**MALARIA IN AFRICA DATA SCIENCE PROJECT, GROUP Z**

1. **Introduction**

The “Malaria in Africa” dataset provides a wealth of information related to malaria incidence and various factors influencing its spread in African countries. This documentation presents a comprehensive data science project that aims to analyze and understand the patterns, trends and potential factors impacting malaria prevalence in the African region.

1. **Dataset description**

The dataset contains the following columns;

* “**Country Name**”: The name of the African country
* “**Year**”: The year of data collection.
* “**Country code**”: A unique code representing each country.
* “**Incidence of malaria (per 1000 population at risk)**”: This metric calculates the rate of malaria cases in relation to the population at risk. It’s expressed as the number of malaria cases per 1000 individuals in the population who are at the risk of contracting malaria. The “population at risk” typically refers to those living in areas with active malaria transmission. This measure helps standardize and compare the malaria burden between countries with different population sizes.
* “**Malaria cases reported**”: This refers to the total number of malaria cases officially documented and reported by health authorities or healthcare facilities in a specific country or region. It is a raw count of malaria cases and does not take into consideration the size of the population or the level of risk. The reported cases may be influenced by factors such as healthcare infrastructure, surveillance systems, and reporting practices.
* “**Use of insecticide-treated bed nets (% of under-5 population)**” : This provides a measure of the percentage of children under the age of 5 who have access to and use insecticide-treated bed nets in a given country.
* “**Children with fever receiving antimalarial drugs (% of children under the age of 5 with fever)**”: This provides a measure of the percentage of children under the age of 5 receiving antimalarial drugs.
* “**Intermittent preventative treatment (IPT) of malaria in pregnancy (% of pregnant women)**”: This a measure of the percentage of pregnant women who receive intermittent preventative treatment for malaria in a given country. This indicator is specifically focused on pregnant women as they are at the risk of malaria-related complications and adverse outcomes during pregnancy.
* “**Rural population (% of total population)**”: Percentage of total population living in rural areas.
* “**Rural population growth (annual %)**”: Annual growth rate of the rural population.
* “**Urban population (% of total population)**”: Percentage of the total population living in urban areas.
* “**Urban population growth (annual %)**”: Annual growth rate of the urban population.
* “**People using at least basic drinking water services (% of population)**”: This indicator is a measure of the percentage of the total population using basic water services as well as using safely managed water services. Basic drinking water services is defined as drinking water from and improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes, protected dug wells, protected springs, and packaged or delivered water.
* “**People using at least basic drinking water services, rural (% of rural population)**”: This indicator is a measure of the percentage of the rural population using basic water services as well as using safely managed water services. Basic drinking water services is defined as drinking water from and improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes, protected dug wells, protected springs, and packaged or delivered water.
* “**People using at least basic drinking water services, urban (% of urban population)**”: This indicator is a measure of the percentage of the urban population using basic water services as well as using safely managed water services. Basic drinking water services is defined as drinking water from and improved source, provided collection time is not more than 30 minutes for a round trip. Improved water sources include piped water, boreholes, protected dug wells, protected springs, and packaged or delivered water.
* “**People using at least basic sanitation services (% of population)**”: This indicator is a measure of the percentage of the total population using at least basic sanitation services, that is, improved sanitation facilities that are not shared with other households. It encompasses both people using basic sanitation services as well as those using safely managed sanitation services. Improved sanitation facilities include flush toilets, piped sewer systems, septic tanks or pit latrines, ventilated or improved pit latrines, compositing toilets or pit latrines with slabs.
* “**People using at least basic sanitation services, rural (% of rural population)**”: This indicator is a measure of the percentage of the rural population using at least basic sanitation services, that is, improved sanitation facilities that are not shared with other households. It encompasses both people using basic sanitation services as well as those using safely managed sanitation services. Improved sanitation facilities include flush toilets, piped sewer systems, septic tanks or pit latrines, ventilated or improved pit latrines, compositing toilets or pit latrines with slabs.
* “**People using at least basic sanitation services, urban (% of urban population)**”: This indicator is a measure of the percentage of the urban population using at least basic sanitation services, that is, improved sanitation facilities that are not shared with other households. It encompasses both people using basic sanitation services as well as those using safely managed sanitation services. Improved sanitation facilities include flush toilets, piped sewer systems, septic tanks or pit latrines, ventilated or improved pit latrines, compositing toilets or pit latrines with slabs.
* “**Latitude**”: The latitude coordinate of the country’s geographical location.
* “**Longitude**”: The longitude coordinate of the country’s geographical location.
* “**Geometry**”: Geographical geometry data for mapping purposes.

1. **Project Objectives**

The main objectives of this data science project are as follows;

* Identify countries with the highest and lowest malaria incidence rates, aiding in targeting interventions and resources.
* Compare the burden of malaria across different African countries, helping prioritize regions for preventative measures and healthcare infrastructure.
* Highlight countries with a high number of reported cases, indicating areas requiring immediate attention
* Assess the effectiveness of malaria prevention campaigns and programs promoting bed net usage among children under 5.
* Identify countries with low bed net coverage, targeting them for interventions to increase usage and reduce malaria transmission.
* Evaluate the efficiency of healthcare systems in diagnosing and treating malaria in children with fever.
* Identify regions with low rates of antimalarial drug administration, indicating potential gaps in healthcare services.
* Examine the coverage and implementation of Intermittent preventative treatment (IPT) for pregnant women, crucial for preventing malaria-related complications during pregnancy.
* Identify regions with lower IPT coverage, requiring targeted efforts to improve maternal and child health outcomes.
* Understand the distribution of malaria incidence levels across African countries, identifying patterns and potential outliers.
* Determine the key factors contributing to the high number of incidences in the country with the highest incidences.
* Evaluate the effectiveness of the existing malaria control and prevention interventions in the country with the highest incidences.
* Investigate the prevalence of drug resistant malaria strains in the country with the highest incidences.
* Assess the level of insecticide resistance among malaria vectors in the country with the highest incidences.
* Examine the capacity of the health care systems in the country with the highest incidence.
* Analyze trends in malaria incidence over time in the country with the highest incidence to assess the progress of malaria control efforts
* Explore key factors that have contributed to the low malaria incidences in countries with the lowest incidences.
* Assessing the impact of geographical location on Malaria Incidence.
* Understanding the relationship between population distribution and malaria incidence.
* Identify any significant correlations between different numeric variables related to malaria
* Provide a visual representation of the malaria burden across African countries, highlighting malaria hotspots.
* Analyze how access to basic sanitation and drinking water services influences malaria incidence.
* Evaluate the availability of basic sanitation and drinking water services in the country with the highest incidences.
* Assess healthcare infrastructure.

1. **Results**
2. **Discussion**
3. **Conclusion**