

## **USE CASE: CONFLICT EVENTS IN AFRICA FROM 1997 TO 2017**

### **Problem**

The number of fatalities arising from conflicts in Africa are alarming and we do not know the factors that may be the most influential in driving these numbers. Africa continues to witness series of conflict across the length and breadth of the continent. Data reports by The Armed Conflict Location & Event Data (ACLED) contain variables that reveal the necessary information regarding the conflicts in Africa from 1997 to 2017. As part of the information contained in the report, 47 African countries have been studied to obtain the number of fatalities recorded in every conflict event. There are nine distinct events in the data set. Apart from the events, the sources where the conflict data were gleaned from as well as the number of actors involved in each conflict event were also reported. However, it is important to answer the question: what relationships exist between the number of fatalities and the events, actors, and sources of reporting in the dataset? If there are any relationships, which factors are most significant and which should be ignored?

### **Solution**

To address this problem, I proposed to determine the relationship between the number of fatalities per country, and the number of actors per country, or the number of conflict events per country, or the number of sources per country. For this activity, I set my variables as follows:

- Y variable = number of fatalities
- X variable 1 = number of events
- X variable 2 = number of actors
- X variable 3 = number of sources

From the original dataset, apart from the number of fatalities, other variables are of string data type and categorical. To determine the values, I pulled the sum of the distinct count of each X variable by country. By testing the total count of the X variables against the Y variable, I will be able to determine the correlation between the variables. I built a linear regression model that tested the relationships and eventually used the Akaike Information Criteria to determine the X variable with the most significant relationship.