MAKERERE UNIVERSITY.

COLLEGE OF COMPUTING AND INFORMATION SCIENCES.

BSE 2302

DATA SCIENCE PROJECT.

MALARIA IN AFRICA.

GROUP B

STUDENT NAME	STUDENT REG NO	STUDENT NO
TUMUHAISE MICHAEL	21/U/19954/PS	2100719954
WALUSIMBI SILVER	21/U/12608/PS	2100712608
NANSERA ENA	22/U/06869/EVE	2100706869
NANGOSHA TREVOR	21/U/1591	210701591
NATUKUNDA CATHERINE	21/U/05595/EVE	2100705595

1. INTRODUCTION TO THE DATASET.

The prevalence of malaria in Africa has been a longstanding public health challenge, affecting millions of individuals every year. As part of the effort to understand and address this critical issue, we have acquired a comprehensive dataset named "Malaria in Africa." This dataset encompasses various socio-economic, demographic, and health-related indicators for multiple African countries, spanning different years. Our objective is to conduct a thorough data cleaning and processing analysis to gain valuable insights into the factors associated with malaria incidence and its impact on communities across the continent.

The dataset comprises essential information, such as the incidence of malaria per 1,000 population at risk, reported malaria cases, and the use of insecticide-treated bed nets among the under-5 population. Additionally, it contains indicators related to the administration of antimalarial drugs to children under the age of 5 with fever and intermittent preventive treatment for pregnant women. We also have data on the usage of safely managed drinking water services and sanitation facilities, both in rural and urban settings.

2. KEY OBJECTIVES FOR ANALYZING THE DATA.

- Analyzing the reported number of malaria cases to understand the burden of malaria in different countries and over time.
- Exploring the trends and patterns of malaria incidence by identifying countries with high and low malaria incidence rates.
- Analyzing Infrastructure for healthcare and disease prevention.
- Analyzing the distribution of rural and urban population and seeing its relationship with malaria cases reported.
- Visualizing the relationship between population and malaria cases reported.
- Visualizing the geographical distribution of malaria incidence.
- Analyzing bed net usage for children under 5.
- Evaluation of antimalarial drug administration for children under 5.

3. FEATURES ANALYZED.

- Incidence of malaria (per 1000 population at risk): Analyzing this column is to understand the number of people at risk of getting malaria in each country.
- Malaria cases reported: Analyzing was to understand the number of malaria cases per country in the particular year.
- People using at least basic sanitation services: Analyzing this column was to understand the percentage of the population using basic sanitation services.
- People using at least basic drinking water services: Analyzing the column was to understand the percentage of the population not having access to basic drinking water services.

- Rural population (%of total population) and Urban population(%of total population): Analyzing the columns was to see the relationship between the population distribution and the malaria cases reported.
- Geometry,Latitude and Longitude: Analyzing the columns was to come up with a map of the malaria incidences for each particular country.
- Usage of insecticide treated bed nets for children under 5: Analyzing the column was to understand the percentage of children under 5 using insecticide treated bed nets.
- Drug administration for children under 5: Analyzing the column was to understand the percentage of children receiving anti-malarial drugs

4. PROCESS AND TECHNIQUES.

a. Data Cleaning

Data cleaning is the process of identifying and correcting errors, inconsistencies, and inaccuracies in a dataset to ensure that it is accurate, reliable, and suitable for analysis. Data cleaning is a crucial step in the data preparation phase of any data science or data analysis project. It helps improve the quality of the data, reduces the risk of making erroneous conclusions, and ensures that the results obtained from the analysis are more meaningful and trustworthy.

Data Cleaning begins with loading the Malaria in Africa dataset in order to show the different columns that are in it.

Before cleaning

	Name		Code	population at risk)	reported	under-5 population)	children under age 5 with fever)	pregnancy (% of pregnant women)	water services (% of population)	services, rural (% of rural population)	services, urban (% of urban population)	services (% of population)	services, rural (% of rural population)	
0	Algeria	2007	DZA	0.01	26.0	NaN	NaN	NaN	NaN	NaN	NaN	18.24	19.96	
1	Angola	2007	AGO	286.72	1533485.0	18.0	29.8	1.5	NaN	NaN	NaN	NaN	NaN	
2	Benin	2007	BEN	480.24	0.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
3	Botswana	2007	BWA	1.03	390.0	NaN	NaN	NaN	NaN	NaN	83.96	NaN	NaN	
4	Burkina Faso	2007	BFA	503.80	44246.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	
589	Togo	2017	TGO	278.20	1755577.0	69.7	31.1	41.7	NaN	NaN	NaN	NaN	NaN	
590	Tunisia	2017	TUN	NaN	NaN	NaN	NaN	NaN	92.66	NaN	NaN	78.12	NaN	
591	Uganda	2017	UGA	336.76	11667831.0	NaN	NaN	NaN	7.07	4.46	15.70	NaN	NaN	
592	Zambia	2017	ZMB	160.05	5505639.0	NaN	NaN	NaN	NaN	NaN	46.25	NaN	NaN	
593	Zimbabwe	2017	ZWE	108.55	467508.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	

594 rows × 27 columns

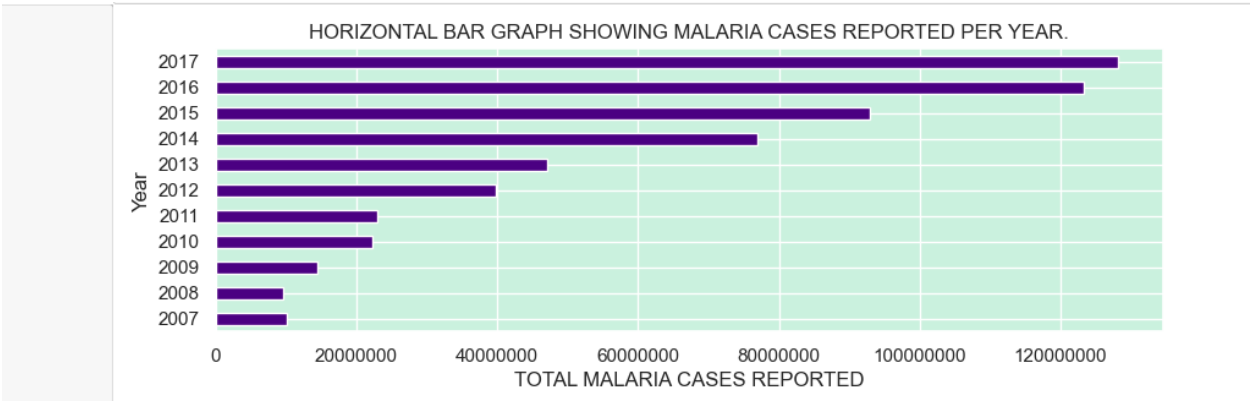
After cleaning

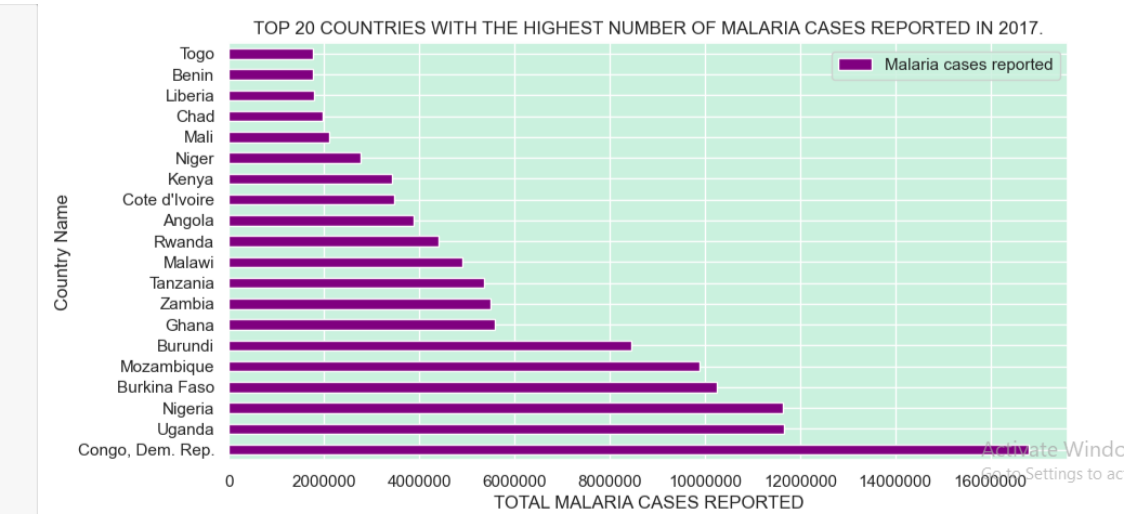
	Country Name	Year	Country Code	Malaria_Incidence_per1000_at-risk	Malaria cases reported	Use_of_treated_bed_nets(% under 5)	Children_with_fever_receiving_antimalarial_drugs(% under 5)	Preventi
0	Algeria	2007	DZA	0.01	26	0.0	0.0	
1	Angola	2007	AGO	286.72	1533485	18.0	29.8	
2	Benin	2007	BEN	480.24	0	0.0	0.0	
3	Botswana	2007	BWA	1.03	390	0.0	0.0	
4	Burkina Faso	2007	BFA	503.80	44246	0.0	0.0	
...	...	...	...	...	...	...	...	
589	Togo	2017	TGO	278.20	1755577	69.7	31.1	
590	Tunisia	2017	TUN	0.00	0	0.0	0.0	
591	Uganda	2017	UGA	336.76	11667831	0.0	0.0	
592	Zambia	2017	ZMB	160.05	5505639	0.0	0.0	
593	Zimbabwe	2017	ZWE	108.55	467508	0.0	0.0	

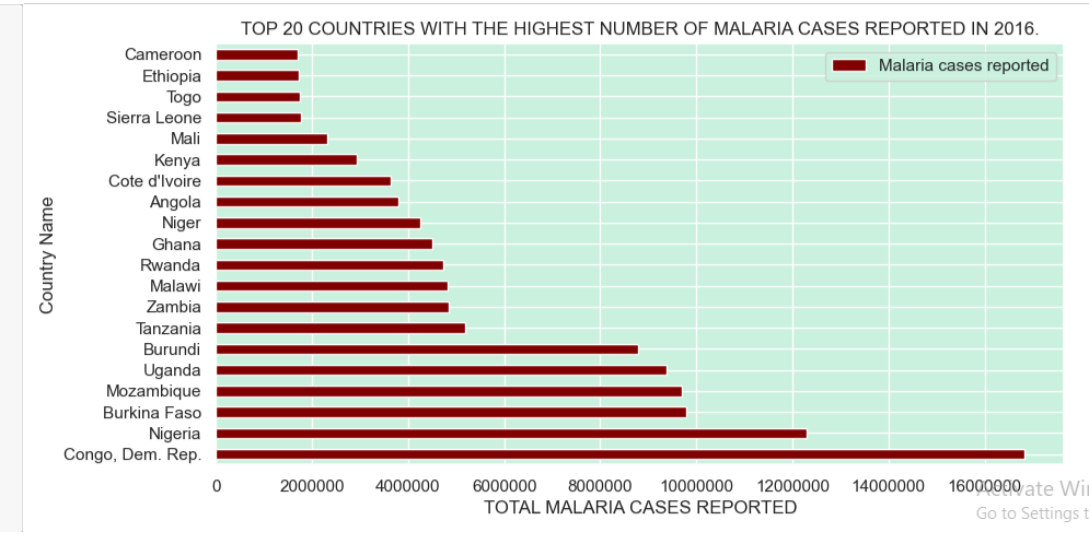
594 rows × 28 columns

b. Data Visualization

Analyzing the reported number of malaria cases to understand the burden of malaria in different countries and over time: The MalariaAfricaDataset has data from 2007 to 2017, for the different countries in Africa, and this objective is to show the number of cases that were recorded during that time, and to pick out the year with the largest number of cases and lowest.

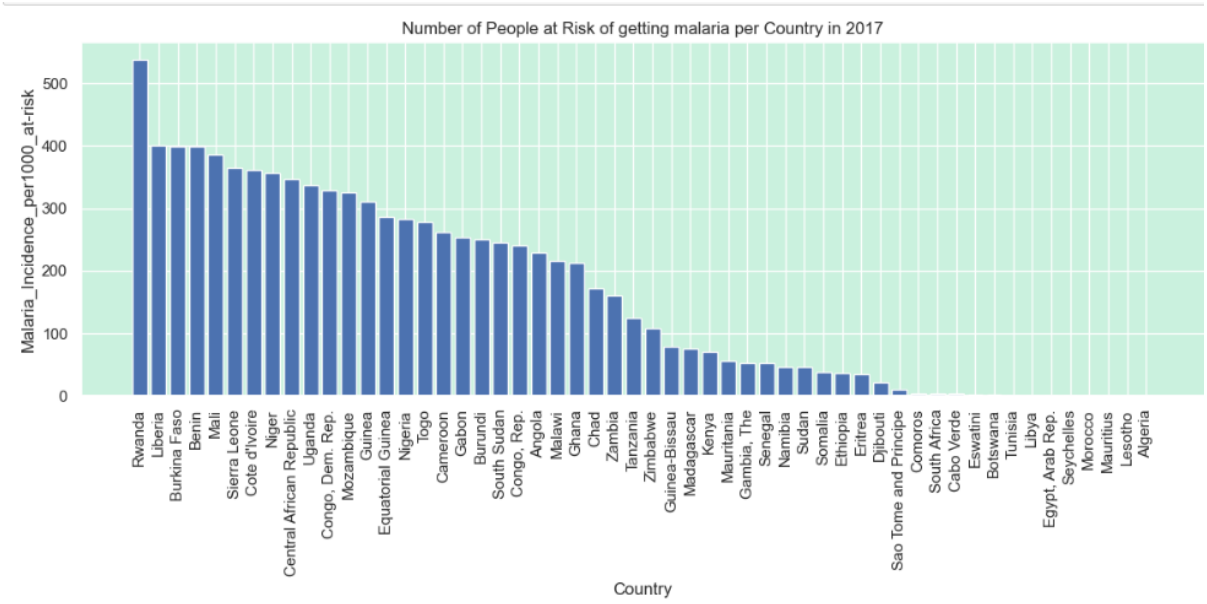




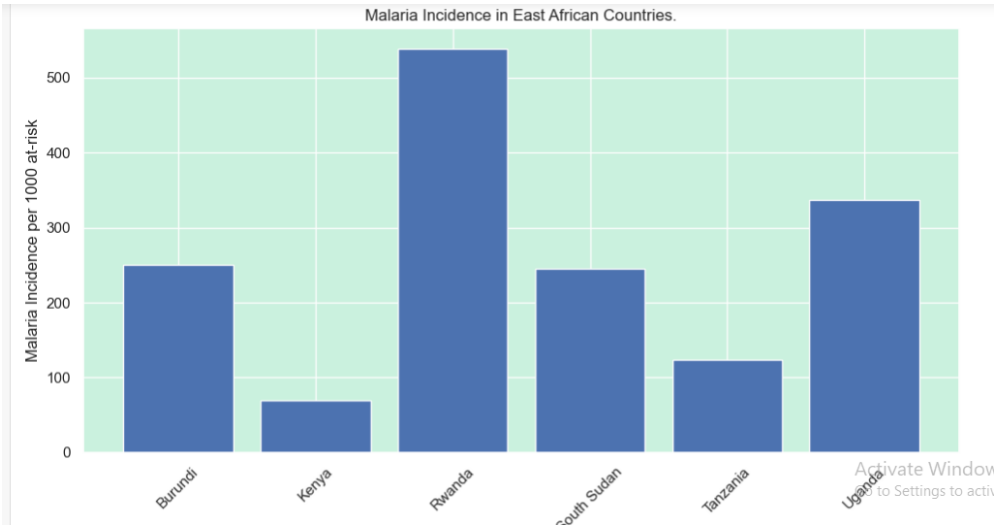


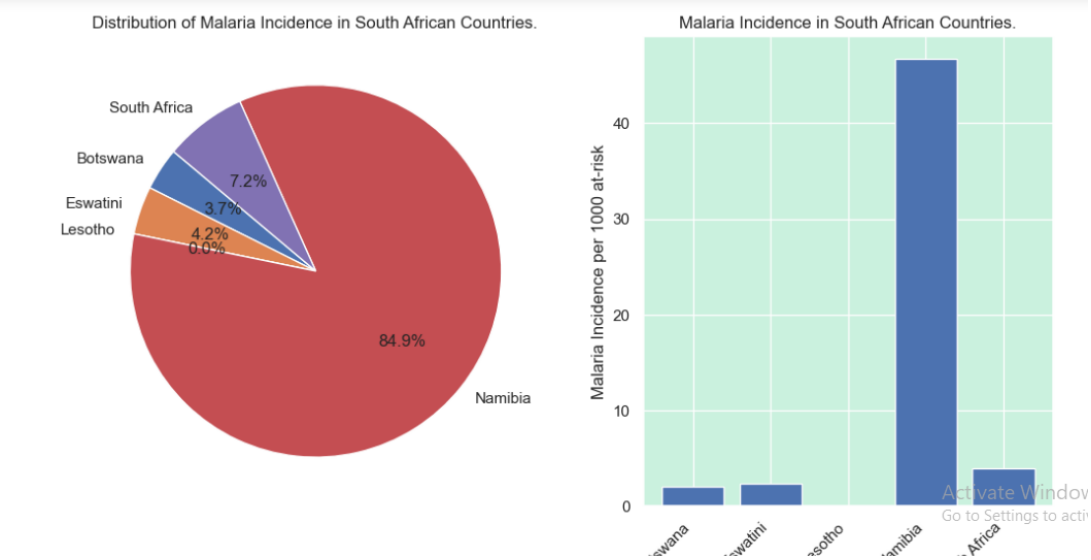
The above horizontal bar graphs show the countries with the highest Malaria cases in 2017 and 2016.

Explore the trends and patterns of malaria incidence by identifying countries with high and low malaria incidence rates: This objective focuses on the number of Malaria Incidence rates column to show the number of incidences in the different countries in the year of 2017

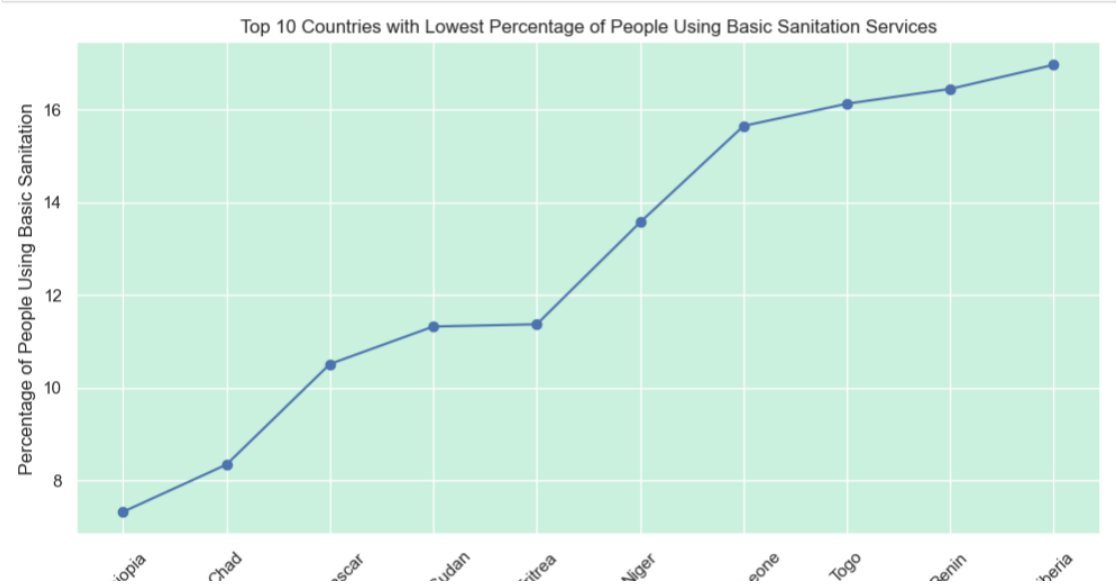


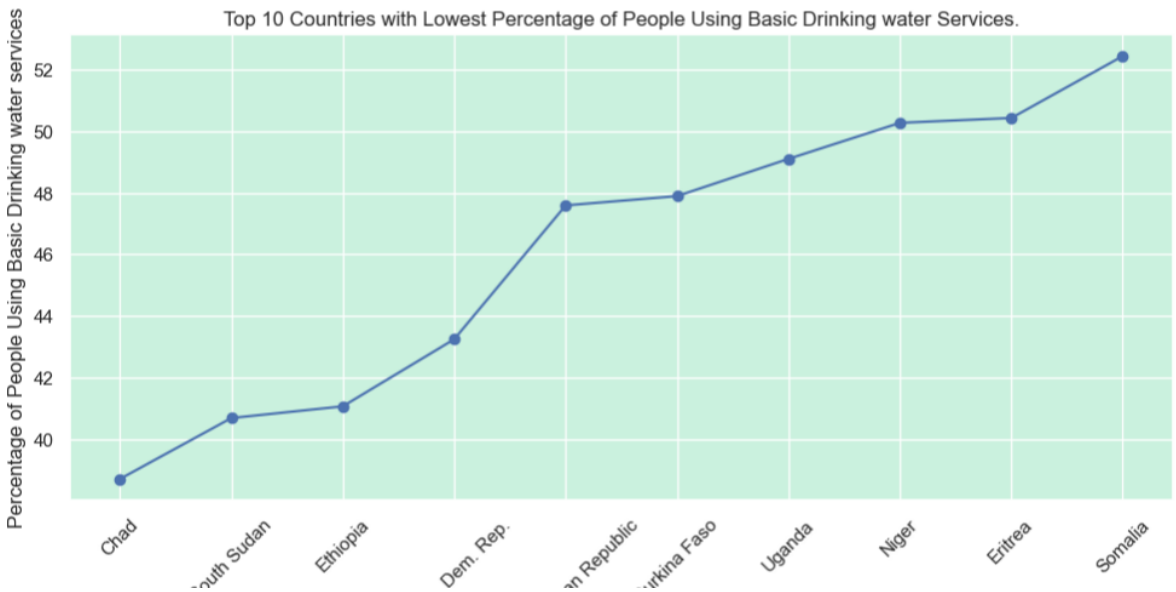
Further visualizing the malaria incidence for the different sections of Africa.

1. 

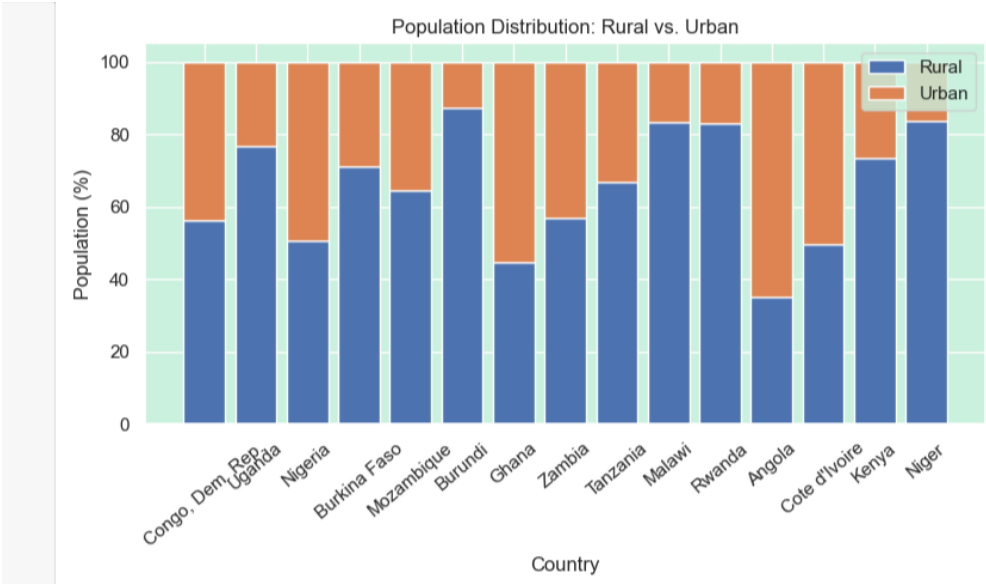


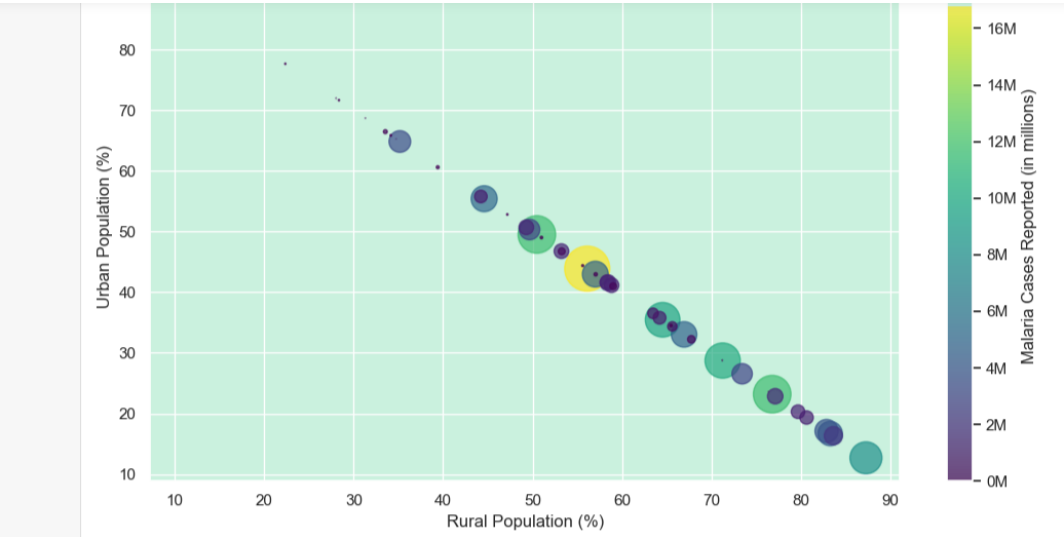
Analyzing Infrastructure for healthcare and disease prevention. This is done by analyzing the sanitation and drinking water column which is “People_using_basic_sanitation(% Total), People_using_basic_drinking_water(% Total)” and it is accompanied with the “Country Name” column to show the countries with the highest and lowest levels of sanitation in a specific year.





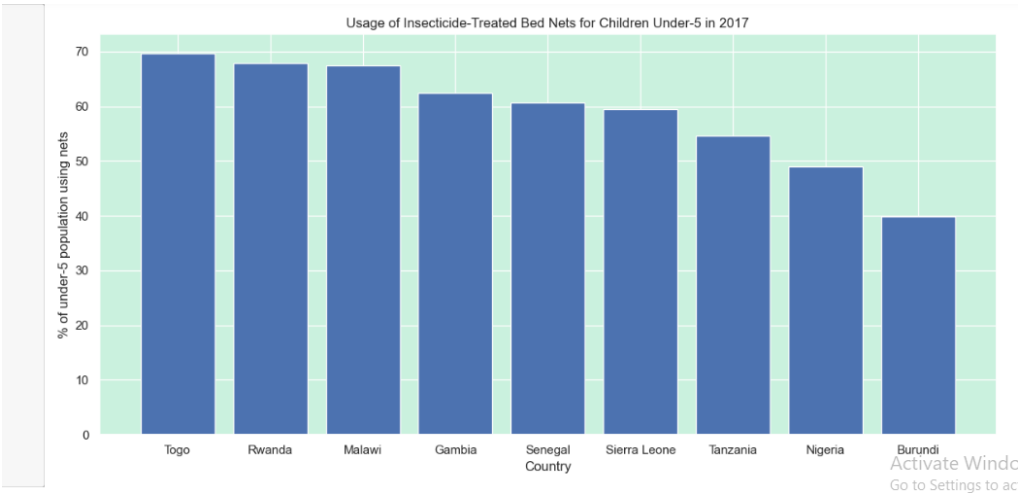
Analyzing the distribution of rural and urban population and seeing its relationship with malaria cases reported: This is to show which region is affected most by the malaria cases, and to compare the population between the two, and to do this the following columns are analyzed "'Rural population (% of total population)', 'Urban population (% of total population)'" accompanied with "Country Name" column to show both the rural and urban population in different countries in a specific year.

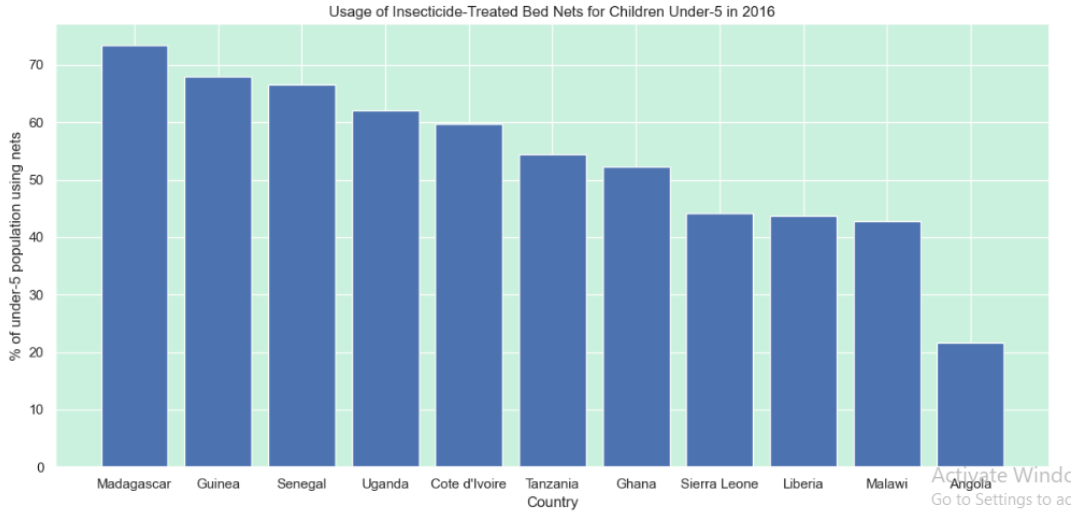




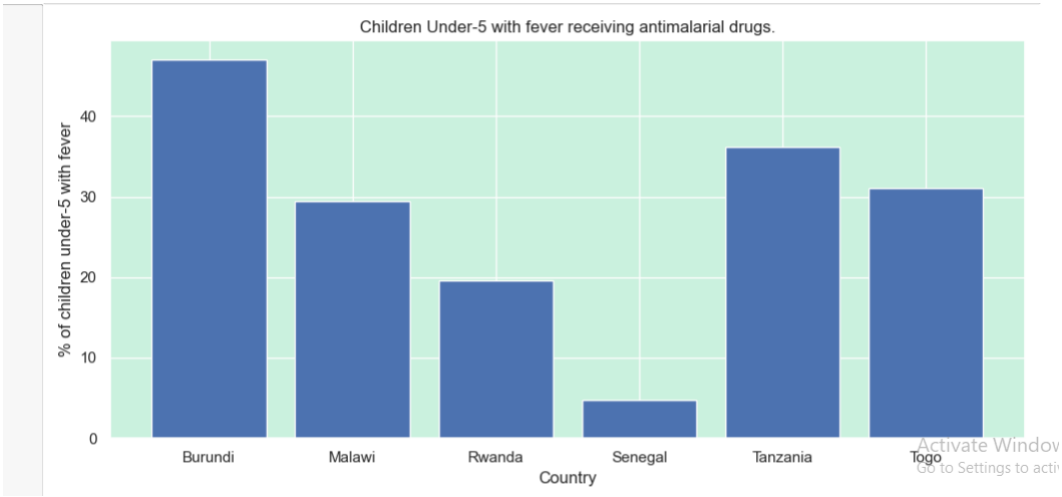
Visualizing the geographical distribution of malaria incidence.

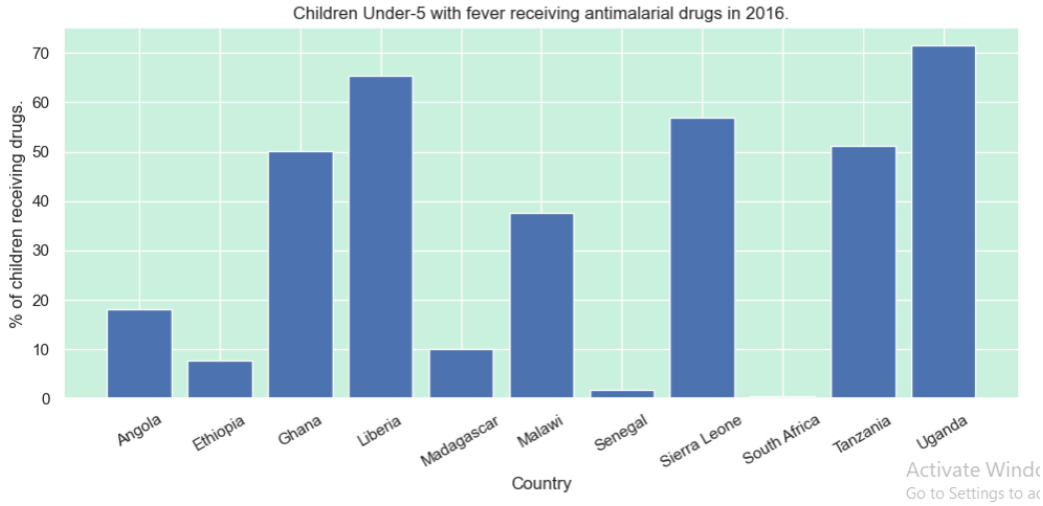
Analyzing bed net usage for children under 5: This is to analyze the Use_of_treated_bed_nets(% under 5) column and get their visualizations for the different countries.





Evaluation of antimalarial drug administration for children under 5: This is done to Filter out % of children under-5 with fever receiving anti-malarial drugs per country and check the percentage of children receiving Antimalarial Drugs.





Conclusion.

There is generally an increase in the number of cases for all the years.

Analysis shows that there was a decrease in usage of insecticide treated bed nets in 2017 as was in 2016. This could have been the spark of the higher malaria cases in 2017.

Analysis shows that administration of antimalarial drugs in 2017 is a bit lower than that in 2016. This explains why malaria cases in 2017 were higher, children with malaria spread the disease to other children since they were not being treated.

In conclusion, the data cleaning and processing journey undertaken in this analysis has transformed the "Malaria in Africa" dataset into a valuable resource for understanding and addressing the challenges posed by malaria in the African context. The enhanced data quality, combined with the potential for impactful analysis, positions this dataset as a catalyst for evidence-based actions to reduce malaria's burden and improve the well-being of populations across the African continent.