Research Summary for Kenyan Election Conflicts

Fordham University – Yiting Cai, Minglu Sun, Chuanze Cai

**Methodology:**

**1. Data collection:** Population and household data were collected from the Kenya national bureau of statistics, event information data from the Gdeltproject database, conflicts, and fatalities data from UCDP Datasets and ACLED datasets, crawled tweets of Kenya from Twitter, timeline pages of news media official accounts from Facebook, and news from The Star website by Python.

**2. Data preprocessing:**  Sentiment analysis was run on tweets and news from Facebook based on text analytics dictionaries, including Harvard General Inquirer, PANAS-X Terms, and Agitation-Dejection Categories. Afterwards, the whole dataset was divided into two dimensions: time series and counties, and then each column was normalized into a dataset.

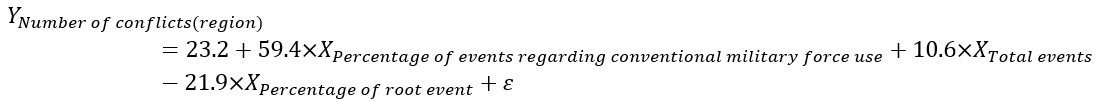
**3. Model building:** R programing ran multiple linear regression model.

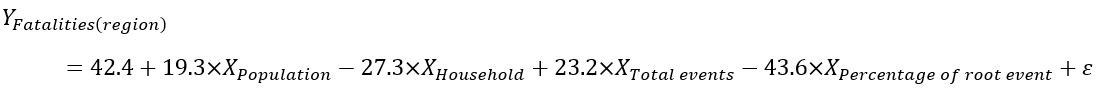
**4. Result evaluation:** Results were conducted on R-square, P-value, Variance Inflation Factor, and Durbin-Watson test to evaluate and adjust the final model.

**5.Data visualization:** Plotted filled maps, symbol maps, trend lines, and combination charts were utilized to visualize exploratory data and to predict the results.

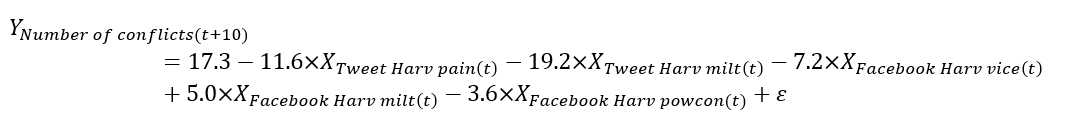
**Results:**

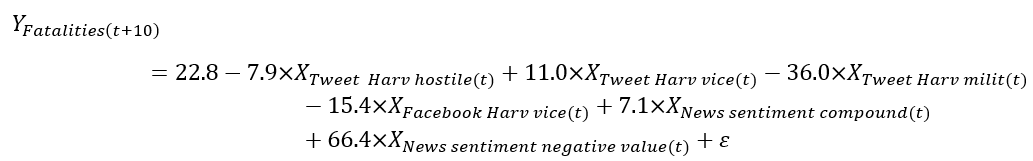
In terms of county, the conflicts will most likely occur in Nairobi (data analysis concludes the total count of conflict in Nairobi to be 82 and 107 people will lose their lives in the conflicts). However, Elgeyo-Marakwet will be the safest province since both counts in conflicts and fatalities are predicted to be 0.





When predicted based on time, another multi-linear regression model was built along with the conflict count for ten days after. According to the predictive analytics, the weekend before election has the highest possibility for conflicts and fatalities. The predicted count of conflicts after election are 1(Aug 8th),0(Aug 9th),4(Aug 10th),1(Aug 11th),3(Aug 12th),2(Aug 13th),2(Aug 14th).





**Appendix - predicting results visualization**

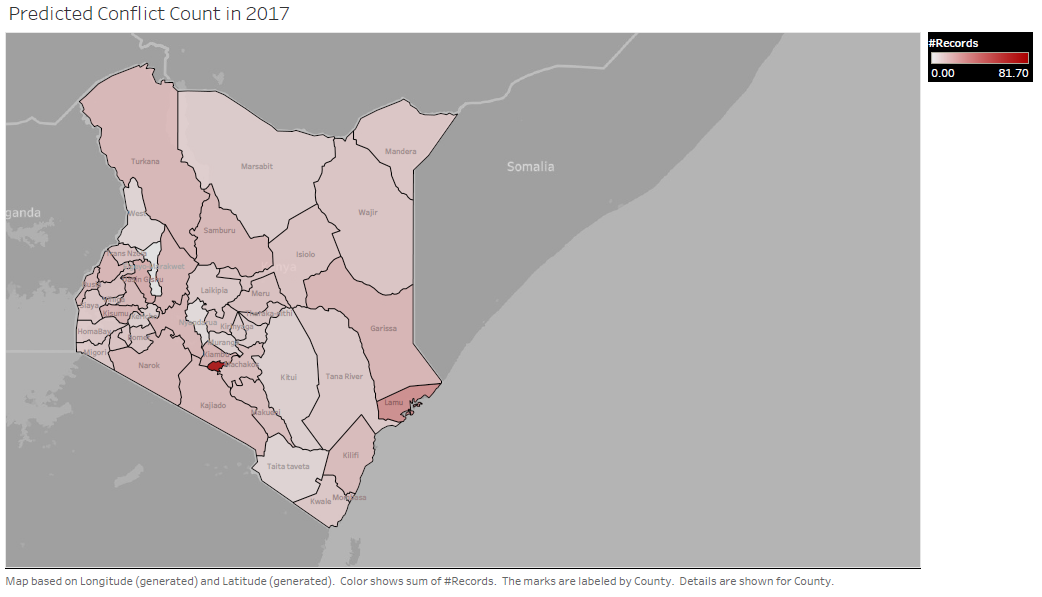


Figure 1: Filled map of predicted conflict counts

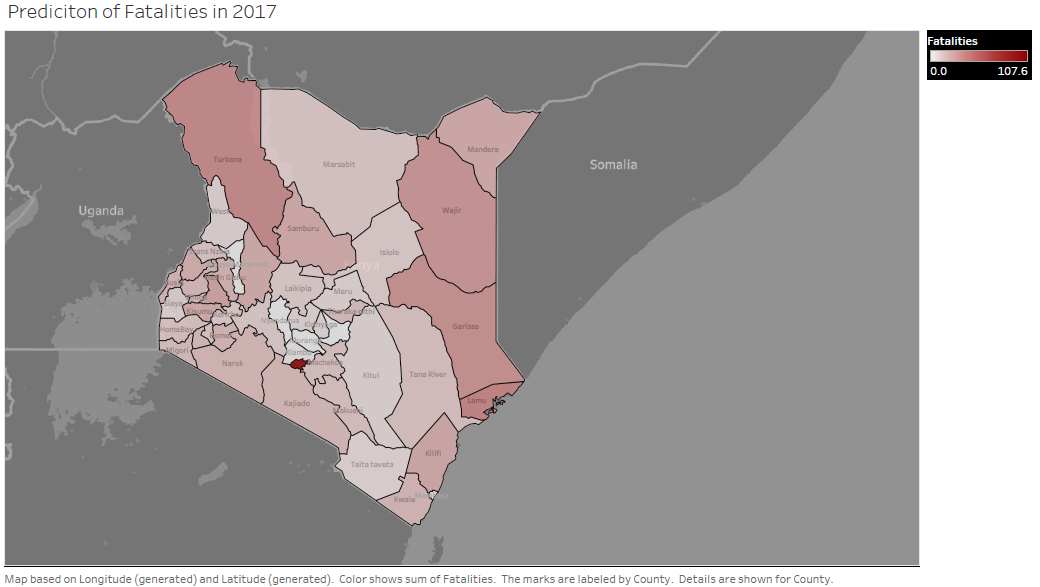


Figure 2: Filled map of predicted fatalities counts

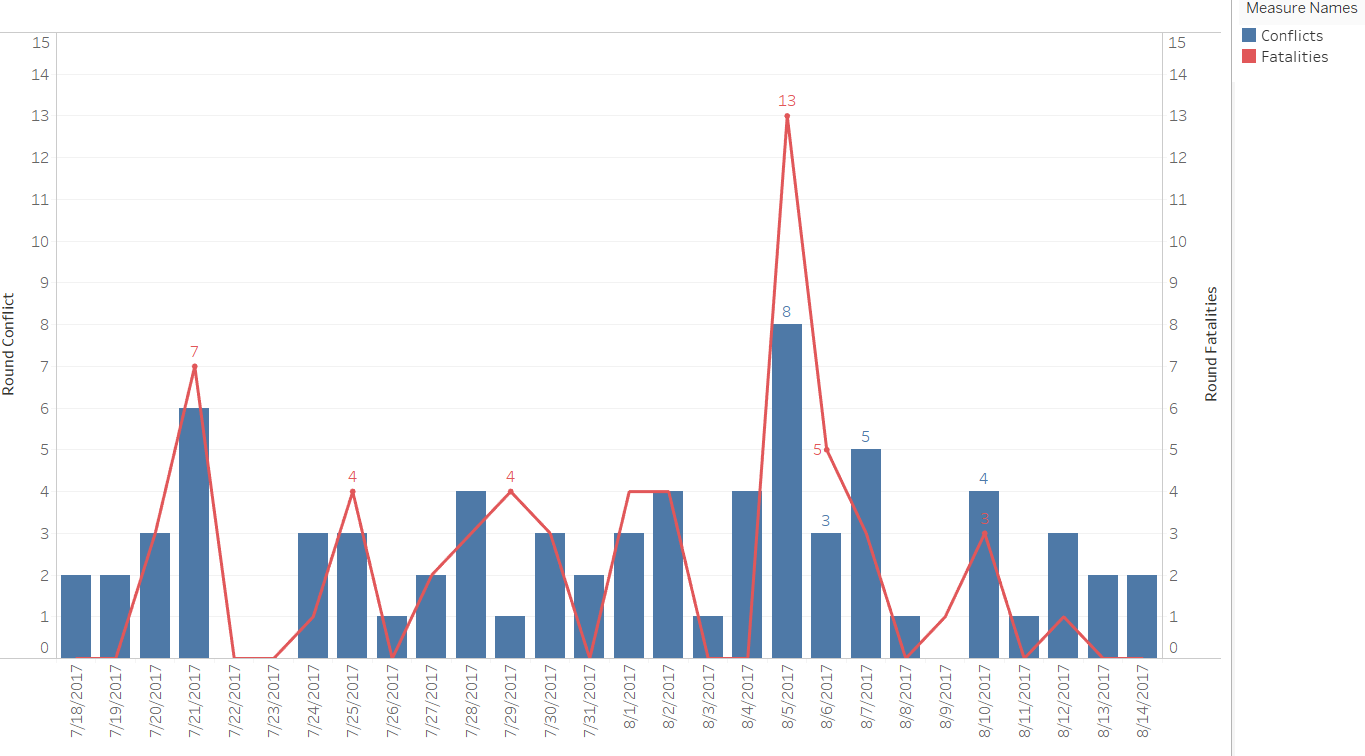


Figure 3: Combined chart of predicted conflicts count and fatalities count by time series.