# Codveda Level1(Basic) Task 3 Exploratory Data Analysis (EDA) Report

This report presents the findings from the Exploratory Data Analysis (EDA) performed on the malaria dataset. The purpose of EDA is to understand the underlying structure of the data, identify patterns, detect anomalies, and highlight important relationships that can guide further modeling and decision-making.

The analysis focused on four major aspects: **descriptive statistics, regional distributions, correlation analysis, and temporal trends**.

## 1. Descriptive Statistics

To begin, descriptive statistics were computed using the .describe() function. This provided a quick overview of measures such as **mean, median, minimum, maximum, standard deviation, and quartiles**.

* **Malaria Cases (Min, Median, Max):** The statistics showed significant variation across regions and years. Some areas recorded very low case numbers, while others had extremely high values, pointing to disparities in malaria burden.
* **Malaria Deaths (Min, Median, Max):** Similar trends were observed in deaths, with some regions reporting close to zero fatalities, while others recorded very high mortality counts.

The wide range in both cases and deaths highlights the unequal impact of malaria globally. Outliers (regions with exceptionally high cases or deaths) suggest hotspots where interventions should be prioritized.

## 2. WHO Region Distribution

The dataset was grouped by **WHO Region** to compute frequency counts of cases and deaths.

* Regions like **Africa** showed disproportionately higher malaria cases and deaths compared to other regions such as the Americas or Europe.
* This aligns with global health statistics, where sub-Saharan Africa is known to bear the highest malaria burden.

Malaria is not evenly distributed worldwide. Certain regions carry the majority of cases and deaths, emphasizing the need for region-specific strategies and policies.

## 3. Correlation Analysis

A **correlation matrix** was constructed for key features such as:

* Median, minimum, and maximum number of malaria cases.
* Median, minimum, and maximum number of malaria deaths.
* Strong positive correlations were observed between cases and deaths. For instance, countries or regions with high case counts generally reported higher death counts.
* The correlation matrix was visualized using a **heatmap**, making these relationships easier to interpret.

This suggests that **reducing malaria cases directly reduces deaths**. Effective prevention and treatment strategies that lower incidence rates will consequently lower mortality.

## 4. Regional Analysis

The dataset was grouped by **WHO Region** to calculate the **total number of malaria cases and deaths**.

* Africa accounted for the majority of global malaria cases and deaths.
* Other regions such as Southeast Asia also showed significant numbers, while regions like Europe and the Americas reported very few cases.

Regional aggregation highlights **where resources and interventions should be concentrated**. It also provides a comparative view of how different regions are coping with malaria control.

## 5. Temporal Trends

Cases and deaths were aggregated by **year** to identify patterns over time.

* Line plots showed whether malaria cases and deaths have been **increasing, decreasing, or fluctuating**.
* For some years, downward trends indicated progress in malaria control programs. However, in certain periods, spikes in cases and deaths were observed — possibly due to outbreaks, reporting changes, or health system challenges.

Monitoring malaria trends over time is crucial to evaluate the effectiveness of intervention strategies. Consistent decreases point to successful policies, while increases highlight the need for renewed efforts.

## 6. Heatmap Visualization

To complement correlation analysis, a **heatmap** was used for visual interpretation.

* Darker shades in the heatmap represented stronger correlations between malaria cases and deaths.
* This visual approach provided a quick way to spot relationships that may not be immediately obvious from raw numbers alone.

Visualization confirms that the data contains **strong linear relationships** among case and death variables, strengthening the evidence from the statistical analysis.

# ****Summary****

* **Malaria is highly uneven across regions**, with Africa carrying the heaviest burden.
* **Cases and deaths are strongly correlated** — controlling incidence directly impacts mortality.
* **Trends over time** show both progress and setbacks, highlighting the dynamic nature of malaria control.
* **EDA provided actionable insights** that can guide future steps in modeling and decision-making, such as predicting malaria trends, identifying at-risk regions, or evaluating the effectiveness of interventions.