UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

Project

A Big Data Analytics Architecture for Urban Transportation System:

A Case for Nairobi City

By

Justin Muchiri Njeru – P54/11125/2018

Supervisor

Dr. Elisha Abade

A research project submitted in partial fulfilment of the requirements of the award of Master of Science in Information Technology Management (Msc. ITM) of the University of Nairobi

May, 2020

DECLARATION

I hereby declare that this project is my original work and has not been submitted for examination in this university or elsewhere for an award of any other degree

Justin Muchiri Njeru
P54/11125/2018

Sign: ---------- *Date:* ----------

This research project has been presented for examination with my approval as the University Supervisor

Dr. Elisha Abade
School of Computing and Informatics

Sign: ---------- *Date:* ----------

DEDICATION

To Wambura, Wanjiru, Njeru, Githinji and Nyambura

ACKNOWLEDGEMENT

This paper and the research behind it would not have been possible without the exceptional support of my supervisor, Dr. Elisha Abade. Special Appreciation to Prof. Waema, Prof. Wagacha, Dr. Ayienga, Dr. Mwaura, Madam Ochukut and Mr. Moturi of School of Computing and Informatics fraternity for the insightful comments offered during my presentations and consultation meetings. I acknowledge the inputs of at the University of Nairobi.

Finally, it is with pleasure that I acknowledge the contributions of transportation agencies in Kenya: NTSA, KURA, KeNHA and Transport department in Nairobi City Government for on time submission of questionnaires and helpful interviews.

ABSTRACT

Urbanization in Nairobi City has witnessed rapid growth over the last 10 years, with the current 2019 census indicating a city populace of about 4.4 million dwellers (KNBS, 2019). This has contributed to transportation challenges, due to constrained public transportation infrastructure. Despite of the additional road networks and provision of road bypasses in Nairobi city, the traffic jam has not reduced either.

The objectives of this research are to find components necessary to establish architecture for Big Data Analytics for Urban Transportation System. The research findings wants to identify whether the big data analytics should be embraced in Nairobi city to enhance route mapping, navigation planning and point to point destination planning. The fact that Nairobi is a hub of major innovations like iHub, Ma3Route, C4D Lab is also advantageous to the researcher. There is also 4th and incoming 5th Generation Network (5G) and Home/Business Fibre link coupled to many internets wireless access points.

City travelers, find themselves in congested and jammed route due to lack of accurate and timely traffic insights of how the traffic is expected on certain routes to better plan and adjust accordingly. Policy Makers, Researchers, Systems Developers and Investors are also to benefit from the research findings.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

AI	Artificial Intelligence
BRTS	Bus Rapid Transport System
CA	Communication Authority
CPI	Consumer Price Index
DOI	Diffusion on Innovation
EU	European Union
GDPR	General Data Protection Regulation
GPS	Global Positioning Satellite
ICT	Information Communication and Technology
ICTA	Information Communication and Technology Authority
IEA	Institute of Economic Affairs
IEEE	Institute of Electrical and Electronics Engineers
IFC	International Finance Corporations
ILO	International Labour Organization
ITF	International Transport Forum
ITS	Intelligent Transportation System
JICA	Japan International Cooperation Agency
KNBS	Kenya National Bureau of Statistics
NTSA	National Transport and Safety Authority
NIST	National Institute of Standards and Technology
PSVs	Public Service Vehicles
SDGs	Sustainable Development Goals
TOE	Technology-Organization-Environment
UAPI	Urban Areas Performance Index
UN	United Nations

DEFINITION OF TERMS

Big Data is a term applied to data sets whose size or type is beyond the ability of traditional relational databases to capture, manage and process the data with low latency. Big data has one or more of the following six characteristics: high volume, velocity, variety, veracity, Volatility or Validity.

Big Data Analytics is the use of advanced analytic techniques against very large, diverse data sets that include structured, semi-structured and unstructured data, from different sources, and in different sizes from terabytes to zettabytes.

GMCS: Global Mobile Consumer Survey reports comprises of analyzed data from thousands of mobile subscribed users and it is focused on consumer behaviors, trends, and sentiments on mobile phone products and services across the globe.

IEEE: (Institute of Electrical and Electronics Engineers) is the world's largest technical professional organization dedicated to advancing technology for the benefit of humanity. IEEE and its members inspire a global community through its highly cited publications, conferences, technology standards, and professional and educational activities.

ITF: The International Transport Forum is an intergovernmental organization with 54 member countries. It acts as a think tank for transport policy and organizes the Annual Summit of transport ministers. ITF is the only global body that covers all transport modes.

ITS: Intelligent Transportation System (ITS) is an advanced application which aims to provide innovative services relating to different modes of transport and traffic management and enable users to be better informed and make safer, more coordinated, and 'smarter' use of transport networks.

KeNHA: Kenya National Highways Authority is a state corporation, established under the Kenya Roads Act 2007 with the responsibility for the management, development; rehabilitation and maintenance of class A, B and C roads in Kenya.

KURA: Kenya Urban Roads Authority (KURA) is a statutory body established by the Kenya Roads Acts, 2007.KURA is responsible for the management, development, rehabilitation and maintenance of all public roads in cities and municipalities except where these roads are categorized as national roads.

KRB: Kenya Roads Board is a State Corporation established by the KRB Act No.7 of 1999. KRB's mandate is to oversee the road network in Kenya and coordinate the maintenance, rehabilitation and development funded by Road Maintenance and Levy Fund.

Mobility: Mobility is having transportation options that you can count on to get you where you need to go. It is the ability of individuals or goods to move through a transportation network.

NTSA: National Transport and safety Authority was established through an Act of Parliament; Act Number 33 on 26th October 2012, to harmonize the operations of the key road transport departments and help in effectively manage the road transport sub-sector and minimizing loss of lives through road crashes.

Rate of adoption: is defined as the relative speed at which participants adopt an innovation. Rate is usually measured by the length of time required for a certain percentage of the members of a social system to adopt an innovation.

SDGs: Sustainable Development Goals were established in September 2015 by the UN General Assembly and adopted as the 2030 Agenda for Sustainable Development that includes 17 Sustainable Development Goals (SDGs).

CHAPTER ONE: INTRODUCTION

Urban Transportation Systems are very important in city life; however, in developing countries the transportation systems are characterized by congestion, chaos and gridlocks. A massive research effort has been devoted to this field of study in the recent past. Therefore, this paper is devoted to find out Intelligent Transportation Systems (ITS) that is informed by analyzing big data in Nairobi city. In the recent past, studies in the Intelligent Transportation Systems (Bekiaris, 2017), (International Transport Forum, 2019) and many others have found that expansion of road networks does not solve the traffic situations.

The modern cities need a paradigm shift, therefore to tackle the traffic situation in cities and upcoming urban areas. Big Data Analytics is what this research paper propositions to develop a framework that can be used by the City Planners and National Agencies of Transport Management to map, manage and plan an Intelligent Transport System. The main driver of this research is the high rate of Mobile penetration with geo-locative apps and great strides information sharing embedded in social apps. New possibilities that have emerged on how to collect and analyze data in real-time have emerged and are now being embraced in other sectors.

In the recent past, general transport feed specification (GTFS, 2019), has been in center stage of researchers to enable public transport players and agencies to transmit data that is consistent and can be accessed by a number of devices (Sarah, 2015). However, the developing countries have found it a challenge to use the GTFS because of a varied of reasons; the traffic pattern is incompatible with those of developed countries and the level of technology is not at par as those of developed countries. The transportation players, however need to have a framework of traffic systems that is compatible with people, road usage and drivability behaviors of the local citizens.

Analysis of big data has profound impact on the way data is used for City planning, traffic management, accident reporting dashboard, predicting flow of traffic, enhanced travel plans and distribution of social amenities (Zhu et al., 2019). Further, the Government of Kenya (GoK) recognizes the transport sector as a facilitator of rapid economic growth. Transportation systems are imperative for cities economic growth because a number of services interdepend on the transportation systems (UN Habitat, 2016).

1.1 BACKGROUND

Traffic congestion is a major challenge for city dwellers in Nairobi and across many cities in developing countries. There is great loss of productivity for commuters who have to spend over three hours in traffic. Public safety is not assured for most travelers due to high road carnage and pollution. The problem of urban traffic and congestion is as resort of population influx into urban areas (KNBS, 2019). Population growth is a worldwide problem that contributes to difficulties in urban transportation systems. It has been confirmed by a number of researchers (Thakur, 2019), (Bekiaris, 2017), (Lee, 2019) and many others that it is difficult to improve the traffic system performance by using the traditional control methods. Transportation includes movement of people and goods in rails, air, water and roads. This research will focus more on urban transportation systems on roads and intends to incorporate analyzing of data sets from various sources including social media, CCTV, traffic sensors to better manage transportation sector in Nairobi city.

The quarterly sectoral reports by (Communications Authority, 2019), indicated that mobile use has increased to (106.8%) per cent and Mobile Internet Services rose by (2.5%) per cent. Cell phones are a significant element of the versatile Internet scene and access to web by Kenyans with 46.8 million subscriptions. The mobile platform provides quantitative and qualitative data sources that by use of efficient data sharing through existing social media can contribute to big data for Analyzing traffic in Nairobi city.

This study focuses on Big Data Analytics unlike many other studies that have been conducted using the fuzzy logic system, General Transit Feed Systems (GTFS) and neural network to control the road intersections. Generated data from mobile users from open sharing platforms is analyzed and forms an important insight to better understand traffic patterns and better plan for routes as well as give live traffic flows on an open to public web based platforms. The insights from the analyzes inferenced from this research can be used to develop frameworks and specifications for Intelligent Transportation Systems by use of Big Data Analytics and come up with traffic live feeds, patterns and traffic dashboard for responsive, effective, safer and secure transportation systems.

1.2 PROBLEM STATEMENT

The just concluded 2019 population and housing census in Kenya indicate that the population living in towns and especially Nairobi City as 4.4 million, 2.4 million for Kiambu County respectively. In cognizant that Nairobi city has five major metropolitan urban areas; Machakos, Kajiando, Kiambu, Murang'a and Nakuru (KNBS, 2019). All these metropolitan towns have exponential growth in population that has put constrain to the existing road infrastructure and other social amenities like; water, sewage, housing, and medical facilities (UN Habitat, 2016).

To effectively and efficiently manage urban transportation, the National and County Government must develop the requisite urban planning and management capacities. The urban planning players must use technology fit to plan for today and the future. The research has found that the infrastructure effort has not yielded much in effectively combating transportation challenges that comes with congestion, road carnage, pollution and safety of pedestrian in a report by UN on Safety Road Week (2019).

Nairobi traffic congestion is ranked top one in Africa cities (Numbeo, 2019). Lack of aggregated data and information to public is a leading contributor to many transportation challenges in Nairobi City. Travelers need accurate and timeline insights of how traffic is expected on certain routes to better plan and adjusts to any situation appropriately (Zhu et al., 2019). The study proposes use a solution that analyzes big data as a paradigm ways to visualize, control, report and predict traffic systems. Real-time information that is open source accessed can great reduce traffic congestion before it happens and change the driver experiences on roads and inform proper planning of itineraries by most travelers.

Lack of harmonized architectures to harness enormous data generated on various devices like; smart phones, GPS tracking, CCTV mounted on roads and geo-tagged information on social media is what this research wants to contribute. Therefore, this research has identified gaps in specialized architectures for big data analytics in transportation systems. The concept is to come up with architecture that can be used by policy makers, researchers, investors and system architectures to actualize big data analytics systems for harnessing publicly available data to provide an intelligent transportation system in Nairobi and other cities in developing countries.

1.3 OBJECTIVES OF THE STUDY

1. **Identify important components, attributes, specifications and guidelines for successful Architectures for Big Data Analytics for Urban Transportation System.**

2. **Develop architecture for big data analytics for Urban Transportation Systems.**

3. **Identify the methodological issues associated with implementing Big Data Analytics for Urban Transportation Systems**

4. **Identify the ecosystem for Nairobi Metropolitan Urban Transportation Systems and its challenges.**

5. **Develop adoption strategy for development and implementation of architecture for Big Data Analytics for Urban Transportation System.**

6. **To propose recommendations for further research activities on Big Data Analytics**

1.4 RESEARCH QUESTIONS

PART 1

a) What are the components making a reliable architecture for Big Data Analytics?

b) What are the standards or guidelines to be followed for guiding policy makers, systems developers on architecture for Big Data Analytics?

c) Identify architectures that have been developed successfully to deal with Big Data?

d) What data sources are available in Nairobi Metropolitan that can make infusion of Big Data?

e) What are the factors influencing development of the architectures?

PART 2

a) What is the rate of adoption for related technologies in data sharing in regards to traffic incidences?

b) What is the rate of responses on masses in online engagement?

c) What are the data repositories of publicly available information of transport related incidences, alerts and reports?

d) Are the transportation players aware of how the data from various sources could be harnessed and play an important role in decision making, travel planning and traffic routing in Nairobi City?

PART 3

a) What are the expected performance indicators of improved urban traffic management system?

b) What are the methodologies approaches for developing architectures in big data technologies?

c) What are the possible outcomes for Big Data Analytics in transportation systems?

d) What are the associated risks of implementing Technologies Big Data Analytics?

PART 4

a) What are the aspects to be measured in Big Data Analytics for Transportation Project?

b) What are the key performance indicators in socioeconomic impact for an Urban Transportation System that is driven by Big Data Analytics?

c) Is there feasibility study in relation to the Urban Transportation Systems through use of Big Data Analytics, in the recent pasts?

PART 5

a) What are the transportation issues, faced by commuters and drivers in Nairobi Metropolitans on weekly scenarios in their travelling programs?

b) What are the perception and awareness of the public on data privacy and does it affect the trustworthiness to share location information?

c) What are the networks making the Nairobi Metropolitan Urban Transportation System?

d) What is the demographic representation of people and modes of transport in the Metropolitan of Nairobi Area?

1.5 RESEARCH GOALS

The research has identified five important objectives that will bring about the aspect of understanding the current situation and what opportunities lies in both research and development for architecture for urban transportation systems. The goals of this research were met by analyzing the relevant literature, analyzing data and reports with an aim of understanding Big Data Analytics and intelligent transportation systems.

Through quantitative and qualitative analysis the researcher arrives to decision making backed up by data inferenced from this study. The ultimate goal is to develop architecture for Big Data Analytics for urban transportation system. Further the researcher has created awareness of important aspect of using the available data in mobile, social media and GPS devices to better manage traffic systems through analyzes of disaggregated datasets for better decision making and real-time information dissemination.

A window of opportunity has been opened for furtherance in research of Big Data Analytics applications in other disciplines including Transportation, Health, Insurance, Supply Chain, Education, Natural Language Processing, Entertainment and Robotics.

1.6 SIGNIFICANCE OF THE STUDY

People are generating more data and much faster (Bekiaris, 2017); the amount of data available for researchers in our world todays has exponentially grown in discipline of health, security, education, manufacturing, aerospace, engineering and other areas. The report by IBM on "*Better Roads are paved with Big Data Analytics*" (2015) showed that, big data is at central area of study for most researchers, systems developers, policy makers and investors in transportation industries. In this paper, big data technologies have been reviewed and investigated on their applications specifically related to traffic systems in Nairobi City. The inference of this study will lead to development of architecture for big data promoting development of application solutions for data exchange on traffic patterns and better manage traffic systems in Nairobi and other major cities in East Africa that has common traffic problems.

The Policy makers will use the findings to develop sets of improvements and standards to govern the transportation sector. The City County could also use the findings to plan, reorganize city routes and establish a mass transit needs assessment. To foresee the eventual fate of Nairobi City Transportation, the big data analytics has proposed different models and implant designs that would investigate the present transportation solutions and anticipate the future system situations. This would advance decision making process and map required resources for future city projects.

The study has found that the rush hour of Nairobi traffic systems which is characterized by congestion and chaos due to private vehicles and public commuting services, spends 3 to 4 hours daily on a stretch of not more than 40 kilometers two and from the central business district (IBM, 2015). Unfortunately, other international transportation models have failed to work in Nairobi scenarios due to cultural, historical planning challenges and economic aspects (Klopp et al., 2015). Analyzes of heterogeneous data offers a sure step to understand, automate and provide insights for decision making in urban transportation systems (Robert et al., 2016) and (Lee, 2019).

CHAPTER TWO: LITERATURE REVIEW

2.0 INTRODUCTION

There exists substantial literature on Big Data Analytics; however the specific information onto the transportation systems remains a challenge. The study has reviewed information on big data analytics, urban transportation systems, intelligent transportation system and reports on future of transportation in urban areas.

Researcher is devoted to establish architecture for Big Data Analytics for Urban Transportation Systems. The reviewed literature, which is divided into seven sub chapters has analyzed information of the recent study in areas of big data, future transportation systems and reported case studies for Nairobi Metropolitan transportation system.

Several techniques are used to control traffic in different cities; traffic lights systems, fuser logic and the analysis has shown their limitations (Daiheng, 2016). Effects of legislation of Laws, Guidelines, Standards and Principles for protection of data, is an area of concern to both road users and policy makers. This study has reviewed the applicability of the relevant principles in promotion of the research and development of the Big Data Analytics Architecture.

2.1 BIG DATA ARCHITECTURE

Big data architecture is the overarching system used to receive and process enormous amounts of data from various sources and it's often referred to as "big data" (NIST, 2019). The blueprints of big data analytics maps the architecture based on the processing activities in an input output scenario of Transmission, Reception, Storage, Manipulation and Retrieval.

Benefits

The Associated benefits of adopting this kind of technology ensure real-time access of important data and information is always available. Whereas, using the Big Data Technologies like Hadoop can hasten decision making process with critical support for reliable information with eased storage requirements (Silva et al, 2018), is developed on a open source platforms. Data visualization and streaming components ensures availability of data real-time and future projections can be plotted for early warnings and preparedness.

Predicting the impending requests makes effective planning. In urban planning and transportation systems, predicting population growth and managing social amenities is key.

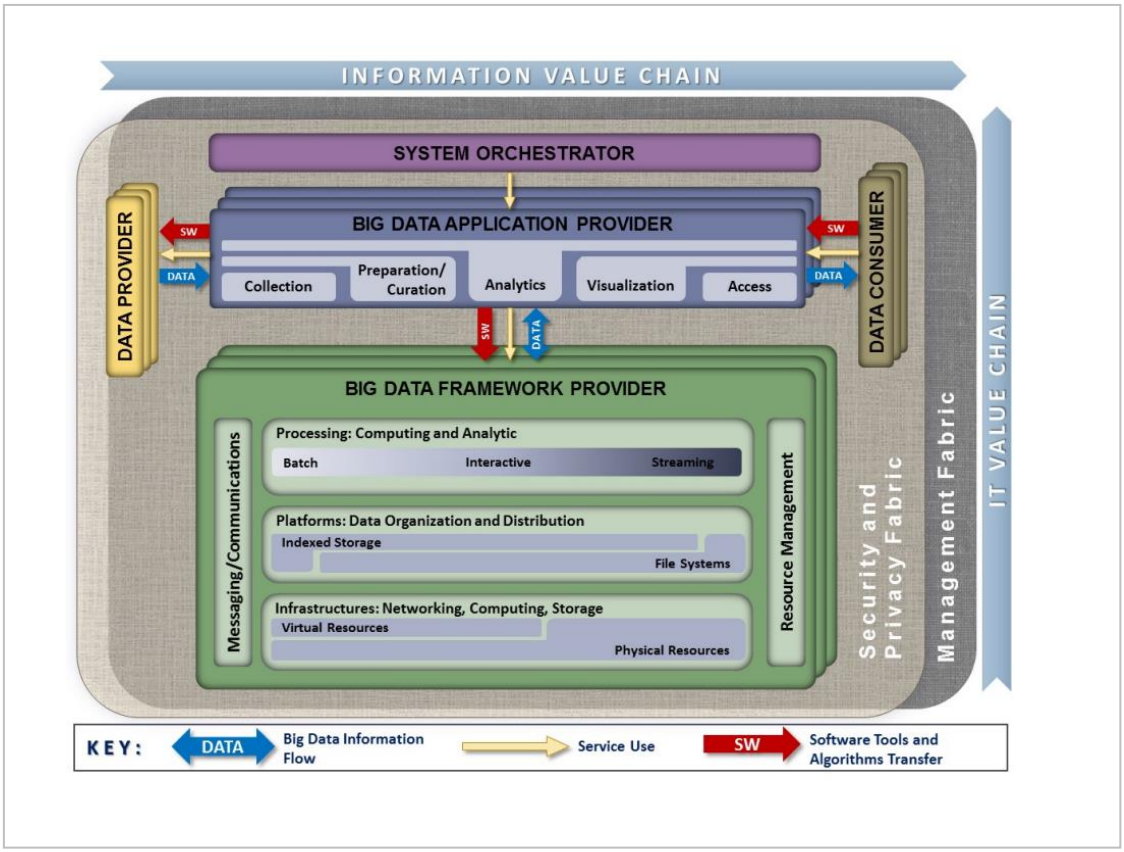


Figure 1: Big Data Reference Architecture; Source: NIST 2019

The diagram illustrates the information value chains and data interdependencies in big data architecture. National Institute for Standard and Technologies credits the above diagram.

Data sources: this is where all the components for data collection are anchored. The sources generating the big data comprises of manned sources and unmanned sources. These sources includes: Social media pages, CCTV live Cameras, Systems logs and GPRS location data from speed gadgets mounted on vehicles.

Real-time messaging: is created to allow information penetration to external devices in real-time. No storage is required for this scenario.

Data store: comprises of unstructured and structured data storage commonly referred to as a data lake. Unstructured types of data storage comprises of images, videos, audios, texts, logs and html types of files. Cloud based infrastructure like Hadoop has established framework for data storage complete with file system called Hadoop File Systems (HFS).

Batch processing: before data can be presented or stored for big data analytics, encapsulation is done to clean and make data presentable. Batch processing combines multiple file system in a process to filter, aggregate or convert data using reliable computing power in a cloud infrastructure offered as Platform as a Service (PaaS).

Reporting Tools: after ingesting and processing various data sources, tools to analyze the data is required. Frequently used tool is BI (Business Intelligence) tool and it requires professionals in data science expertise to explore and make use of available data.

Automation: the computing devices are automated to make repeatable workflows so that big data is orchestrated through auto-feeding of data and batch processing without human intervention (Microsoft, 2018). Automation reduces the human interference to the processes of acquiring data, loading it to processors and finally deriving the insights in a complete repeatable process.

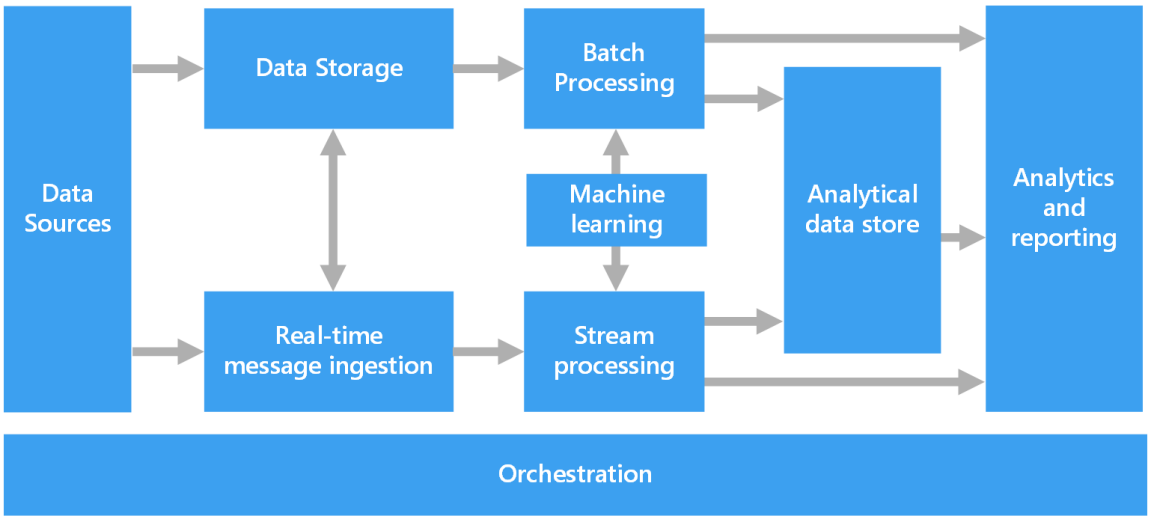


Figure 2: Components of big data architecture; Source: Microsoft Inc.

Challenges

The possibilities of big data architecture are to predict, promote data sharing real-time for optimal decision making for an organization. Prediction of Transportation trends and incidences can be made possible, but it is not without its challenges. National Institute of Standards and Technology (NIST, 2019) has published some of the considerations:

a) That, data volume can be extremely large and overwhelm the storage capacity

b) That, data velocity can bring about speed requirements for an elastic cloud computing platforms, unnecessarily expensive for certain applications

c) That, data validity could bring issues of data disposition and short life times for data needed for real-time analytics.

d) Unreliable data analysis can make wrongful decision making process. Therefore, duplicate data and missing information has to be done at pristine level to avoid any form of noise in big data.

e) Exponentially growing data can make the storage of data requirement storage space to increase with potential rise in storage costs. When not properly provision the storage capacity can pause a big challenges for this type of projects.

f) Cyber security poses a threat to provisioning the big data analytics. The number of attacks on these kind of platforms has exponentially gone up with devastating outcomes of data lose or compromised information systems. Mitigations to prevent potential penetration into the system must therefore be deployed to protect big data from fraudsters and hackers

2.2 BIG DATA ANALYTICS

The information today is the most valued assets. Analyzes of big data has totally changed and revolutionized the way of life. In the recent past, researchers have immersed on fact finding impacts, applications, and trends in big data analytics. Its applications in Commerce, Banking, Healthcare, Education, Manufacturing, Insurance, Media and Transportations are quite substantive (Ajah, 2019). Literature in Big Data Analytics (BDA) in transportation is scarce especially with application in context of Nairobi City or any of East African Countries' Traffic Systems.

The Transport sector is in perfect advantage point for technological growth especially in big data analytics. DHL, a global logistics and transportation company that provides logistics to massive flow of goods and at the same time creates vast data sets of logistics data defines transportation as a "practical arithmetic". The logistic data contains of millions of shipments every day with information on; origin and destination, size, weight, content, and location are all traced across global delivery networks (Martin et al, 2015).

There is huge untapped data on transportation sector that big data analytics can leverage on for improved logistical planning, prediction of traffic flow and providers have realize that BDA is a game changer. BDA can also provide competitive advantage to logistic

companies because of optimization of customer information and globalization of availability of network.

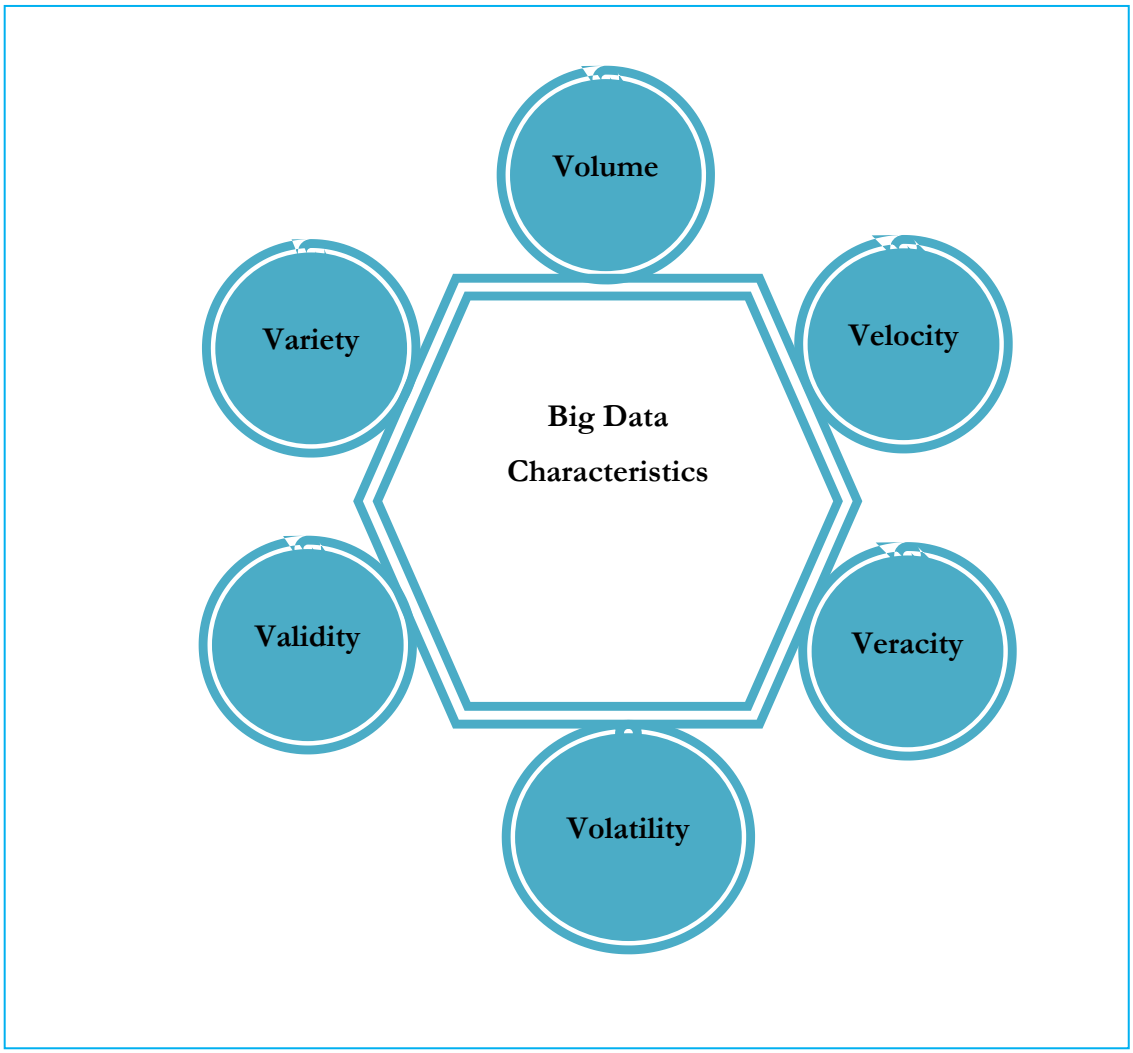


Figure 3: Characteristics of Big Data

Originally according to IBM the first 3Vs were derived for big data characterized by size/volume, variety and speed/velocity. More Vs have been added which are the Veracity that denotes the data source reliability, Validity for the correctness of the information and Volatility which denotes the shelve life of the data received. All these 6Vs are important aspect of collection and analyzing data (Silva et al, 2018). BDA is a process according to Acharjya (2016) is in applications like Machine Learning, Data Science, Data Mining, Predictive Analytics and Texting Analyzing (Gandomi, 2015).

Machine Learning: is a subset of artificial intelligence that trains machines on how to learn using natural language processing techniques. Robotics industry and driverless automobiles are on higher gear of advancement to machine learning.

Data mining: has been applied to resolve complex business equations. The software's in data mining are useful in helping users sift through all the hectic and messy data and pinpoint what is relevant and make use the information to predict possible outcomes.

Data analytics: involves use of mathematical methods to gain information from data

Descriptive analytics involves instances where data is described to make understanding based on analysis of past, present and projections of future.

Diagnostic analytics is part of analysis of data to investigate the causes and efects.

Predictive analytics involves outcome to predict future based on analysis of present and past datasets. Concepts are designed to provide mathematical solutions to make inferences of data relationships in different timings and map a future outcome.

Prescriptive analytics uses a mathematical model to prescribe data from the available sets

Statistics: conclusions of inferred data are drawn based on the simulation constructs of models designed to analyze datasets of various types to establish character and behavior of certain scenarios.

Simulation: is a real-world imitation of complex systems based on models which can evaluate numerical data to establish a statement of behavior of scenario.

Optimizations: involves combination of independent variables and objectives related to variables to bring optimal function values.

In cognizant that the Big Data Analytics generates immense data is a major setback. The challenges facing the industries while implementing big data analytics are to overcome certain characteristics of bog data. The data privacy has also emerged as the regulation on GDPR enacted by the Parliament of Kenya (2018) comes into effect. Security concerns are also ongoing concerns where there are possible data breach and the complexity of storage facilities.

There are different software's options frequently used to analyze structured big data. NoSQL is one of the applications available for analyzing and is characterized by data that is mostly available in soft state and consistent. MongoDB and TerraStore are both classified as NoSQL-related products that are used for "document-oriented applications" (Mukherjee, 2016). Corporations like Google, IBM, Yahoo, Microsoft, Netflix and

Amazon have developed there specific software's to deal with Big Data Analytics. Open Source solutions to Big Data are; Mahout, MOA, R Project, Vowpal, Wabbit, Pegasus, GraphLab Create and MLLib (Pedro, 2015). The open source distributed processing architecture s used in analyzing Big Data are; Hadoop, Spark, Storm, Flink and H2O (Pedro, 2015).

Table 1: Open Source Big Data Platforms Comparison

Features	Mahout	MOA	Wabbit	R Project	PEGASUS	GraphLab	MLLib
Programming Paradigm	Parallel Computing	Serial Computing	Parallel Computing	Serial Computing	Parallel Computing	Parallel Computing	Parallel Computing
Operating Systems	Windows, Linux Mac OS X	Windows, Linux Mac OS X	Windows, Linux Mac OS X	Windows, Linux Mac OS X	Windows, Linux Mac OS X	Windows, Linux Mac OS X	Windows, Linux Mac OS X
Software Requirements	Hadoop / Spark, JDK 1.6.x, Maven 3.x,	JDK 1.6	none	none	Hadoop, Apache Ant, JDK 1.6.x Python 2.4.x GnuPlot 4.2.x	64 bit OS	Spark
Programming Languages	Java	Java	C, C++	R, S, C, C++ Fortran	java	C++, Python	Java, Scala, Python
User Interface	N/A	GUI, CMD, Java API	N/A	R Studio Graphical Interface	N/A	GUI via GraphLab Canvas	N/A
Data Types	All	All	All	all	Graphs	Graphs	All
Available Algorithms	Recommendation on mining, Classification, Clustering and others	Classification Regression Clustering and others	Own single algorithm	undefined	Page rank, Random Walk with Restart	Belief Propagation, Gibbs Sampling or co-EM	Classification, Regression, Clustering
Scale of Supported Datasets	Up to Petabytes	Few megabytes	Up to Terabytes	Upto gigabytes	Up to Petabytes	Up to Petabytes	Up to Petabytes

From the table above, we draw the following conclusion; that the platform and support for data set is important while choosing any of the platforms. For companies with large infrastructure will require parallel computing establishment of either; Mahout, Wabbit, Pegasus, GraphLab or MLLib. It is evident that all the platforms are cross platform and can run on any operating system. The platforms that need heavier system requirements are Mahout and Pegasus though all the platforms are open source and thus free for users.

2.3 ARTIFICIAL INTELLIGENCE IN TRANSPORTATION

Artificial Intelligence (AI) in Transportation is conventional to give instruments that permits taking care of issues in vehicle and their communication (air, street, railroad and water transport) and is utilized in regions, for example, Real time transport overseeing, design, activity, time calendar and organization of strategic frameworks and cargo transport. Intelligent transportation systems (Mahashreveta, 2019) was introduced to bring and bridge gaps in management of traffic systems through listening of technology devices and making informed decisions on routes and planning of travelers (Daiheng, 2016). ITS objectives are to increase safety and security of road users in urban areas.

Cooperative Intelligent Transport Systems (C-ITS) will enable connected vehicles to openly broadcast relevant warnings (International Transport Forum, 2019). In addition the International Transport Forum (ITF) reported that it will be possible for connected vehicles to share system performance in real time and throughout the entire road network. This is expected to benefit all road users, connected or not. Network managers could use data from short range messages as a rich source of information, where road-side receivers are installed. The ITF found that new technologies could offer solutions but also pose problems.

A negative influence of Big Data impacts on road safety, for instance, drivers might be distracted from driving by new services, from navigation and infotainment to Cooperative Intelligent Transport Systems (C-ITS) alerts. The report further warns of biases in harnessing Big Data for Road Safety, whereby there is a lot of excitement and professionals need to acknowledge a number of pitfalls (Abduljabbar, 2018). Therefore, it is imperative for transport players to govern where, how and when to use these dynamically changing technologies to make a speedy development that will relieve jamming and make travel time more reliable to the road users.

Manual reports written by individuals can have a delay in spotting incidents and prompt quicker decision making. The algorithms for incident detection were first implemented using statistical techniques. Over time the algorithms have been refined for accurate image detection. "Efficient Multiple Model Particle Filter (EMMPF) was developed to sense incidents on highways using computer-generated data and field data (Abduljabbar, 2018).

Predictive modeling in Artificial Intelligence has been developed for transportation systems to provide travel information in timely manner. Advance traveler information

systems are emerging backed by historical data. In the recent past, research focused for short-term flow prediction by using feedforward neural network. Effortlessly, there is social media engineering tactics in use to give road users travel information about; incidents, accidents, congestion situation and informative security & safety bulletins through Twitter, Google and Facebook.

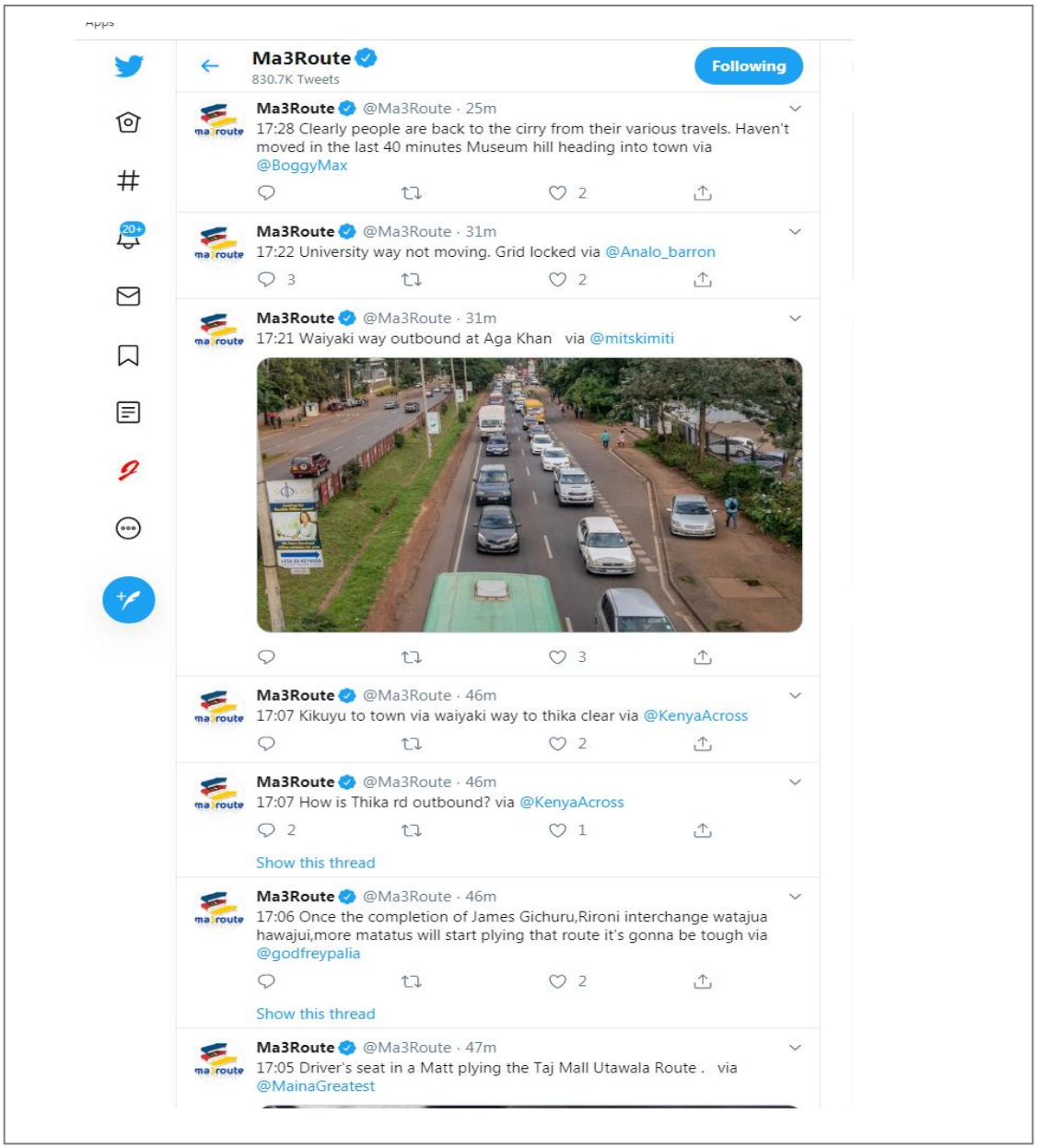


Figure 4: Twitter Traffic updates - Credits: @Ma3Route

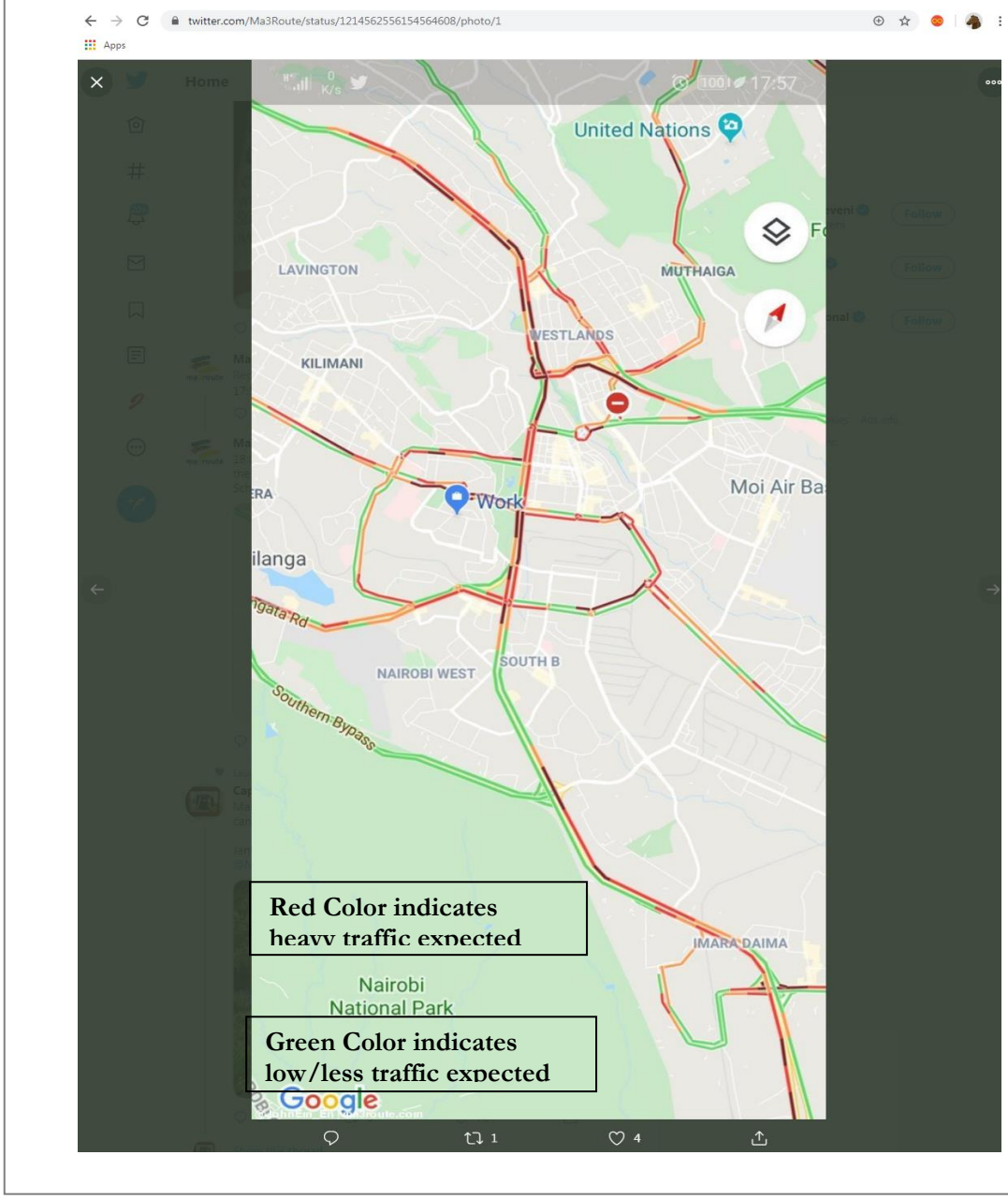


Figure 5: Google Maps indicating traffic Situation for a section of Nairobi City

However it has limitations as the AI instruments and their control are suitable for intricate and differentiated transport systems. The disadvantages of this model are that it requires huge hardware support, learning process can take long time and there are no methodologies for neural network architecture and related functions. AI is already having a profound impact on the way interact with world around us, it is already disrupting the way we move people and goods. AI offers new solutions to transport challenges by making affordable solutions using the existing market ready platforms like Google, Twitter and

mobile apps. A white paper by International Finance Corporation (IFC) on How AI is making Transport Safer, Cleaner, More Reliable and Efficient in Emerging Markets (2019) elicit some fresh ideas on the impact and endless possibilities once it is implemented.

In advanced economies, there are significant applications of AI; Small scale Autonomous bus trials have been implemented in Finland, China, Singapore, Norway, Sweden and France, Conde and Twinn (2019). Ride hailing and sharing platforms like Uber, Bolt, Olli and Taxify uses AI for driver and ride matching, route optimization and driver onboarding. The ride hailing apps have already disrupted the transport sector in their easy offering of services.to single out Uber which is now synonymous with taxi, and has reached over 700 cities around the world has mastered the penetration by offering more services like; Ride, Eat, Freight, Bike and Fly (Khosrowshahi, 2019).

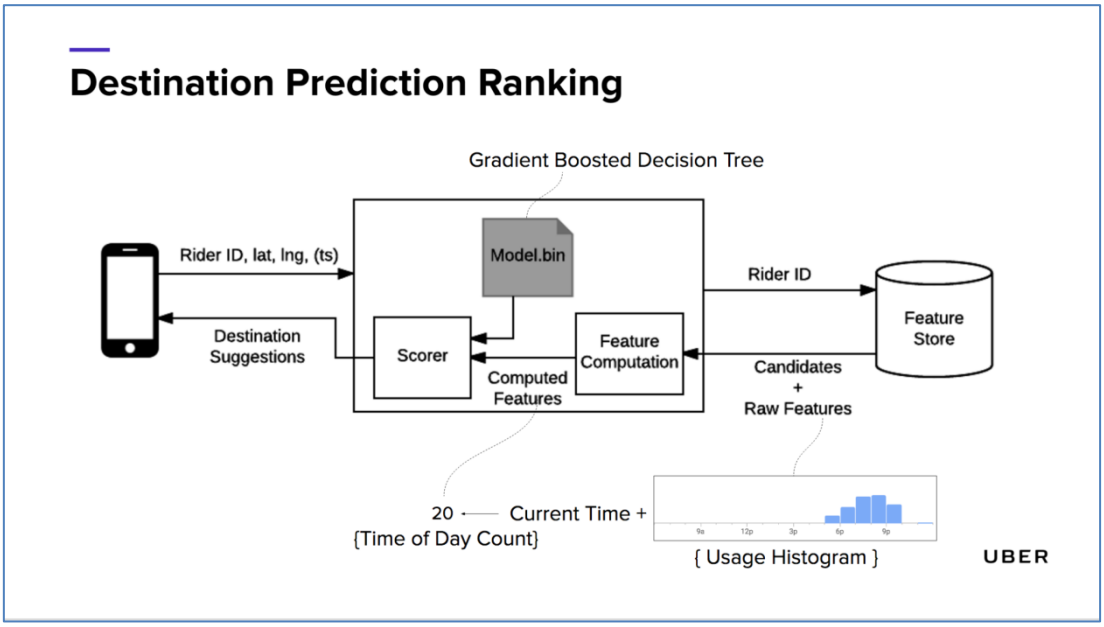


Figure 6: Uber Engineering Model with Machine Learning and (AI)

During Uber Engineering's first Machine Learning Meetup on September 12, 2017, Principal Engineer (Waleed, 2017) discusses how machine learning extracts raw data and turns it into actionable knowledge about the physical world to improve user experiences on the Uber platform. Examples of Uber Maps' machine learning applications include destination prediction and computer vision for map optical character recognition using Michelangelo Machine Learning platform developed as Cloud Service by Uber (Hermann, 2017).

2.4 ADOPTION OF BIG DATA ANALYTICS

In the recent past, research in adoption of big data analytics is predominance in three philosophies; Technology-Organization-Environment (TOE), Technology Innovation Theory (TIT) and Diffusion of Innovations (DOI) across domains of applications (Baig, 2019). TOE is influenced by Technology, People, Organization and the Environment. In this paper we have reviewed literature and reports on uptake of big data across domains and disciplines (Columbus, 2016). Important insights have been reviewed generally indicating a sharp increase of adoption with increase of 59% in the year 2018 from the previous 17% in 2015, getting a Compound Annual Growth Rate (CAGR) of 36% (Forbes, 2018).

Technology Organization Environment (TOE) Theory: is a process by which a firm adopts and implements technological innovations which is influenced by the technological context, the organizational context, and the environmental context.

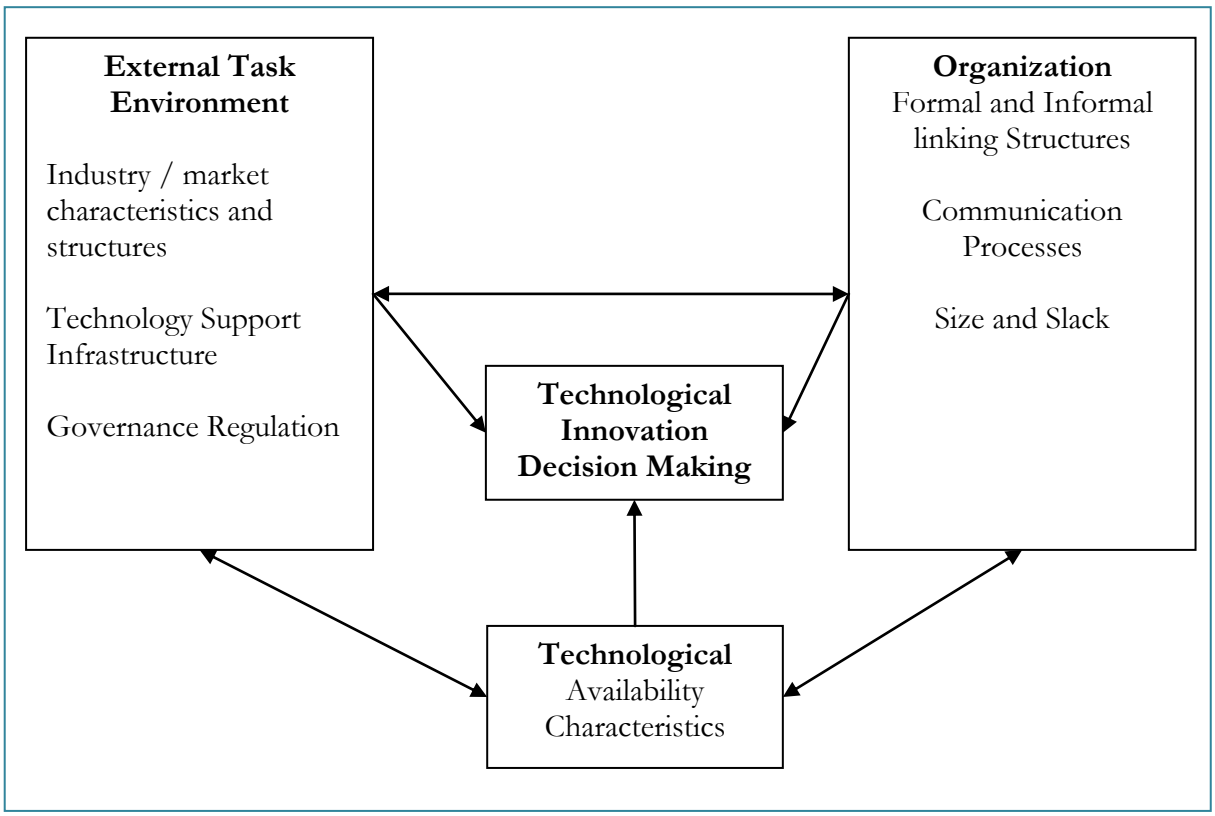


Figure 7: Schematic diagram of TOE Framework

The framework developed by DePietro, Wiarda and Fleischer in 1990 combines Technology, People and Organization in a context to establish dependences and relations. The three elements as elaborated in the figure above, influences the way a firm sees the

need for, searches for, and adopts new technology. This framework reviews technologies relevant to the organization in an environment contextualized of size, structure, industry, microeconomic, macroeconomic and competitive advantages.

Diffusion on Innovation (DOI) theory is still relevant today since its development by E.M Rogers in 1962. Adoption of innovation in social systems does not happen simultaneously and therefore there are stages across the social aspect of people. The group of people will therefore behave different in adapting to idea, product or an innovation and are grouped by Rogers in (Wayne, 2019) into; early adopters, early majority, late majority, and laggards. Each group has certain characteristics as elaborated below:

Innovators: These are venturesome and interested people into the innovation or an idea. These are also risk takers and are often the first to embrace innovation or new ideas.

Early Adopters: represents opinion leaders and enjoy leadership roles in making change opportunities to embrace change with little information to appeal to the masses.

Early Majority: These groups adopt new ideas and innovations before the average person but inquisitive of knowing how the innovation works. To appeal to this group a lot of convincing is done through success stories.

Late Majority: This group waits for the early majority to try the innovation since they are skeptical about the new idea.

Laggards: this group consists of highly conservative kind of people and to appeal to this group includes provisioning them with statistics or force of circumstance to change into the new innovation.

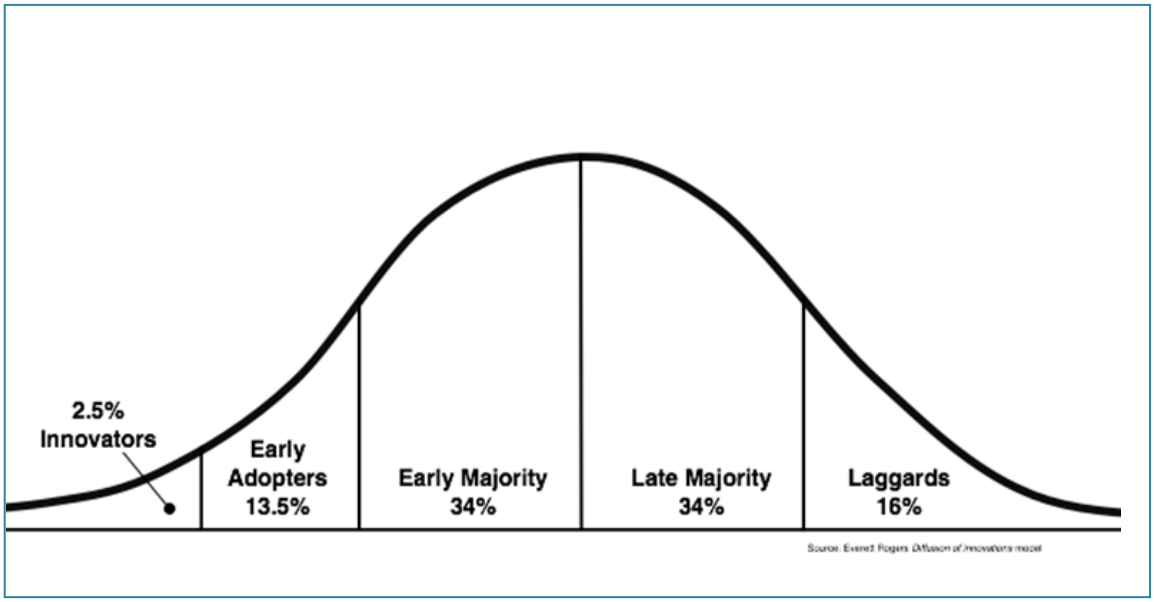


Figure 8: The Diffusion of Innovation Curve (Rogers, 1962)

There are several limitations of Diffusion of Innovation Theory, which include the following:

a) DOI was not developed to explicitly apply to adoption of new behaviors or innovations.

b) It does not foster a participatory approach to adoption of a Big Data Analytics in this research context.

c) It works better with adoption of behaviors rather than cessation or prevention of behaviors.

d) It doesn't take into account an individual's resources or social support to adopt the new behavior (or innovation).

This theory has been used successfully in many fields including communication, agriculture, transportation, public health, criminal justice, social work, and marketing. In transportation the theory explains the acceleration of the use of technology in certain sector of development and the technology. Use of Taxi Hailing apps on mobile phones explains the DOI theory in context to Kenya's Population the rest of the world. The use of Mpesa and other technology can also be attributed to explain the adoption rate of technology and innovation in Nairobi.

The statistics shows the biggest barriers to the adoption of big data among corporations as of 2019, is lack of organization alignment and the agility at 40.3% (Statista, 2019). Respondents suggested that big data adoption was held up by a Lack of organizational alignment or agility and lack of understanding data as an asset at 30%. In this survey conducted in 2019 indicates cultural resistance 23.6%, technology solutions 15.2% and executive leadership at 7.5%.

Hypothetically, the case study of Nairobi City of becoming a data-driven organization in transport management, five (5) milestones and best industry practices will need to be deployed. The five milestones developed by Micro Focus in the article "*Becoming a mature, data-driven organization*" (2017) are; (i) Nascent (ii) Pre-Adoption (iii) Early Adoption (iv) Corporate Adoption, and (v) Visionary/Mature stage.

Table 2: Summary of Big Data Analytics Adoption Stages

Adoption Stage	Milestones to Achieve	How to get there
Initial / Nascent	• Understand available big data • Harness relevant big data • See the value in analyzing information and rich data sources	• Transformation Workshops • Benchmark on similar projects • Researchers' Conference in ITS and Big Data
Pre-Adoption	• Understand the value of faster analytics and faster time to insight • Ask questions in forums on big data analytics and get answers • Extract value from unstructured data sources	• Develop roadmaps and master plans • Establish data discovery services • Conduct open forums on big data • Engage relevant stakeholders in Transport sector through conferences and meetings
Early Adoption	• Enable legacy solutions to keep pace with big data implementations • Leverage a phased approach to tackle key vs. future duplications • Gain executive approvals and buy-ins • Integrate scattered and siloed data and data sources • Establish an efficient data governance model and robust process for big data capture, curation, validation and retention • Gain clarity about tools, technology and necessary skills	• Conduct assessment need • Use case definition consulting • Proof of concept and pilot test implementation • Develop specifications • Architecture / design services • Set budget and conduct public participation forum • Develop policy and regulations on data governance, privacy and sharing of data procedures • Analytics solution • Develop data lifecycle management • Deploy and test security strategies • Establish organization structures for the entire implementation process
Corporate Adoption	• Implementing, operationalizing and supporting real-time big data analytics • Overcome data security concerns and fears of exposing data • Integrate with existing big data investments (NoSQL, SQL, Hadoop Open Source, etc) • Coordinate effectively across the business on big data initiatives • Gather resources and	• Establish hybrid data management services • Conduct integration services • Consult widely on the ecosystems • Establish institutions for Big Date center of excellence • Pre training and sensitization for all levels of users and participants • Promote awareness through established media channels

Adoption Stage	Milestones to Achieve	How to get there
	models for big data analytics initiatives	
Mature / Visionary and Continuous Improvement	Execute big data analytics program for Intelligence in transportation systems using a highly tuned infrastructure with well-established data governance strategies • Better manage costs and understand big data growth • Optimize the resources to effectively compete successfully using enhanced decision-making capabilities • Move analytics closer to the big data sources	• Big Data Cloud Accelerators • Solution management Services • High-Performance Data Ingestion Services • Predictive, Descriptive, Simulations and Data mining Solutions • Machine Learning solutions

Common best practices like; commitment, competency and a culture that embraces change will make the overall achievement of the five adoption stages.

2.5 COST BENEFIT OF AN INTELLIGENT TRANSPORTATION SYSTEMS

In cognizant that the costs of implementing an intelligent transportation systems depends on the city subjects, the scope of specifications, design architecture and the levels of technology to be used. The design process would ultimately guide the policy makers and investors in arriving at the optimum life-cycle costs that include capital and operating costs (McQueen, 2017). In the recent pasts, researchers have advocated for a service evolution approach to determine the scope of such undertakings.

This paper uses a paradigm shift in the cost benefits analysis, where a detailed scope of undertaking is established through a guided framework that provides high-level guidance to influence the investment and modernize city transport. The limitation of this research is to establish important components for guiding specifications development architecture for consideration. In the investigation we found that the underlying key benefits for such an undertaking were primary grouped in; safety, efficiency and user experience.

2.6 NAIROBI TRANSPORT SYSTEM

Nairobi City connects all major urbans; Kiambu County, Murang'a County, Embu County, Nyeri County, Machakos County, Kajiando County and Nakuru County. The City has higher mobile subscriber convergences than any other part of the country. Nairobi is a hub

of major innovations like iHub, Ma3Route, C4D Lab and enjoys network infrastructure provided by both 4th Generation Network (4G) and Home/Business Fibre link. The main players in telecommunications are Safaricom, Airtel, Jamii, Zuku, and Telkom. The low cost of handsets, coupled with manageable data packages which goes as low as $ 0.20 USD is a catalyst for mobile usage and on the go data access. According to KNBS (2018), the city of Nairobi is a technology driven and hot bed of major researches in technology sphere.

According to the International Labour Organization (ILO, 2017) the Matatu industry is Kenya's leading mode of public transport. Matatu operators run branded type buses, mini buses and Nissans across the country. It is therefore, a growing sector with presence in both cities, urban and rural areas. There are over 65,000 Matatu Operators in Nairobi as per the Sacco Societies Regulatory Authority (SASRA, 2019). The private transport sector provides public service transportation in and is comprised of drivers, conductors and office staff of about 350,000 people. The sector also offers opportunities to indirect jobs in the informal sector of mechanics, car wash, restaurants and insurance companies.

The National Transport and Safety Authority (NTSA, 2018), indicates there are about 200,000 Public Service Vehicles operating in Nairobi City on 150 routes to metropolitan areas. However, as the city suburbs expand due to high influx of people migrating to the city, resources and infrastructures like roads are strained and drivers diverts from their designated routes to evade traffic. This causes strains on adjacent routes, with ripple effects to major highways.

2.7 DATA SHARING & PRIVACY CONCERNS

Approaches to development of systems architectures assumes that the ideal technology solution will be guided on data sharing and privacy concerns. Therefore, it is paramount to establish key objectives to be observed and evaluated throughout the planning and design process to ascertain policies of data are breached. In the General Data Protection Regulation (GDPR) by (EU, 2016) as explained by (Dove, 2018), it is very clear on how to protect data submitted and generated by users. The rights to forgo data that represent data subjects, need to be defined and anchored in the consent for users and data subjects to consent or not before participating in any data collection.

The researcher has established a success factor to be concerned is to address the issues of how certain information about users is likely to be handled. The limitation of this research

is to develop a general "code of conduct" for architectures and embed it in the strategy for Big Data Analytics for transport solutions. The aforementioned factors will influence how the users are likely to respond willingly to share and accept datasets of location and incident reporting for analysis in developing an intelligent transport system (Korbel, 2019).

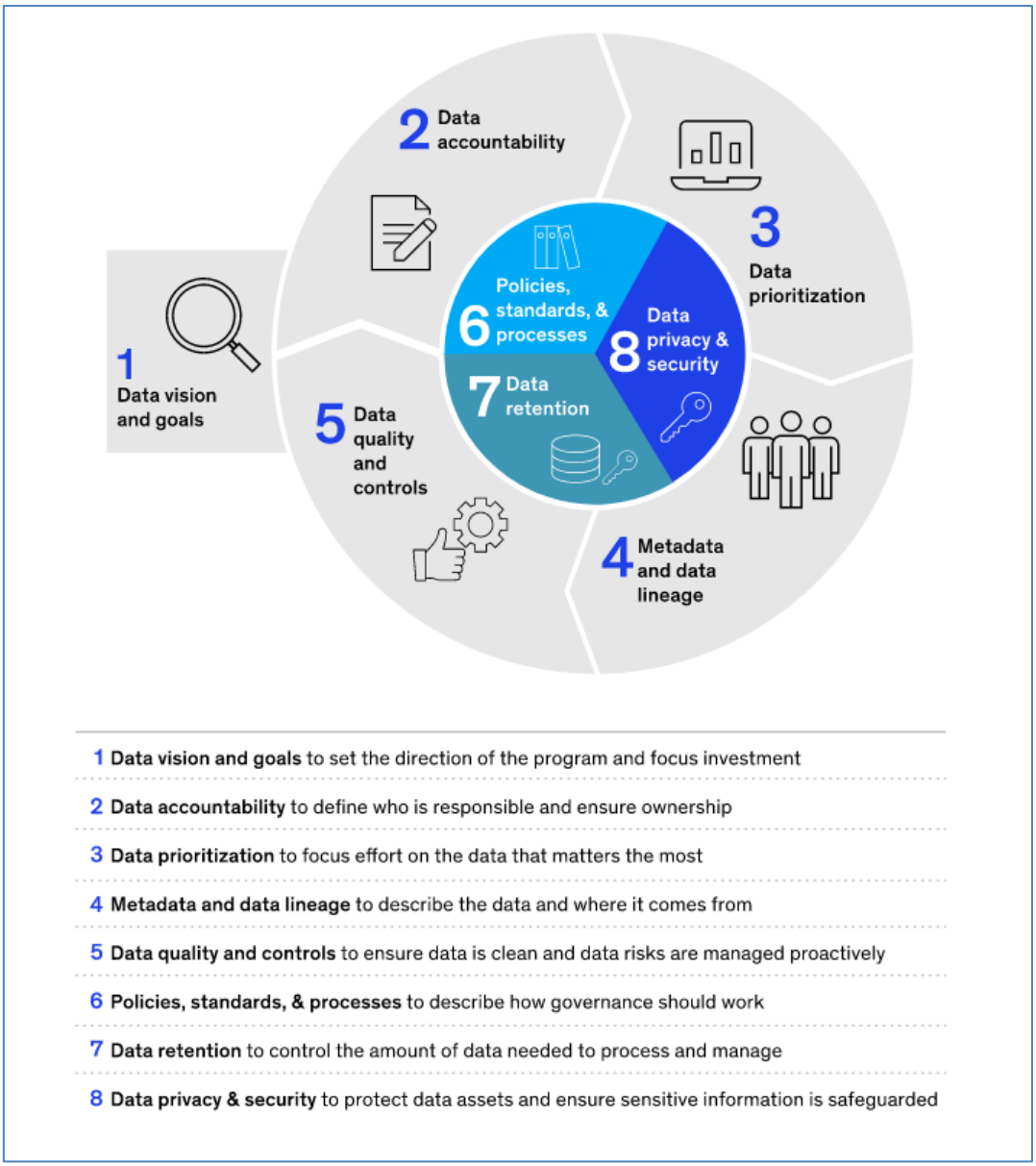


Figure 9: Critical Requirements for effective data governance

There is criticism that GDPR imposes a heavy burden in today's information oriented society. Privacy concerns have higher likelihood in affecting how data is to be acquired, stored, and disseminated and for how long it should be kept. To avoid this bottleneck a smart approach has been proposed to avoid data subject that personifies or defines a persona. However, it must be noted that the GDPR main aim is to balance the need to protect personal data with other important interests, including freedom of expression and

information. In that regard, the researcher has treated personal data with respect and therefore need not to be very concerned about the GDPR.

Though the data about personal attributes like; names, age, sex, county of residents, suburbs, level of education, work, social media handles, number of children, single or married could make the best case of this study, it could trigger compliance issues to the data privacy and sharing concerns. Therefore, this research has adapted to minimization and anonymization of data collected from the data subjects (Rik, 2019) as defined in the GDPR. This will absolve the researcher from many responsibilities and possible litigations.

2.8 THE FUTURE OF TRANSPORTATION

A number of industries are now being disrupted. The transport sector has already been redefined with taxi hailing apps like Uber, Taxify, Bolt and Little cabs. The Transportation industry is being redefined through disruptive technologies. User mobility is now being defined on basis of efficiency, safety and concerns of health, whereby more commuters are using bicycles and walking as way to keep healthy and fit. The population trends have been on steep increases in urban areas. The just concluded 2019 census indicates halve of the Kenya's population lives in urban areas (KNBS, 2019). Constrains to public amenities is an ongoing concerns that the City management continues to address with limited resources (UN Habitat, 2016). The future of transportation for Nairobi and across many developing cities is driven by technology (Victor, 2017).

A report by United Nations indicated that half of humanity lives in cities today and by 2050, two-thirds of the global population will be urban dwellers (UN Habitat, 2016). This mass exodus of people to urban will result into straining infrastructure systems in urban areas like roads, water and public amenities. It is evident that social impact of competition does not affect food, water, security but also movement of people and goods. The report by the United Nations Conference on SDG, entitled "The Future we want", pays special attention on planning and development of cities to promote economy, social and environmental conscious societies. Further it indicated that data is not adequately detailed, documented and harmonized, or sometimes not available and therefore impacts the quality of decision making.

The City Prosperity Index (CPI) which was developed by (UN Habitat,, 2019) is a proportion of six variables comprising of; Infrastructure, Productivity, Quality of Life, Equity, Environmental Sustainability and Governance. CPI was developed by UN Habitat

to monitor framework of Sustainable Development Goal 11 (UN Habitat, 2016). The tool is tries to rates on a global scale the systemic view of a given city. A key indicator for a sustainable and resilient city is measured based on percentage of city. The Nations accorded and committed to enhance inclusive and sustainable cities as per the SDG 11 commitments (UN Habitat, 2016).

Big data analytics gives important insights on projection of needs by city dwellers in terms of food, water, public space, transportation systems, safety and security of people and assets. World Health Organization (WHO, 2019) found that road carnage remains three times higher in low-income countries than in high-income countries. According to the report by UN during the yearly held safety road week (2019), half of the world's motor vehicle fleet is located in low and middle income countries, 90% of road traffic injuries and deaths occur in these countries. Kenya has grappled in reducing the number of fatalities on her roads by; new Transport and Safety regulations, enforcement of laws and deployment of country wide safety measures for all road users. None, of the initiative seems to give impact in reducing the fatalities in Kenyan Roads.

2.9 OPPORTUNITIES FOR IMPROVEMENTS

In Kenya and around the world the transport facilitates movement of passengers and goods. Mobility is now replacing the transport meaning in the world of technology. Transportation efforts to build more infrastructures for roads need to incorporate the mobility aspect to ensure both motorized and non-motorized are incorporated in provisioned mobility services. However, the infrastructure upgrades has always considered improving the motorized ways. The new technology driven and whereby Data Analytics is applied considerations to establish secure walkways, bikeways and physically challenged persons have designated facilities along the road networks. Mobility is highly advocated term in the new generation and supports by the UN on Sustainable Development Goals objectives for an all-inclusive society (UN Habitat, 2016) will give a paradigm shift in approaching transport as an enabler to mobility.

The institutions established by the Kenyan Government to manage transport sector are faced with challenges of budget and unsustainable growth to enable transport and safety of its citizen. Road transport is a lucrative sector that employs more than 500,000 people mostly youths and it is estimated to be worth billions of Kenya Shillings (SASRA, 2019). It is a sector that cannot be ignored despite of many shortcomings (Mukabanah, 2020). In a Concept paper "Creation of National Institute for Transport Studies", Mukabanah's

proposal is overdue. This research also benchmarked such establishment in USA, Sri Lanka and Japan. The Transport Institute plays important roles in knowledge creation, research and development of transport solutions.

The future of transport is in mobility. In the recent past technology is in center stage to provide mobility concepts. In Kenya the State Corporations Act (Cap. 446) through a Gazette notice No. 17 of 2017, established the Nairobi Metropolitan Area Transport Authority (NAMATA, 2017) to effectively address Traffic Problems affecting the five Counties which are Suburbs to Nairobi City. Further Nairobi has rolled out a pilot testing of road junction decongestion system at Kilimani Nairobi dubbed "Intelligent Transportation System", (Huawei, 2017) which can be scaled up to benefit other areas affected by traffic congestions in Nairobi and other towns. Such efforts has created Acts of Parliament to enact concerted efforts can be channeled easily and efficiently through such kind of national institute of transport. Establishment of committee to access the global trends and apply cases for Nairobi can also add value to the research and development of frameworks and architectures for guiding software developers to come up with bespoke systems that have unified standards for interoperability.

2.11 CONCEPTUAL MODEL

To tests whether the objectives will be met by this research a conceptual model was adopted with independent variables comprising of; costs, technology, infrastructure, data, performance, skills and scalability. The controls used were adopted from the McKinsey's 7S model. The model visualizes why certain organizations or countries would implement a certain technology and succeed while others fail. These 7S moderating variables are: System, Skills, Staff, Style, Structure, Strategy and Shared Values. Despite that the McKinsey's model originated from the practice as opposed to theory, it has been proved as the ultimate model of testing the organization to achieve its objectives (Gokdeniz, 2017). It also represents the significance of interrelationships that exists among all the seven variables. However, the new information age requires for the model to be adjusted catering for emerging trends and needs (Oya, 2017).

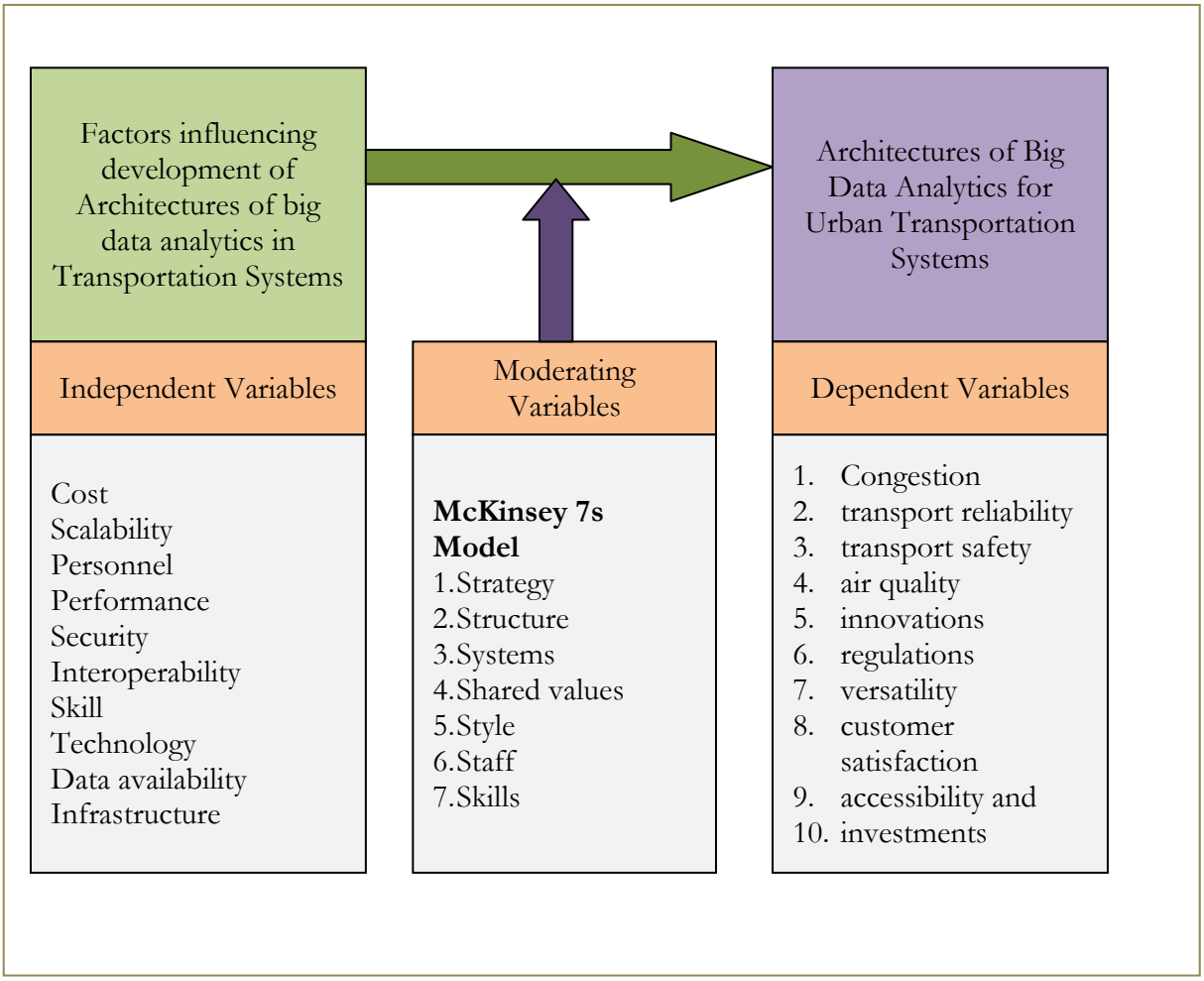


Figure 10: Conceptual Model

2.10 CONCEPTUAL FRAMEWORK

This paper attempts to develop a conceptual framework based on the analysis of big data for an intelligent transportation solution. The proposed framework will be tested on qualitative analysis of surveys to selected organizations; NTSA, NAMATA, KURA, NPS, KeNHA and City County of Nairobi Transport Department. Private organizations that are mainly transport players like "Matatu Sacco's" have also been interviewed.

The theoretical framework diffused from the literature documents is in premise that Big Data Analytics has become a technology innovation that has matured and its now being applied in a number of disciplined subjects amongst them transport industry (Forbes, 2018). The concept tries to interpret how data from various sources will be sourced, categorized, analyzed to provide insights of live feed information on traffic, predict patterns and futuristic provide an intelligent transportation systems.

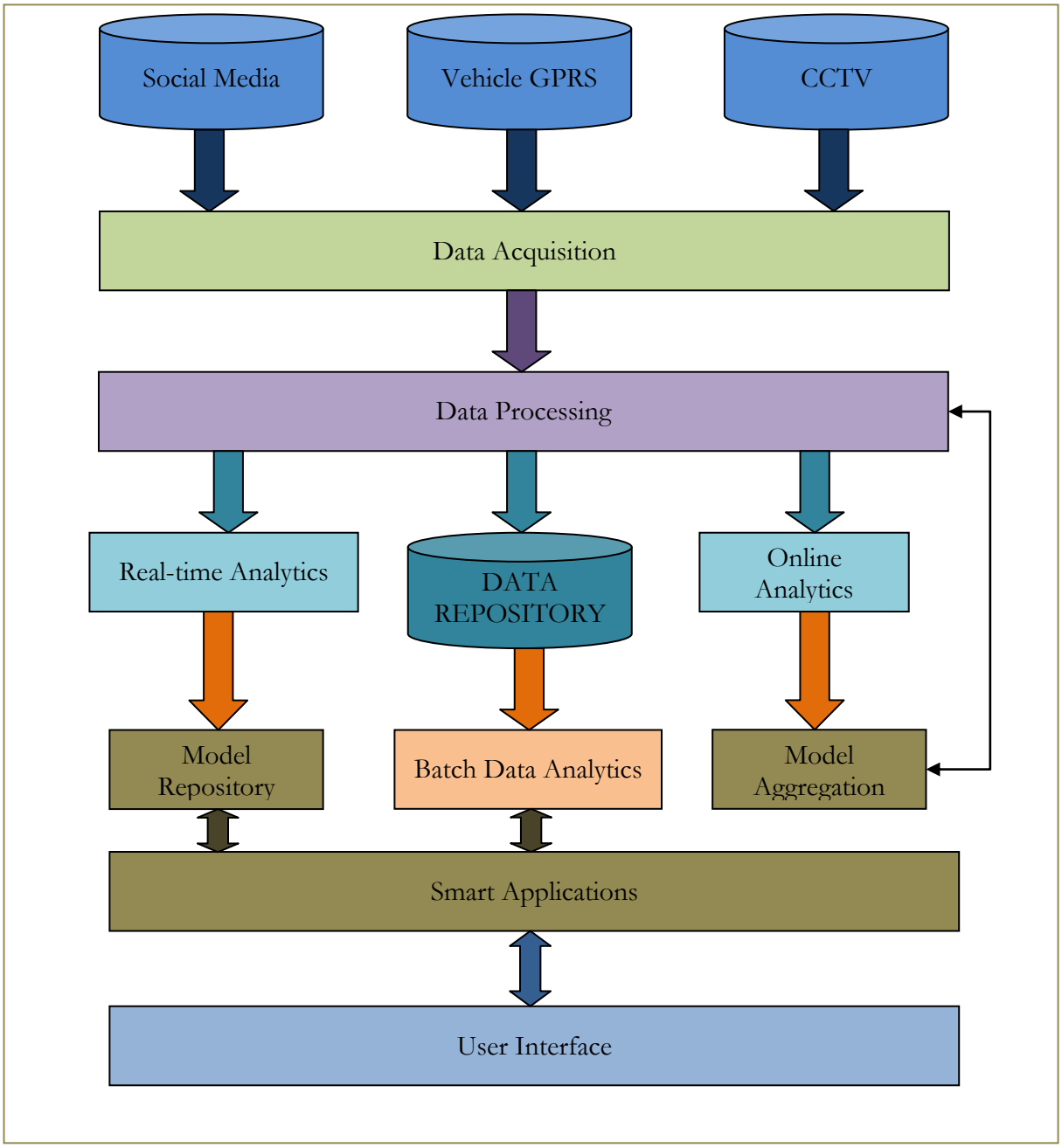


Figure 11: Conceptual Framework

The research assesses guidelines, techniques and principles applied to successfully implement intelligent transportation systems in big cities. Kenya as a developing country, can harness some of the techniques and principles through which to develop policies and guidelines in transportation sector. Lack of frameworks and elaborate architectures are a major contributor to lack of development of applications, systems and software's for Big Data Analytics. System Developers need guidelines and frameworks to enable them tap into available data for development of systems that solve traffic problems, through

provision of real-time data and on time notifications of incidents, accidents or safety information.

Assessing social economic value of Big Data Analytics is an important aspect of knowing whether the venture is feasible to both investors, policy makers on researchers in this field (Hoti, 2015). Methodologies have provided key performance indicators and provide important score weighting evaluators to better assess the impact to the society in aspect of social and economic measurements. This architecture has also identified important implementation stages that Nairobi City can utilize to achieve an intelligent transportation system through analyze of Big Data.

The future growth of the architecture has to be design in way that it brings more solutions with expansion to provide sufficient services covering divergent areas for it to be feasible (Hofman, 2015). A complex system of this nature requires a secure environment to operate in. the security and authenticity of data has to be assured to win stake holders approval.

2.11 CONCLUSION

Data is being captured in all domains of society's life, yet it is crystal clear it is not utilized. It is also clear that the pace of growth of highways does not keep up with that of vehicles on our roads. Road transportation is increasingly suffering from over congestion, leading to decreasing safety and satisfaction among road users. The future of transport lies in the disruptive nature of technology. Big Data Analytics is one of them. The mobility industry is an all-inclusive approach to movement and accessibility of people and goods.

To achieve an intelligent transportation system in Nairobi, a more robust, deployment of techniques and standards should be establish to set a precedence of future mobility solutions. This paper advocates for more research, with believes that the Big Data analytics has matured enough for deployment and penetration to solving the transport crisis in Nairobi Metropolitan, and therefore it should be pursued and invested in by the transportation players. The establishment of National Transportation Institute emerges as core objectives for enabling the best practices, formulation of guidelines, policies, standards and benchmarking to better manage transportation sector in Kenya.

CHAPTER THREE: RESEARCH METHODOLOGY

3.0 INTRODUCTION

This study sought to develop architecture suitable for analyzing big data that is designed for transportation systems. Further the research has been designed to find the necessary components that are necessary for establishing the architecture through a conceptual model that establishes ten (10) success factors of an urban transportation system: congestion, transport reliability, transport safety, air quality, innovations, regulations, versatility, customer satisfaction, accessibility and investments.

3.1 RESEARCH METHOD

The research proposes the case study with qualitative and extensive quantitative research design. The data primary data sources identified are the commuters, policy makers and agencies operating the public transport systems in a sampled size of 400 participants out of 4 million city dwellers. The online form of questionnaires and phone apps will be used primarily to collect data and analyze using the statistical package tools like the SPSS and Excel.

The study will be limited to four months of data collection and analysis. The primary source of data is Stratified Sampling while the secondary source of data will be by judgmental and convenience to get both probabilistic and non-probabilistic data. Interviews are expected to be conducted during the data collection and key subjects to be interviewed are the NTSA, KeNHA, the department of Transport in Nairobi City, The Traffic Commandants in Kenya National Police Service, Kenya Urban Roads Authority (KURA) and Operational Managers of sampled SACCOs in Nairobi.

3.2 RESEARCH DESIGN

The design of this proposal will bring out a case study with qualitative and quantitative information to shed light on inputs towards developing a suitable architecture for analyzing big data in transportation system. The cause and effect being the source of data that is homogenous and participants are from diverse regions and background. When researching on the main players in the transportation industries. A different set of design will be applied to gather qualitative information, while getting information from the commuters a quantitative approach will be deployed to get stratified data or random sampling to get the feeling and expectations across various routes of transportation.

Architecture for analyzing big data is informed by several construct of the research design; congestion, transport reliability, transport safety, air quality, innovations, regulations, versatility, customer satisfaction, accessibility and investments. The research will mainly focus on the Public Transit Sector which comprises of 46% while 39% for walking, 1% bicycling and 13% on Private car (Deloitte, 2017).

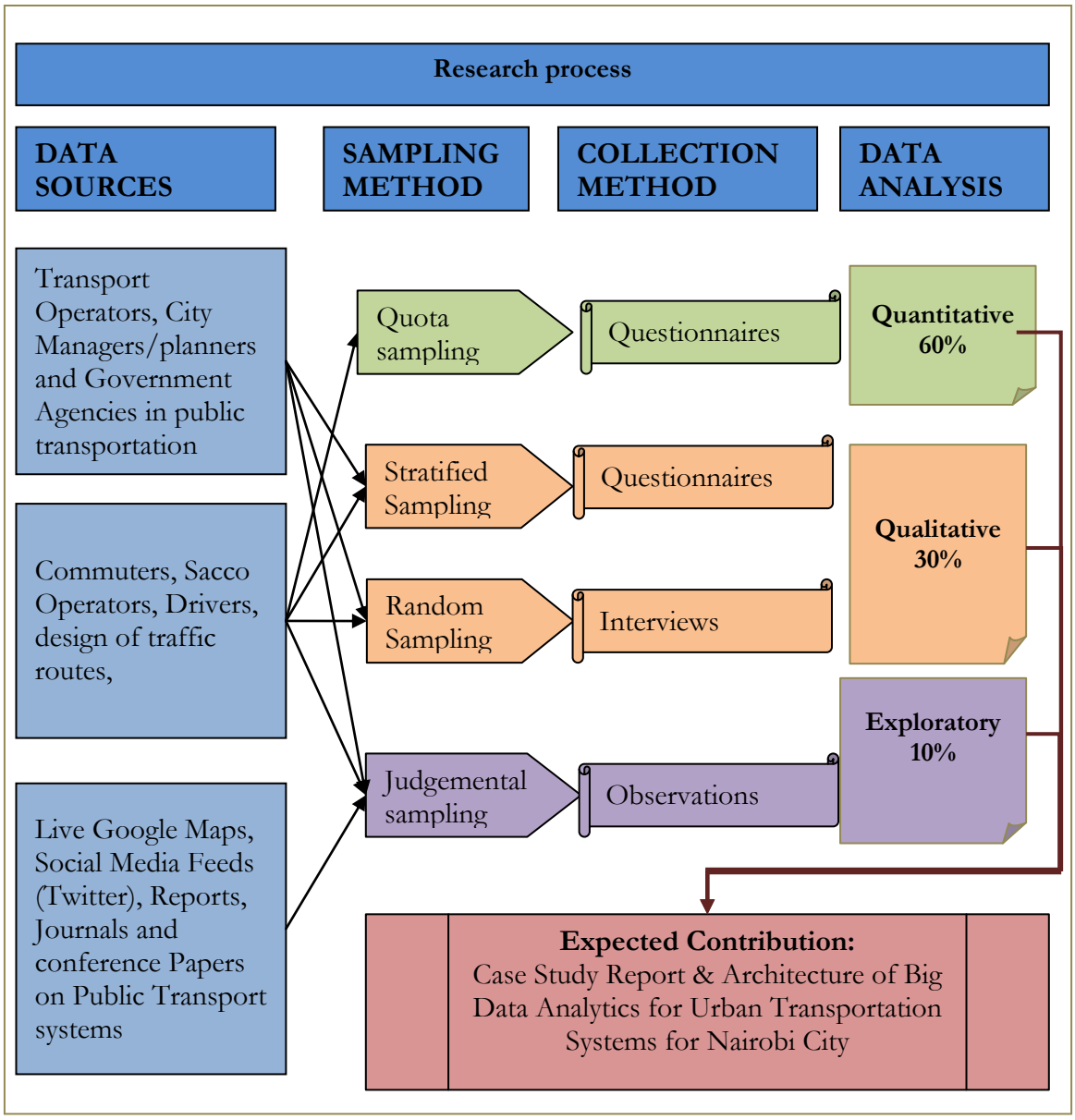


Figure 13: The Research Design

In this proposal, quantitative research method has been given a higher quota due to its objectivity, replicability, and generalizability of findings. Integral to this approach is the expectation that the case will collect reliable data. The researcher will enumerate findings that can be tested, replicated and confirmed as a single source of truth.

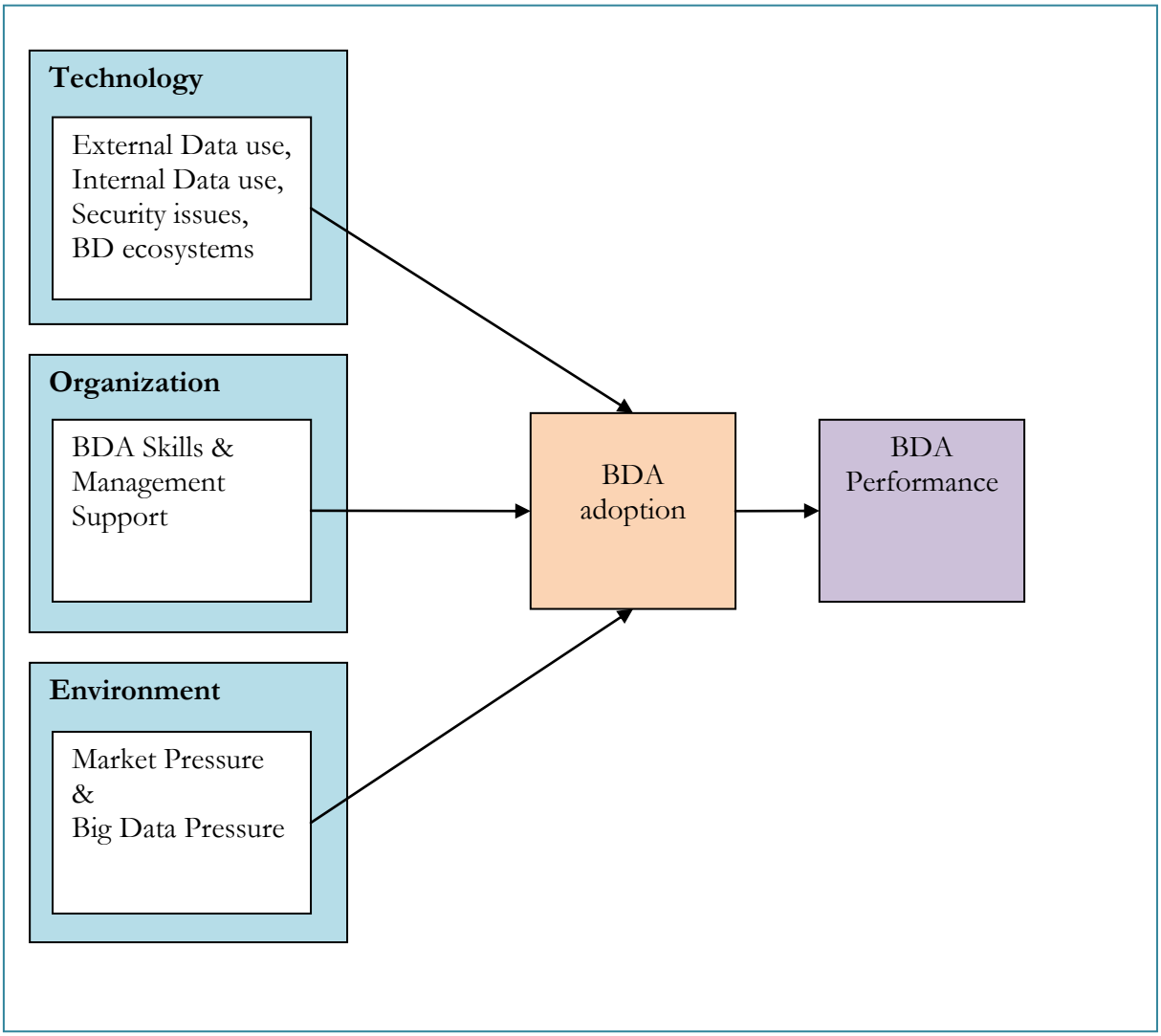


Figure 14: Research Model adopted from (TOE)

3.4 DATA COLLECTIONS METHOD

The research entails conducting interviews to related and selected transport players in Nairobi. The researcher has proposed to reach the participants comprising of personnel in National Police Service are mandated to manage traffic situation in Nairobi City including; KeNHA, KURA, NTSA, and Traffic Department of Nairobi County. Qualitative data analysis will be achieved with use of online questionnaires. The entire data collection process is elaborated in six parts;

The questionnaire that will be used for this research will have six parts;

Part 1: This will relates to general information relating to the respondents; residence location, place of interview or location the questionnaire was filled in and the time the questionnaire was administered.

Part 2: Will concern with the route of the transport, from location to the destination, the fare amount, the Sacco the transport operates and passenger capacity.

Part 3: Deals with transport players, other Transport Companies/Sacco Operators in the same route and size or number of vehicles plying the route.

Part 4: The technology of innovation that commuters, operators and drivers have used in the last few months to get updates of the traffic situation. Enlist some of the likes and dislikes of drivers on using the specified route. Collect general observations and comments on the improvement of the traffic situation for the specified route.

Part 5: Will be addressed by private car, walking and bicycling participants, indicating the route and reasons for the chosen mode of transport. The amount spent per month on transport and general question of the innovation they think can improve the traffic situations.

Part 6: This part will try to measure the reliability and safety of the traffic system in the city, measure the level of awareness in various traffic rules and regulations, measure the awareness of certain technologies used in transportation like social media, and google maps.

In this research three types of interviews will be used; focused group, unstructured and structured interview. The interviews will form part of the qualitative data that will be later analyzed and hypothesized to get deep insights of both descriptive and exploratory study.

3.5 DATA ANALYSIS METHOD

The data that was received from the observations and interviews formed part of qualitative information while the data from questionnaires was analyzed using the statistical package tool: SPSS and Excel. The three groups which form the primary source of data were analyzed to establish if there are common factors, by first cleaning the data to get rid of biases.

Since data was received from various sources a total of 50 questionnaires were analyzed to get three common elements and patterns; frequency tables, correlation and standard deviation. Data was transcribed, organized, coded and validated in a process to get qualitative insights of the interview data. Descriptive and Constructive approaches were used to get inferences, patterns and themes to support the hypothetical questions. Computer aided analysis were used to get matches and patterns in textual format.

Google key words analysis also played an important role of getting frequencies of certain words in this research. All the respondents questionnaire were uploaded through a google excel documents for analysis and important insights of search keywords were analyzed to make inferences in the participants wording during the survey

The data analysis method is intended in this research to accurately identify the user centric design for architecture for big data analytics. The needs assessment, user expectations, performance indicators and organization attributes were all captured in the questionnaires and it's upon the research to layout a road map of interpreting each segment of findings using readily available tools.

CHAPTER FOUR: RESULTS AND DISCUSSIONS

4.0 INTRODUCTION

The results of the online server were received in excel format. The total numbers of received respondents were 40 out of targeted survey of 50. The corresponding excel format were exported into tableau format file to analyze the data into visualization tables. This was necessary to determine the weight scaling of the corresponding Linkert scale of some of the questionnaires. The discussions of the results are therefore provided for group scaling rather than the individual Reponses through data aggregation.

The dependent variables were analyzed in a priority order starting with; 1.Congestion, 2.Transport reliability, 3.Transport safety, 4.Air quality, 5.Innovations, 6.Regulations, 7.Versatility, 8.Customer satisfaction, 9.Accessibility and 10.Investments.

4.1 RESPONSE RATE

The questionnaires were designed for general public and for specific State department agencies mandated with public transport and urban roads management. Out of 100 target participants of the sampled population size of 4 million city dwellers, 40% did submit the fully responsive survey questionnaire. This however was not without the challenges due to impending COVID-19, pandemic challenges of social distancing, lockdowns of offices and cessation of county of Nairobi.

Review therefore was made to reduce the sample size to **50** participants and therefore the overall response rate is **80%.** Situations that made the deliberate review of the sample size were that the questionnaire were biased to only participants with mobile phones or laptops and can fill the questionnaire without any assistance by them and submit it only. Initial design of this research sought to publicize questionnaires in different modes; print media, online and mobile based assisted data collection. So that to enhance quality, inclusivity and participation.

4.2 DEMOGRAPHIC CHARACTERISTICS

The research sought to understand urban transportation challenges faced by the city dwellers in these regions of Nairobi Metropolitan: Nairobi, Kajiado, Kikuyu, Ruai, Machakos, Thika, and Ngong. This is to enable the research understand the roads with high urban populations.

Majority of city dwellers have lived and worked in Nairobi for 5 to 10 years, with a majority indicating 11 to 20 years, which is a strong indication of attachment to their living places. This was noted about **36%** of the respondents. This is also an indication of popularity of the youth group representation in this questionnaire (KNBS, 2019). The researcher also sought to understand how they access and disseminate traffic related incidents, alerts and accidents and which media is used to circulate information.

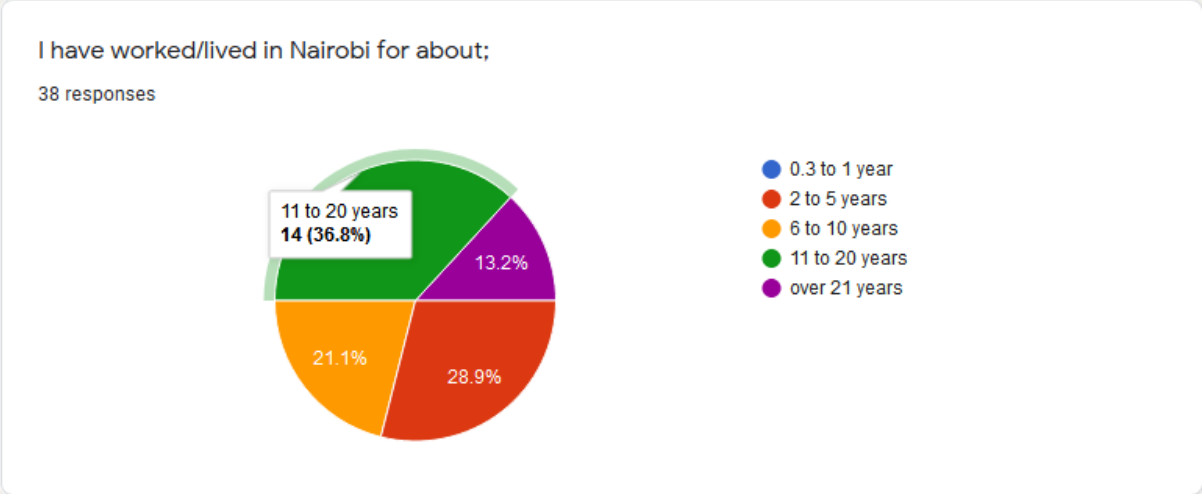


Figure 15: Demographics of age groups

The demographic representation of people sectors of economy of mainly two groups; Public Service workers = 39.5% and private company's 36.8%. The group with self-employed and freelancers had 13.2% which is an indicator of irregular working hours and irregular travelling patterns.

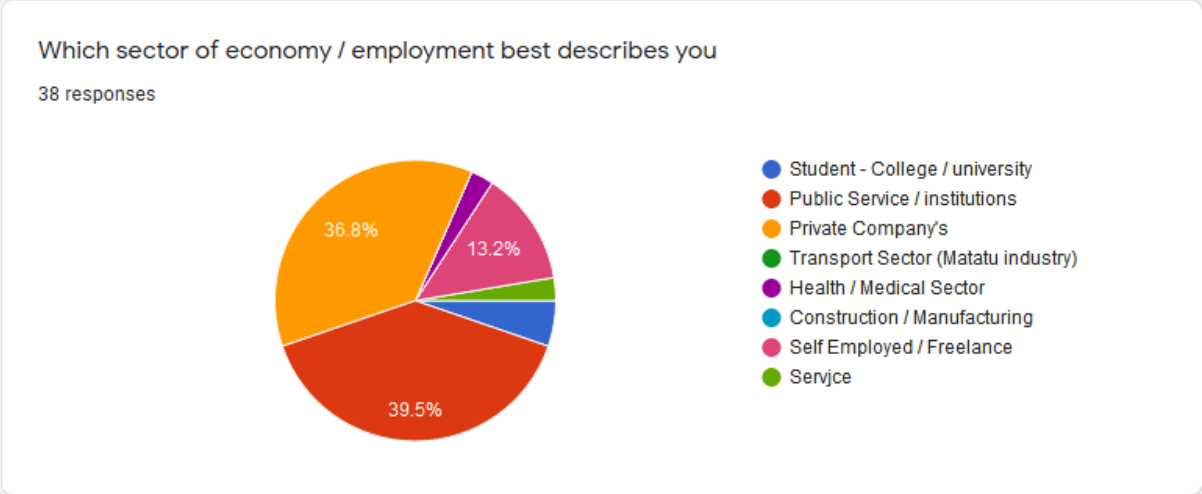


Figure 16: Representation of the participants' area of economy

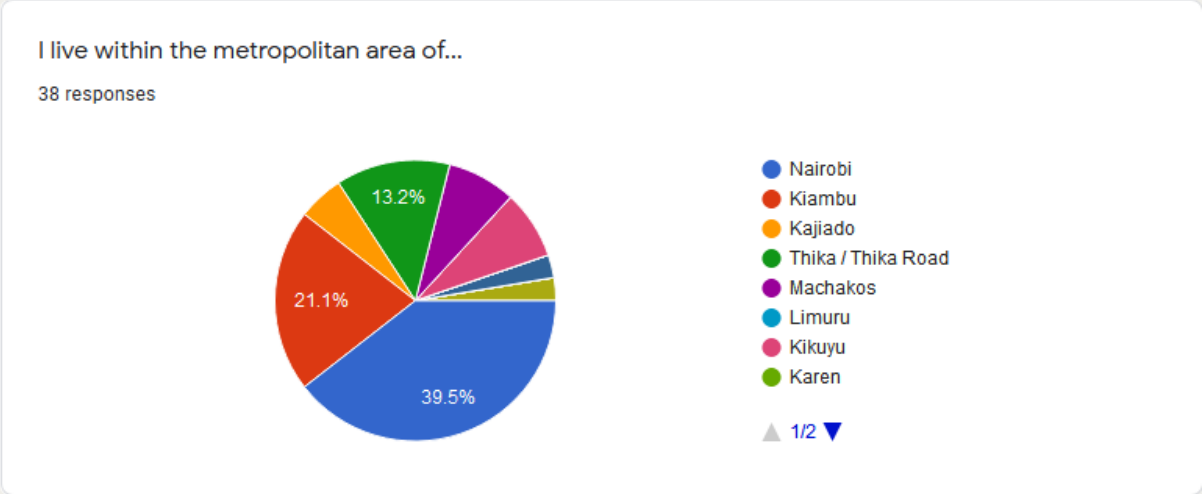


Figure 17: Representation of Nairobi Metropolitan urban areas

Results on Preferred mode of Transport in Nairobi Metropolitan

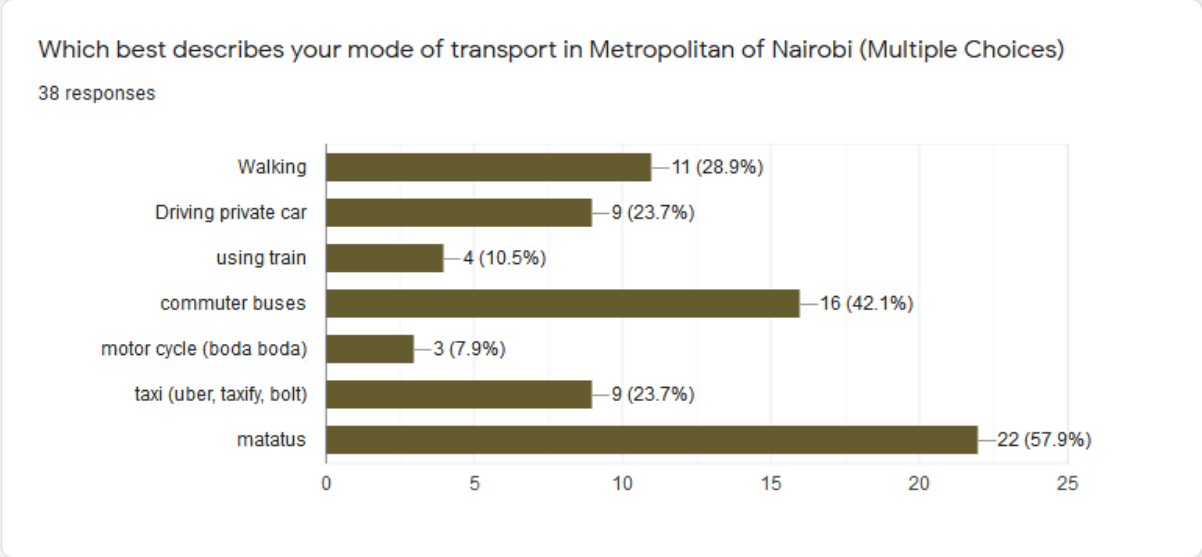


Figure 18: preferred mode of transport in Nairobi Metropolitan

Whereas the mode of transport shades a clear picture of the transportation in the metropolitan area, the "Matatus" has high preferences taking 57.9% of the overall. Whereas it is expected last mile mode of transportation for many city dwellers is by walking private cars and taxi takes a combination of 31.6%. This is a great insight in the modeling and channeling of the critical transportation information.

The inference of this study has also showed the customer satisfaction ratings for the transport agency. In rating of 1 to 10 where ten is the highest, 73.7% rated their mode of transport less than 5/10. There is clear indication of dissatisfaction of the transportation in relation to Matatu which translated highest as the mode of transport.

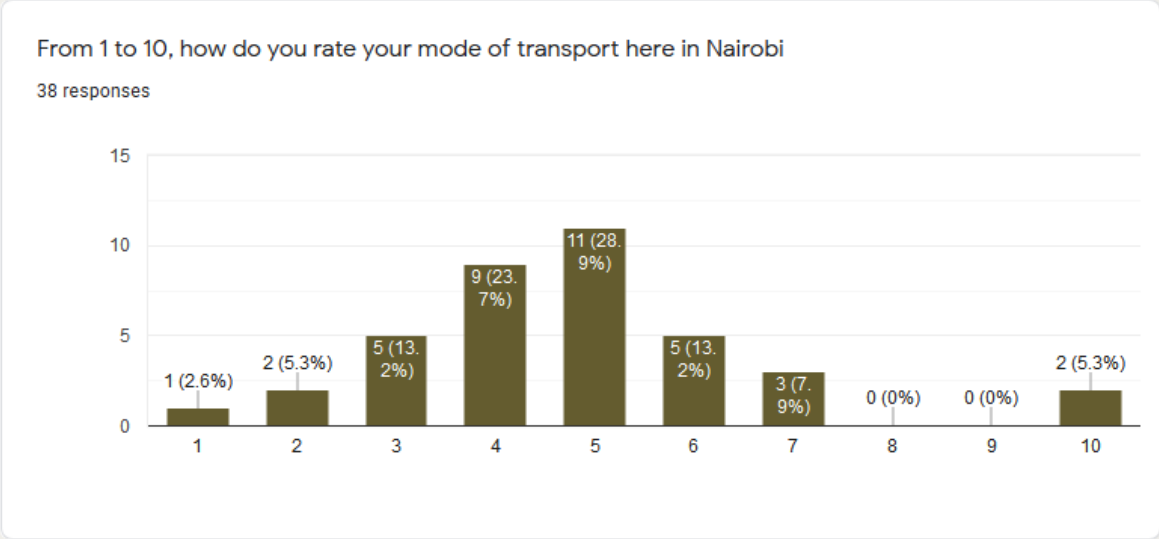


Figure 19: Customer Satisfaction Ratings based on their mode of transport.

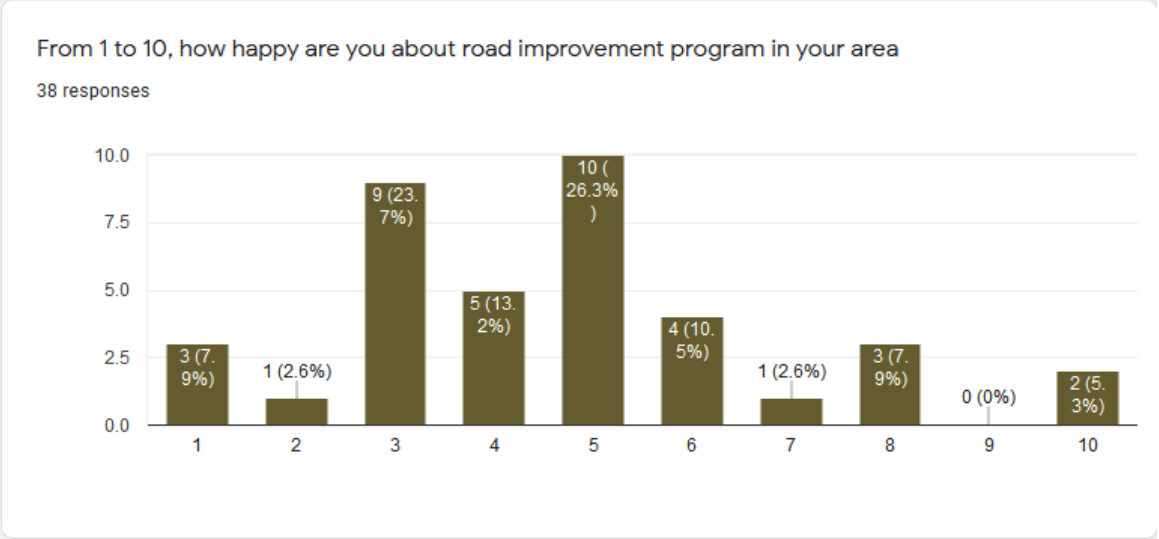


Figure 20; Rating of the road improvement program in the respective urban areas

The study has revealed that the improvement program by various transportation agencies in the urban areas under review has not been rated satisfactorily as indicated in figure 19. The satisfaction of the improvement programs to road users indicated 26.3% rating of above 6-10 scoring, a clear indication of dissatisfaction to the improvement programs.

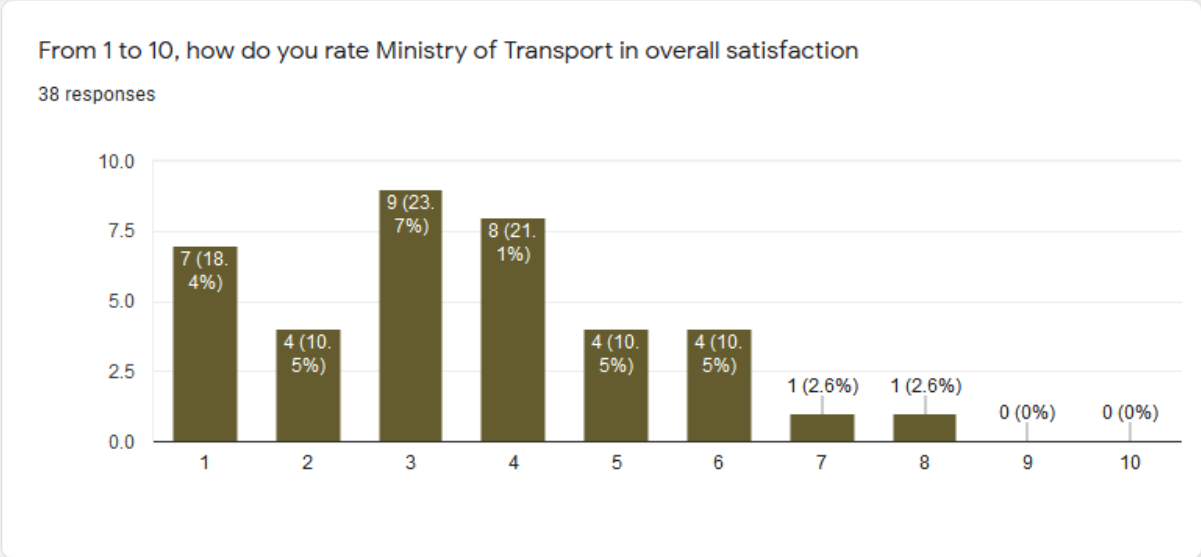
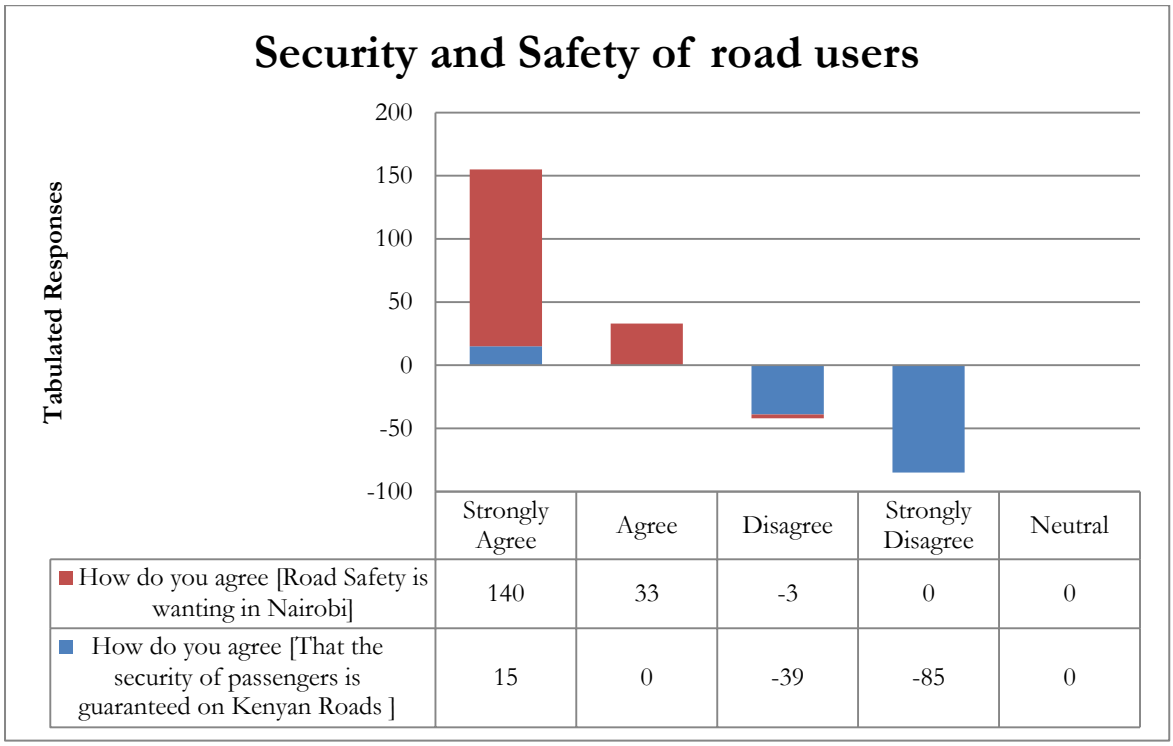


Figure 21: Satisfaction rating for the transport agency

Transport reliability, Security and Safety, Congestion of roads and air quality can be attributed to the results of this questionnaire on analysis of aggregated individual repossess.

The table below gives an aggregated safety and security concerns results;



	Strongly Agree	Agree	Disagree	Strongly Disagree	Neutral
How do you agree [Road Safety is wanting in Nairobi]	140	33	-3	0	0
How do you agree [That the security of passengers is guaranteed on Kenyan Roads]	15	0	-39	-85	0

Figure 22: Responses on Security and Safety

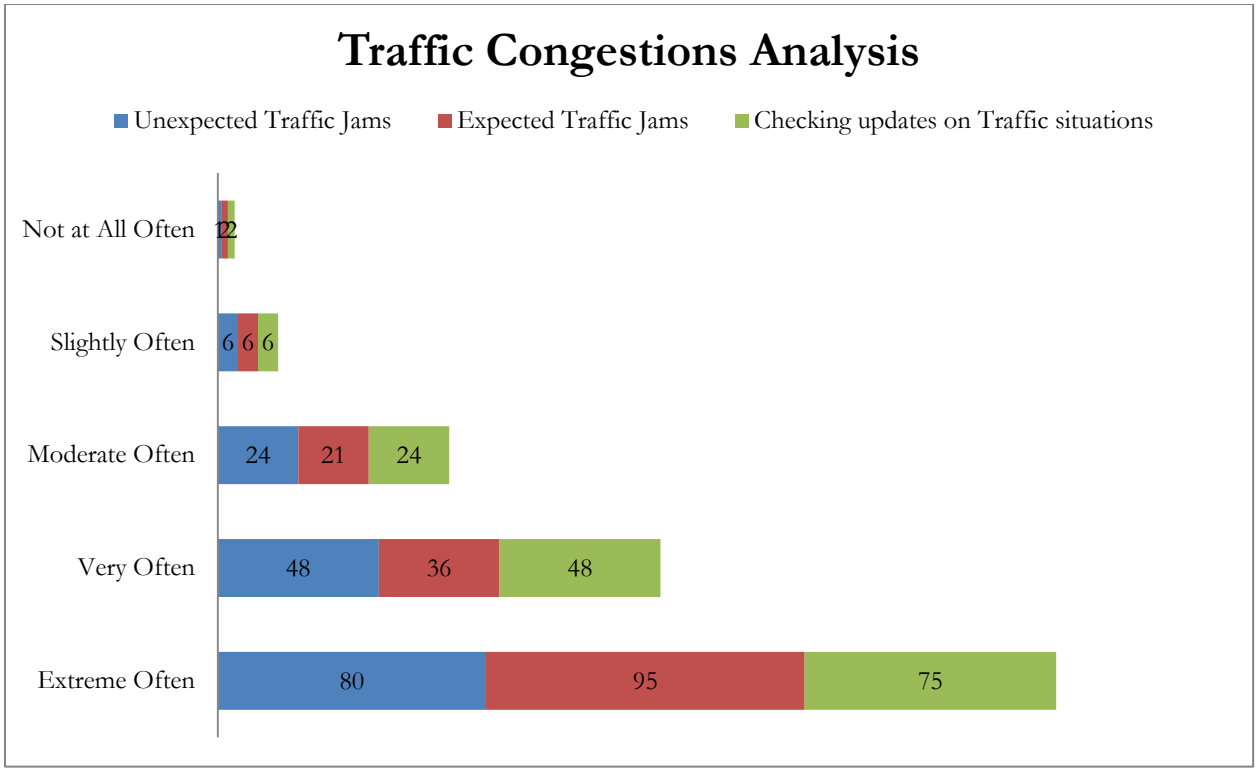


Figure 23: Traffic Congestion Analysis

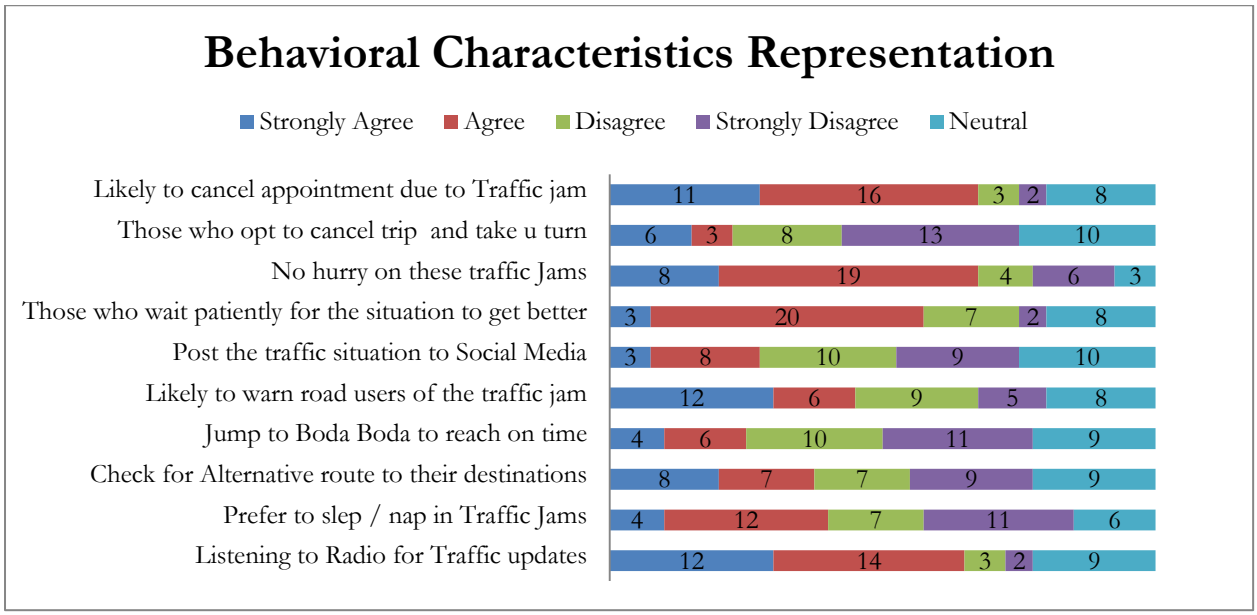
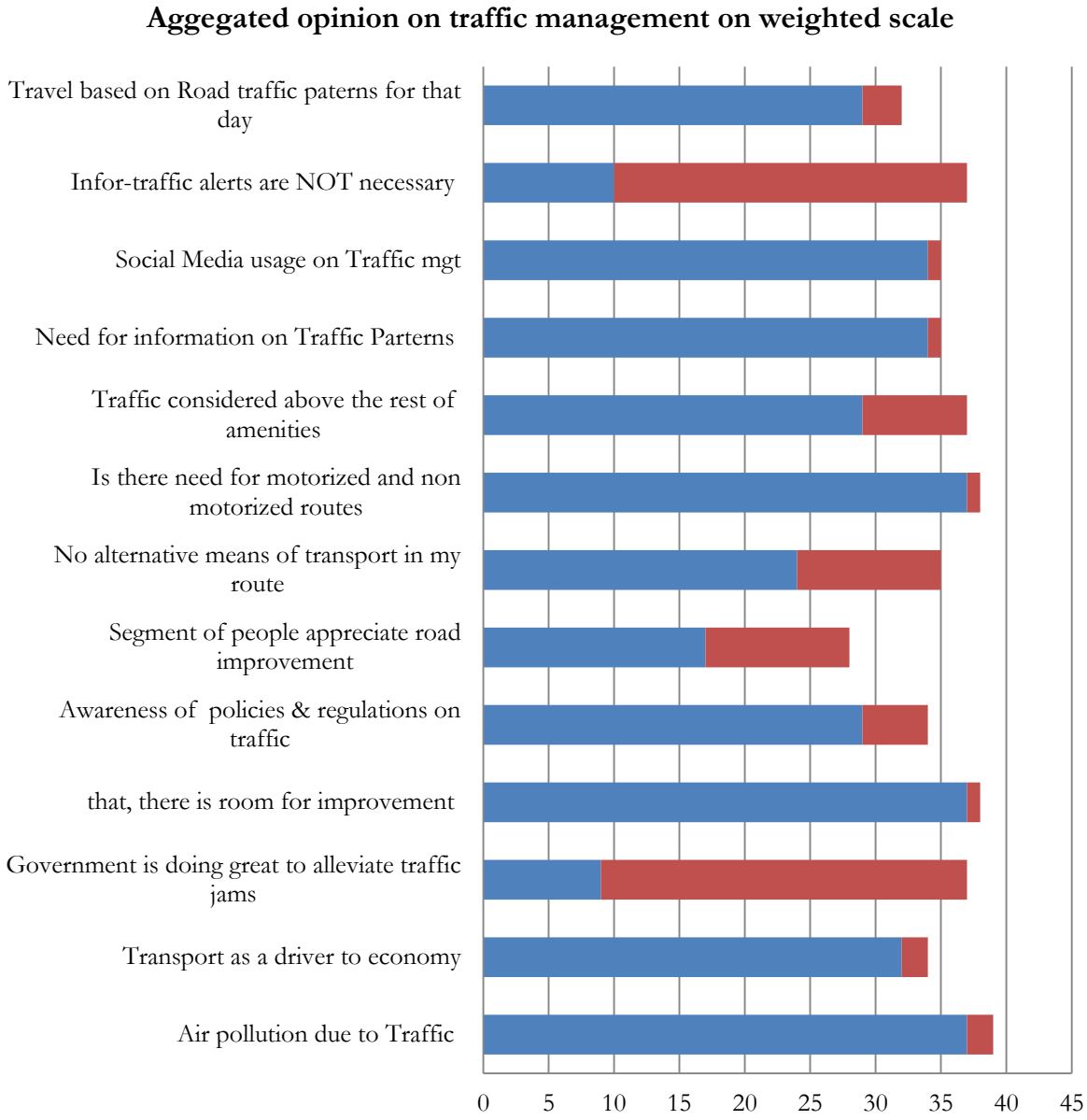


Figure 24: Behavioral representation of road users

To get the opinion on traffic management a Linkert scale was introduced to the research questionnaire with scale of 5 to 1 on how the respondents agreed to certain statements about the congestion, policies, regulations, air pollution, info-traffic updates and use of social media.

The results are shown in the figure below:



	Air pollution due to Traffic	Transport as a driver to economy	Government is doing great to alleviate traffic jams	that, there is room for improvement	Awareness of policies & regulations on traffic	Segment of people appreciate road improvement	No alternative means of transport in my route	Is there need for motorized and non motorized routes	Traffic considered above the rest of amenities	Need for information on Traffic Parterns	Social Media usage on Traffic mgt	Infor-traffic alerts are NOT necessary	Travel based on Road traffic paterns for that day
Agreed	37	32	9	37	29	17	24	37	29	34	34	10	29
Disagreed	2	2	28	1	5	11	11	1	8	1	1	27	3

Figure 25: Aggregated opinions on road users on traffic management

The aggregated opinions in the above table listed areas of concerns in traffic management and responses from the participants are as follows in summary:

a) That traveler is cautiously planning for their journey based on traffic patterns of that day. Therefore, provision of reliable source of traffic information updates will change the situation by providing a central source of reliable traffic updates for the road users.
b) Information on traffic has been listed as necessary by road users
c) That, Social media as the central focal point for accessing traffic updates as well as socializing app has been appreciated and adopted by over 90% of the respondents.
d) Need for motorized and non-motorized means of transportation is a requirement for ensuring mobility thrives in Nairobi city. This is a consideration of foot paths, bicycle lanes and in physical road planning.
e) Over 90% of the respondent's indicated that there is need to improve road infrastructure while indicating lower rating that government has done less to their expectation on road improvement programs.
f) Contribution of air pollution can be attributed to traffic congestion according to the results of the survey. This is a great concern for policy makers in matters of environment and health.

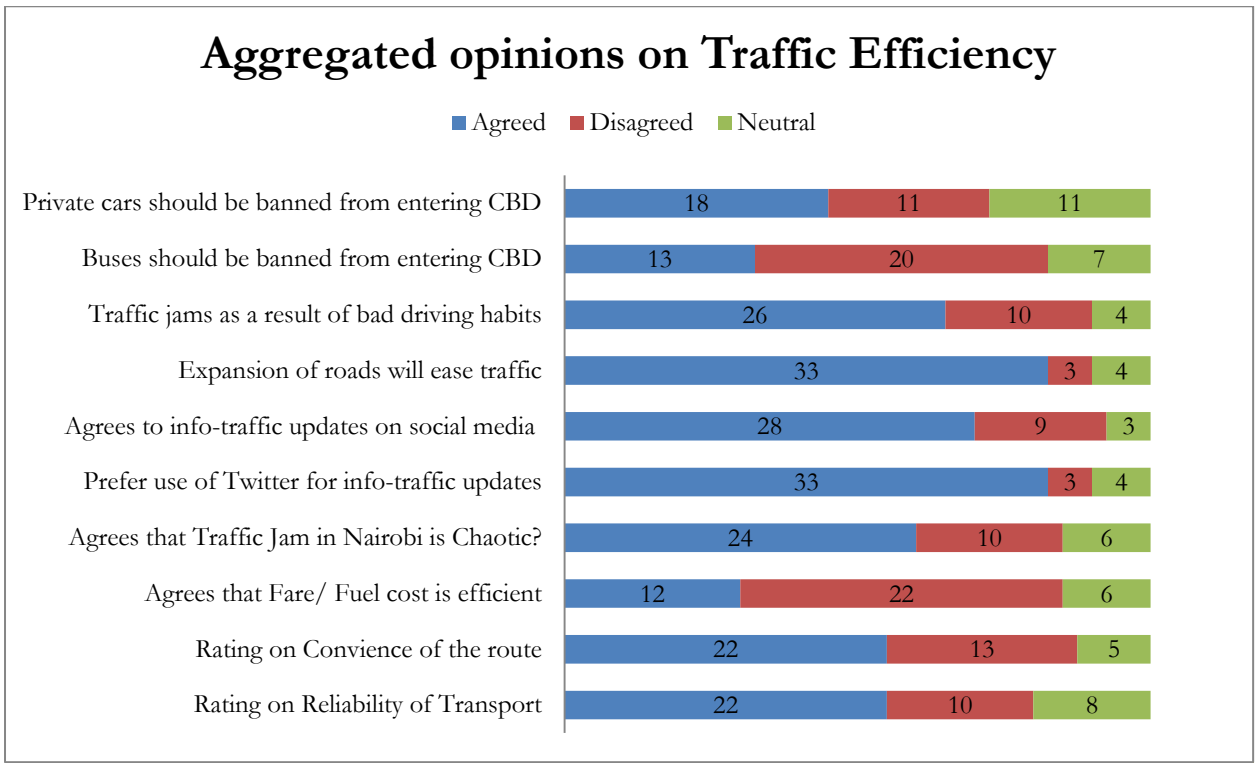
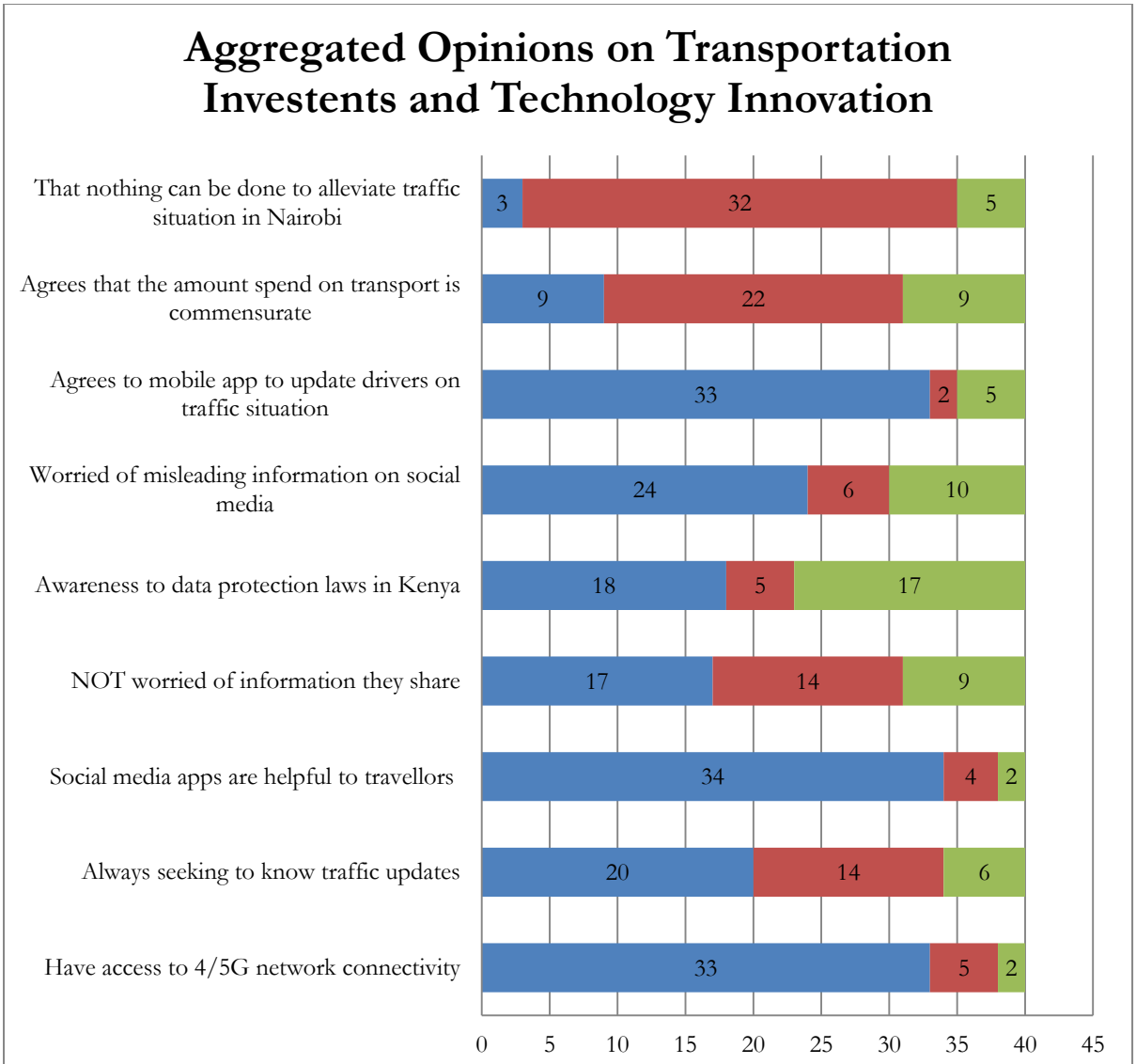


Figure 26: Aggregated opinions on Traffic Efficiency



	Have access to 4/5G network connectivity	Always seeking to know traffic updates	Social media apps are helpful to travellors	NOT worried of information they share	Awareness to data protection laws in Kenya	Worried of misleading information on social media	Agrees to mobile app to update drivers on traffic situation	Agrees that the amount spend on transport is commensurate	That nothing can be done to alleviate traffic situation in Nairobi
Agreed	33	20	34	17	18	24	33	9	3
Disagreed	5	14	4	14	5	6	2	22	32
Neutral	2	6	2	9	17	10	5	9	5

Figure 27: Analysis of aggregated opinions on transportation and technology

Table 3: Regression Analysis

Independent Variables – X Axis	Dependent Variables – Y Axis
Cost of Transportation	Transport Reliability
Scalability	Congestion
Personnel	Transport Reliability
Performance	Transport Safety
Security	Air Quality
Interoperability	Innovations
Skill	Regulations
Technology	Versatility
Data Availability	Customer Satisfaction
Infrastructure	Accessibility & Investments

To understand the history of Nairobi City Transportation system for future planning, a regression analysis is introduced to cast a probability of current and future trends. This analysis of relationship of independent variables and dependent variables through a formula:

$$\boldsymbol{y = a + bX1 + cX2 + dX3 + e}$$

y:	dependent variable
x_1,x_2,x_3:	independent variable
a:	intercept
b,c,d:	slopes
e:	residual (error)

Table 4: Regression Statistics

Regression Statistics	
Multiple R	1
R Square	1
Adjusted R Square	1
Standard Error	3.98E-17
Observations	14

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	1	2.07E-17	4.83E+16	4.1E-195	1	1	1	1
Disagreed	-1	5.34E-17	-1.9E+16	3.6E-190	-1	-1	-1	-1

The relationship of transport costs, insecurity, air pollution due to traffic was plotted against the reliability and customer satisfaction and insights indicated that they are indirectly proportional. The use of social media, access to 4/5G network and car radio traffic updates analysis as the emerging technologies associated to traffic updates relates to traffic congestion, and therefore the users prefer to get updates of the road situation as updated by other road users.

Regression analysis of how users can benefit from access to traffic updates from social media to enable them plan there route accordingly indicates is measured to establish if there are gaps in information provided on social media. A 63% of the respondents indicated having access to social media and 4G network and 72% of this group trusted the information shared on the platforms. Route planning measures are emerging with 52% and are related to the choice of road user's preferences to access devices for updates on certain routes.

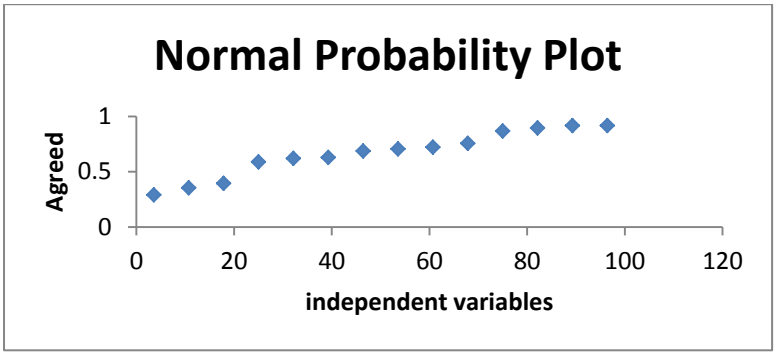


Figure 28: Regression Analysis

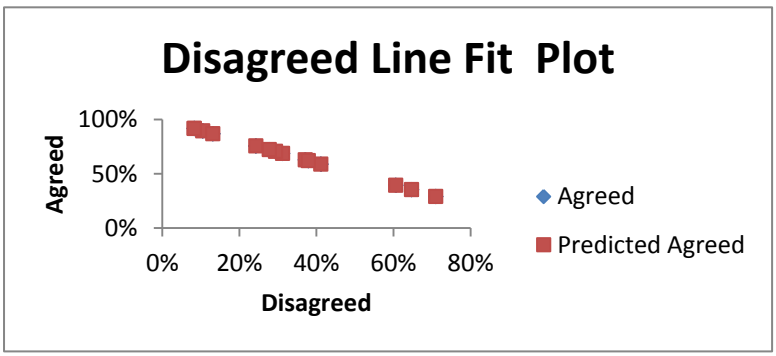


Figure 29: Predicted Analysis

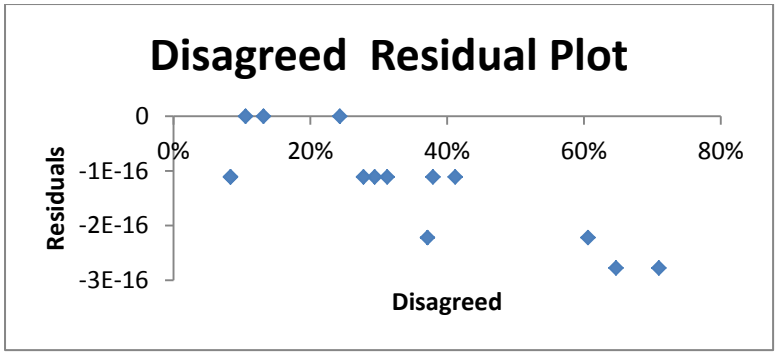


Figure 30: Residual / Error Plot

4.3 TRAVEL PATERNS IN NAIROBI METROPOLITAN

The research intended to understand aspect of timings for wakeups, starting the journey, time spent on the route and arrival times in respective work place. 65.8% arrives at their respective work place between 7:00am to 8:00am. The correlation of long hours spent on the journey can be attributed to the fact that 68.4% indicated to have been wake from 5: am to 6:00 am during the normal working days. The average time spent on the road from home to work is one (1) hour.

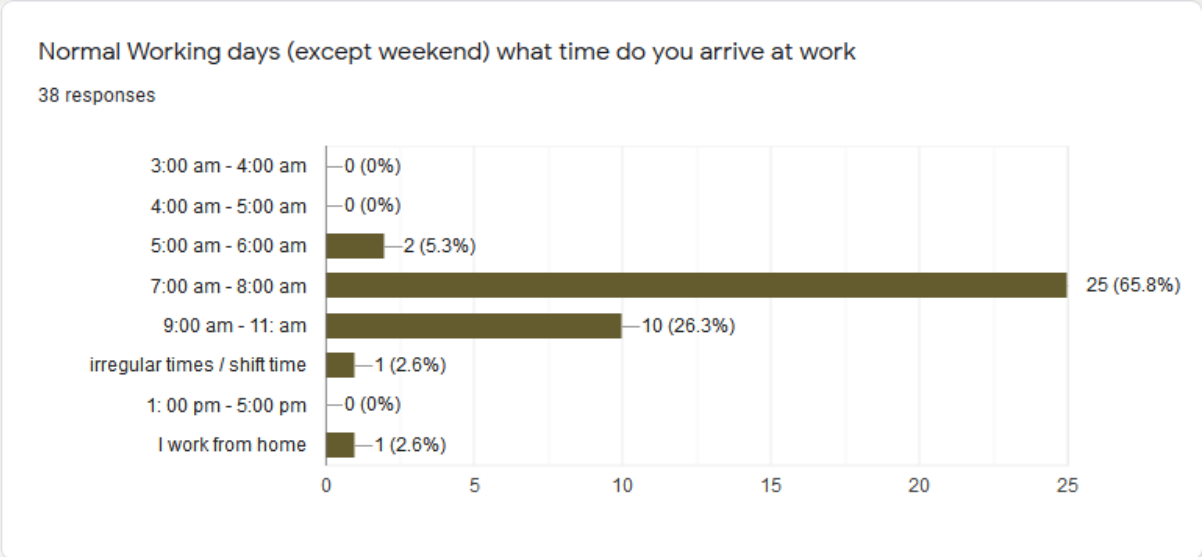


Figure 31: Estimate of timings of arrival at work

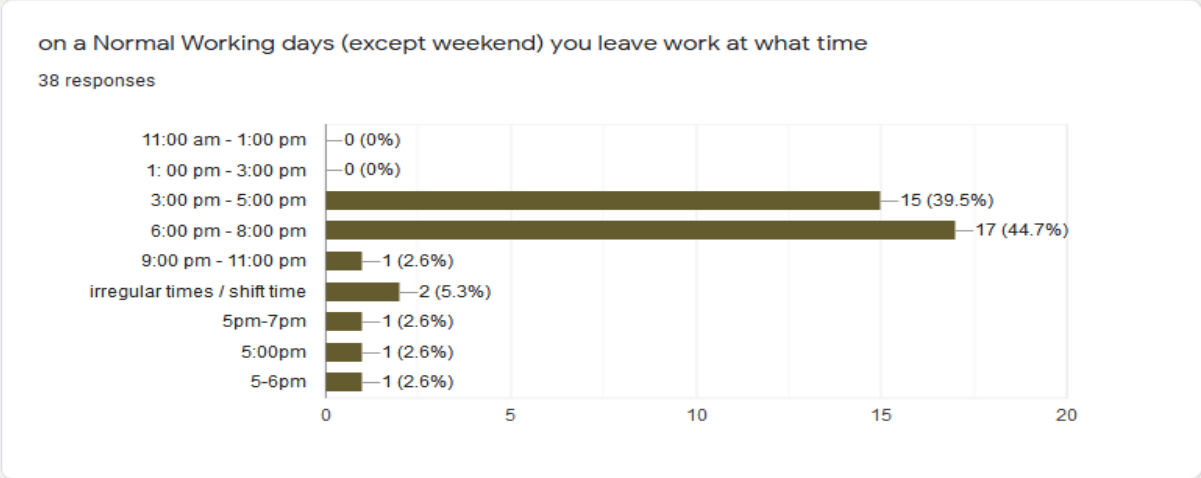


Figure 32: Estimate of timing for leaving work place

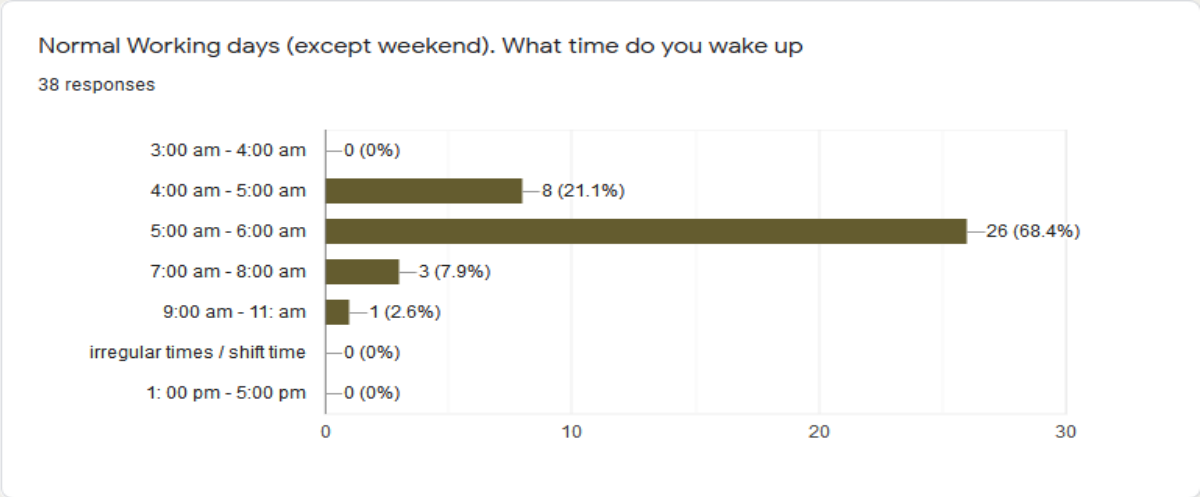


Figure 33: Responses on time for waking up

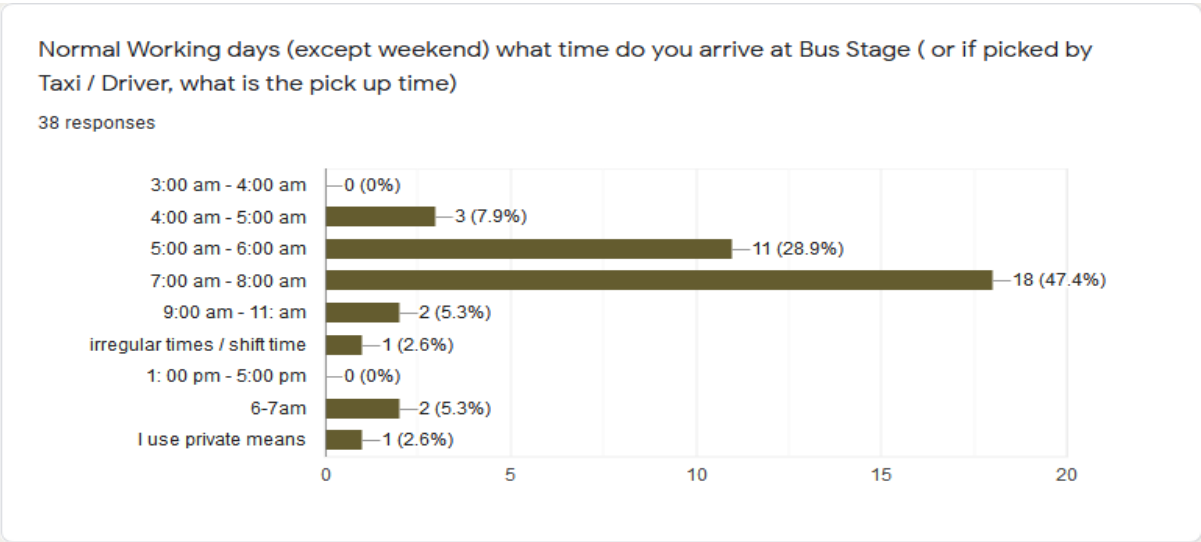


Figure 34: Estimate of times the participants start the journey

4.4 CHALLENGES FACED IN DATA COLLECTION

Online Questionnaires have limited administration provisions, where participants decide when to fill in the forms. Follow ups on participants, is therefore key to ensure successful data collection. The main challenges faced were;

1. Whereas, online questionnaire is highly portable and easily distributed through social media applications like Twitter, WhatsApp and Facebook. Some participants did not respond to all the required questions in the form citing technical challenges.

2. Not all the participant filled in the forms at the first recipients of the link to the questionnaire, therefore follow ups had to be done to remind.

3. Due to anonymity some criteria to clean data has to be put in place to avoid clutter and noise data being submitted.

4. Suspicions of participants filling the questionnaire superficially were life, despite of making two sections of the questionnaires

5. The open ended questionnaires generated a lot of information that requires extra technical expertise to run search keyword analysis. This was not anticipated in the initial research design.

6. Late submission of the online questionnaires meant not all forms will be analyzed and therefore while accounting for the overall results, may bring disparities in the percentages of the results received and the actual results on the online dashboard.

4.5 PREVALENCE OF BIG DATA ANALYTICS

The insight received from this research confirms that participants generate information related to transportation systems in various forms and shared across networks. Big Data is characterized by enormous amount of data that is generated in different forms and structures. This research found that the participants check on traffic updates via; radio, twitter, WhatsApp and Facebook platforms. While car radio plays informative roles of informing drivers about traffic situations and rated with 47.4%, Google Live map has indicated a 26.3% followed by Twitter at 12.5%. The homogeneity of the data shared across platforms plays an important role on understanding the prevalence of Big Data in this research.

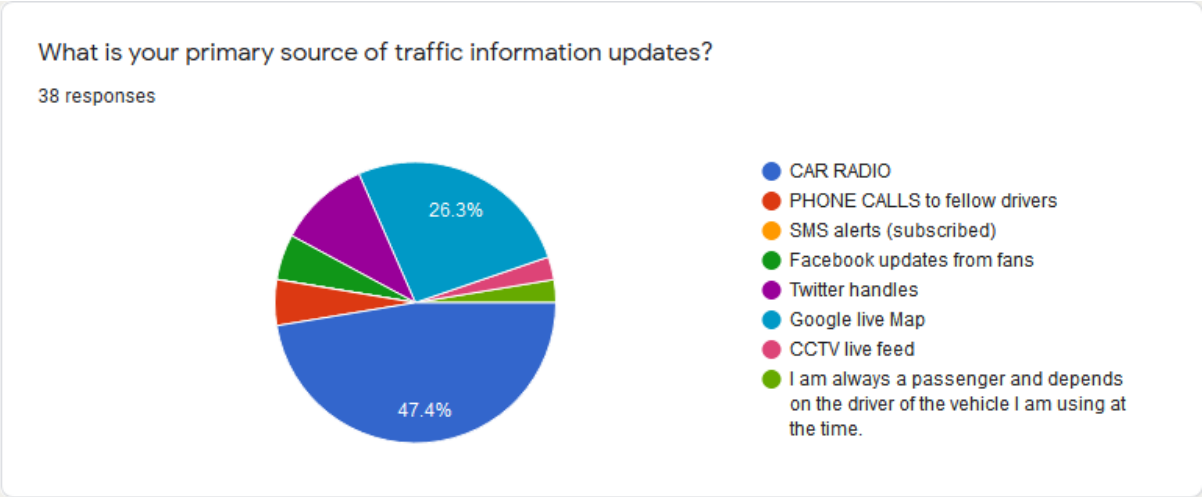


Figure 35: Traffic updates information dashboard

To accurately develop user centric architecture of Big Data Analytics for Urban Transportation systems, the concerned must understand the sources of data, types of medium used to transmit data and the storage requirements. For instance, in this research we found that audio media has high prevalence. Also real-time information is preferred with zero latency is at 47.4%, and associated with radios.

The specification for architecture does not look at the inputs only but also the data output. This is an indicator of successful, accepted architecture will require to perform in certain ways, whereby the users dictates how to receive updates, alerts, incidents report related to traffic. The intensity of how technology has assisted in planning itinerary of most travelers in the city, Google live map is also a technology of choice with 26.3%, an indication of adoption rate of technology in the phase of mature stage.

An efficient architecture of Big Data Analytics for Urban Transportation requires an understanding of existing technologies that create data, receive data types and synthesis data. The research found these types of data association; text, audio, picture, video, log files, video, and live audio & visual streaming. These types constitute structured and unstructured data types.

4.6 COMPONENTS MAKING BIG DATA

From this research finding we found structured data types and unstructured comprising of Photos, text, voice, video, reports, web pages and log files. These components are the main source of data for the architecture of the big data to manage transportation systems. The mapping of components and sub-components extracted from the research findings are as per the table below.

Table 5: Mapping of components making big data

Data Flow Stage	Components	Existing Technologies
Data Source	Images, Video, Audio, Social Media, Web pages, Documents, Email, log files and reports	Mobile, Cameras, Laptops, CCTV, Radio, Webserver, Traffic Sensor
Data Collection	Stream Processing	Apache Spark, Extra hop
	Batch Processing	Hadoop Framework, Apache Sqoop
Data Extraction	Data Store	Data Lake
	Data Discovery	Spark, Hadoop, Storm, RapidMiner, Mahout, Orange, Weka, Datamelt, KEEL, SPMF, Rattle
	Data ingestion	ETL, Apache Kafka, Hive, Apache Spark, Apache Pig, Apache NiFi Middleware, .NET, J2EE, COBRA, SOAP, WSDL
Analytics	Data Analysis	SAP Hana, Birst, SAP Lumira Cloud, Tibco, Hazecast, Bime
Service	Application	Smart Mobile Apps
	Models	Histograms Conceptual model
	Visualization	Visual maps Route mapping
	Live Streaming	Web pages Live video stream url

4.7 DISCUSSIONS

The results show that components of making a Big Data Architecture are readily available. These components making the architecture of big data analytics are effort free in terms of gathering and sourcing of information. The data comes in all forms; pictures, text, log files, videos, audio and SQL files. The architecture development will also establish data source communication paths, data storage, processing and real-time data streaming of the Big Data. Some of the reviewed technologies in Big Data Analytics that are open source can be the good starting point for any developer to customize and establish this kind of architecture.

CHAPTER FIVE: CONCLUSION AND RECOMMENDATIONS

5.0 INTRODUCTION

Bid Data Analytics Architectures are going to shape the way data is received, processed, disseminated and visualized. The complexity of Big Data Analytics requirement is that specialized systems are developed to deal with enormous data exchange and data storage requirements (Charles, 2018). This research has proposed a Big Data Analytics Architecture for urban transportation systems and found out the necessary components required for data acquisition, technologies required for the data processing and an analysis of available open source architectures that can be transformed to fit into urban transportation systems.

5.1 ACHIEVEMENT OF THE STUDY

The achievements' made on this study can be derived from the six objectives that was set and has helped in shaping the mission of this research. The main highlight of the achievement is the realization of existence of data that is being generated veraciously at large scale by city dwellers on social media. In full realization of the availability of open source architectures in big data analytics have open more windows of opportunities to researchers' and systems developers to create more specific solutions for urban transportation systems.

The achievements as follows for each objective:

1. **Identify important components, attributes, specifications and guidelines for successful Architectures for Big Data Analytics for Urban Transportation System.**

   These components include; data sources, data storage, data processing and data livestreaming. The attributes of each data source can be structured or unstructured data types: i.e. file system, pictures, text and SQL. The specification of architectures was realized through research questionnaires whereby user centric developments of specifications were given great preference. Both the literature review in chapter three and research questionnaires in chapter our brought about the information of developing a successful architecture for big data analytics.

2. Develop architecture for big data analytics for transportation systems.

The proposed conceptual framework was tested with the research results and a final architecture presented as follows:

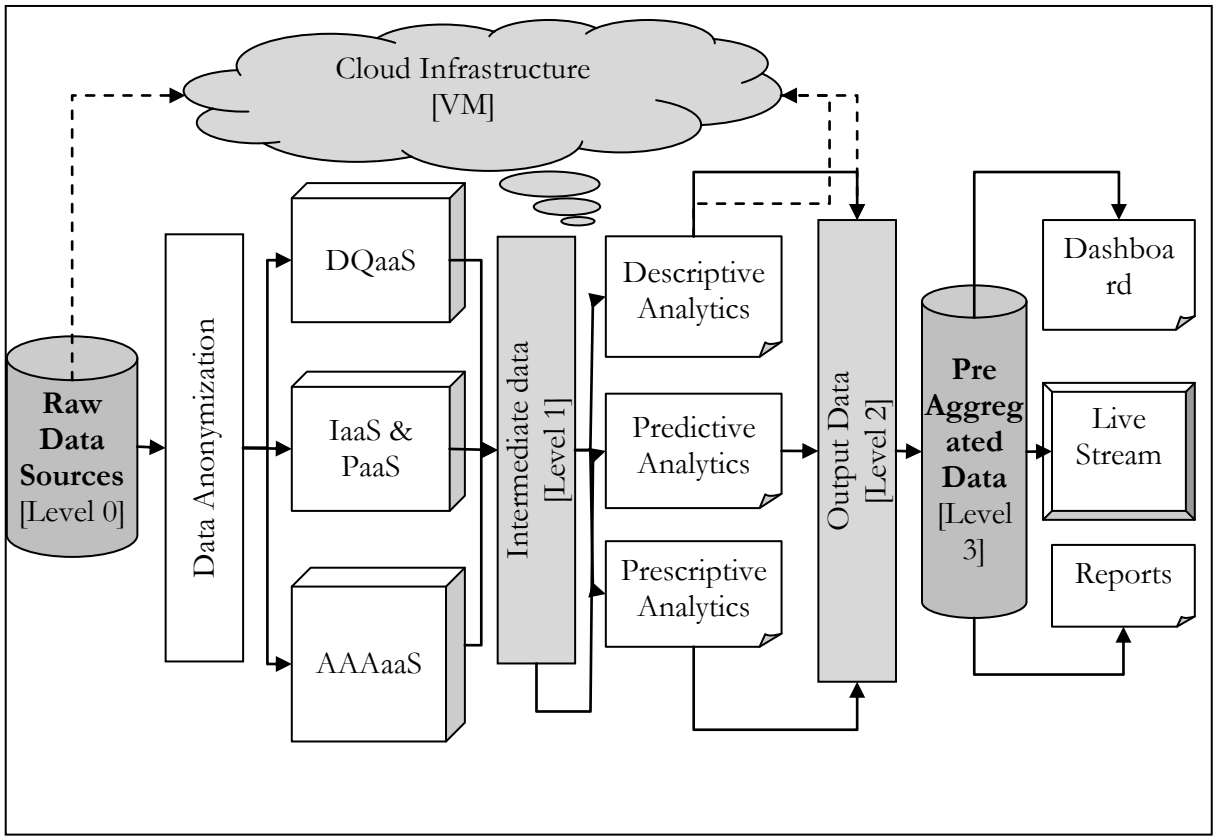


Figure 36: A Big Data Analytics Architecture

This architecture highlights three levels of information interchanges and utilizes three types of cloud infrastructure namely: Data Quality as a Service (DQaaS), Infrastructure and Platform as a Service (IaaS and PaaS) and Authentication, Authorization and Accounting as a Service (AAAaaS). This architecture encapsulates data to give predictive, descriptive or prescriptive analytics. Final aggregated data can be live streamed, dashboard alerts or given as periodical reports.

3. **Identify the methodological issues associated with implementing Big Data Analytics for Urban Transportation Systems**

   Some of the methodological issues identified in this research are based on the technology, innovation, acceptance and strategy. Urban transportation in Nairobi Metropolitan is characterized of players in diverse field comprising of six agents in government and two associations in public. The coordination of all the government agencies and private owned transport enterprises stands out as main challenge. However, the issues associated with implementation can be dealt with through establishment of a National Transport Institute and utilization of McKinsey's 7S model (Strategy, Style, Structure, Staff, Skills, Shared values & Systems).

4. **Identify the ecosystem for Nairobi Metropolitan Urban Transportation Systems and its challenges.**

   The ecosystem for Nairobi Metropolitan comprises of private institutions running the mass public transportation and regulated by government agencies comprising of; NTSA, City County, KeNHA, NaMATA and KURA. Public participation in in matters of development and improvement of transport sectors are tenants of the Kenyan Constitutions. However, this research has identified weakness in the institutions and confirmed with road user satisfaction rating which is below average. The last mile of ecosystems were also measured and rated in terms the efficiency, conditions, lighting and security. The dependent variables used to measure the indicators were: Congestion, Transport Reliability, Transport safety, Air quality, Innovations, Regulations, Versatility, Customer satisfaction, Accessibility and investments.

5. **Develop adoption strategy for development and implementation of architecture for Big Data Analytics for urban transportation System.**

   This research has found that the components necessary to developing architecture are prevalent in the public despite that there is little correlation of the technology behind to majority of users. The mobile penetration in the masses and adoption of social media usage is also high in Nairobi (Deloitte, 2019). Technology-Organization-Environment (TOE) framework of Tornatzky and Fleischer has given insight in understanding the ecosystems of adoption in big data analytics whereas giving outlook as to why despite of the technology some sectors remain behind in adoption

of the technology. The three forces (People, Innovation and the Environment) are an important indicator of measurement and consideration when developing the strategy for adoption.

Further, the evaluation of organization in TOE framework has brought about introduction of a finer tool developed by McKinsey to align organization to reach their objectives as popularly known as McKinsey's 7S model. This model can be introduced to access and align the organization implementing the big data analytics.

5.2 LIMITATIONS

The research was limited to the scope of road, highways and foot path as the commonly means used in Nairobi. Therefore, water, railways and air were not reviewed as the alternative means. Therefore, the research focused on road users, the challenges and the prevalence usage of social media and mobile phones to communicate on incidents, status and situation of the roads. The interviews section of this research was only dedicated to the government transportation agencies, and therefore to strike a balance more independent bodies like UN Habitat, Manufacturers, and Road Contractors.

Further, this kind of research required more travelling to the various routes, yet there was limited time and budget to gather information. Therefore the researcher resorted to sampling method as a representation of the targeted response group.

5.3 RECOMMENDATIONS

This research despite of the limitations focused on Nairobi Metropolitan in validation of the construct of Technology-Organization-Environment (TOE) in realization of the extent of use of social media, mobile and radio as sources of data by city dwellers. The results achieved were focused on developing big data analytics architecture for urban transportation systems. The open source architectures analyzed in the literature review can be adopted as the foundation of building a transportation system, like Apache Spark.

In cognizant of people related issues in implementation, this research further recommended establishment of National Transport Institute with mandate to coordinate, formulate policies, guidelines and conduct dedicated research of various technology that can be adopted in the transportation sector. Further, more research in the field of big data and intelligent transportation systems should be conducted.

5.4 CONCLUSION

Big Data Analytics is hot topic in diverse field of study around the globe. It is in constant advancement and adoption in large scale has only been noted in large corporation to ease business and realize new approaches when dealing with data. Netflix, Uber, Amazon and IBM have invested billion dollars to advance and bring competitive edge in the business. Transportation industry is among the drivers of the economy in most developed and developing countries, yet little advancement in technology has been done in Nairobi to alleviate woes in the sector. This research recommends development of architecture for big data analytics for urban transportation to better get data on movement and usage of infrastructure in Nairobi metropolitan.

In summary, the volume of data in the world is increasing exponentially. By some estimates, 90% per cent of the data in the world has been created in the last two years, and it is projected to increase by 40% per cent annually. A large share of this output is "data exhaust," or passively collected data deriving from everyday interactions with digital products or services, including mobile phones, GPS devices, and social media. Data is growing because it is increasingly being gathered by inexpensive and numerous information-sensing and mobile devices. Better usage of this data can only be achieved through development of architectures to consume the data produced and produce real-time analytics, periodical reports and visualization of our world from the transportation sector perspective.

BIBLIOGRAPHY

Abduljabbar, R. D. (2018). Applications of Artificial Intelligence in Transport: An Overview. In D. o. Engineering, & S. U. Technology (Ed.). MDPI. doi:http://dx.doi.org/10.3390/su11010189

Acharjya, P. a. (2016). A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools. *International Journal of Advanced Computer Science and Applications, VII*(2).

Acharjya, P. a. (2016). A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools. *International Journal of Advanced Computer Science and Applications (ijacsa)*, 7 - 14.

Ajah, I. a. (2019, June). Big Data and Business Analytics: Trends, Platforms,Success Factors and Applications. *Big Data Cognitive Computing*, 32. doi:http://dx.doi.org/10.3390/bdcc3020032

Baig, M. S. (2019, August 24). Big data adoption: State of the art and research challenges. *56*(6). doi:https://doi.org/10.1016/j.ipm.2019.102095

Barrachina et al., J. G. (2015). A V2I-based real-time traffic density estimation system in urban scenarios. *Wireless Personal Communications*, 259-280.

Bekiaris, L. N. (2017). Highway traffic state estimation per lane in the presence of connected vehicles. Transportation research part B: methodological. 106.

Charles, P. B. (2018, February). BIG DATA – CONCEPTS, ANALYTICS, ARCHITECTURES – OVERVIEW. *International Research Journal of Engineering and Technology (IRJET) , V*(2). Retrieved May 21, 2020, from https://www.irjet.net/archives/V5/i2/IRJET-V5I231.pdf

Columbus, L. (2016). Roundup Of Analytics, Big Data & BI Forecasts And Market Estimates. *Forbes Magazine Online*. Retrieved January 9, 2020, from https://www.forbes.com/sites/louiscolumbus/2016/08/20/roundup-of-analytics-big-data-bi-forecasts-and-market-estimates-2016/#5e78fad6f218

Communications Authority. (2019). *Sector Statistics, Third Quarter Sector Statistics report for the Financial Year 2018 - 2019*. Nairobi: CAK. Retrieved September 29/09/2019, 2019, from https://ca.go.ke/wp-content/uploads/2018/12/Sector-Statistics-Report-Q1-2018-2019.pdf

Daiheng, N. (2016). Traffic Sensing Technologies. *Science Direct*. doi:https://doi.org/10.1016/B978-0-12-804134-5.00001-5

Deloitte. (2017). *Deloitte City Mobility Index*. Nairobi: Deloitte.

Deloitte. (2019). *Global Mobile Consumer Survey*. Nairobi: Deloitte Kenya. Retrieved December 20/12/2019, 2019, from https://www2.deloitte.com/content/dam/Deloitte/ke/Documents/technology-media-telecommunications/Deloitte_GMCS_Report_The_Kenyan_Cut_August_2019.pdf

Dove, E. (2018). The EU General Data Protection Regulation: Implications for International Scientific Research in the Digital Era. *The Journal of Law, Medicine & Ethics*. doi:10.1177/1073110518822003

EU. (2016). General Data Protection Regulation. Retrieved 10 4, 2019, from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN#d1e1404-1-1

Forbes. (2018). *Big Data Analytics Adoption Soared In The Enterprise In 2018*. Forbes. Retrieved January 10, 2020, from https://www.forbes.com/sites/louiscolumbus/2018/12/23/big-data-analytics-adoption-soared-in-the-enterprise-in-2018/#38730e14332f

Gandomi, A. a. (2015). Beyond the Hype: Big Data Concepts, Methods and Analytics. *International Journal of Information Management, 35*(2), 137-144. doi:https://doi.org/10.1016/j.ijinfomgt.2014.10.007

Gokdeniz, I. (2017). Strategic Assessment based on 7S McKinsey Model for a Business by Using Analytic Network Process (ANP). *International Journal of Academic Research in Business and Social Sciences, 7*. doi:10.6007/IJARBSS/v7-i6/2967

GTFS. (2019, 08 05). *General Transit Feed Specification*. Retrieved August Monday, August 05, 2019, 2019, from https://gtfs.org/

Guilherme, G. P. (2017). An Architecture for Big Data Processing on Intelligent Transportation Systems. *An application scenario on highway traffic flows*, 4-9.

Hermann, J. a. (2017, September 5). *Meet Michelangelo: Uber's Machine Learning Platform*. Retrieved from Uber Engineering: https://eng.uber.com/michelangelo/

Hofman, W. (2015). Data collection architecture for Big Data - a framework for a research agenda. *Proceedings of the 6th Workshop on Enterprise Interoperability, Nîmes, France.* Soesterberg. Retrieved April 14, 2020, from http://ceur-ws.org/Vol-1414/paper2.pdf

Hoti, E. (2015). The technological, organizational and environmental framework of IS innovation adaptation in SME. Evidence from Research over the last 10 years. *International Journal of Business and Management, III*(4), 14. doi:10.20472/BM.2015.3.4.001

Huawei (Producer). (2017). *Huawei Intelligent Transportation Solution* [Motion Picture]. Retrieved January 9, 2020, from https://www.youtube.com/watch?v=Gbp-AsQwynI

IBM. (2015). *Better Roads are paved with Big Data Analytics.* IBM Research Editorial Staff Report. Retrieved 08 24, 2019, from https://www.ibm.com/blogs/research/2015/07/better-roads-are-paved-with-big-data-analytics/

IEA. (2017). *Kenya Urban Areas Performance Index Report 2017.* Nairobi: Institute of Economic Affairs 2017. Retrieved from https://www.ieakenya.or.ke/downloads.php?page=1536232616.pdf

IFC, a member of the World Bank Group. (2019, November). How Artificial Intelligence is Making Transport Safer, Cleaner, More Reliable and Efficient in Emerging Markets. (M. L. Twinn, Ed.) *Fresh Ideas About Business in Emerging Markets*(75). Retrieved January 7, 2020, from https://www.ifc.org/wps/wcm/connect/7c21eaf5-7d18-43b7-bce1-864e3e42de2b/EMCompass-Note-75-AI-making-transport-safer-in-Emerging-Markets.pdf?MOD=AJPERES&CVID=mV7VCeN

ILO. (2017). *Operations/Human Resources Manual for Matatu Saccos/Companies.* International Labour Organizations.

International Transport Forum. (2019). *New Directions for Data-Driven Transport Safety.* ITF Corporate Partnership Board.

Kauser, A. a. (2016). A Survey on Big Data Analytics: Challenges, Open Research Issues and Tools. *(IJACSA) International Journal of Advanced Computer Science and Applications, 7*(II). Retrieved January 6, 2020, from https://thesai.org/Downloads/Volume7No2/Paper_67-A_Survey_on_Big_Data_Analytics_Challenges.pdf

Khan, S. C. (2018). Non-Real-Time Transportation Applications: Potential Use of Connected Vehicle Data and Data Infrastructure Requirements. *Journal of Infrastructure Systems*(I), 25.

Khosrowshahi, D. (Producer), & Dara Khosrowshahi, C. -U. (Director). (2019). *An Operating System for Everyday Life* [Motion Picture]. Youtube. Retrieved December 23, 2019, from https://www.youtube.com/watch?v=NvNpZuC7RhM

Kim Hee, N. M. (2018). *Leveraging Big Data for Managing Transport Operations.* European Union . Frankfurt: Goethe-University Frankfurt. Retrieved 2019, from https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5bc2154c4&appId=PPGMS

Klopp et al., J. W. (2015). Leveraging Cellphones for Wayfinding and Journey Planning in Semi-formal Bus Systems: Lessons from Digital Matatus in Nairobi. *In Planning Support Systems and Smart Cities*, 200-241.

KNBS . (2018). *Mobile Penetration in Kenya.* Nairobi: Kenya National Bureau of Statistics.

KNBS. (2019). *2019 Kenya Population and Housing Census.* Nairobi: KNBS. Retrieved December 31/12/2019, 2019, from https://www.knbs.or.ke/?wpdmpro=2019-kenya-population-and-housing-census-volume-i-population-by-county-and-sub-county&wpdmdl=5615&ind=qpst0Ic7REF_8RI26gl5KA39hjmcXumdCrndPG9dMKmg_gZE6CqPMT8zy_TGhc1P

Korbel, M. S. (2019, December). *Enabling a digital and analytics transformation in heavy-industry manufacturing.* (M. &. Company, Ed.) Retrieved January 08, 2020, from McKinsey & Company: Enabling a digital and analytics transformation in heavy-industry manufacturing

Lee, Z. e. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. *IEEE Transactions on Intelligent Transportation Systems.*

Mahashreveta, C. (2019, 08 08). *What is Intelligent Transport System and how it works?* Retrieved August Wednesday, 08-09-2019, from https://www.geospatialworld.net/blogs/what-is-intelligent-transport-system-and-how-it-works/

Martin et al, J. L. (2015). *A DHL perspective on how to move beyond the hype.* Troisdorf: DHL Customer Solutions & Innovation.

McQueen. (2017). Benefit and Cost Estimation for Smart City Transportation Services. In M. Bob, *Big Data Analytics for Connected Vehicles and Smart Cities* (p. 280). Artech House, 2017. Retrieved January January 02, 2020, 2020, from https://books.google.co.ke/books?id=gKg0DwAAQBAJ&lpg=PA225&dq=main%20objectives%20of%20big%20data%20analytics%20projects%20in%20transportation&pg=PA218#v=onepage&q&f=false

Micro Focus. (2017, February 18). *Journey through the five stages of Big Data Analytics Adoption.* Retrieved January 6, 2020, from Micro Focus: https://www.microfocus.com/media/infographic/big_data_analytics_services_journey_through_the_five_stages_of_big_data_adoption_infographic.pdf

Microsoft. (2018, December 02). *Big Data Architectures.* Retrieved March 25, 2020, from Data Architecture Guide: https://docs.microsoft.com/en-us/azure/architecture/data-guide/big-data/

Milan, K. L. (2015). Fuzzy Logic in Traffic Engineering: A Review on Signal Control. In A. E. Wang (Ed.), *Studies in Fuzziness and Soft Computing* (p. 14). Copyright © 2015 Milan Koukol et al.

Mukabanah, E. (2020, 1 4). *Kenya Private Sector Alliance, KEPSA.* Retrieved from Kenya Bus Service: http://kenyabus.net/KEPSA%20Concept%20on%20Transport%20Institute.pdf

Mukherjee, S. a. (2016, February). Big Data – Concepts, Applications, Challenges and Future Scope. *International Journal of Advanced Research in Computer and Communication Engineering.* Retrieved October 20, 2019, from https://www.ijarcce.com/upload/2016/february-16/IJARCCE%2015.pdf

NAMATA. (2017, February 17). The Nairobi Metropolitan Area Transport Authority . *Orders and Acts, 1*(I), p. 75. Retrieved January 5, 2020, from http://namata.go.ke/images/Documents/EXECUTIVE-ORDER-NaMATA.pdf

NIST. (2019, March 18). NIST Big Data Interoperability Framework. *NIST Big Data Public Working Group Definitions and Taxonomies Subgroup, 6*(3), 3-20. doi:https://doi.org/10.6028/NIST.SP.1500-6r2

NTSA. (2018). *Public Transport and Safety Report.* Nairobi: National Transport and Safety Authority.

Numbeo. (2019). *Traffic Index 2019, Mid-Year Report of 2019.* Numbeo. Retrieved 01 05, 2020, from https://www.numbeo.com/traffic/rankings_current.jsp

Oya, Z. a. (2017). An Imagination of Organizations in the Future: Rethinking McKinsey's 7S Model. *DOI: 10.4018/978-1-5225-1656-9.ch006*. doi:DOI: 10.4018/978-1-5225-1656-9.ch006

Pedro, A. a. (2015). A comprehensive overview of open source big data platforms and frameworks. *International Journal of Big Data*, 15-33. Retrieved November 18, 2019, from https://www.researchgate.net/profile/Jorge_Bernardino/publication/310621654_A_comprehensive_overview_of_open_source_big_data_platforms_and_frameworks/links/5834298308ae004f74c78a33/A-comprehensive-overview-of-open-source-big-data-platforms-and-frameworks.

Prem, P. S. (2018). Intelligent Transportation System – A Critical Review. *International Journal of Civil Engineering and Technology (IJCIET)* . Retrieved 12 02, 2019, from http://www.iaeme.com/MasterAdmin/UploadFolder/IJCIET_09_06_047/IJCIET_09_06_047.pdf

Rik, C. G.-J. (2019). Why and how we should care about the General Data Protection Regulation. 1357. doi:https://doi.org/10.1080/08870446.2019.1606222

Robert et al., B.-A. W. (2016). Big data analytics for transportation: Problems and prospects for its application in China. In 2016 (Ed.). IEEE. doi:http://dx.doi.org/10.1109/TENCONSpring.2016.7519399

Sarah, W. A. (2015). The digital matatu project: Using cell phones to create an open sourcedata for Nairobi's semi-formal bus system. *Science Direct*, 1-28.

SASRA. (2019). *The Sacco Societies Regulatory Authority, The Sacco Societies Act (No. 14 of 2008).* Nairobi: SASRA.

Silva et al, T. &. (2018). Big Data Analytics Technologies and Platforms: a brief review. Rio de Janeiro (Brazil): Conference: LADaS - Latin America Data Science Workshop. Retrieved January 07, 2020, from

https://www.researchgate.net/profile/David_Araujo9/publication/329239211_Big_Data_Analytics_Technologies_and_Platforms_a_brief_review/links/5c069db392851c6ca1fd5b30/Big-Data-Analytics-Technologies-and-Platforms-a-brief-review.pdf

Statista. (2019, January 24). *Biggest challenges to big data adoption among corporations in the United States and worldwide, as of 2019*. Retrieved from Statista: https://www.statista.com/statistics/742983/worldwide-survey-corporate-big-data-adoption-barriers/

Taha, A. M. (2018). Traffic Simulation System Based on Fuzzy Logic. Procedia Computer Science.

TEDx (Producer), & Chin, R. (Director). (2020). *Smart Cities: Sustainable Urban Mobility-on-Demand* [Motion Picture]. Youtube Channel. Retrieved January 05, 2020, from https://www.youtube.com/watch?v=YHSu4f4fcSI

Thakur, A. a. (2019). Internet of Vehicles Communication Technologies for Traffic Management and Road Safety Applications. *Wireless Personal Communications, Vol III*(I), 3-19.

The Parliament of Kenya. (2018). *The Data Protection Bill.* Ministry of ICT. Nairobi: The Parliament of Kenya. doi:http://kenyalaw.org/kl/fileadmin/pdfdownloads/Acts/2019/TheDataProtectionAct__No24of2019.pdf

The UN Safety Road Week. (2019). *By 2020, halve the number of global deaths and injuries from road traffic accidents.* UN.

UN Habitat. (2016). *Sustainable Cities and Communities; A guide to assist National and Local Governments to monitor and report on Sustainable Development Goal – 11 Indicators.* United Nations (UN Habitat).

UN Habitat. (2016). *Sustainable Development Goals (SDG).* UN Habitat.

UN Habitat,. (2019, 08 05). *The Tools of the City Prosperity Initiative.* Retrieved August Monday, 2019, from http://cpi.unhabitat.org/tools-city-prosperity-initiative

Vesal et al., M. A. (2018). Quantitative analysis of probe data characteristics: Coverage, speed bias and congestion detection precision. *Journal of Intelligent Transportation Systems*, 103. doi:10.1080/15472450.2018.1502667

Victor, M. M. (2017). Beyond Traffic: Trends and Choices 2045 - A National Dialogue About Future Transportation Opportunities and Challenges. In G. M. Shaheen, & G. Meyer (Ed.), *Disrupting Mobility* (1st ed.). Springer International Publishing. Retrieved January 03, 2020, from https://books.google.co.ke/books?id=3y_XDQAAQBAJ&lpg=PP1&pg=PA2#v=onepage&q&f=false

Waleed, K. (Producer), & Waleed, K. (Director). (2017). *Machine Learning & Uber Maps* [Motion Picture]. Retrieved January 08, 2020, from https://www.youtube.com/watch?v=OAGxzVX3I34

Wayne, W. L. (2019, September 9). *Behavioral Change Models.* Retrieved from Boston University School of Public Health: http://sphweb.bumc.bu.edu/otlt/MPH-Modules/SB/BehavioralChangeTheories/BehavioralChangeTheories4.html

WHO. (2019). *Global Safety Week.* World Health Organization.

Zhengbing H. et al, J. H. (2019). Vehicle sensor data-based transportation research: Modeling, analysis,. *Journal of Intelligent Transportation Systems, 23*(II), 109. Retrieved November 24/11/2019, 2019, from https://www.tandfonline.com/doi/pdf/10.1080/15472450.2019.1586335?needAccess=true

Zhu et al., F. R. (2019). Big Data Analytics in Intelligent Transportation Systems: A Survey. *in IEEE Transactions on Intelligent Transportation Systems, Vol IV*.

APPENDICES

APPENDIX I

INTERVIEW GUIDE

Survey Questionnaire: - NaMATA

FOREWORD:
This research intends to gather qualitative information from the transport agencies involved in planning, controls, provisions, regulations, administration and policy development of transportation systems in Nairobi City.

This questionnaire is being conducted by Justin Njeru and supervised by Dr. Elisha Abade as part of research project submitted in partial fulfilment of the requirements of the award of Master of Science in Information Technology Management (Msc. ITM) of the University of Nairobi.

BACKGROUND:
Urban Transportation Systems are very important in city life; however, in developing countries the transportation systems are characterized by congestion, chaos and gridlocks. A massive research effort has been devoted to this field of study in the recent past.

Therefore, this paper is devoted to establishing architecture for big data analytics for urban transportation system in Nairobi city. In the recent past, studies in the Intelligent Transportation Systems (Bekiaris, 2017), (International Transport Forum, 2019) and many others have found that expansion of road networks does not solve the traffic situations.

INTRODUCTION:
The Nairobi Metropolitan Area Transport Authority (NaMATA)was established via an executive order by His Excellency President Uhuru Kenyatta on Thursday 9th February 2017, under the enabling provisions of the State Corporations Act. NaMATA covers the counties of Nairobi City, Kiambu, Kajiado, Machakos and Murang'a.

NaMATA is mandated on policy development that provides a comprehensive and dynamic platform for addressing the challenges in the transport sector that have bedeviled the Metropolitan Area. The Authority aims at formulating a sustainable integrated public transport strategy based on the development of a sustainable urban mobility plan that will be the basis for the orderly and structured development of the proposed Metropolitan Area mass-transit system, which incorporates both bus rapid-transit and commuter rail.

ABOUT THE QUESTIONNAIRE:
This questionnaire will take a few minutes to complete, and all the responses will be treated with utmost confidentiality. Any documents helpful to this research can be shared via email: justin.njeru@uonbi.ac.ke or WhatsApp: +254 765 895 167
* Required

1. Which areas best describes NaMATA's mandate in Urban Transportation Systems (select all relevant) *

Check all that apply.

- [ ] Motorized Drive Ways
- [ ] Non Motorized Path ways
- [ ] Light Train
- [ ] commuter buses
- [ ] Policy Development
- [ ] Taxi Regulations
- [ ] Matatus
- [ ] Regulations
- [ ] Safety and Security
- [ ] Urban Planning
- [ ] Other: _____

2. What are your thoughts on technologies like Big Data Analytics. if you have heard of it: [paragraph]

_____

_____

_____

_____

_____

3. What are the priority projects under way for the metropolitan in NaMATA's jurisdiction *

Check all that apply.

- [ ] Road Expansion
- [ ] Technology Driven Projects
- [ ] City / Urban Planning
- [ ] Restoration of Order
- [ ] Sensitization Programmes
- [ ] Data Collection
- [ ] Public Participation exercise
- [ ] Policy Development
- [ ] Other: _____

4. In incidents reporting and traffic flow monitoring what would the agency consider most? *

Check all that apply.

- [ ] Police Radio Call
- [ ] Twitter Handle
- [ ] Facebook updates
- [ ] Traffic Marshalls
- [ ] Phone / Telephone Calls
- [ ] Helicopter
- [ ] other Transport Agencies
- [ ] Other: _____

5. In your View What could be ailing Nairobi Transport Systems [paragraph] *

_____

_____

_____

_____

_____

6. How do you agree each of the statements: Strongly Agree - 5 and Strongly Disagree = 1, Neither = 0 *

Check all that apply.

	Strongly Agree	Agree	Neither Agree / Disagree	Disagree	Strongly Disagree
Conduct of Drivers contribute to Urban Transportation Challenges	☐	☐	☐	☐	☐
Data Driven Decision Making is vital to alleviate traffic problems	☐	☐	☐	☐	☐
There is need to invest more on Data Analysis Technologies to assist in understanding our City	☐	☐	☐	☐	☐
Social Media Plays a vital segments of Nairobi Travelers for receiving and sharing updates	☐	☐	☐	☐	☐
Contribution of Technologies cannot be ignored in Urban Development and especially roads	☐	☐	☐	☐	☐
That Architectures for receiving Data on traffic incidents are increasingly becoming important	☐	☐	☐	☐	☐
Coordinating with other government agencies and other parties for the development and operation of transport infrastructure could be challenging.	☐	☐	☐	☐	☐
Planning of the city has factored future expansions	☐	☐	☐	☐	☐
The issue of mobility involves more than transportation	☐	☐	☐	☐	☐

That, there is need for local solutions to manage transport solutions in Nairobi	☐	☐	☐	☐	☐
That, the agency finds the data on Social Media highly dependable	☐	☐	☐	☐	☐
That Agency could benefit on Big Data Analytics for Transportation Systems in Nairobi for decision making and real-time traffic monitoring	☐	☐	☐	☐	☐
That, Big Data Analytics is mature technology for adoption	☐	☐	☐	☐	☐
The agency has flexible budget to cater for technology improvement	☐	☐	☐	☐	☐
That, Agency has versatile personnel to undertake more research in Big Data Analytics	☐	☐	☐	☐	☐

7. In your Assessment What are the expected Social Economic Effects of Public Transport Strategy [paragraph] *

8. How would you consider these sources of data for decision making on Policy, Planning and Budgeting in your strategy: rated: 2 = Highly Likely and 0 = less likely *

Check all that apply.

	Highly likely	likely	neither	unlikely	Highly Unlikely
Social media	☐	☐	☐	☐	☐
Policy Briefs	☐	☐	☐	☐	☐
Research Papers	☐	☐	☐	☐	☐
Mainstream media	☐	☐	☐	☐	☐
Congress Papers	☐	☐	☐	☐	☐
Agency Reports	☐	☐	☐	☐	☐

Agency Opinion / Analysis

In your past analysis , which modes of transport contributes to the traffic jams / congestion in Nairobi : Rated 10 = Highest and 1 = Lowest

9. Buses *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Lowest	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Highest

10. Private Cars *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Lowest	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Highest

11. Matatus (9/14 Seater Vans) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Lowest	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Highest

12. Trucks / Pickups *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Lowest	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Highest

13. In terms of Prioritization Which route(s) is/are being focused on by the agency for improvements ... *

Check all that apply.

- [ ] Nairobi
- [ ] Kiambu
- [ ] Kajiado
- [ ] Thika / Thika Road
- [ ] Machakos
- [ ] Limuru
- [ ] Kikuyu
- [ ] Karen
- [ ] Ngong road
- [ ] Ruai / Eastern bypass
- [ ] Naivasha / Nakuru
- [ ] Murang'a / Embu / Nyeri

Other: ☐ _____

14. In your own opinion, *

Mark only one oval per row.

	Strongly Agree	Somewhat Agree	Neither Agree Nor Disagree	Somewhat Disagree	Strongly Disagree
Air quality has been affected by motor vehicles					
Economy depends on transportation systems					
That, despite of more routes and bypasses, Traffic jams has not subsidized					
There is need to have inclusive means of transport catering for motorized and non motorized					
That information is key in understanding traffic patterns, route planning for persons and also for transportation agencies					
Social Media can bring information necessary to plan route efficiently					
Road traffic alerts are NOT necessary for commuters / drivers					
That, City Transport Routing program ought to be Data Driven					
That, there is sufficient data generated from mobile phone users to help city planners and traffic controllers					

agencies if captured and disseminated properly

15. As an Agency What is your primary source of traffic information updates / incidents ? *

Mark only one oval.

- ◯ CAR RADIO
- ◯ PHONE CALLS to fellow drivers
- ◯ SMS alerts (subscribed)
- ◯ Facebook updates from fans
- ◯ Twitter handles
- ◯ Google live Map
- ◯ CCTV live feed
- ◯ Other: _____

16. What Innovations / Technologies for Transportation Systems are being implemented or are in pipeline that the researcher can highlight in the report: [paragraph] *

APPENDIX II:

CONSENT FORM

Table 6: Survey respondents consent form

☐	I..... voluntarily agree to participate in this research study.
☐	I understand that even if I agree to participate now, I can withdraw at any time or refuse to answer any question without any consequences of any kind.
☐	I understand that I can withdraw permission to use data from my interview within two weeks after the interview, in which case the material will be deleted.
☐	I have had the purpose and nature of the study explained to me in writing and I have had the opportunity to ask questions about the study.
☐	I understand that participation involves filling of research questionnaires and interviews.
☐	I understand that I will not benefit directly from participating in this research.
☐	I agree to my interview being audio-recorded.
☐	I understand that all information I provide for this study will be treated confidentially.
☐	I understand that in any report on the results of this research my identity will remain anonymous. This will be done by changing my name and disguising any details of my interview which may reveal my identity or the identity of people I speak about.
☐	I understand that disguised extracts from my interview may be quoted in the said academic research paper or published in the University of Nairobi Journals.
☐	I understand that signed consent forms and original audio recordings will be retained in at University on Nairobi Electronic Repository.

- I understand that a transcript of my interview in which all identifying information has been removed will be retained for one year from the date of panel presentation of the research findings by the student.
- I understand that under freedom of information legalization I am entitled to access the information I have provided at any time while it is in storage as specified above.
- I understand that I am free to contact any of the people involved in the research to seek further clarification and information.

Names, degrees, affiliations and contact details of researchers

---

Signature of research participant

---------- ----------

Signature of participant *Date*

I believe the participant is giving informed consent to participate in this study

---------- ----------

Signature of researcher *Date*

APPENDIX III:

SURVER QUESTIONNAIRE

Survey Questionnaire: - Transportation System in Nairobi

This research will mainly focus on the Public Transportation System in Nairobi Metropolitan to exam these areas:

1. congestion
2. transport reliability
3. transport safety
4. air quality
5. innovations
6. regulations
7. versatility
8. customer satisfaction
9. accessibility and investments.
10. Safety and Security

* Required

1. Which best describes your mode of transport in Metropolitan of Nairobi (Multiple Choices) *

Check all that apply.

- [ ] Walking
- [ ] Driving private car
- [ ] using train
- [ ] commuter buses
- [ ] motor cycle (boda boda)
- [ ] taxi (uber, taxify, bolt)
- [ ] matatus

Other: [ ] _____

2. I have worked/lived in Nairobi for about; *

Mark only one oval.

- ◯ 0.3 to 1 year
- ◯ 2 to 5 years
- ◯ 6 to 10 years
- ◯ 11 to 20 years
- ◯ over 21 years

3. Normal Working days (except weekend). What time do you wake up *

Check all that apply.

- ☐ 3:00 am - 4:00 am
- ☐ 4:00 am - 5:00 am
- ☐ 5:00 am - 6:00 am
- ☐ 7:00 am - 8:00 am
- ☐ 9:00 am - 11: am
- ☐ irregular times / shift time
- ☐ 1: 00 pm - 5:00 pm

Other: ☐ _____

4. Normal Working days (except weekend) what time do you arrive at Bus Stage (or if picked by Taxi / Driver, what is the pick up time) *

Check all that apply.

- ☐ 3:00 am - 4:00 am
- ☐ 4:00 am - 5:00 am
- ☐ 5:00 am - 6:00 am
- ☐ 7:00 am - 8:00 am
- ☐ 9:00 am - 11: am
- ☐ irregular times / shift time
- ☐ 1: 00 pm - 5:00 pm

Other: ☐ _____

5. Normal Working days (except weekend) what time do you arrive at work *

Check all that apply.

- [ ] 3:00 am - 4:00 am
- [ ] 4:00 am - 5:00 am
- [ ] 5:00 am - 6:00 am
- [ ] 7:00 am - 8:00 am
- [ ] 9:00 am - 11: am
- [ ] irregular times / shift time
- [ ] 1: 00 pm - 5:00 pm

Other: [ ] _____

6. on a Normal Working days (except weekend) you leave work at what time *

Check all that apply.

- [ ] 11:00 am - 1:00 pm
- [ ] 1: 00 pm - 3:00 pm
- [ ] 3:00 pm - 5:00 pm
- [ ] 6:00 pm - 8:00 pm
- [ ] 9:00 pm - 11:00 pm
- [ ] irregular times / shift time

Other: [ ] _____

7. Which sector of economy / employment best describes you *

Mark only one oval.

- ◯ Student - College / university
- ◯ Public Service / institutions
- ◯ Private Company's
- ◯ Transport Sector (Matatu industry)
- ◯ Health / Medical Sector
- ◯ Construction / Manufacturing
- ◯ Self Employed / Freelance
- ◯ Other: _____

8. I live within the metropolitan area of... *

Mark only one oval.

- Nairobi
- Kiambu
- Kajiado
- Thika / Thika Road
- Machakos
- Limuru
- Kikuyu
- Karen
- Ngong road
- Ruai / Eastern bypass
- Naivasha / Nakuru
- Murang'a / Embu / Nyeri
- Other: _____

9. How do you agree *

Mark only one oval per row.

	Strongly Agree	agree	neither agree/disagree	disagree	Strongly Disagree
that my mode of transport is convenient and reliable					
the route to my workplace is convenient and timely					
the fare i use / fuel is considerate					
that transport sector is chaotic for my daily routine					
somewhat i spend considerable time on traffic					
i use phone call to find out traffic situation on the way					
i use social media (twitter) to get traffic updates					
i listen to radio in the vehicle for traffic updates					
that finding out traffic situation helps me navigate for alternative route					
that getting traffic updates can be very helpful in planning your route					
that there are alternative routing in and out of town to my house					
that sharing my traffic situation with others helps others not to get into similar problems					
that i have always managed to arrive to my destination on					

time despite of traffic situations					
that expansion of roads will ease traffic					
traffic jams in CBD are as a result of bad driving habits					
that buses should be banned from entering CBD					
that private cars should be banned from entering CBD					
that walk paths will eliminate traffic jams					
that bicycle lanes can moderate traffic jams being alternative mode of transport					
That the security of passengers is guaranteed on Kenyan Roads					
Road Safety is wanting in Nairobi					

10. To understand your opinion on these statements, and measured as: 5= "extreme often" if frequency is five times in a week, 4= "very often" for four times in a week, 3= "moderate often" for 3 times in a week, 2= "slightly often", and 1= "not at all often" for occurrence of once in a week *

Mark only one oval per row.

	extreme often	very often	moderate often	slightly often	not at all often
How often do you get stack unexpectedly in traffic jams	◯	◯	◯	◯	◯
How often do you get into the traffic jams and expecting it	◯	◯	◯	◯	◯
How often do I check on phone for traffic updates	◯	◯	◯	◯	◯
How often do I listen to preferred radio station for traffic updates bulletins	◯	◯	◯	◯	◯
How often do I call to find out what happening on the traffic route I am headed to	◯	◯	◯	◯	◯
How often do I change my route because of traffic situations (for drivers)	◯	◯	◯	◯	◯
How often do I find my route diverted due to traffic ahead	◯	◯	◯	◯	◯
How often have I considered devices (Phone, Tablet) as a good companion for checking traffic and routing in the City	◯	◯	◯	◯	◯
How often has the information on traffic situation from social media (Twitter, Facebook) be reliable	◯	◯	◯	◯	◯
How often do you feel unsafe in our roads	◯	◯	◯	◯	◯
How often do you feel your security and your belongings	◯	◯	◯	◯	◯

11. In your your own opinion, *

Mark only one oval per row.

	Strongly Agree	Somewhat Agree	Neither Agree Nor Disagree	Somewhat Disagree	Strongly Disagree
Air quality has been affected by motor vehicles					
Economy depends on transportation systems					
Government has done enough to alleviate traffic jams					
that, there is room for improvement as long as we want efficient transportation					
I am aware of the government policies and regulations on traffic					
that majority of city dwellers do appreciate the transport mode in their areas					
that there are no alternative means of transport in my route (bike lanes, walk paths)					
There is need to have inclusive means of transport catering for motorized and non motorized					
That before i change my living area, my consideration is more of Traffic pattern in that route rather than availability of water,					

security, closeness to work area.					
That information is key in understanding traffic patterns, route planning for persons and also for transportation agencies	◯	◯	◯	◯	◯
Social Media can bring information necessary to plan route efficiently	◯	◯	◯	◯	◯
Road traffic alerts are NOT necessary for commuters / drivers	◯	◯	◯	◯	◯
my travel plan is organized based on the traffic lead times; rush hour, etc	◯	◯	◯	◯	◯

12. What is your primary source of traffic information updates? *

Mark only one oval.

- ◯ CAR RADIO
- ◯ PHONE CALLS to fellow drivers
- ◯ SMS alerts (subscribed)
- ◯ Facebook updates from fans
- ◯ Twitter handles
- ◯ Google live Map
- ◯ CCTV live feed
- ◯ Other: _____

Skip to question 13

Page 1 of 2

Technology plays a vital role in transportation industry either used to warn drivers of congested routes, incidents and accidents. This information is reliable for decision making by traffic controllers/ agencies. For drivers who find the information key to planning their route is also time saving.

Therefore, how does information get collected, shared and is it dependable for drivers, commuters and transport agencies.

13. Which mode do you use to share traffic updates / incidents *

Mark only one oval.

- ◯ RADIO CALL
- ◯ PHONE CALLS to fellow drivers
- ◯ SMS
- ◯ Facebook page
- ◯ Twitter hashtag
- ◯ Not applicable (NEVER SHARED UPDATES)
- ◯ Other: _____

14. How do you agree with these statements : 5= Strongly Agree 4= Agree 3= Neither Agree nor Disagree 2= Disagree 1= Strongly Disagree

Mark only one oval per row.

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
that I Have a phone i use to check on traffic updates					
that i have access to 4/5G network connectivity					
that i always seek to know traffic updates					
social media is part of life; spending time on it while on traffic					
I am NOT worried of information i share on social media					
I am aware of data protection laws in Kenya					
I am worried of misleading information on social media					
That a mobile app to update drivers on traffic situation in our roads will play an important role in city dwellers					
that the amount i spend on transport is commensurate to the services I get					
that nothing can be done to alleviate traffic situation in our roads					

15. From 1 to 10, how do you rate your mode of transport here in Nairobi *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Worst	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Best

16. Describe what you feel should be changed to improve traffic flow in Nairobi *

17. Describe your worst experience in traffic jam (if any)

18. What do you normally do when you find stack in traffic jam *

Mark only one oval per row.

	Strongly Agree	Agree	Neither Agree Nor Disagree	Disagree	Strongly Disagree
Listen to radio / music					
Sleep / nap					
Look for alternative route to my destination					
jump to boda boda ride					
call my friends to warn them of the traffic jam					
post the situation of the jam to social media					
fan myself and wait for the situation to improve					
wait patiently (hakuna matata)					
take a U turn or cancel the trip					
call my employer / supervisor or my client that I sense I will be late					

19. From 1 to 10, how happy are you about road improvement program in your area *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Less Happy											Extreme Happy

20. From 1 to 10, how do you rate your last mile to reach your house *

Mark only one oval per row.

	Best	Good	Bad	Worst
Road Condition (paved, earth,tarmacked)				
Security of users				
Safety of pedestrian				
Accessibility				
Congestion / Jam				
Lighting Condition				
Foot paths				
Cleanliness				
Drainage System				

21. From 1 to 10, how do you rate Ministry of Transport in overall satisfaction *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Bad											Excellent

22. From 1 to 10, how do you rate National Transport and Safety Authority (NTSA) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Bad											Excellent

23. From 1 to 10, how do you rate Kenya Urban Roads Authority (KURA) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Bad	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

24. From 1 to 10, how do you rate Kenya National Highway Authority (KeNHA) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Bad	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

25. From 1 to 10, how do you rate Nairobi Metropolitan Area Transport Authority (NaMATA) *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Bad	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

26. From 1 to 10, how do you rate Transport Department of City County *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Bad	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

27. From 1 to 10, how do you rate your Matatu SACCO / transport company servicing your route *

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Bad	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Excellent

28. Your route is serviced by which Matatu Sacco / Transporter?

_____

29. What innovations / improvements that you feel would lessen traffic jam situation in Nairobi and would like the Transport Ministry to consider *

_____

_____

_____

_____

_____

Thank you for your responses!
Click SUBMIT button

APPENDIX IV:

SUMMARY OF THE RESULTS

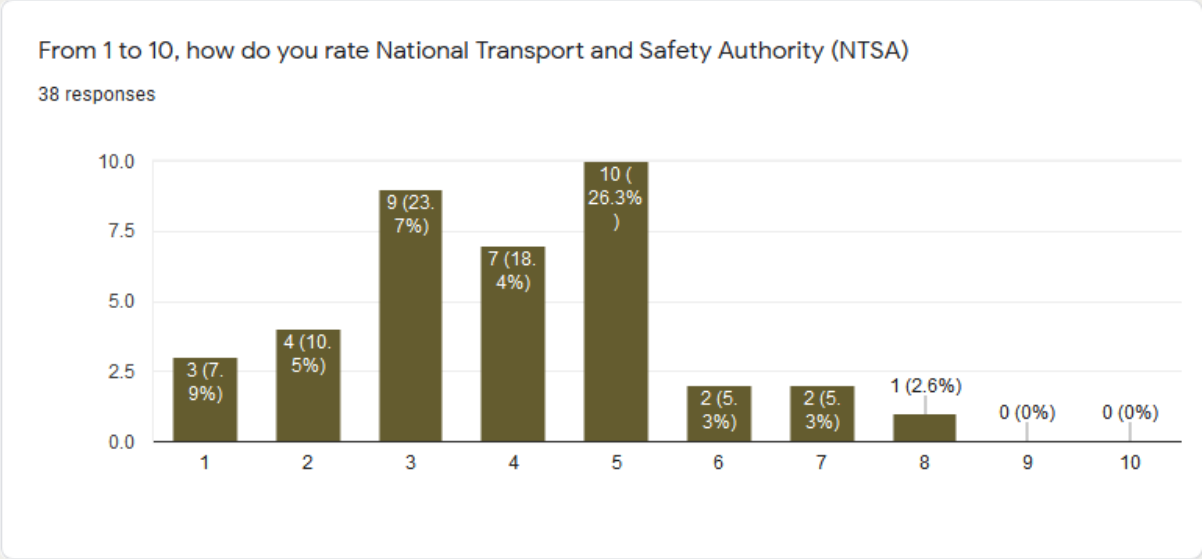


Figure 37: Satisfaction rating for the NTSA by participants

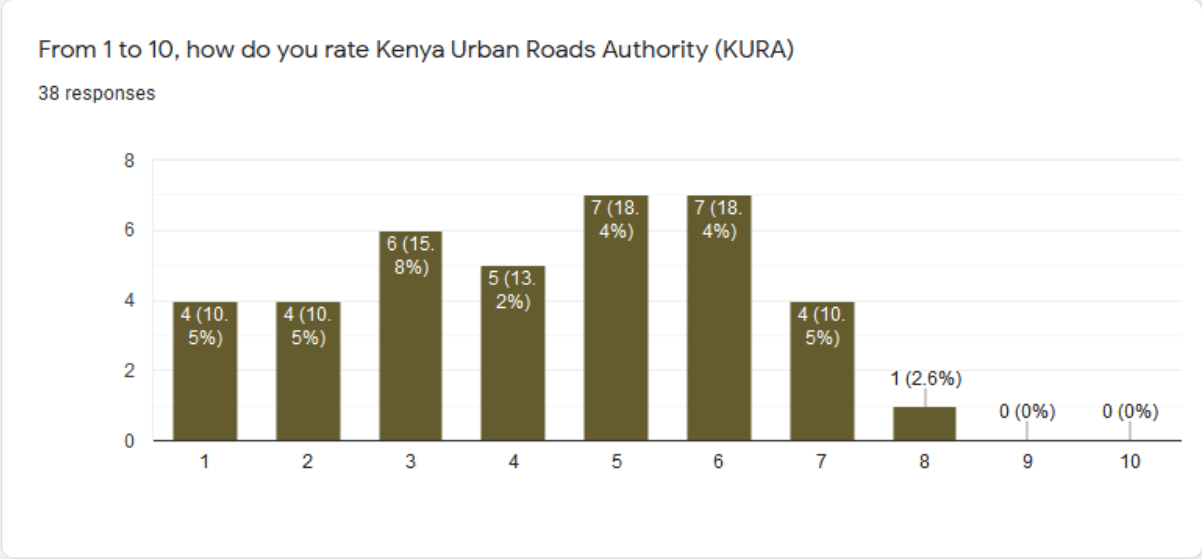


Figure 38: Satisfaction rating for KURA

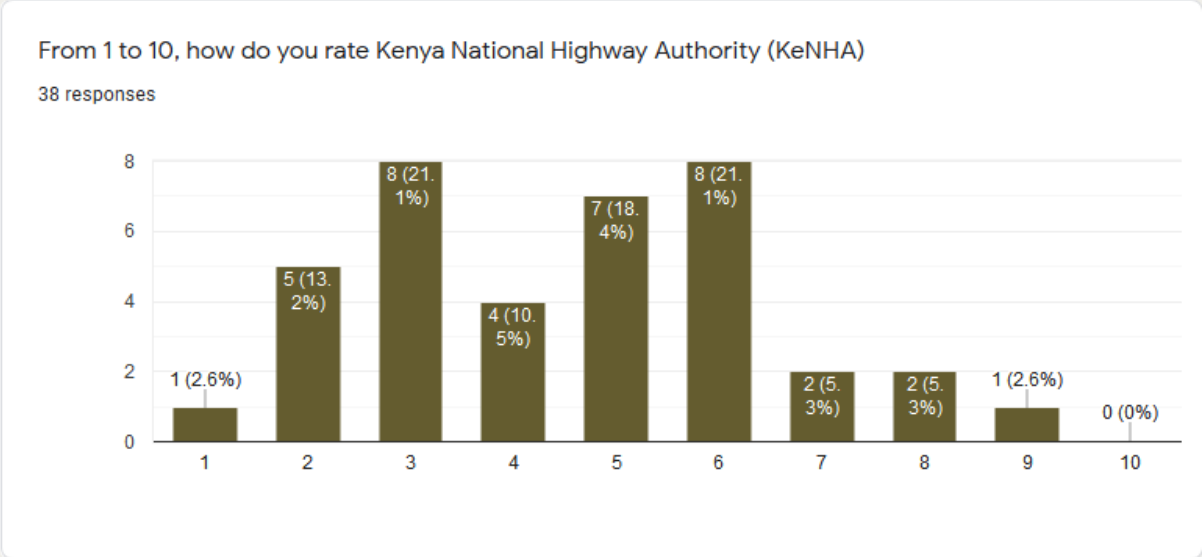


Figure 39: Satisfaction rating for KeNHA

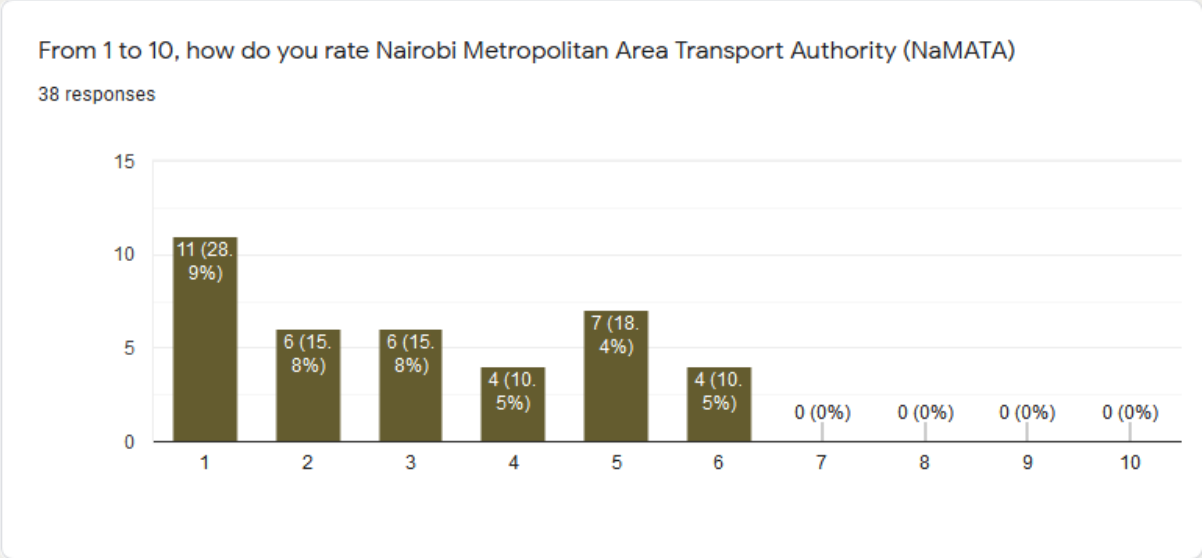


Figure 40: Satisfaction rating for NaMATA

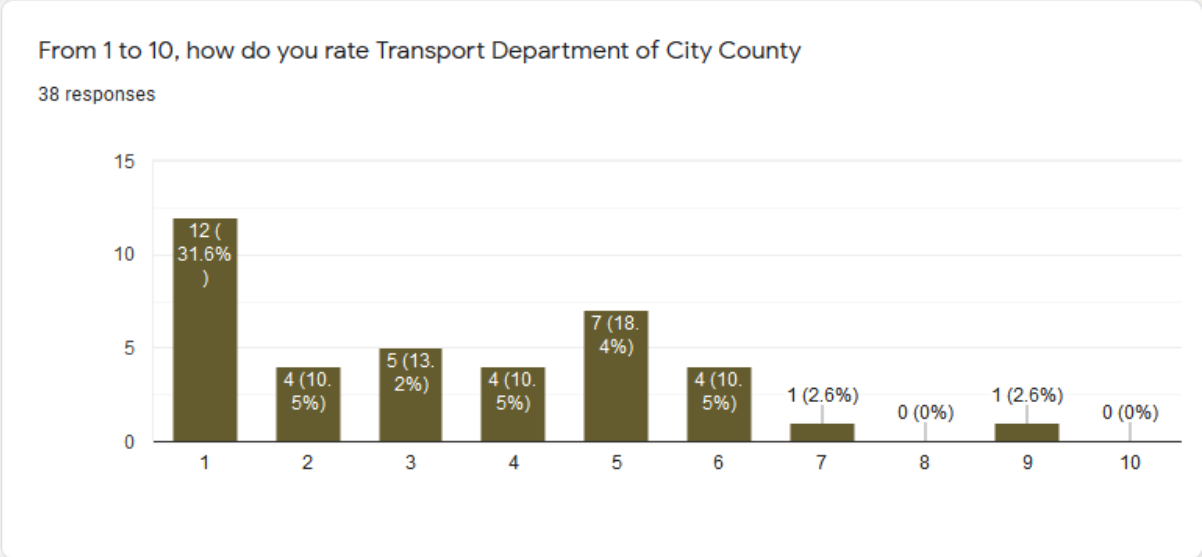


Figure 41: Satisfaction rating for the transport department for County of Nairobi

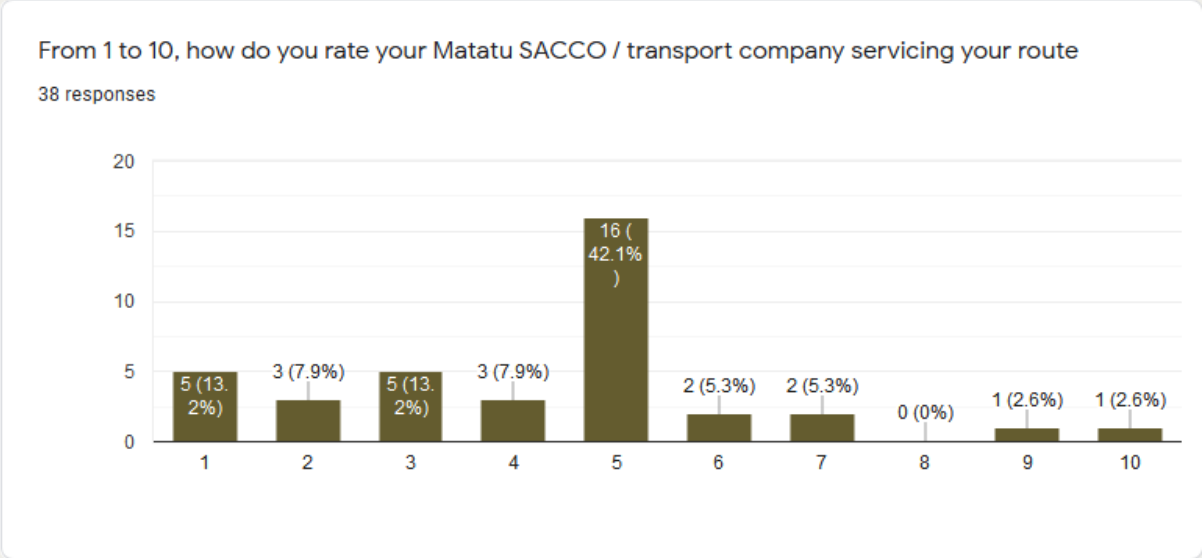


Figure 42: Satisfaction rating for respective Transportation Companies / SACCO