Capstone 2 Proposal

Problem Statement

The United States has suffered from thousands of terrorist attacks over the last several decades. Considering the numerous tragic deaths and billions of dollars worth of property damage resulting from these incidents, the social, political, and economic impact of terrorism is of major concern to several agencies and institutions throughout the country. ***Is it possible to draw insights from previous incidents that strengthen the overall prevention, preparation, and response to specific threats with greater success?***

Within six weeks, I will offer insight into the patterns of terrorism threats (both foreign and domestic) within six major cities throughout the continental US. This will primarily involve predictions of overall likelihood, attack type, expected targets, and projected damage/casualties based on incidents recorded over the last 70 years. The resulting models will lay groundwork for further investigation into questions of likelihood and observable patterns at the global scale.

Context

The University of Maryland (in conjunction with Pinkerton Global Intelligence Services) has recorded 3121 incidents of foreign and domestic terrorism in the United States since 1970. According to primary sources, these incidents have resulted in over 400,000 deaths and billions of dollars worth of property damage. To address this major social, political, and economic issue, certain agencies such as the Department of Homeland Security (DHS) exist primarily to train and deploy terrorism prevention and response teams according to expected threats throughout the country.

Successful terrorism prevention and response requires data-driven insights and approaches based on several factors related to previous incidents. This is a tremendous amount of data beyond the scope of intuition and observation – a series of predictive models is necessary to identify the likelihood of terrorist attacks and their related features within specific cities.

Criteria for Success & Deliverables

Success for this project is defined by the following:

1) The production of a series of models to predict overall likelihood, attack type, expected targets, and projected damage/casualties in six major US cities based on incidents recorded over the last 70 years.
2) A detailed written report of insights based on the models.
3) A slide deck presentation of insights, suggestions, and opportunities for further exploration for DHS Executives.

Scope of Solution Space

The scope of training and testing data for this project is bound to observations recorded in the Global Terrorism Database (GTD). Though the GTD is the most comprehensive unclassified source of terrorist incident records in the world, it certainly has its limits and may lack clarity for certain events or even lack records altogether.

Moreover, the extent of modeling for this project will be sufficient to provide helpful insights for further discussion and exploration. This project will not involve deploying a model to automate any processes or track success over time.

Concerning the scope of analysis, this project will include seven large cities in various areas of the country as a baseline approach to the problem. ***Thus, insights developed based on modeling apply only to these cities and no others.*** Exploration of more cities is possible pending the success of this project. The seven cities are chosen based on population (within 50 largest US cities), location (diverse geographic representation of the continental US), and social impact (relatively well-known among the general populace). The cities are as follows:

New York City, Chicago, Miami, Houston, Los Angeles, Washington DC, Seattle.

Constraints

The actionable insights derived from this project will contain many implications that must be considered and validated through other studies, e.g., first-responder training methods and technology.

Terrorist incidents do not occur in a straightforward, routinely manner. This project will detect existing patterns and weaknesses of the defense infrastructure of seven cities, but specific predictions are subject to error and uncertainty.

The GTD is extremely comprehensive – too comprehensive for the timeline of this project. I will need to remove ~70 features from the given dataset to make it malleable enough for a six-week analysis, which may negatively affect modeling power.

Stakeholders

The solutions generated from this project will provide meaningful insight to the following:

> Security: Pinkerton Global Intelligence Services, Department of Homeland Security
> Liability: Insurance agencies, Large financial institutions, Academic institutions, Hospitals

Data Sources

Global Terrorism Database
Wikipedia - general cities data, political campaign calendars
Data.gov - geospatial information (govt. buildings, academic institutions, religious organizations, etc.)