DATA REPORT

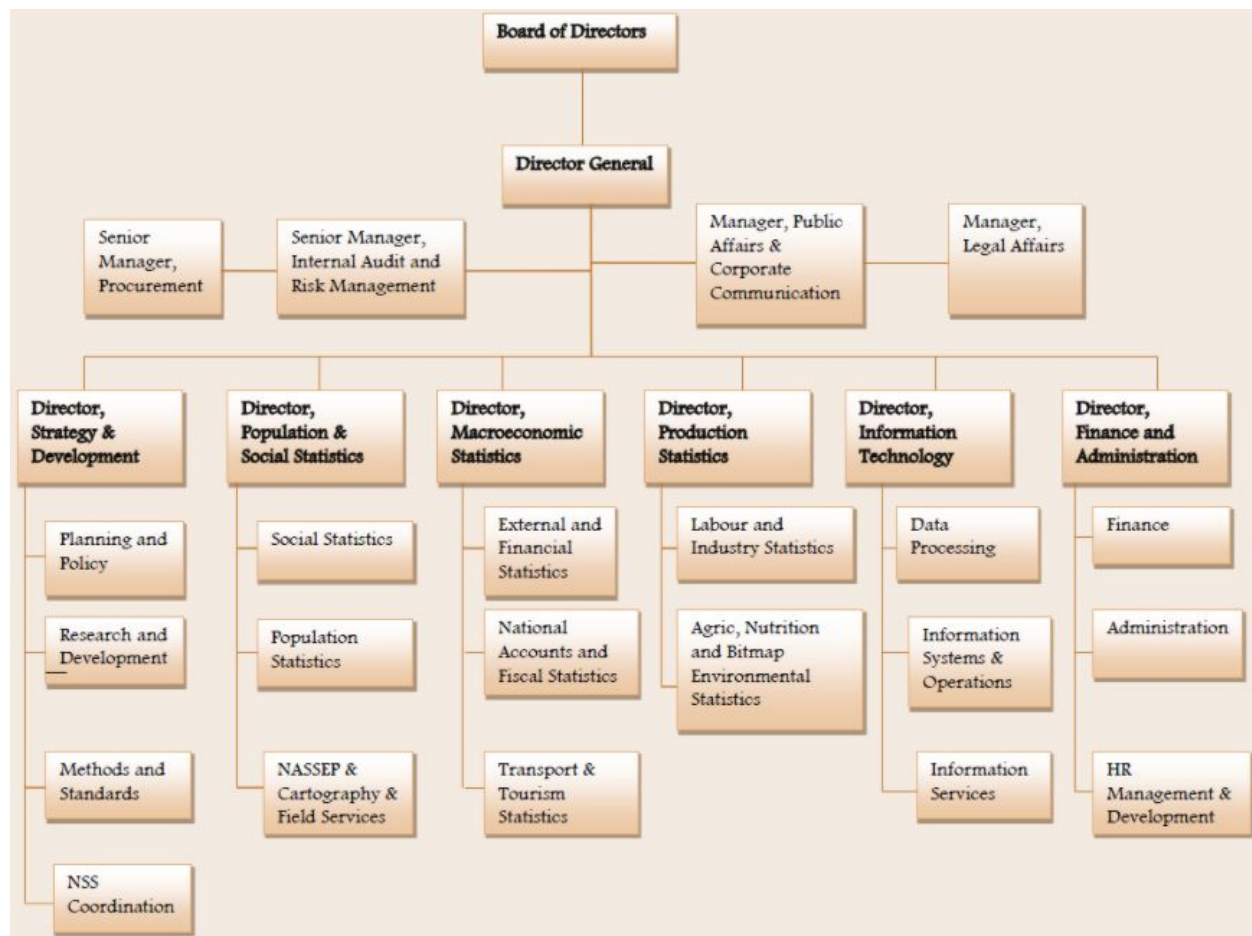## 1. BUSINESS UNDERSTANDING

1.1. Business Background

1.1.1. Organization Structure:

The following diagram is a representation of the organization structure of KNBS:

1.1.2. Census management structure:

Various committees were established and operationalized at both national and county levels to manage the census process. These committees drew their membership from various stakeholders namely; government departments and agencies, private sector, learning institutions, religious organizations, and Non-Governmental Organisations (NGOs).The National Census Steering Committee spearheaded the implementation

process and provided policy direction such as approving all the census guidelines, strategies and mobilization of resources. Coordination of the processes was undertaken by the National Census Coordination Committee.

The Technical Working Committee was responsible for all the technical aspects of the census, which included development of the instruments and ensuring adherence to quality assurance guidelines. At the county level, the County Census Committees (CCCs) and Sub-County Census Committees (SCCCs) were established to oversee implementation of the 2019 census activities, such as recruitment of personnel, procurement of training venues and logistical arrangements for transport and security at their respective levels. These committees were instrumental in rallying public participation for the census.

1.1.3. Sectors that will be affected by the project.
The following are areas that will be affected by the census project:
- Education
- Agriculture
- Housing
- Religion
- ICT
- Water and Sanitation
- Employment
- Ethnic/Nationality Background

1.1.4. Describing the Problem Area
The main problem here is in achieving the country's development agenda such as the Big Four and Vision 2030, and other global initiatives including the Sustainable Development Goals (SDGs). In order to achieve this, a national census is done in response to the demand for statistical information every 10 years to enable planning and allocation of resources.

1.1.5. Current Solution
The current solution as of now is that a count of persons takes place every 10 years on a *de facto* basis with the midnight of 24th/25th August as the reference point. The censuses have been implemented in accordance with the United Nations (UN) Principles and Recommendations for conducting population and housing censuses.
An advantage of this is that:

- Nowadays, more variables/aspects are put into consideration when taking a count which make it easier to address these problems.
- The use of technology has also been implemented during the count, making it easier to capture data efficiently and effectively.

The challenge here is that:
- the census data has been released in PDF format. Therefore it does not allow for it to be reused for visualisations that simplify understanding for everyone.
- It also does not allow people to overlay the data with other kinds of data for fresh insights that would help in the achievement of the Sustainable Development Goals (SDGs.)
- There is also uncertainty in the genuineness of the respondents during the count.

1.2. **Business objectives**

Our main focus is on the recent census that took place in 2019 with particular detail on the socio-economic characteristics of the population.

The **main objective** of the 2019 KPHC(Kenya Population and Housing Census) was to collect information on the size, composition, distribution and socio-economic characteristics of the population.

The specific objectives were to ascertain the following:
- Population size, composition, and spatial distribution;
- Levels of fertility, mortality and migration;
- Educational attainment;
- Household composition;
- Rate and pattern of urbanization;
- Size and deployment of labour force;
- Distribution of persons with disability;
- Housing conditions and availability of household amenities; and
- Agricultural indicators to inform the creation of an agriculture sampling frame.

This information will be used in:
- planning, budgeting and programming for important services;
- future policy formulation, resource allocation;
- creation of administrative and political units;
- monitoring and evaluation of programmes and projects;
- research;
- development of a master household sampling frame;
- development of geo-spatial database; and

- benchmark for agricultural census/surveys.

1.3. Business Success Criteria
The census is usually deemed to be successful if:
- The above named objectives are met and
- Valuable predictions can be made from the data collected about issues such as: mortality rate, level of education etc.

1.4. Assessing the Situation
1.4.1. Resource Inventory

- *Research Hardware Resources* - Laptops/PCs will be used for research and analysis.
- *Data Sources and Knowledge Stores* - The data from the Census is available from the KNBS website in .pdf format. There weren't any security issues when obtaining the data.
- *Personnel* - The data was collected by enumerator officers. The visualizations will be done by Team Decoders, members of the DCS5 Moringa Core Program.
- *Software*s used
    - ★ I love Pdf used to convert the data from pdf format to excel format
    - ★ Tableau and Data studio used for various visualizations
    - ★ Excel for the editing of the tables to be used in visualizations

1.4.2. Requirements, Assumptions, and Constraints
There were no requirements towards obtaining the data.
Assumptions:
- the data can be assumed to be of quality since the enumerators had initially been trained. Also data was collected electronically.
- the data has some missing values indicated by the "Not Stated" column.
Constraints:
- Our major constraint is on the format the data came in(.pdf) and converting it to the appropriate format.

1.4.3. Risks and Contingencies
There is not a great deal of immediate risk. It is necessary that the project is delivered within the expected time frame of 3 weeks. It might however take us some time to

compile the data. We might also not be able to analyze all the socio-economic factors as listed in the data. We nonetheless, aim to cover as much as we can to deliver what is expected and within the time frame.

1.4.4. Terminologies
- *Household*: Refers to a person or group of persons who reside in the same homestead/compound but not necessarily in the same dwelling unit, have same cooking arrangements, and are answerable to the same household head.
- *Group Quarters:* For the purposes of this census, group quarters refer to persons enumerated in collective living quarters such as army barracks and hostels in learning institutions.
- *Household Head:* This is the most responsible/respected member of the household (as at the census night) who makes key decisions in the household on a day to day basis, and whose authority is recognized by all members of the household.
- *Dwelling Unit:* This is a place of abode or residence occupied by one or more households, usually with a private entrance.
- *Enumeration Area (EA):* This is a designated area with an average of 100 households, but may vary from 50 to 149 households depending on the population density, terrain, and/or vastness of the area concerned. An EA may be a village, group of villages or part of a village, and is so delineated to be conveniently covered by an enumerator.
- *De Facto Census:* One that enumerates all persons depending on where they spend (or are found on) the census night.
- *Age:* The number of years/months that a person has lived as at the census reference night
- *Area:* The extent of a surface enclosed within a specified boundary, usually expressed in square kilometres.
- *Density:* Refers to number of persons per square kilometre, obtained by dividing the enumerated persons by the size of the land area in a given area, expressed in square kilometres.
- *Sex:* Refers to the biological condition or characteristics of being male, female or intersex.

1.4.5. Cost-Benefit Analysis
The cost incurred by KNBS was quite high for the data collection. It involved procurement of new devices for the count, training personnel. KNBS received support from the government and development partners to implement the Census, which cost

about Kshs. 18.5 billion. The benefits of the census go a long way towards: creation of jobs, planning, allocation of resources, provided the data which is used for analysis and research in other sectors. The data was also used for making the 2020/2021 budget where we saw certain counties were given more funds towards their sustainability and development, rehabilitation and improvement of existing infrastructure.

1.5. <u>Data Mining Goal</u>

- Our main data mining problem is visualization of the education attainment as a cluster of the socio-economic factors with precise reference to gender(male/female)
- Building an interactive dashboard of the 2019 population and housing census with a focus on Socio-Economic characteristics.

1.5.1. Data Mining Success Criteria

This project will be considered a success if great visualizations are built to represent the census data with respect to the different Socio-Economic characteristics.

1.6. <u>Project Plan</u>

Phase	Time	Resources	Challenges faced
Business Understanding	Week 1: 2 days(Mon,Tue)	All analysts	- Finding the right visualization software
Data Understanding	Week 1:	All analysts	- Data sourcing & conversion
Data Preparation	Week 1:		
Analysis	Week 1 &2:		
Evaluation	Week 2&3:		
Deployment	Week 3:		

Initial assessment of tools and techniques
The following is a list of the tools we used for this project:

- *I LovePDF* to convert the original data to excel format.
- *Microsoft Excel* to store the compiled data
- *Tableau/Data Studio/ Power BI* for visualization

2. <u>DATA UNDERSTANDING</u>

2.1. Data Collection

KNBS did a census where they had enumerators and field officers who collected data in a primary form across the counties in the country. This allowed them to perform various analyses that included distribution of population by age, sex and administrative units, population by county and sub county and social-economic characteristics.

Some of the challenges they faced during data collection was; some of the sub-locations had no enumerated population on the census night for various reasons, in some instances they were boundary disputes and poor network connectivity hence slowed down the pace of mapping and in some areas hampered real-time transmission of data to the central server.

Need for further analysis enabled us to use the primary data they had provided in their website as our secondary data, where we came up with a visualization of the socio-economic characteristics.

2.2. Data Description and exploration

The data was encrypted prior to transmission and was backed up in off-site locations. The data was processed using CSPro, Stata and SPSS software. In addition, validation checks were done to ensure that all EAs, aligned to administrative boundaries, were accounted for in the dataset. Outputs were generated based on administrative and geo-political units.

The data was collected according to the various counties in the country and according to the gender of the participants.

The data provided by KNBS was in pdf format and had around **45 ** tables that gave various social-economic characteristics of the population in Kenya.

Some of the characteristics that were observed were the highest level of education reached, ethnicity, activity status of labour force, distribution of households, waste disposal method and disability status.

The data was collected in numeric form as it was a count of the population in the country.

The data we used did not have any anomalies despite the numerous counts of tables that we encountered.

3. <u>DATA PREPARATION</u>

3.1. Data selection

The dataset we obtained had a total of 101 tables and we had to select the relevant tables for our analysis. We settled on the 8 tables listed below as they were essential for our analysis.

Tables used
- Table 2.14 Source of drinking water
- Table 2.2 School Attendance
- Table 2.25 Agriculture
- Table 2.24 Livestock & Fish
- Table 2.36 Ownership housing
- Table 2.26 Disability
- Table 2.30 Religion
- Table 2.31 Ethnicity

3.2. Data cleaning & integration

The census data provided an easier integration of editing once it was converted to excel, each table as shown above from the table list was edited into a new excel file and named according to the table.

There were no missing values or errors in the data that we used. In addition to that, there were no coding inconsistencies in the data hence the smooth transition into data analysis. We however had to rename some columns to make sense of the data collected and also to make the data used from the tables easily understood and more direct.

3.3. Constructing New Data

Some of the tables had to be split into different sheets for proper and in-depth analysis. An example is separating the counties from the sub counties table. This aided is the visualization of the difference between the trends in counties and sub counties concurrently and see how each performs.

With our data, we did not need to create any base models or perform modelling as it was unnecessary for our analysis. The data we had was direct and could easily be interpreted with the univariate, bivariate and multivariate plots of the summaries.

4.   RECOMMENDATIONS

Insights from tableau

Kilifi is seen to have a lower population of students transitioning from pre-primary all the way to secondary. Hence a recommendation to make a consistent transition from pre-primary to secondary to avoid the large rates of dropouts. On the other hand areas like Bungoma and Kakamega have a high density of individuals who are in school.

Counties like Turkana, Narok, Garissa, Mandera, Wajir, Kilifi and Meru have experienced a high number of the importance of education avoiding their olden days of pastoralism where cattle bucolic was important to attending school but now they are beginning to give education more importance despite the age.
Each county boasts itself for a certain trait where Kakamega and Bungoma are doing well in terms of agricultural labourship. Hence the notion of if you want hardworking people you can find them in these two counties. Nevertheless, an assumption of farming being the most common among most households in kenya. Where they're either practicing crop or livestock production or even both concurrently.

Irrigation

The flow of water from rivers is not enjoyed by all the counties hence counties like Isiolo,and Tharaka-Nithi are quite dry, hence practice irrigation. A specification of either drip, sprinkler or central pivot irrigation wasn't noticeable since the findings were only recorded as either taking irrigation but not the form of irrigation. This is supported by most rivers flowing through Meru county.

The county of Kirinyaga boasts for its large production of rice. This is done from the harvesting of rice from mwea sub county where mwea irrigation scheme is located. The streams flowing through the irrigation region feed the land with water that is used for irrigation. On the other hand Machakos was issued 1.9 billion for irrigation and the project is in continuum and has improved the livelihoods of the people in this county.

Being favoured by the Delmonte farms , Kiambu county is the leading producer of pineapples in kenya. This is due to the favouring man made reservoir in the farm used for irrigations mostly during the dry season.

Aquaculture

The link besides provides the history of aquaculture and fisheries in Kenya. (History)

Garissa,Wajir and Mandera -due to their harsh climatic conditions its land isn't suitable for crop farming hence, huge parts of it is used for livestock keeping.

Aqua guarantees returns in terms of food & economy hence should be favoured in the counties that there's proximity to a water source.
The county of Homabay, Bungoma, Busia and Kisumu since their proximity to the lake they're encouraged to practise aquaculture to avoid overfishing from the Lake, also when the lake starts to experience water hyacinth.

Crop production and Farming

The county of Nakuru and Meru due to them being in the highland areas experience relief rainfall which is suitable for crop production. While the counties of Homabay, Siaya, Bungoma, Kakamega and Busia experience convectional rainfall due to proximity to Lake Victoria. Machakos and Makueni county practice irrigation and currently there are many projects ongoing in the two counties

Fishing

Due to the proximity to Lake Victoria, Homabay, Kisumu, Migori, Siaya and Busia counties are the leading counties practicing large scale fishing. Other counties like Turkana have the advantage of Lake Turkana hence practice fishing and on the coastal beaches Kilifi, Tana River, Kwale and Lamu county are close to Indian Ocean hence making it easy for them to access the ocean and practice fishing.

Livestock

Due to the harsh experienced climatic conditions in the following counties West Pokot, Narok, Mandera, Wajir, Garissa and Samburu they practice pastoralism at an extensive nature. Though some individuals have started to conform to the modern ways of living, these households are practicing crop production through farming due to the suitable climatic condition experienced for at least a short period of time in these regions.

Kitui is the leading county known for goats and sheep rearing. While Kilifi county is known for dairy cattle but mostly the indigenous breeds.

Comparison between total pop. & tot. Pop practising agriculture

The total population from a generated calculated field of the general forms of agriculture were calculated as a single entity and calculated but a county like Kiambu the total of those practicing agriculture calculated as a single field are 1.9M persons and those practicing agriculture from a calculated field are 1.1M and Makueni County has a count of 430,656 while those practicing agriculture is 1M. On the other hand, Machakos has a high number compared to total practicing the agriculture from a calculated field this is due to the fact that each agricultural practice is counted on its own even if it is being done by just one person

Livestock

Cattle and chicken are the common reared animals in kenya. At least 13,005,664 indigenious cattle and 30,320,632 indigenious chicken are reared. This translates to most kenyans preferring the keeping of indigenous breeds of livestock and chicken. This can be contributed by the fact that they're cheaper to keep and maintain hence common in most homesteads. Most homes are either rearing chicken or livestock, but only few in Nairobi and Mombasa county can actually rear both.

Education

At least 40.7% of Kenyans have attended school and are still in the learning intuitions from the entire population, 26.4% have completed learning up to the tertiary level while at least 16.3% have never attended school hence a small population is illiterate. 15.7% of the entire population are drop out and 0.9% can't recall if they attended or didn't attend school.
On the other hand, from the entire population 19.98% females and 20.67% males have attended school and are still in the learning intuitions. 12.93% females and 13.5% males left school after completion of studies, 8.9% females and 7.37% males have never been to school while 8.4% females and 7.3% males have left school before completion.

Nevertheless, 29.54% of the population are in the rural areas and 11.1% of the population are in urban areas who have attended school and are still in the learning intuitions. 13.8% of those in rural areas and 12.7% of those in urban areas left school after completion, 13.5% of those in rural areas and 2.7% of those in urban areas have never been to any learning institution hence illiterate and 11.6% of those in rural areas and 4.04% of those in urban areas are drop outs.

At least the whole 18.4% who attended pre-primary attended primary and secondary as well.Due to primary education being free there is an increase to 56.4% of students who attended primary school. 37.3% dropped out while transitioning to secondary relating to 0.7% increase of those who are able to complete from pre-primary to secondary.

Kilifi county has more drop outs in the secondary level compared to any other counties.

Housing

Rented property is the common among most kenyans, rent individual is the commonest with at least 910 thousand persons while the lowest is the rented county government property with at least 19 thousand properties being let out.

Kenyans prefer their own constructed homes to inherited properties. This is due to the fact that at least 6.9 million constructed personal homes to the number of inherited homes which is around 242 thousand. On the brighter side developed properties ready for sake have a market amongst the first time home owners who have at least purchased 146 thousand properties.

ICT

20.6 million of individuals aged 3 years and above owned a mobile phone. Out of this, 10.4 million are women while 10.2 million are men.

The data also shows that 22.6 percent of individuals aged 3 years and above used the internet while 10.4 percent used a computer.

Only 4.3% of Kenyans aged 15 and above buy goods and services online. This would translate to 1.2 million people or 2.5% of the total population. Out of this, 550,116 are women while 698,944 are men.

Turkana, Tana River and Marsabit record the lowest number of people having mobile phones. This may be attributed to their low population.

Nairobi, Kiambu and Nyeri have the highest number of mobile phone users. This may be due to the fact that there is reliable network connection in these areas.

On connectivity, a lot still needs to be done to get more people online. The rural areas are the ones that have a very limited connection, hence a need to focus on them. Digital inclusion is something Kenya will need to keep doing for the next few years.

E-commerce is still in the infancy stage. As more and more people get online, the number of people buying goods and services online will also increase. This can be promoted by various ecommerce platforms that are helping to bridge the gap between online and offline shopping.

It is important to note that the figures might have changed during the COVID season due to the government health directives. People have been forced to transact, shop and learn online to avoid contact with other people.

Amenities

Tana River County comprises several areas of forest, woodland and grassland. It is also home to the mathenge forestry explaining why it has the highest number of charcoal users.

Reducing water scarcity in Turkana County is the main objective of the local government. To achieve this, it is investing in desalination of the water in the Lotikipi aquifer. "The water in the Napuu aquifer, which serves the city of Lodwar and its surroundings, is safe to drink, but the Lotikipi aquifer is larger and the boreholes around Lake Turkana are both salty and fluoridated. More counties in Kenya could emulate them to minimize scarcity of water in their region.

BBI and Census

The BBI document was brought forth by the kenyan government to amend issues that had and have been affecting kenyans over the years. While the document was officiated after the 2019 census, we believe that it will impact the socio-economic factors of the kenyan

economy as a whole. As a result, here are some of the projections made in the document that are yet to be actualised:

- Devolution
  - The devolution of resources by the government to the kenyan citizens is projected to serve the citizens and bring the national resources closer to them.
  - With the devolution in place, there is the projected integration of schools and the delocalization aspect. This is to say, schools will have students from all over the country and the students will be able to interact and learn from their fellow students irrespective of their ethnicities. This also will allow the teachers to widen their scopes and experiences as they will not be limited to teaching in their localities. The devolution process will allow the teachers to interact with their fellows too.
- National ethos
- Lending to priority sectors
- Youth empowerment
  - The BBI aims at creating new employment opportunities for the youth through innovation and entrepreneurship. With this projection, the country is looking at a bright future with energetic and brilliant-minded individuals who will be impactful to the economic development of the county.
- Safety and Security
  - A nation whose residents have no sense of security is a failed nation. The BBI aims at providing the kenyan citizens with security in every sector, be in a safe society, food security and even the security of the resources by the government for the citizens.
- Inclusivity
- Shared prosperity
  - A country that is economically imbalanced will never prosper, tht is something that history can attest to. The BBI projects an economy free of perverse incentives that are against the innovation, growth and creation of jobs to supplement the economic system.
- Ethnic antagonism and competition
  - The document dtrafted is aimed at creating harmony between different ethnicities in the nation. It projects a higher percentage of peace in the country.
  - The aim of the BBI is to listen to the voices of all Kenyans, not just those with more socioeconomic advantages living in big cities, but also the rural

folk from the coast to Lake Turkana and Lake Victoria regions. Hopefully, the BBI report will answer what has been elusive for far too long; what is it that the people really want and how do they want it done. In the final scheme of things, the census and the BBI should promote people-centred leadership. Leaders need to appreciate that leadership is sacrifice, a privilege, a calling and not an avenue for self-aggrandisement and wealth-creation.

5. CONCLUSION

➔ There's a higher population of females owning and using mobile phones though more males are connected to the internet compared to females. (At least 9% of females are connected to the internet while 25% of males are connected to the internet)
- More sensitization on the importance of internet usage amongst the female population should be conducted
- Outreach programs to enhance and educate females on how to access the internet and use it should be conducted

➔ On average though terrazzo is the most common type of flooring material in kenya, Nairobi, Kiambu, Mombasa, Kajiado, Machakos, Uasin gishu, Kisumu and Nakuru have more than 10% of their entire population adapting to the new form of flooring, i.e. the tiles.
- More sensitization should be done in the counties of Baringo, Bomet, Bungoma and Busia should be done by the urban planning to sensitize on the importance and durability of tiles to other forms of flooring.
- Though locally available, counties like Turkana, Mandera, Wajir, Tana River, Samburu and Garissa still conform to their olden ways of using sand as their flooring material. This might be due to the lack of proper housing because they have semi-permanent due to their ways of living and consistent shifting in search of food and pasture for their animals.

➔ Despite Nairobi, Kiambu, Kajiado and Mombasa counties being the largest consumers of grid electric power, their adaptation to solar power is almost less than 0.1%.

➔ Nairobi is the largest county that uses the main sewer for its waste disposal, while counties like Laikipia, Kisumu, Nyeri, Bomet and Kakamega among other counties greatly rely on covered pit latrines for their waste management.

- These counties should ensure that more resources are put in place to ensure that their dwellers are connected to the main sewer line for waste management.
- Due to the lack of permanent structures in the counties of Turkana, Samburu, Tana River, Marsabit, Wajir and Pokot there's a large usage of open bush as their mode of waste management. These counties should be pumped in resources from the central government to ensure at least enough free public toilets are located near homesteads of their dwellers.