**To:** Dedrone engineer team

**From:** Longsheng Xie

**Subject:** Memo – DS Internship Test 2024 - Longsheng Xie

**Date:** 04/08/2024

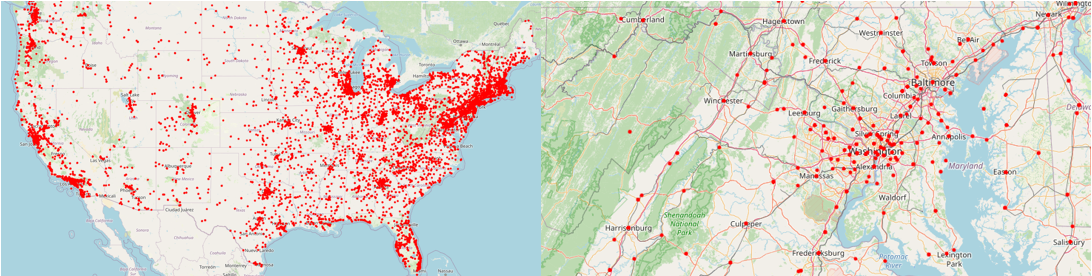
This analysis aims to explore a dataset containing information about protest activity in the United States, including time, location, and other relevant details. The initial step involved loading the data into a Pandas DataFrame to enable detailed examination and manipulation.

I then checked the amount of missing data in the dataset. There are significant missing values in columns such as actor2 and assoc\_actor\_2, indicating that these fields might not be consistently reported across all events. The admin3 column is entirely missing, which may suggest that it is not relevant to our analysis. Actor2, assoc\_actor\_2, and admin3 columns were deleted from the dataset.

And we also need to check the type of data. Data types seem appropriate for most fields, but the event\_date should be converted to a datetime format for any time-series analysis. Therefore, I used pandas to convert the event\_date to datetime format.

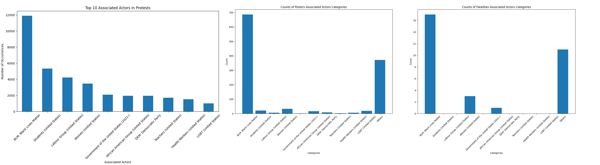
The time series plot displays the number of protests over time. While specific trends and periodic patterns are not distinctly visible in this high-level view, there appears to be variability in the frequency of protests. I also created a visualization showing the geographical distribution of protests across states, as well as the types, sub-types, and source scale of protest events. We can learn that most of the demonstrations were peaceful and reported by local media. The majority of these demonstrations occurred in California, New York, and Florida, states with relatively large populations and areas.

The parade's geography was unclear to us as we only knew that certain states were represented. Therefore, I borrowed from https://www.kaggle.com/code/chandancharchitsahoo/visualize-maps-using-folium/notebook#Choropleth-Maps- and used folium to map the coordinates of all parades on the US map. Most protests are concentrated near major cities with convenient transportation access. Moreover, the number of protests in the eastern United States is greater than that in the west, which may be related to the fact that the political center of the United States is tilted towards the east.

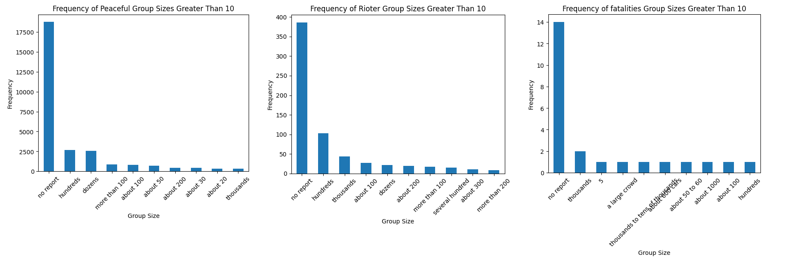


I then also checked the ratio of rioters in California, New York, and Florida and found that California has a slightly higher ratio than the national average, New York is almost the same, and Florida is well below the average.

Next, I split the associate actors, studied the main reasons for all protests, compared the rioters with the associate actors of protest with fatalities, and found that BLM accounted for a slightly higher proportion.



Finally, I separated the sizes reported in the notes and compared the protest sizes of peaceful and rioters. I found that the number of people marching during the rioter protest was significantly higher than that of the peaceful protest. To confirm this finding, a study was conducted on protest size with casualties. It was found that the number of people marching was significantly higher than in other cases.



**Key Insights**

Through the above analysis, we found that the number of protests does not exhibit a significant pattern over time, which may be due to the insufficient volume of our data.

In states with large populations and developed countries such as California, New York, and Florida, the number of people protesting will be significantly greater than in other states. It was also found that most of the protests in the United States took place near major transportation routes, highlighting the importance of convenient transportation as a prerequisite for protests.

The main participants in the protests were Black Lives Matter activists, students, labor representatives, and women. Among the rioters and the protests resulting in casualties, BLM appears to be the primary driving force. Of course, this is most likely due to some unfortunate events that occurred during 2020-2022.

Finally, when the number of people protesting is too large, the possibility of dangerous situations resulting from the protests increases significantly. Therefore, when the above conditions are met, people should treat the protests with great caution.