PROJECT PROPOSAL

- **TITLE**: Temporal Dynamics and Mortality Burden of Lassa Fever in Nigeria: A Descriptive and Inferential Time-Series Analysis.
- **EXECUTIVE SUMMARY**: This study assessed weekly surveillance data for Lassa fever to understand trends, seasonality, and case fatality rates. Analyses indicate stable weekly case counts, strong seasonal variation by month, and consistent mortality over the years.
- **PROBLEM STATEMENT**: In Nigeria, Lassa fever remains a recurring public health concern with outbreaks reported annually. Despite ongoing surveillance by the Nigeria Centre for Disease Control (NCDC), there is a limited statistical assessment of temporal trends, seasonal patterns, and mortality burden using longitudinal surveillance data. Understanding whether Lassa fever incidence exhibits significant seasonal variation, long-term trends, or changes in case fatality rate (CFR) is essential for improving preparedness, outbreak forecasting, and resource allocation. This study seeks to statistically evaluate the temporal dynamics of Lassa fever in Nigeria using weekly surveillance data.
- **ANALYTICAL OBJECTIVES:** This analysis aims to;
    - Describe the weekly trend of confirmed Lassa fever cases over time.
    - Assess whether a statistically significant long-term trend exists.
    - Evaluate seasonal variation in incidence across months.
    - Estimate and analyze the Case Fatality Rate (CFR) over time.
    - Determine whether mortality burden differs significantly across years.
    - Present findings in a manner understandable to non-technical audiences.
- **DESCRIPTION OF DATA SOURCE**: The dataset consists of weekly Lassa fever surveillance data compiled from reports published by the NCDC (Nigeria Centre for Disease Control). The key variables include;
    - Week start date – Start date of epidemiological week.
    - Week end date – End date of epidemiological week.
    - Epi year (epidemiology year) – Year of reporting.
    - Epi week (epidemiology week) – Epidemiological week number.
    - Suspected cases – Reported suspected cases.
    - Confirmed cases – Laboratory-confirmed cases.
    - Probable cases – Probable cases (no lab confirmation).

- Deaths – Number of deaths reported.
- Time structure; weekly observations, multi-year time series, National-level aggregation.

This dataset represents officially reported surveillance data and reflects confirmed epidemiological records during the specified time period.

- **PROPOSED METRICS:** The following metrics will be computed to assess the public health burden and temporal behavior of Lassa fever;
    - Descriptive Metrics such as;
        - Mean Weekly Confirmed Cases
        - Median Weekly Cases
        - Standard Deviation of Weekly Cases
        - Annual Total Confirmed Cases
        - Annual Total Deaths
        - Peak Outbreak Week
        - Rolling 4-week Moving Average
    - Derived Public Health Metrics
        - Case Fatality Rate (CFR) = Deaths/Confirmed Cases * 100
        - Suspected-to-Confirmed Ratio (indicator of surveillance confirmation efficiency)
- Proposed Statistical Methods:
    - Descriptive Analysis
        - Time-series line plots
        - Rolling mean visualization
        - Boxplots by month
        - Summary statistics
    - Inferential Analysis
        - Independent variable: Time index
        - Dependent variable: Confirmed cases
        - Hypothesis: Null – No significant trend over time

          Alternative – Significant increasing or decreasing trend

        - Seasonal Variation – One-Way ANOVA
        - Grouping variable: Month
        - Hypothesis: Null – Mean weekly cases are equal across months

Alternative – At least one month differs significantly.

- Mortality Burden Variation – ANOVA
    - Compare CFR across years
    - Assess whether the mortality rate differs significantly over time

- **RISKS AND CONSTRAINTS:**
    - Data Limitations – Aggregated national-level data (no state-level disaggregation)
        - Possible reporting delays or underreporting
        - Surveillance bias in suspected case reporting
    - Analytical Constraints –
        - Weekly aggregation may mask daily variation
        - Linear regression assumes linearity and independence
        - ANOVA assumes normality and homogeneity of variance
        - External confounders are not included (climate, intervention policies)
    - Scope Constraints –
        - Study does not include causal modeling
        - No adjustment for population growth
        - No predictive epidemiological modeling

REPORT

DATA CLEANING AND METHODOLOGY

Data Cleaning Steps;

- Converted date columns to datetime.
- No missing value was observed; no imputation required.
- Created derived metrics: rolling mean, CFR.
- Verified ranges, consistency, and data types.

Methodology;

- Descriptive Analysis;
    - Line plots were used for weekly confirmed cases and rolling mean.
    - A bar chart was used for the annual summary.
    - A box plot was used for the monthly summary.
    - A line chart was used to show the CFR trend over the years (yearly mortality assessment).

- Inferential Analysis;
  - Linear Regression: Weekly confirmed cases vs. time index to assess trend.
  - ANOVA: This was used for month-to-month variation to test seasonality and year-to-year variation in CFR.
- Statistical Justification;
  - Linear regression tests whether the trend is statistically significant.
  - ANOVA appropriate for comparing means across multiple groups (months/years).

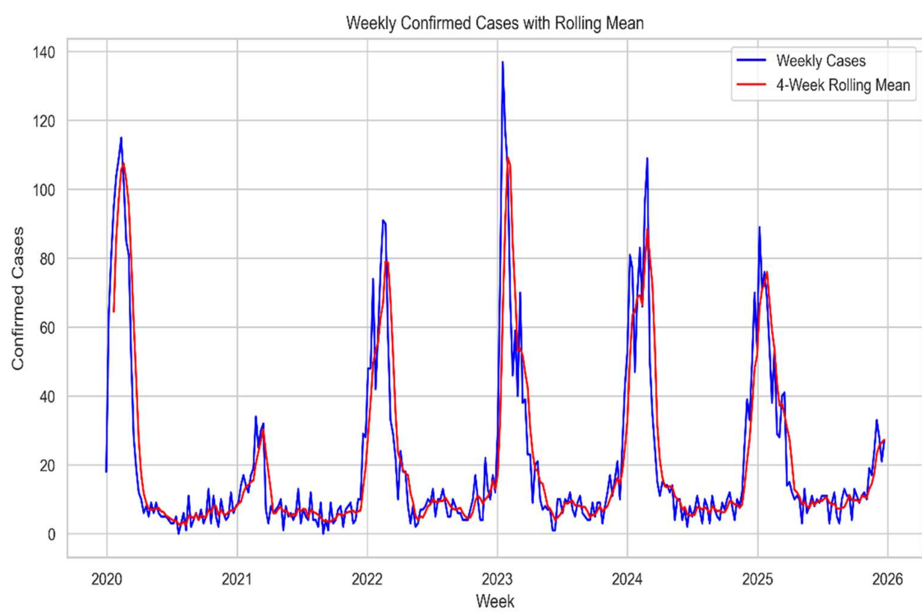
Data Preparation;

- Computed rolling mean to smooth fluctuations.
- Computed Case Fatality Rate (CFR);

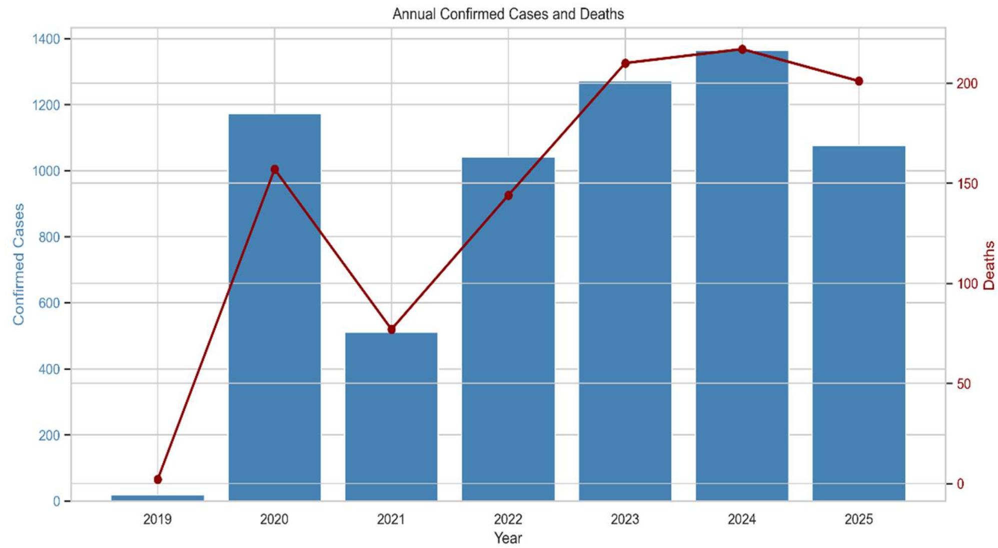
  CFR (%) = Deaths/Confirmed Cases * 100
- Ensured all metrics are numeric and consistent with epidemiological definitions.

RESULTS

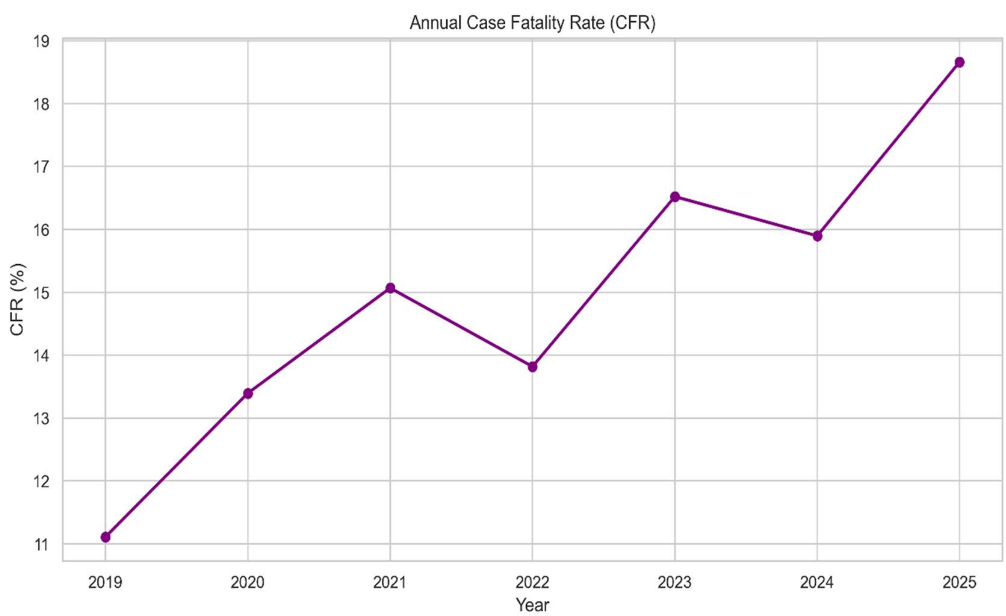
- Descriptive Analysis
  - Weekly Cases & Rolling Mean:



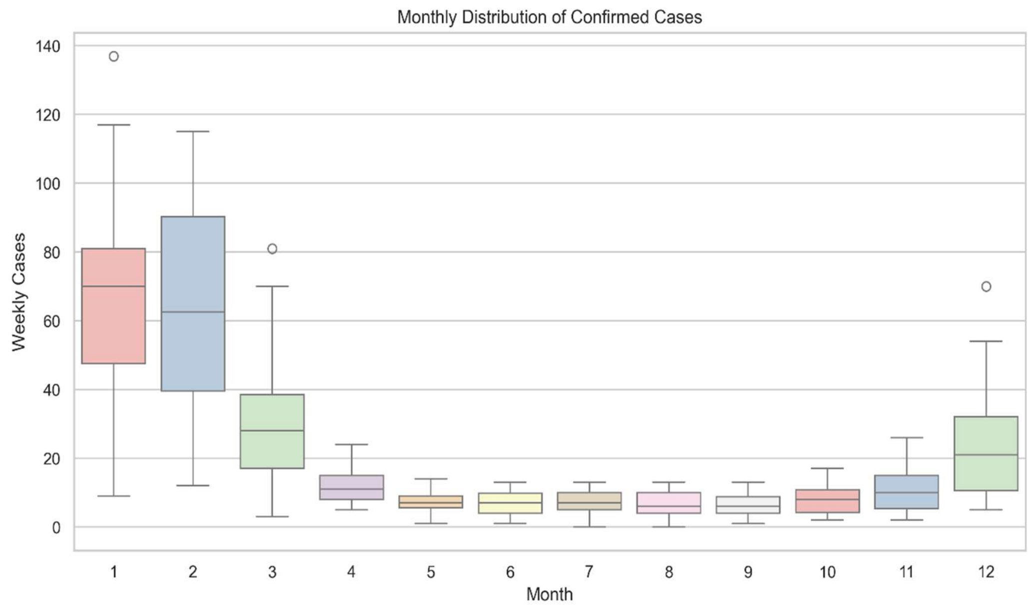
  - Insight - Weekly confirmed cases fluctuate over time without a statistically significant long-term upward or downward trend.
  - Annual Summary with Deaths:



- Insight – This highlights the mortality burden alongside cases.
- Annual CFR:



- Insight – Visual inspection of annual CFR values shows noticeable fluctuations across years.
- Monthly Summary:



- Insight – The monthly distribution of confirmed cases reveals a clear seasonal pattern. The first two months exhibit the highest median case counts and greatest variability, with Month 2 slightly exceeding Month 1, indicating heightened transmission during the period.

- Inferential Analysis:

Analysis	Test	Statistic	p-value	Interpretation
Trend (Linear Regression)	Linear Regression	-0.00036551	0.981881	Increasing/decreasing trend
Seasonality (ANOVA - monthly)	ANOVA	54.8931	2.47037e-65	Some months differ significantly
CFR Variation (ANOVA - yearly)	ANOVA	0.841305	0.538788	CFR differs across years

INTERPRETATION

- Linear regression shows no meaningful change in weekly confirmed cases over time (slope = -0.00036, p = 0.981), indicating stability.

- ANOVA confirms strong seasonal variation in cases across months ($F = 54.893$, $p < 0.001$), suggesting predictable outbreak patterns.
- Yearly CFR did not vary significantly ($F = 0.841$, $p = 0.539$), indicating consistent mortality. This suggests that while CFR appears to vary from year to year visually, the differences are not strong enough to conclude that mortality risk has meaningfully changed over time.

KEY INSIGHTS/SUMMARY OF ANALYSIS

- The summary of the study was to assess temporal trends, seasonality, and mortality burden of reported cases using weekly surveillance data.
- The dataset was carefully checked for missing values, and no imputation was required.
- Derived metrics, including rolling mean and case fatality rate, were computed to capture key epidemiological patterns.
- Visualizations of weekly confirmed cases with rolling mean provided a clear picture of the outbreak trajectory.
- Linear regression analysis indicated that weekly confirmed cases remained essentially stable over time (slope = -0.00036, $p = 0.981$), showing no significant upward or downward trend.
- ANOVA tests revealed strong seasonal variation ($F = 54.893$, $p = 2.47e\text{-}65$), indicating that certain months consistently experienced higher or lower-case counts.
- Yearly, CFR showed no statistically significant variation ($F = 0.841$, $p = 0.539$), suggesting that the mortality burden remained relatively stable throughout the observed period.
- All statistical analyses were chosen to address the research questions: linear regression for trend assessment, ANOVA for comparing groups, and calculation of CFR to quantify severity.
- Charts were directly aligned with analytical claims, enabling the results to be interpreted intuitively.
- Overall, the analysis demonstrates that while confirmed cases fluctuate seasonally, the general trend is stable and the case fatality rate remains consistent.

CONCLUSION & RECOMMENDATIONS

- Weekly confirmed Lassa fever cases remained stable over the observed period, with strong seasonal peaks and consistent mortality.
- Public health authorities can use seasonal patterns for preparing outbreak response, allocating resources, and timing public awareness campaigns.
- Monitoring should continue to detect any changes in trend or severity. Future studies may incorporate more sophisticated time-series models to capture potential autocorrelation and outbreak spikes.