. General G). When choosing the local value spikes as high value clusters in such a distribution, general G statistics are more suitable. The dataset may be examined for clusters since Moran's I reports the parameters as clustering, dispersion, or randomly distributed and globally correlates the feature values with a preset distance band or the average closest neighbor.

3.5 Hotspot Analysis

The resultant z-scores and p-values of a hotspot analysis show you where spatial grouping of features with low or high values takes place. Each attribute is studied using this tool in relative to its surrounding aspects. A characteristic with a high value is interesting even if it doesn't necessarily indicate a statistically significant hot spot. To get a statistically significant hot spot, a parameter must have a high value and be surrounded by other parameters that also have high values. Whenever the localized summation for a function and its neighbors deviates considerably from the expected local sum and by an amount that is too large to be the result of random chance, a statistically significant z-score is generated. The FDR adjustment adjusts statistical significance to take repeated testing and spatial dependence into account.

Figure 9

Calculations for Hotspot analysis

The Getis-Ord local statistic is given as:

$$G_i^* = \frac{\sum_{j=1}^{n} w_{i,j} x_j - \bar{X} \sum_{j=1}^{n} w_{i,j}}{S \sqrt{\frac{\left[n \sum_{j=1}^{n} w_{i,j}^2 - \left(\sum_{j=1}^{n} w_{i,j} \right)^2 \right]}{n-1}}} \quad (1)$$

where x_j is the attribute value for feature j, $w_{i,j}$ is the spatial weight between feature i and j, n is equal to the total number of features and:

$$\bar{X} = \frac{\sum_{j=1}^{n} x_j}{n} \quad (2)$$

$$S = \sqrt{\frac{\sum_{j=1}^{n} x_j^2}{n} - (\bar{X})^2} \quad (3)$$

The G_i^* statistic is a z-score so no further calculations are required.

3.5.1 Hotspot analysis explanation:

The Gi* statistic that was calculated for each function in the dataset is known as z-score. A concentration of high values rises with rising z-scores for statistically significant positive z-scores (hot spot). A clustering of low values increases as the z-scores drops for statistically significant negative z-score.

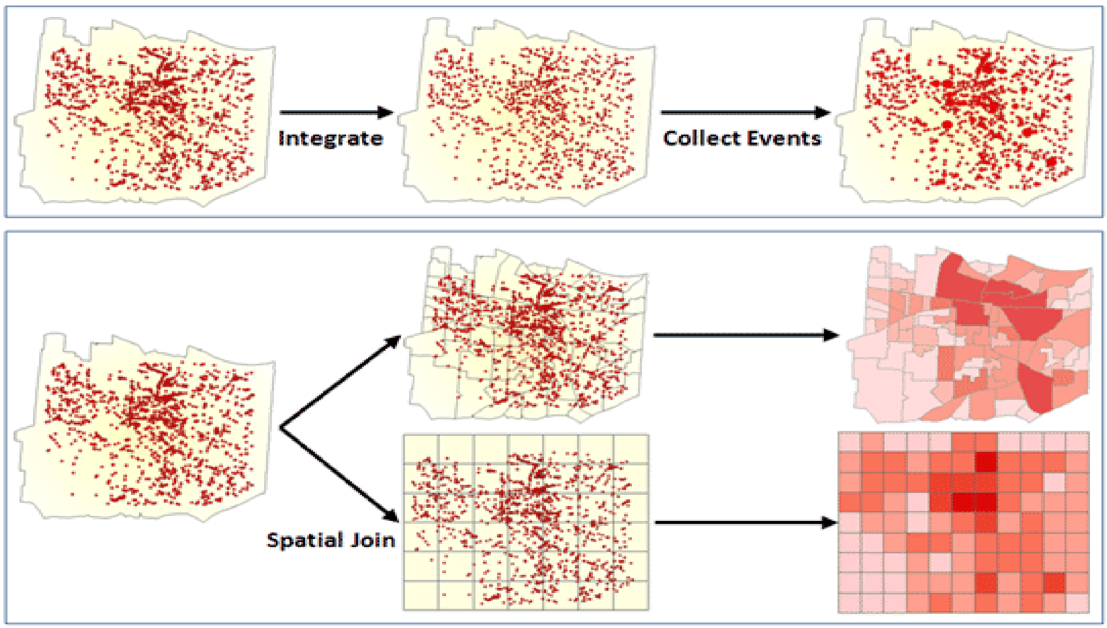
3.5.2 Results of the hotspot analysis

With a z-score, p-value, and confidence level bins for every element in the input feature class, this program builds a completely new feature class (Gi Bin). Only chosen features will be examined if a choice set has been applied to the input parameter class, and only chosen features would be present in the output image. As an alternative, you can use the Integration capabilities with the Collect Events tool to group features that are close to one another or coincident points together. After that, to display how many events or points

were snapped together, a new feature class may be generated with a value at each separate position. Use the ICOUNT field that is
generated as your analysis's input field.

Figure 10

Strategies for Aggregating Incident Data



Strategies for aggregating incident data

CHAPTER IV

Results and Discussion

4.1 Accident intensity and associated causes

Data on accidents for Somalia region was collected from the traffic police's main office in Mogadishu. The data was analyzed to find accidents that only resulted in injuries or fatalities along the study region of Mogadishu. Table 3 below shows the results.
With a total of 7560 incidents, 2021 saw the most accidents ever. This slightly dropped to 2552 accidents in 2020 and 2032 accidents in 2019. The fatal accidents that occurred in 2019 accounted for 26.88% of all fatal accidents that were noted throughout the research period. Moreover, with 33.76% and 39.36% of fatal accidents, respectively, 2020 and 2021 saw a noticeable rise in accident deaths and injuries.

Table 3

The Accident intensity in Mogadishu Region.

Fatalities		**Frequency**		**Injuries**	
Year	**No. of Accidents**	**(F)**	**%**	**Frequency (F)**	**%**
2019	2032	6,101	25.65	660	31.21
2020	2552	7,560	31.77	767	36.26
2021	2976	10,132	42.58	688	32.53
Total	**7560**	**23,793**	**100**	**2115**	**100**

To find the important indications for the particular safety issues at the locations, the gathered data were evaluated. According to Table 2, which shows that speed violations (SPV) and loss of control accounted for around 16.56% and 13.88% of all incidents, respectively, these were the most frequent accident-contributing variables. The most common reason for accidents is interwoven, when drivers run the risk of loss of control of the steering at high speeds. Wrongful overtaking, which is responsible for 13.60% of all accidents, is the third most common accident cause. Less than 10% of all accidents along the study region are attributable to other variables, which is consistent with the yearly report. This underreporting of incidents may be brought on by inadequate security, the lack of modern technology for accident reports and officials' poor reporting standards. Additionally, incidents that occurred in the wee hours of the morning (between 12:00 and 6:00 a.m.) or late at night (between 9:00 and 11:59 p.m.) might not be recorded since officials cannot be everywhere at once.

Table 4

Associated Causes of Road Accidents in Mogadishu Region (2019–2021).

		2019		**2020**		**2021**		**Total**	
No.	**Associated causes**	**F**	**%**	**F**	**%**	**F**	**%**	**F**	**%**
1	Speed Violation	486	16.56	579	17.11	609	16.21	1674	16.77
2	Loss of Control	394	13.88	494	14.60	561	14.93	1425	14.28
3	Tyre Burst	182	6.41	157	4.64	153	4.07	492	4.93
4	Wrongful Overtaking	370	13.60	476	14.15	496	13.2	1366	13.69
5	Dangerous Driving	316	11.14	469	13.84	467	12.4	1252	12.54
6	Route Violation	247	8.70	396	11.70	356	9.47	999	10.0
7	Use of mobile phone While driving	229	8.07	254	7.50	298	7.95	781	7.82
8	Mechanically Deficient Vehicle	137	4.80	218	6.44	277	7.38	632	6.33
9	Brake Failure	108	3.80	172	5.08	169	4.52	449	4.50
10	Others	119	4.19	83	2.45	112	2.98	314	3.14
11	Road Obstruction Violation	90	3.18	38	1.12	112	2.98	240	2.4
12	Fatigue	69	2.44	22	0.65	88	2.36	179	1.81
13	Bad Road	78	2.76	19	0.56	54	1.42	151	1.53
14	Dangerous Overtaking	13	0.47	07	0.20	04	0.10	24	0.26
Total		**2,838**	**100**	**3,384**	**100**	**3,756**	**100**	**9978**	**100**

4.2 Spatial Distribution of Accidents

Datasets which had (x, y) coordinates of accident locations were used to develop the spatial distribution map of the study area. Traffic accident hotspots were analyzed using GIS methods, and the accident sites were shown on a digital map. Using the following GIS Spatial Statistics techniques, analysis of the accident sites and hotspots based on accident frequencies was conducted; -

i. Weighted mean center,
ii. Kernel Density Estimation,
iii. Cluster Analysis,
iv. Hotspot analysis.

The geographic characteristics of the accident locations were demonstrated using KDE, fishnet polygons, Moran's I Statistic and the network spatial weight matrices. The mean center and density study uses all accident data from 2019 to 2021. Both the overall accident data and the annual accident data were utilized for the cluster analysis (Moran's I statistic). The fishnet polygon analysis used the whole accident data, whereas the spatial weight matrix used the complete and yearly accident data, allowing for a comparison of the hotspot study's findings.

4.2.1 Mean Center Analysis

Figure 11 depicts the mean center determination model using ArcGIS. Figure 12 shows the location of the cumulative accident frequency's geographic mean center (2019–2020) as well as the individual years. Figure 11. Model Builder Script of generating Mean Centre.

Figure 11

Mean Centre Location Map

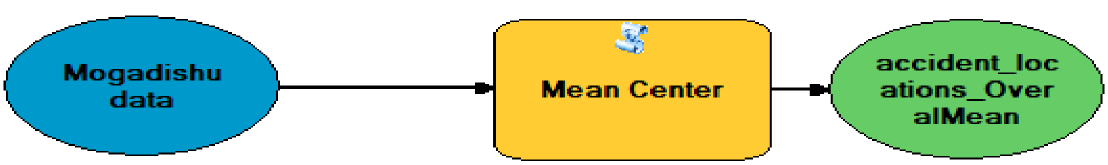
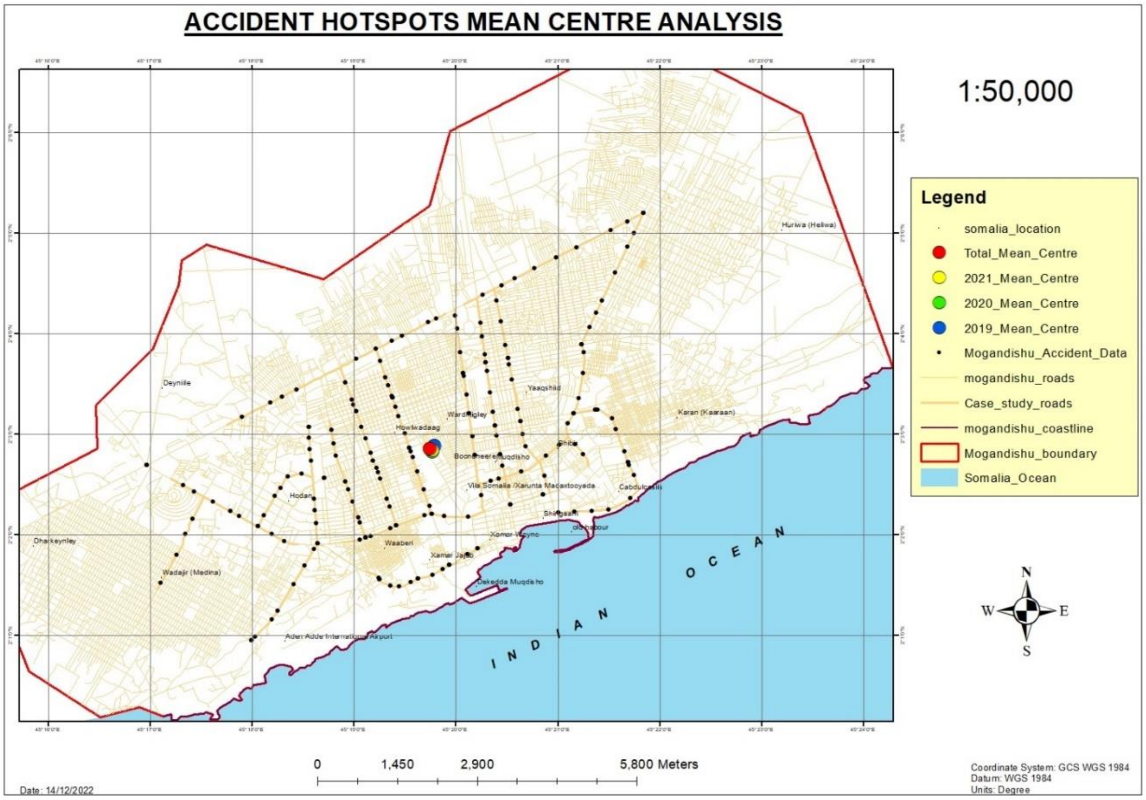


Figure 12

Results for Mean Center Analysis



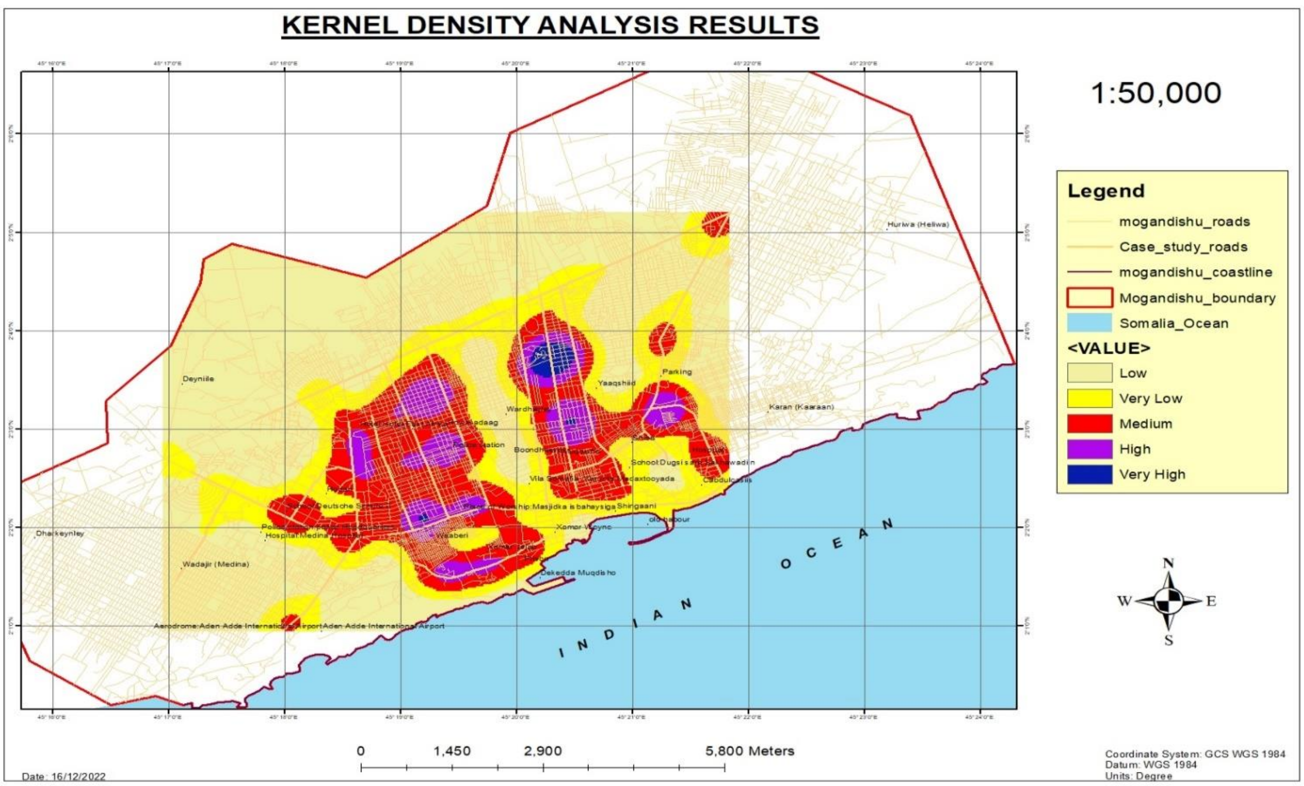
It can be observed that there was a very minimal shift in the concentration of road traffic accidents in the region in 2019, 2020 and 2021. In 2020, the mean center maintained same position as overall/total mean center, hence the reason why it is not visible in figure 12. The overall mean center is located somewhere midway between two minor roads. This suggests that Use of mobile phone While driving and loss of control have more influence in accident occurrence. Since roads appear to be straight, it can be concluded that road curves are not contributory factor in accident occurrence. Locations of mean center analysis generally indicates regions which need more attention than the rest.

4.2.2 Kernel Density Analysis

Figure 13 below displays the KDE surface for the cumulative accident frequencies over the study region. The spatial difference in accident frequency from high to low is easily represented by the density surface. This subjective map, however, cannot be used to assess the statistical significance of the low to high accident rates at various locations around the research zone. In other words, there is no guarantee that a reason for these differences; they might arise from a random process. However, the density surface pertaining to the road network suggests that generally speaking, high-frequency accident locations are related to higher vehicular and pedestrian traffic in significant town centers, as well as road junctions and road bends.

Figure 13

Kernel Density Analysis Classification



4.2.3 Cluster Analysis

With the help of the following spatial statistics tools, cluster analysis has been carried out;-

i. Spatial Autocorrelation (Morans I),

ii. Calculate the Distance Band from Neighbor Count,

iii. Incremental Spatial Autocorrelation,

iv. Cluster and Outlier Analysis (Anselin Local Morans I).

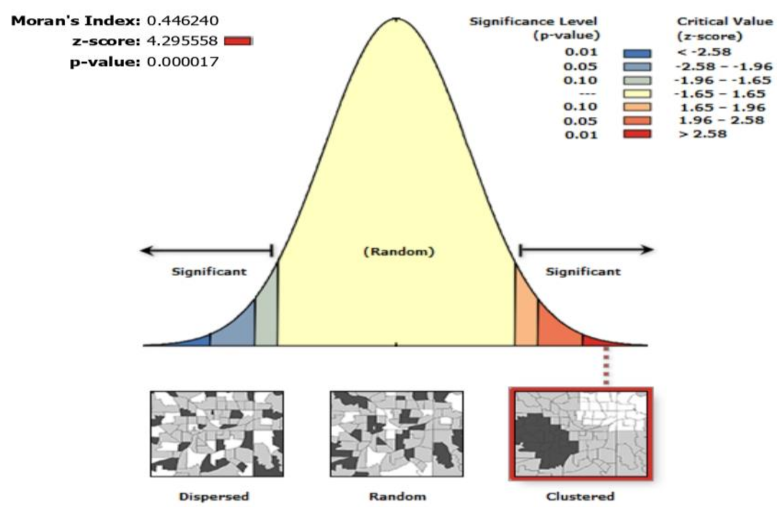
4.2.3.1 Spatial Autocorrelation (Moran's I)

Figure 14

Spatial Autocorrelation Report.

12/16/22, 8:42 PM Spatial Autocorrelation Report

Spatial Autocorrelation Report



Given the z-score of 4.29555765185, there is a less than 1% likelihood that this clustered pattern could be the result of random chance.

Global Moran's I Summary

Moran's Index:	0.446240
Expected Index:	-0.004348
Variance:	0.011003
z-score:	4.295558
p-value:	0.000017

Dataset Information

Input Feature Class:	Mogandishu_Accident_Data
Input Field:	F
Conceptualization:	INVERSE_DISTANCE
Distance Method:	EUCLIDEAN
Row Standardization:	False
Distance Threshold:	594.7842 Meters

file:///C:/Users/CAL/Documents/ArcGIS/MoransI_Result_14132_12372_.html 1/2

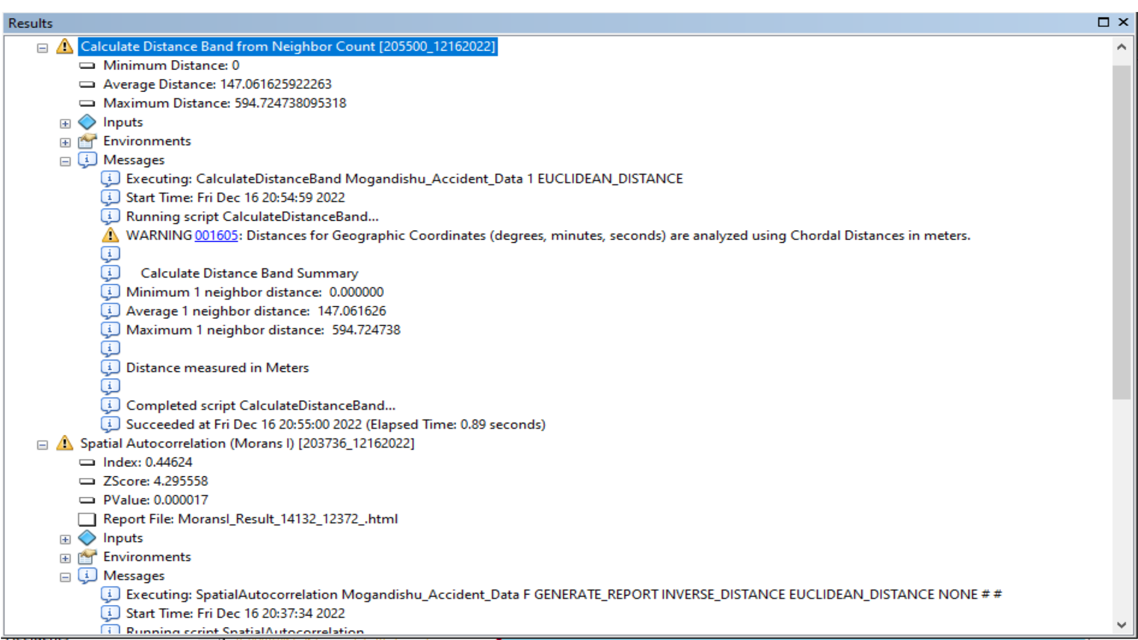
Figure 14 shows graphically the outcomes of the spatial autocorrelation (Moran's I) statistic of the accident data for the overall three years. On the graph, the Z-scores for each year are displayed in comparison to the other years. The clusters have a lower intensity, as shown by the Z-score of 4.295558, which also has a 90% confidence interval and a 10% chance of being the outcome of random processes. However, Figure 14 shows that there was no overall clustering of accidents during the three years, which illustrates the Moran's I finding for the cumulative accidents; the pattern is otherwise random. Additionally, it should be noted that the total data tends to clustering, therefore precise identification of clustered places is required, which is achieved by running the second tool of cluster analysis, which is to calculate the distance band from neighbor count.

4.2.3.2 Calculating the Distance Band from Neighbor Count

This tool determined the minimum distance at which any given accident hotspot location has atleast one neighbour, which resulted to 147.0616m, as shown in figure 15. This distance was significant in computing the next tool, which is Incremental Spatial Autocorrelation.

Figure 15

Minimum distance from neighbour count

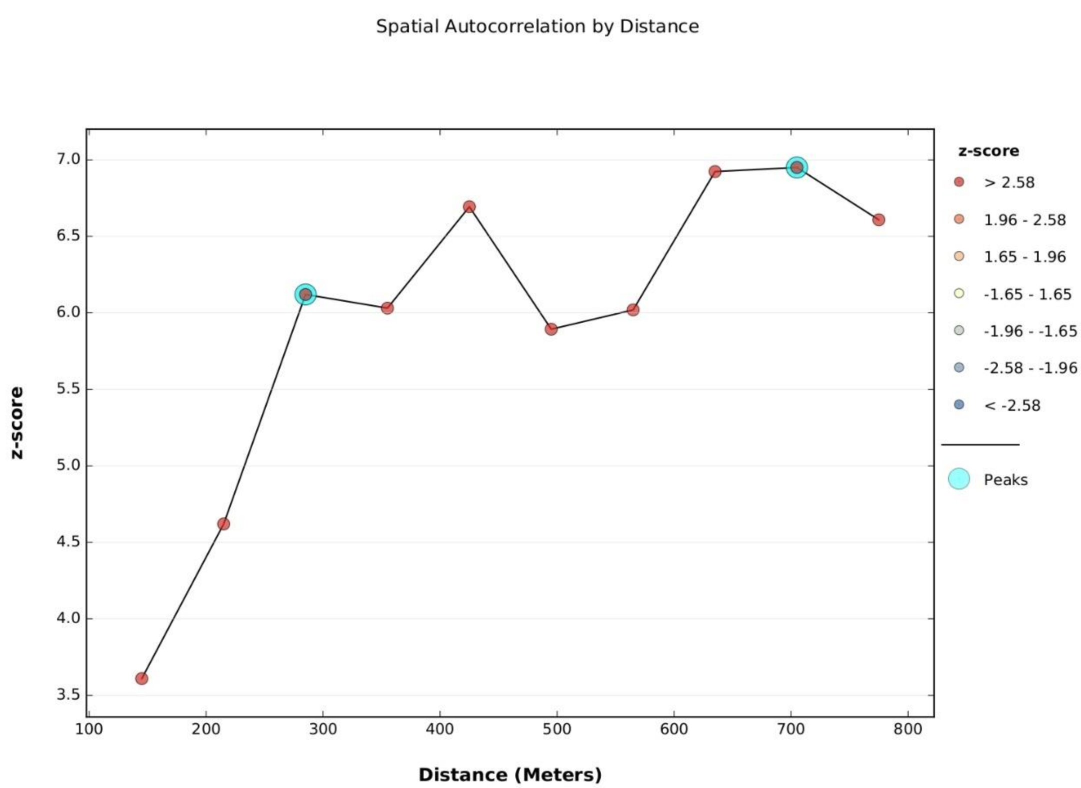


4.2.3.3 Incremental Spatial Autocorrelation

Average distance of 147m for point neighbours was used. By dividing the distance by half and rounding off to nearest hundred, a distance of 100m was adopted for start of incremental distance, with a constant incremental distance of 100m as well. This way, the tool was able to accurately analyse each point, and putting accident frequency into account, determine the locations where there is likelihood of data clustering. Figure 16 gives detailed results in graph format. There are two maximum instances of peak, meaning that we expect two different locations within Mogadishu region where there is accident hotspot clustering/ concentration/grouping. The physical locations of such regions will be revealed by the next and final tool: Cluster and Outlier Analysis (Anselin Local Morans I).

Figure 16

Graph of Incremental Distances Computing Peak Points of Clustering

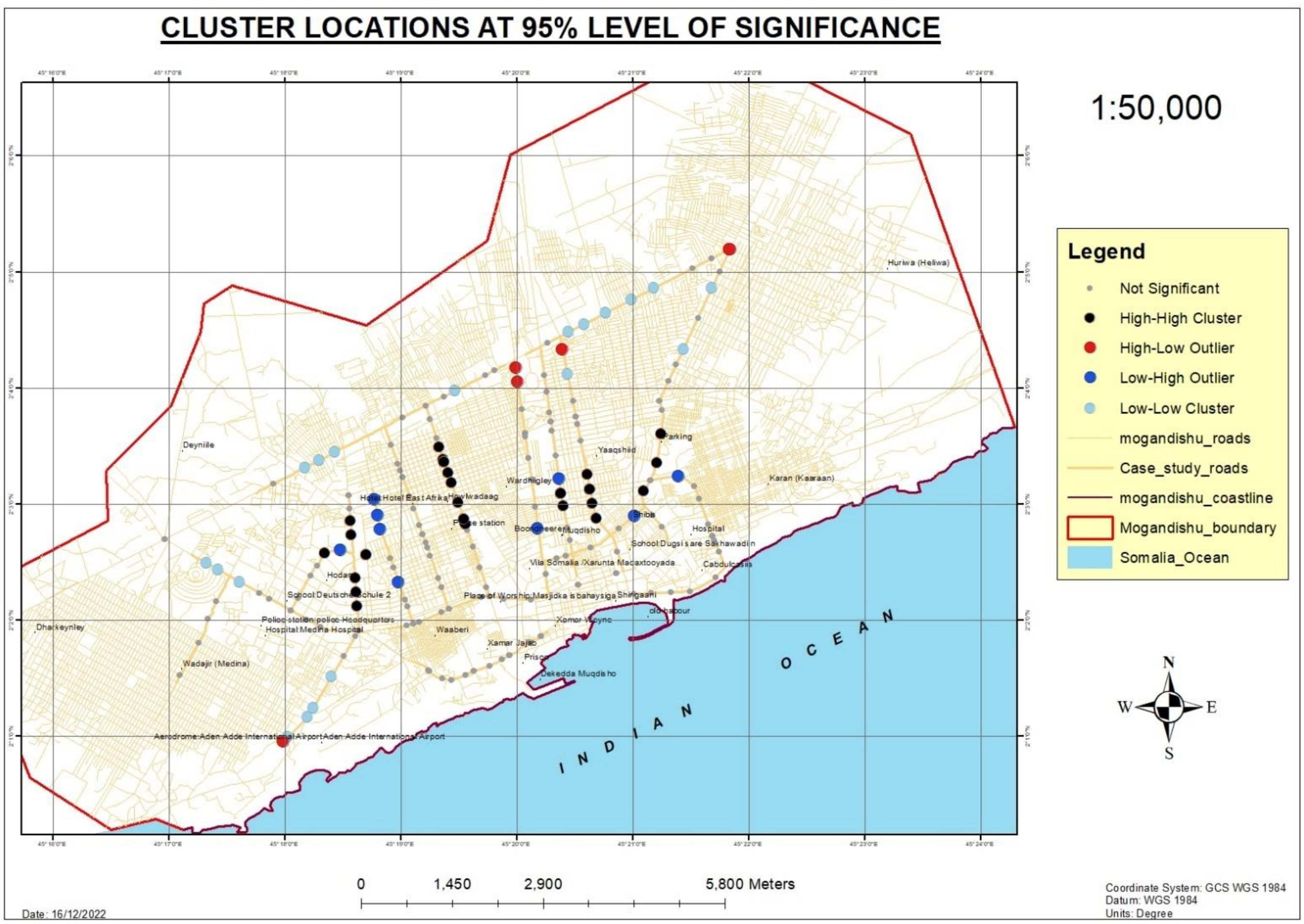


4.2.3.4 Cluster and Outlier Analysis (Anselin Local Morans I)

Being the final step of displaying the zones of clustering in form of a map, it was accurately achieved. Points grouped as High- High Outlier indicate locations with high accident occurrence near each other. Such locations are statistically significant, meaning that they are very far from being random arrangement. Low-Low represents points with low accident frequency near each other.

Figure 17

Cluster Spatial Results

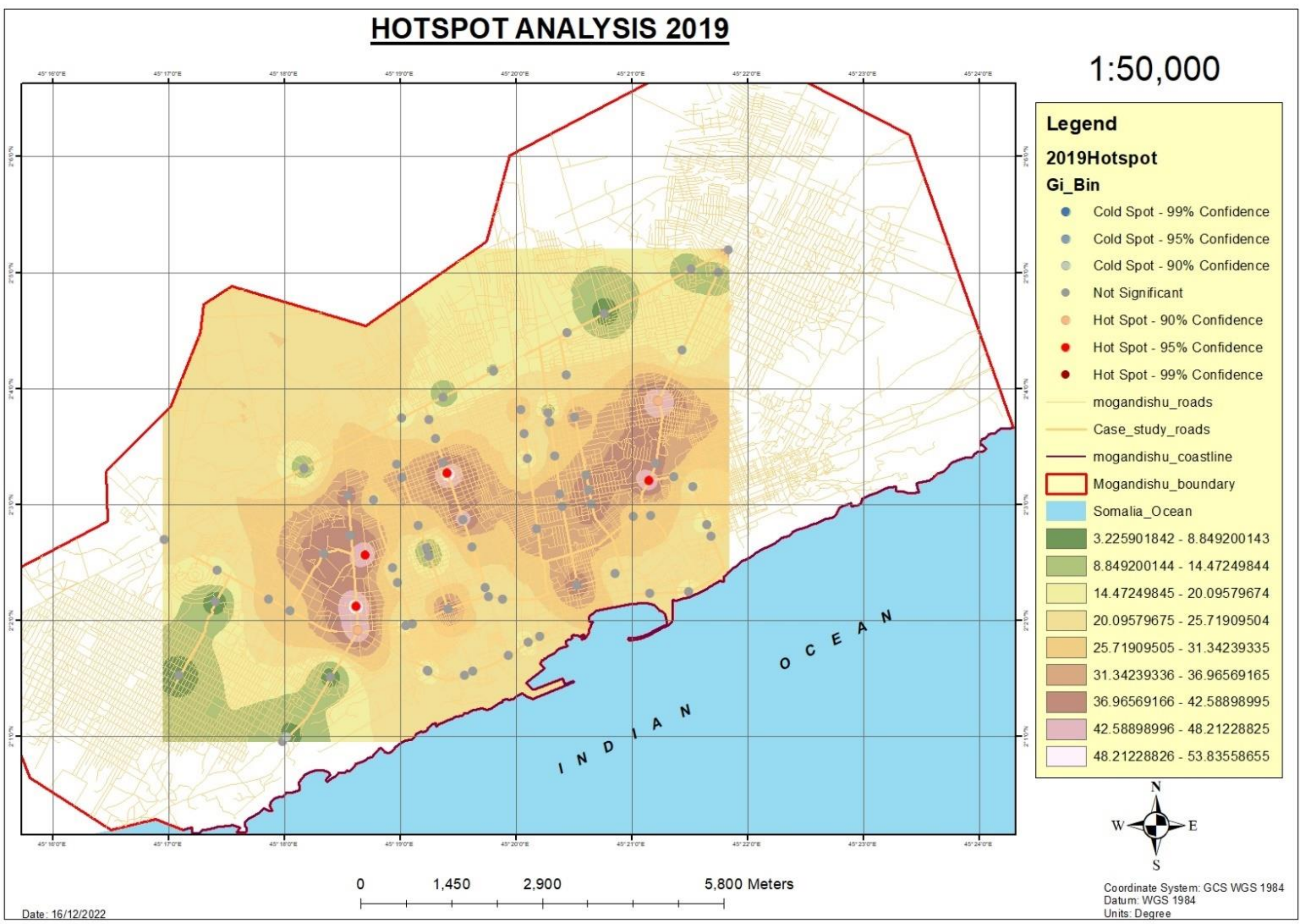


4.2.4 Hotspot Analysis

Figure 18–20, respectively, each year from 2019 to 2021, shows the findings of the hotspot analysis using a network spatial weight matrix. In the Mogadishu region, there appear to be four hotspot spots in 2019, compared to just one in 2020. However, no areas have been identified as accident hotspots for 2021. Additionally, as shown in Figure 21, the total accident record for the three years reflects more hotspot sites. As seen in Figure 18, locations with geometric features, such as junctions, bends, bridges, U-turns, slopes, median barriers, impediments, and roadsides, are where hotspots are most likely to be found.

Figure 18

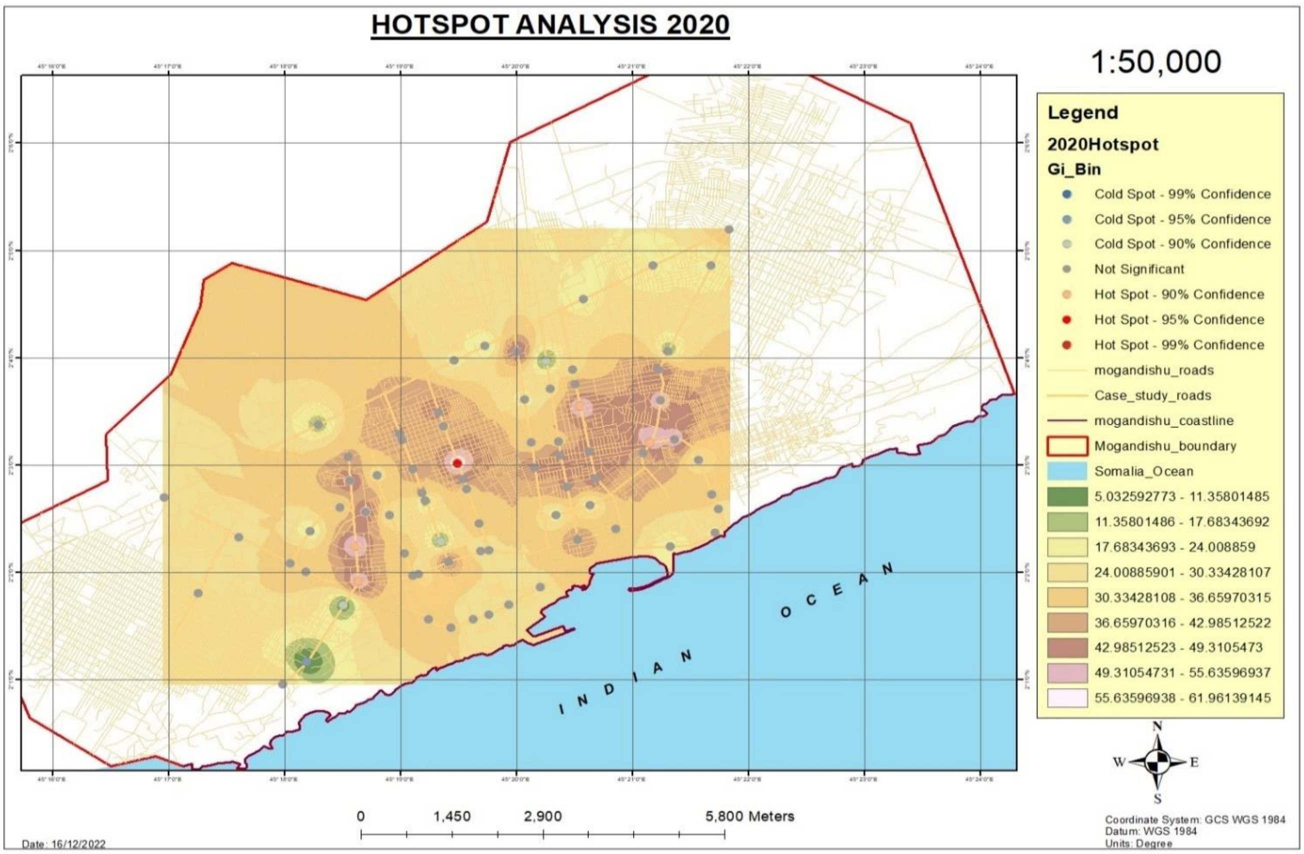
Hotspot Analysis for 2019



The standard deviation measurements for 2019 revealed 4 sites with a Z-score exceeding 1.90 standard deviations based on the accident frequency at each of the 79 locations. These areas make up the top 10 sites for motor vehicle accidents overall that the hotspot analysis indicated to have a Z-score over 1.90.

Figure 19

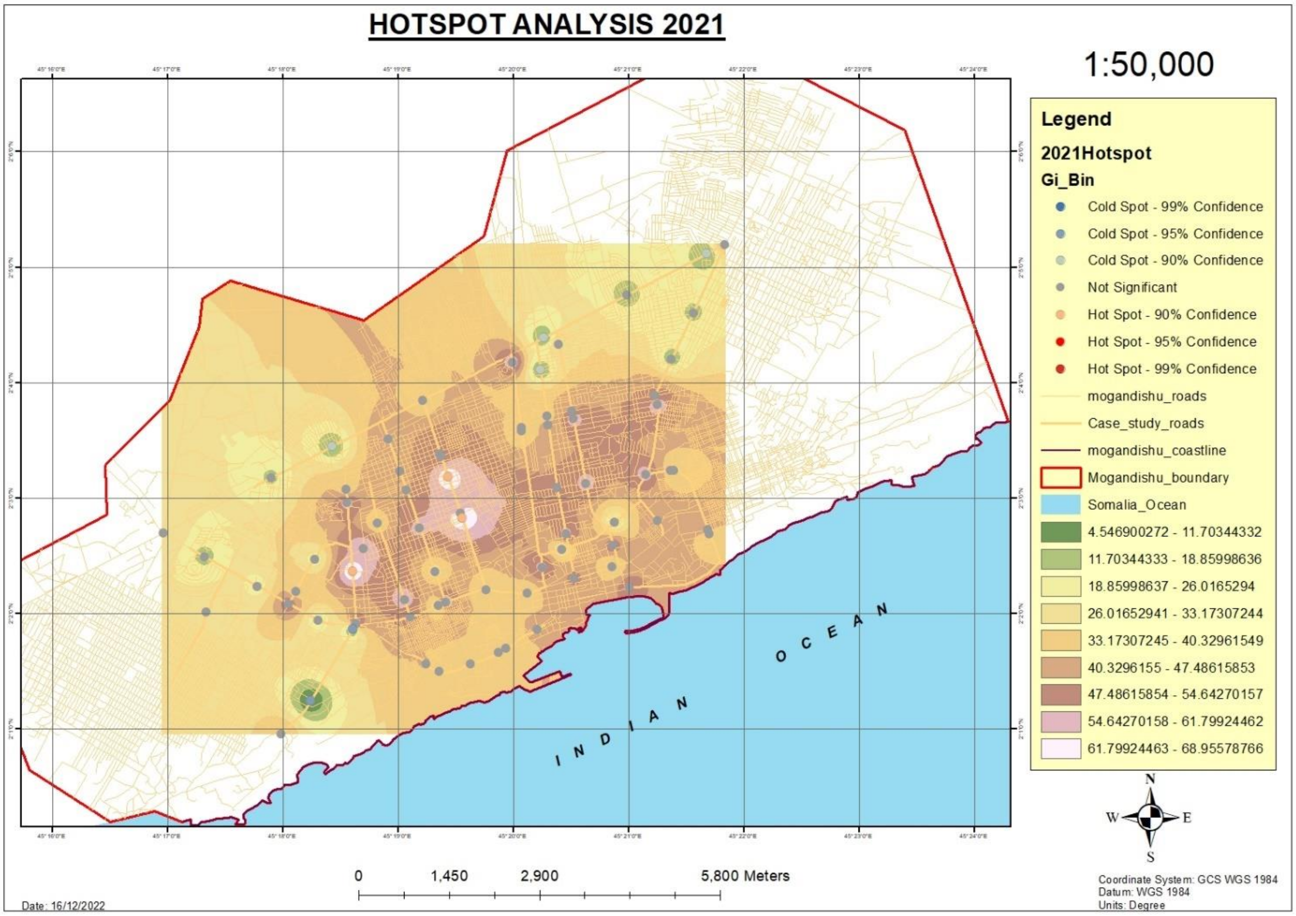
Hotspot Analysis for 2020



Only one location had a Z-score exceeding 1.90 standard deviation in 2020, according to measures of the accident frequency. This represents value significantly higher than the rest of accident locations within the neighborhood. Since the Hotspot location at 95% confidence is outside Mogandishu Central Business District, it can be concluded that human traffic was not a contributory factor of the hotspots, but other factors such curves, intersections, U-turns, bridges, grades, interchanges, median barriers, roadside obstacles, and hilly terrain.

Figure 20

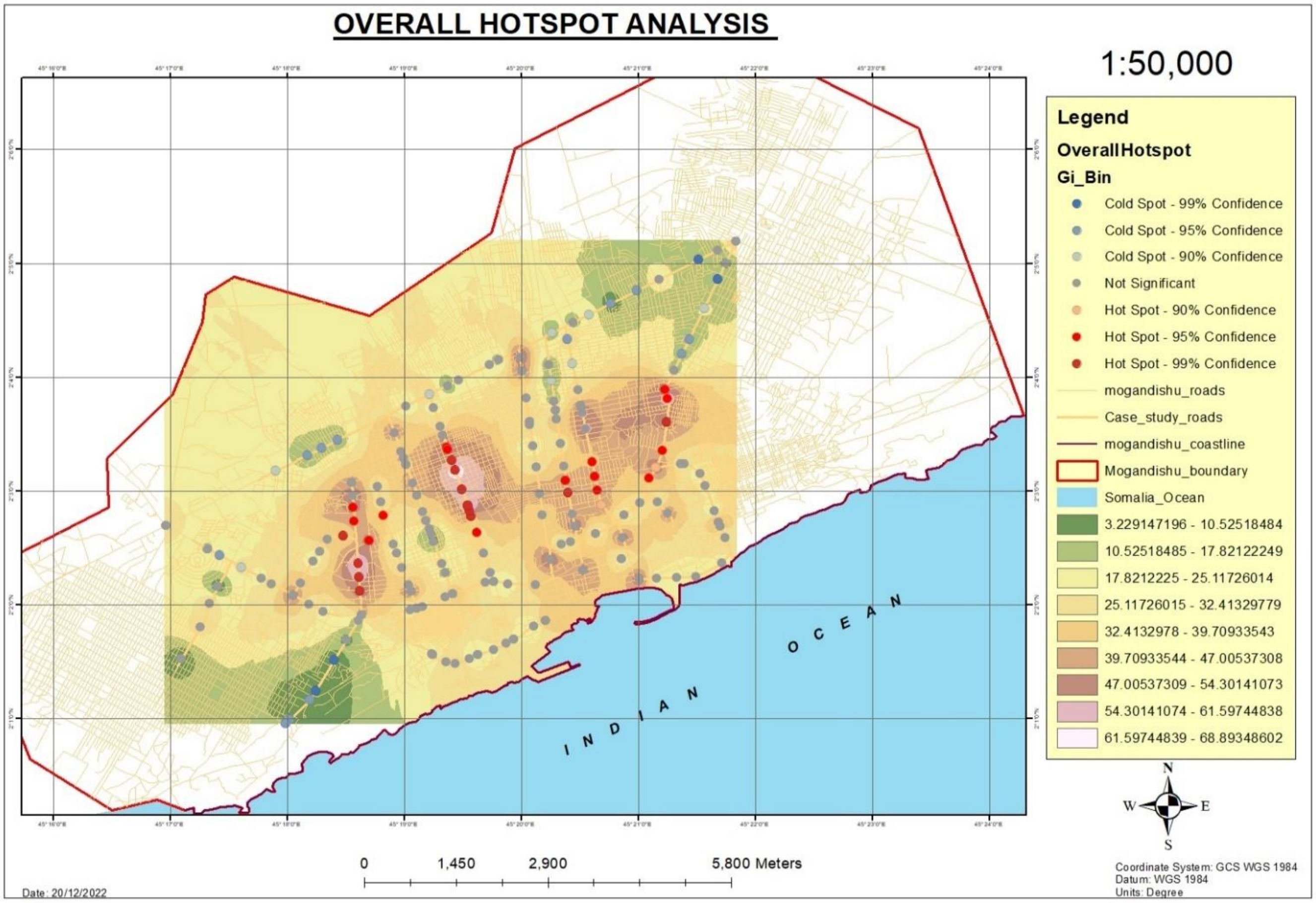
Hotspot Analysis for 2021



The standard deviation assessment discovered zero places in 2021 with a Z-score greater than 1.90 standard deviations. Technically speaking, this indicates that there is a 99% chance that the clustering of traffic accidents is the result of random.

Figure 21

Overall Hotspot Analysis



Highest levels of Hotspots locations were recorded outside Mogadishu Central Business District, through analysis of commulative frequency data of 2019, 2020 and 2021. This can be as a result of loss of vehicle control, lack of traffic lights, limited or no footbridges, among other factors. Human traffic is not likely to be contributing factor in accident occurrence.

Finally, If a decision had to be made based on these data, it would be wiser to start by looking at statistically significant clusters of the high accident frequencies in every year. It's possible that pavement failure, which is a typical occurrence on Somalian roads, and other variables, like reckless driving, the lack of or inadequacy of traffic signals, and vehicle worthiness, are to blame for the shift in accident hotspots over time.

CHAPTER V

Conclusion and Recommendations

Conclusion

Many researches, like the current study, have utilized GIS techniques to locate accidents on a map, identify them and also to analyze the accident hotspots. Both the quantity of accident data records as well as the criteria considered vary greatly. The availability of data has a significant impact on the amount of accident data. Hot Spot Analysis and Kernel Density, followed by Statistical Analysis as well as Optimized Hot Spot, are the most often used spatial techniques for timely accident data. Spatial analysis is a valuable approach for studying accident hotspots because it clusters accident areas based on optimum spatial patterns and is very reliant on the locations of accidents. Most accidents occurred in 2019, followed by those in 2020, and then those in 2021. The majority of accidents happened on roads and in urban areas. For low- and middle-income groups, the cost of surviving a traffic accident is relatively substantial.

The results show that the commercial, residential, and industrial zones in and around the research area's center areas are hotspots for all categories that were analyzed. This is supported by the fact that there are low speed restrictions and significant traffic volumes in certain areas due to the high population density.

More accident features, such as the causes of accidents, the correlation between traffic accidents and the major contributing elements, and the use of machine learning to predict hotspots for other locations are all possible extensions of the current work. Although the scope of this study is restricted to the detection for hotspots in particular years, also it may be used to pinpoint cold spots. The current study may be utilized as a starting point for additional work on the impact of road conditions, time of day, and driving behavior.

Recommendations

Based on this investigation, a few safety precautions are described below:

- Incorporating stakeholders into the planning process: Authorities must ensure community involvement during the plan creation in order to be aware of user demands and take such suggestions into account at various planning stages. There should be a public forum where the general public may freely voice their opinions and share their ideas in the presence of all interested parties at every step of the planning process.
- A community-based program toward awareness: The Central Government, with the assistance of Municipalities, should start some awareness programs, training, in every ward to inform the people of the traffic rules and regulations, individual roles in road safety, and providing people with information on how they can get involved to establish a safer environment for their children. This will increase public awareness of and responsibility for their obligations.

- Incorporating an adaptive traffic signaling system: Because there are no designated traffic police in charge of maintaining traffic in any intersection or road, and because there is no pedestrian signal on any of the roads, when high-speed vehicles and pedestrians attempt to cross the intersection within the same time, severe accidents result.
- Government aid for transportation research: Even though automation in the transportation sector is advancing in the contemporary world, Mogadishu city remains far behind. In Mogadishu, there are no organizations dedicated to promoting road safety. The government should prioritize and invest in the development of intelligent transportation systems (ITS), which may significantly improve the comfort, security, and integration of the nation's road networks.

- Strict law enforcement:

  Whether it is a pedestrian or a vehicle, the government must be extremely strict when it comes to breaching traffic laws and regulations. Drunk driving, excessive speeding, and irresponsible driving should all be strictly prohibited. The road must be cleared of any unsafe cars. It is necessary to operate the mobile court more frequently. For the sake of guaranteeing road safety, all forms of law enforcement personnel who are involved in traffic control must fulfill their duties from their own location.

REFERENCES

Aldegheishem, A., Yasmeen, H., Maryam, H., Shah, M.A., Mehmood, A., Alrajeh, N. and Song, H. (2018) Smart Road Traffic Accidents Reduction Strategy Based on Intelligent Transportation Systems (TARS). Sensors, 18, Article No. 1983.

Ali, A.M., Kamarudin, A., Ahmed, M.M. and Deprizon, S. (2019) A Review of Traffic Accidents and Related Practices Worldwide. The Open Transportation Journal, 13, 65-83.

Ameratunga, Shanthi, Martha Hijar, and Robyn Norton. "Road-traffic injuries: confronting disparities to address a global-health problem." *The lancet* 367.9521 (2006): 1533-1540.

Andrášik, R.; Bíl, M. Clustering of Traffic Accidents with the Use of the Kde+ Method; Transport Research Centre: Brno, Czech Republic, 2015.

Batamag, E. (2020) Cameroon Overloading, One of the Main Causes of Road Causes of Road Accidents: Ivory Coast Tribune.

Blazquez CA, Celis MS (2013) A spatial and temporal analysis of child pedestrian crashes in Santiago, Chile. Accid Anal Prev 50:304–311.

Düzgün, ¸S. Analysis of Point Patterns. Spatial Data Analysis; Middle East Technical University: Ankara, Turkey, 2009.

El-Said Mahmoud, M.Z.; Soon, J.T.; Eng Hie, A.T.; Nurul Amirah ‘Atiqah Binti Mohamad’ Asri Putra, Yok Hoe Yap & Ena Kartina Abdul Rahman. Spatial analysis of road traffic accident hotspots: Evaluation and validation of recent approaches using road safety audit. J. Transp. Saf. Secur. 2019.

Erdoğan, S., Dereli, M.A. & Şenol, H.İ. A GIS-based assessment of long-term traffic accidents using spatiotemporal and empirical Bayes analysis in Turkey. *Appl Geomat* **14**, 147–162 (2022).

Ihueze, C.C. and Onwurah, U.O. (2018) Road Traffic Accidents Prediction Modelling: An Analysis of Anambra State, Nigeria. Accident Analysis & Prevention, 112, 21-29.

Josephine, H., Shanthi, A. and Rod, J. (2012) Burden of Road Traffic Injuries and Related Risk Factors in Low and Middle-Income Pacific Island Countries and Territories: A Systematic Review of the Scientific Literature (TRIP 5). BMC Public Health, 12, Article No. 479.

Khan, G.; Qin, X.; Noyce, D.A. Spatial analysis of weather crash patterns in Wisconsin. In Proceedings of the 85th Annual Meeting of the Transportation Research Board, Washington, DC, USA, 22–26 January 2006.

Khan, M. (2007) Road Traffic Accidents. The Professional Medical Journal, 14, 323-327.
Kundakci E, Tuydes-Yaman H (2014) Understanding the distribution of traffic accident hotspots in Ankara, Turkey. In: Proceedings of the 93rd annual TRB conference, Washington, 12–16 Jan.

Kuo PF, Lord D, Walden TD (2012) Using geographical information systems to effectively organize police patrol routes by grouping hot spots of crash and crime data. Zachry Department of Civil Engineering, Texas A&M University.

Liu, G., Chen, S., Zeng, Z., Cui, H., Fang, Y., Gu, D., et al. (2018) Risk Factors for Extremely Serious Road Accidents: Results from National Road Accident Statistical Annual Report of China. PLoS ONE, 13, Article ID: e0201587.

Matthew, O.A., Rebi, A.M., Rani, M. and Mohd, R.J. (2015) Commercial Bus Accident Analysis through Accident Database. Journal of Transport System Engineering, 2, 7-14.

Olemo, C.D. (2016) Exploring the Major Causes of Road Traffic Accidents in Nairobi County, University of Nairobi, Nairobi.

Plug, Charlotte, Jianhong Cecilia Xia, and Craig Caulfield. "Spatial and temporal visualization techniques for crash analysis." *Accident Analysis & Prevention* 43.6 (2011): 1937-1946.

Private Investigator (2021) The Comprehensive Private Detective Resource.

Road Accidents Statistics Worldwide: Report on Mortality Rate per Country (2019).

Silverman, Bernard W. *Density estimation for statistics and data analysis*. Routledge, 2018.

Wim, W. (2020) Socio-Economic Costs of Road Crashes in Middle-Income Countries: Applying a Hybrid Approach to Kazakhstan. IATSS Research, 45, 293-302.

World Health Organization (2013) Global Status Report on Road Safety. World Health Organization, Geneva.

World Health Organization (2018) Global Status Report on Road Safety. World Health Organization, Geneva.

Zahran, E.-S.M.M.; Tan, S.J.; Yap, Y.H.; Rahman, E.K.A.; Husaini, N.H. A novel approach for identification and ranking of road traffic accident hotspots. MATEC Web Conf. 2017.

Zhou, H.C., Yuan, C., Dong, N., Wong, S.C. and Xu, P.P. (2020) Severity of Passenger Injuries on Public Buses: A Comparative Analysis of Collision Injuries and Non-Collision Injuries. Journal of Safety Research, 74, 55-69.

APPENDICES

Appendix A

A: Global Moran's I Summary

Global Moran's I Summary by Distance

Distance	Moran's Index	Expected Index	Variance	z-score	p-value
145.00*	0.294138	-0.009259	0.007066	3.609255	0.000307
215.00*	0.322425	-0.006849	0.005080	4.619812	0.000004
285.00*	0.368876	-0.005208	0.003736	6.120090	0.000000
355.00*	0.328023	-0.004695	0.003044	6.030798	0.000000
425.00*	0.307517	-0.004425	0.002172	6.693424	0.000000
495.00*	0.235535	-0.004405	0.001658	5.893034	0.000000
565.00*	0.218593	-0.004386	0.001372	6.019170	0.000000
635.00	0.227131	-0.004348	0.001118	6.923346	0.000000
705.00	0.208141	-0.004348	0.000935	6.950002	0.000000
775.00	0.178757	-0.004348	0.000768	6.608078	0.000000

First Peak (Distance; Value): 285.00; 6.120090

Max Peak (Distance; Value): 705.00; 6.950002

Distance measured in Meters

* At least one distance increment resulted in features with no neighbors which may invalidate the significance of the corresponding results.

B. Incremental Autocorrelation Parameters

Incremental Autocorrelation Parameters

Parameter Name	Input Value
Input Features	Mogandishu_Accident_Data
Input Field	F
Number of Distance Bands	10
Beginning Distance	145.000000
Distance Increment	70.000000
Distance Method	EUCLIDEAN
Row Standardization	False
Selection Set	False

Appendix B Ethics Certificate

NEAR EAST UNIVERSITY

SCIENTIFIC RESEARCH ETHICS COMMITTEE

30.11.2022

Dear Abdikarin Isak Hashi

Your project **"Road Accidents in Mogadishu, Somalia: A Geographic Information System Analysis and Assessment"** has been evaluated. Since only secondary data will be used the project does not need to go through the ethics committee. You can start your research on the condition that you will use only secondary data.

Prof. Dr. Aşkın KİRAZ

The Coordinator of the Scientific Research Ethics Committee

Appendix C Turnitin Similarity Report

Mustafa Alas | User Info | Messages (4 new) | Instructor ▼ | English ▼ | Community | ⑦ Help | Logout

turnitin

Assignments | Students | Grade Book | Libraries | Calendar | Discussion | Preferences

NOW VIEWING: HOME > 20213907 ABDIKARIN ISAK HASHI > THESIS

About this page

This is your assignment inbox. To view a paper, select the paper's title. To view a Similarity Report, select the paper's Similarity Report icon in the similarity column. A ghosted icon indicates that the Similarity Report has not yet been generated.

Thesis

INBOX | NOW VIEWING: NEW PAPERS ▼

Submit File

Online Grading Report | Edit assignment settings | Email non-submitters

AUTHOR	TITLE	SIMILARITY	GRADE	RESPONSE	FILE	PAPER ID	DATE
Abdikarin Isak Hashi	Abstract	0%	--	--		2013158216	13-Feb-2023
Abdikarin Isak Hashi	Chapter V- Conclusion	0%	--	--		2013167068	13-Feb-2023
Abdikarin Isak Hashi	Chapter III- Methodology	3%	--	--		2016461919	17-Feb-2023
Abdikarin Isak Hashi	Chapter I- Introduction	8%	--	--		2013158600	13-Feb-2023
Abdikarin Isak Hashi	Chapter II- Literature Review	9%	--	--		2013159047	13-Feb-2023
Abdikarin Isak Hashi	Chapter IV- Results and Discussion	10%	--	--		2013161045	13-Feb-2023
Abdikarin Isak Hashi	Full Thesis	13%	--	--		2013168028	13-Feb-2023