

Forecasting Fatalities in NYC

DSCT Capstone 1 - Project Proposal
Jonathan D. Williams



Context

The Fire Department of the City of New York (FDNY) is the second largest fire department in the world, and the largest within the United States. More than 8.5 million residents--and tourists--within the five boroughs of New York City are protected on a daily basis by just under 15,000 uniformed personnel, which are comprised of: firefighters, EMTs, and paramedics. This agency provides invaluable services to all within the city, but its efforts are often subject to several constraints.

The goal of this project is to develop Machine Learning models that predict whether or not a patient that receives some form of emergency medical services will perish. Further investigation of the data will seek to identify specific neighborhoods within the City of New York that have a high concentration of incidents that result in death.

Audience

Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?

Several stakeholders involved with the policy-making, administration, and delivery of emergency response services throughout the City of New York will benefit from this project. Parties include, but are not limited to:

- NYC Elected Officials (e.g. Mayor, Comptroller, City Council Members)
- NYC Fire Commissioner
- Chief of EMS for New York City
- Emergency Response Personnel (e.g. EMTs, firefighters, police officers, etc.)

Insights gathered from this analysis can be used by various agencies within the municipal government to:

Proactively allocate personnel and resources to high-risk neighborhoods to minimize the number of lives lost

Identify variables in the emergency response protocols that influence the prediction of fatalities, with respect to the developed models

Data

What data are you using? How will you acquire the data?

The main data for this project is the NYC EMS Incident Dispatch Data, which is made publicly available via [NYC Open Data](#). This file contains data that is generated by the EMS Computer Aided Dispatch System, and spans from the time the incident is created in the system to the time the incident is closed in the system. It covers information about the incident as it relates to the assignment of resources and the Fire Department's response to the emergency.

Methodology

Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.

Currently I intend to model this scenario as a supervised, binary classification problem. First, substantial preprocessing will need to be performed on the core dataset. Next, either a singular model--or multiple models--will need to be developed. The primary objective is to train these models that predict whether or not an incident will result in a fatality, based on feature variables in the dataset that affect the outcome of an incident.

Additionally, various EDA techniques can be performed to identify, for instance:

- Neighborhoods in NYC with the highest frequency of incidents that result in fatalities
- The time of year when the most fatalities occur

Deliverables

What are your deliverables? Typically, this includes code, a paper, or a slide deck.

The deliverables for this capstone project will include a final report, a slide deck that summarizes the methodology, results, and other relevant insights gathered. In addition, all Jupyter notebooks that contain the ML models, visualizations, and more detailed information about this project will be uploaded to my GitHub account.