Dataset Link

[Nigerian Road Traffic Crashes (2020- 2024)](#)

Dataset Features

- Quarter
- State
- Total_Crashes
- Num_Injured
- Num_Killed
- Total_Vehicles_Involved
- SPV (Speed Violation)
- DAD (Dangerous Driving)
- PWR (Poor Weather/Road Condition)
- FTQ (Faulty Tyre/Vehicle)
- Other_Factors

Dataset outline

- Dataset Name : Nigerian Road Traffic Crashes (2020-2024)
- Dataset Description : This is data about car crashes in Nigeria over several years. For each state and every few months, it shows:
  - When did it happen ?
  - Where did it happen ?

- How many crashes were there ?
- How many people were hurt ?
- How many people died ?
- How many cars were involved ?
- Why did it happen ?
  - Too fast (speeding)
  - Bad driving
  - Bad weather or bad roads
  - Bad tires or broken car
  - Other reasons
- Number of Records : 518
- Feature Categories :
  - Categorical Features (Label/Type)
    - Quarter
    - State
  - Numerical Features
    - Total_Crashes
    - Num_Injured
    - Num_Killed
    - Total_Vehicles_Involved
    - SPV (Speed Violation)
    - DAD (Dangerous Driving)
    - PWR (Poor Weather/Road)
    - FTQ (Faulty Tyre/Vehicle)
    - Other_Factors
- Target Variable : Num_Killed

What models have been applied in online published research papers based on this dataset?

- Poisson/Negative Binomial Regression – for crash count prediction
- Random Forest/XGBoost – to find key causes (e.g., speeding vs. bad roads)
- Time Series (ARIMA) – for trend forecasting
- Clustering (K-means) – to group high-risk states
- Logistic Regression – for fatal vs. non-fatal classification

Which advanced type model is applicable in these dataset for better results to published thesis papers? :

For a thesis paper, use Bayesian Spatiotemporal Negative Binomial Model. Why:

- Handles count data (crashes are counts)
- Accounts for time (quarters) and space (states)
- Includes uncertainty - gives credible intervals
- Identifies hotspots - finds high-risk states/periods

Result will be

- Maps of crash risk over time.
- Probability that speeding causes more deaths
- Forecasts with uncertainty bounds

Tools: Use R-INLA or PyMC in Python.