

## Data Sampling Strategy

- Time span
  - 2022-2023
    - If we include more tweets before 2022, annotators may lack information about the situation in which the tweets were written.
    - We can get enough data
- Variety of topics (Events): specify events that have escalated the spread of hate and abusive tweets. This might vary from region to region.
- Events:
  - Zamfara massacre where over 200 people were killed between Jan 4 to Jan 6, 2022
  - ASUU strike, which lasted for 8 months, from Feb 14 - 14 Oct, 2022.
  - Abuja-Kaduna train attack on March 28, 2022.
  - Primary elections in 2022.
  - Owo church attack where 40 people were killed in June, 2022.
  - Kuje Prison attack on July 6, 2022.
  - *Abba Kyari arrest , 14th Feb, 2022.*
  - Naira redesign 2023.
  - Killing of Deborah Yakubu on May 12, 2022.
  - General Elections on 25th Feb, and 11th March, 2023.
  - Removal of fuel subsidy 29th May, 2023.
  -
- Lexicon
  - Use lexicon entries to filter tweets that may contain hate and abusive tweets
  - Pilot random samples to see the distribution of labels and adjust the dataset pool combinations based on the distributions.
- Topic modeling with the tweets to select samples for annotation