

Predicting Number of Volunteers for Disaster Relief

Team 321

Vishwajeet Shelar

Aaron D'Souza



NYU



Disasters in the United States

- 766 disasters in the United States since 2010
- Around 100,000 Red Cross volunteers required for assistance during these disasters for the same time period

Importance of Resource Management

- Volunteers are an important resource during the time of a disaster.
- Prediction of volunteers is vital to optimize this constrained resource.

Methodology

- Clean the data set (remove NaN, duplicates)
- Descriptive statistics to understand the model better.
- Convert data to be compatible to machine learning libraries.
- Fit to models and understand accuracy

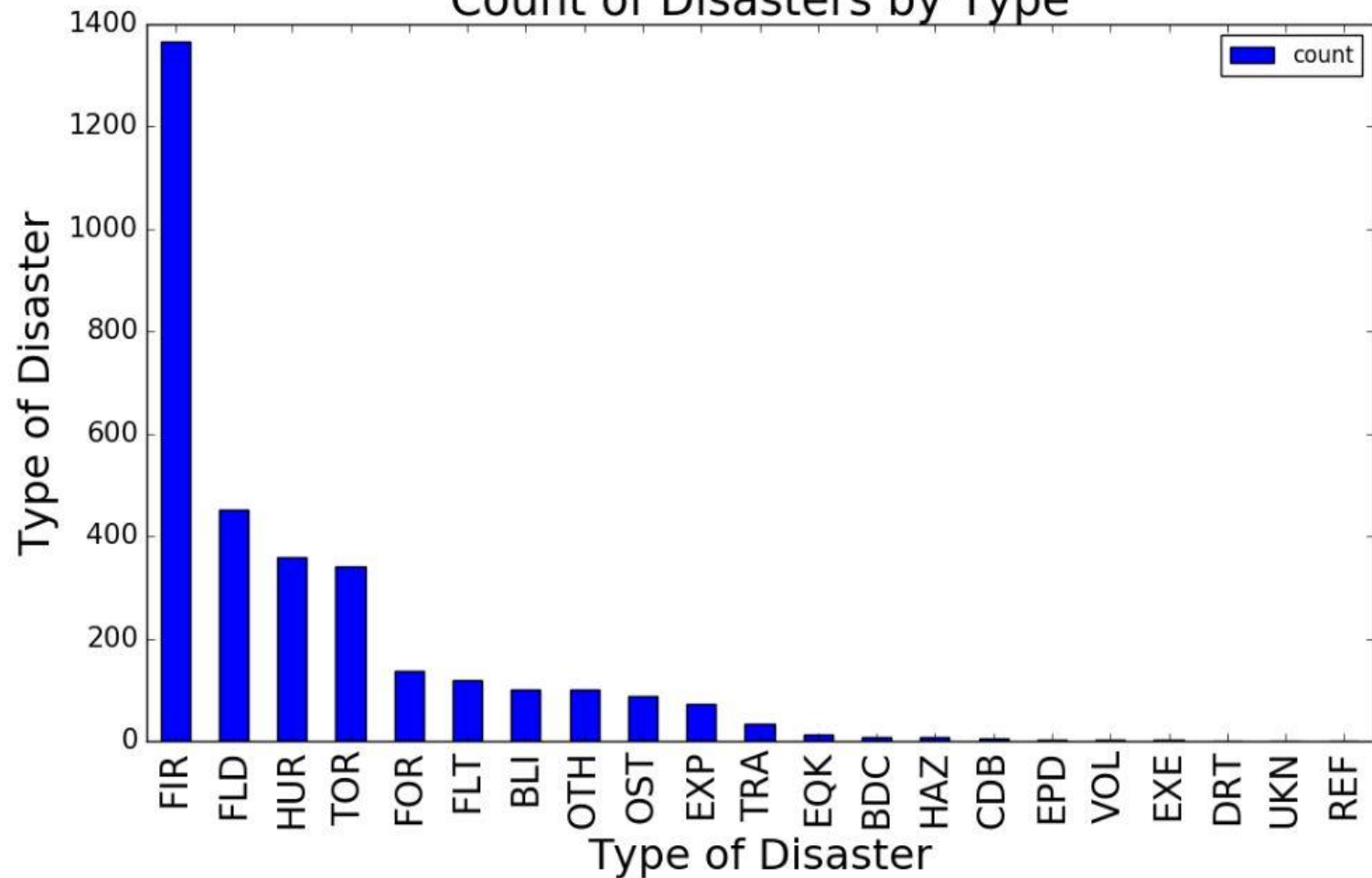
Clean the data set

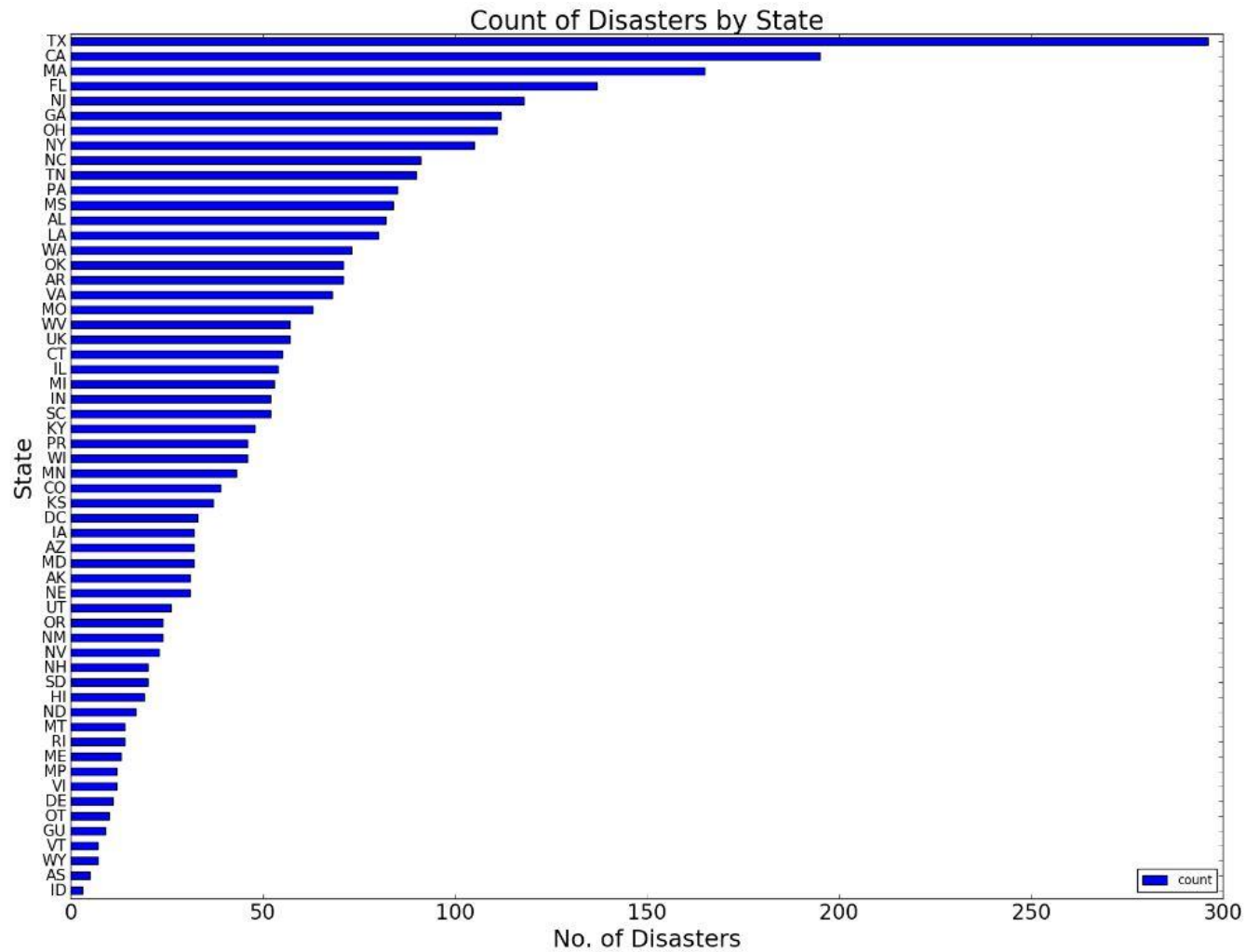
Not many complications with the data set (Phew!!!)

Descriptive statistics on the data

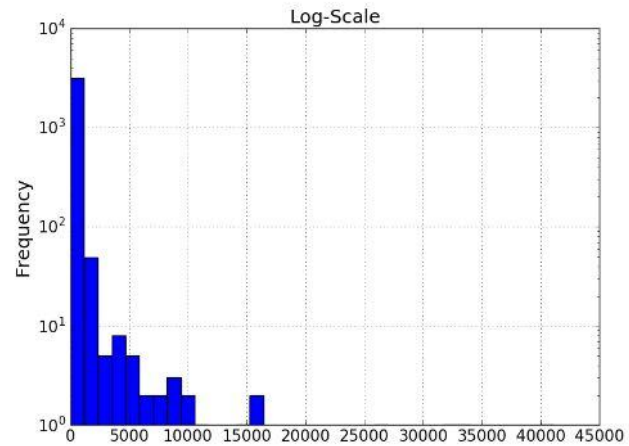
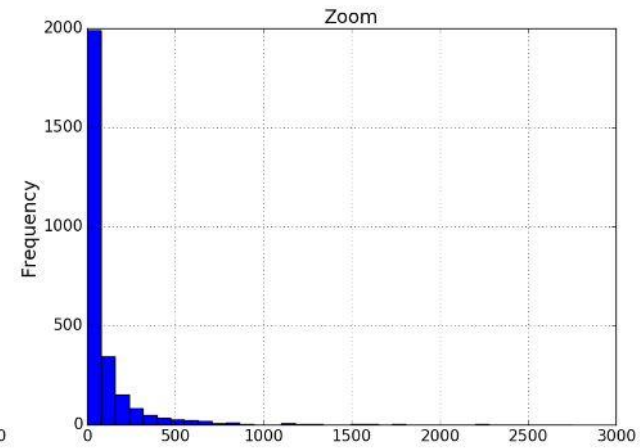
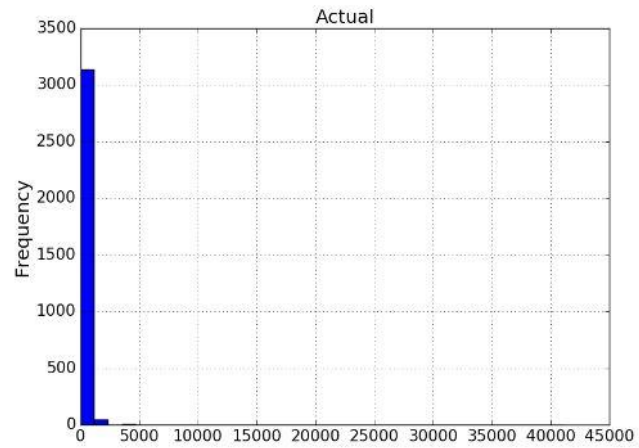
- Needed to understand the data
- Check for correlated data
- Check for outliers

Count of Disasters by Type





Histograms of Total Volunteers required



Convert data to be compatible to machine learning libraries

- Removed outliers (Volunteers greater than zero and less than 20000)
- Created Dummy variables for categories (Ordinal Data)
- Convert volunteer count data into log scale to reduce skewness

Fit to models and understand Accuracy

- Random Forest Regressor
 - Used log of volunteer count
 - Optimized on Max depth
 - Maximum features = square root of Features
 - Accuracy = 80.57 %

Fit to models and understand Accuracy

- Gradient Boosting Regressor
 - Optimized on learning rate
 - Loss is least square regression
 - Maximum features = square root of Features
 - Accuracy 81.80%

Future work

- Feature selection
- Understand spatial attributes such as state in which relief operations are needed.