# Electoral Sentiment Analysis Project Overview

## Introduction

Female political candidates often receive less media attention than their male counterparts, and analyses of public sentiment seldom focus on the gendered dimensions of electoral discourse. This project explores how voters discuss women contesting parliamentary seats in India by harvesting posts from X (formerly Twitter) during the 2024 election season. The study focuses on four princely states of India: Bihar, Kerala, Maharashtra and Telangana. The study compares how local contexts shape conversations about female candidates and highlights regional differences.

## Data and methodology

Posts were collected using a scraper that targeted combinations of female candidates' names and constituency terms. After removing URLs, usernames and stop-words, the text was lemmatised to normalise variants. Each post was assigned a sentiment score using a lexicon-based classifier, which was chosen because lexicons provide interpretable polarity scores without requiring extensive training data. This classifier maps words to continuous sentiment values from −1 (very negative) to +1 (very positive) and aggregates them into an overall score and a categorical label. The resulting dataset shows that most posts have scores around 0.6, indicating broadly positive tone. To probe lexical patterns further, the project generated three-dimensional frequency plots, polar bar charts, word clouds and Zipf distributions for each state. These visualisations were produced using Python and Matplotlib, and the processed data were saved as Excel files for transparency.

## Critical findings and analysis

The sentiment analysis reveals a consistently positive attitude towards female candidates across all four states, with very few negative or neutral posts. In Bihar, the most frequently used words relate to local districts (*Saran* and *Sheohar*) and topics such as development and public welfare. The word cloud for Bihar also contains gendered terms like "daughter" and "sister," suggesting that voters sometimes frame women candidates through familial roles. Kerala's discourse is markedly different: posts frequently mention national political slogans and the prime minister's name, and terms like "constituency" and "vote" feature prominently. This indicates that discussions around female candidates in Kerala are closely tied to national politics. In Maharashtra, campaign-related vocabulary dominates; words such as "campaign," "sabha" and "constituency" have the highest frequencies in polar plots. Word clouds reinforce this focus on formal election activities, highlighting phrases like "Lok Sabha," "meeting" and "taluka," which point to active fieldwork. Telangana's discourse centres on party identity and local geography; the largest words in the cloud are the name of a major political party and the candidate, along with terms such as "election," "Parliament," and district names like Adilabad and Nirmal .

The corpus as a whole obeys Zipf's law: a log-log plot of word rank versus frequency displays a straight-line decline. State-wise analyses of this slope reveal that Kerala has a shallower slope (≈ −0.70) than Bihar (≈ −0.82), Telangana (≈ −0.91) and Maharashtra (≈ −0.99), indicating greater lexical diversity in Kerala's posts and a more concentrated vocabulary in Maharashtra. Collectively, these findings show that although social-media sentiment towards women candidates is overwhelmingly positive, the content of these discussions varies significantly: Bihar's discourse revolves around development and familial imagery, Kerala's integrates national narratives, Maharashtra's emphasises campaign mechanics and Telangana's foregrounds party affiliation. Such differences underscore how regional political cultures shape the way gender and politics intersect online. https://github.com/JonSnow016/Sentiment-Analysis.git

GitHub repository for code and data: https://github.com/JonSnow016/Sentiment-Analysis.git