Gun Violence in the US. Unsupervised Learning Application

by Sumaira Afzal, Viraja Ketkar, Murlidhar Loka, Vadim Spirkov

Abstract To do

Background

Objective

The objective of this research is to ...

Data Analysis

The data set used for this research contains 260k of gun violence incidents in the US between January 2013 and March 2018. The data has been soursed from Kaggle.

Originallly the data set was uploaded to Kaggle from Gun Violence Archive (GVA) Web site gunviolencearchive.org. This is a not for profit corporation formed in 2013 to provide free online public access to accurate information about gun-related violence in the United States. GVA will collect and check for accuracy, comprehensive information about gun-related violence in the U.S. and then post and disseminate it online.

Data Dictionary

Column Name	Column Description
incident_id	Incident ID
date	Date of crime
state	State
city_or_countyCity	City/county of crime
address	Address of the location of the crime
n_killed	Number of people killed
n_injured	Number of people injured
incident_url	URL regarding the incident
source_url	Reference to the reporting source
incident_url_fields_missing	TRUE if the incident_url is present, FALSE otherwise
congressional_district	Congressional district id
gun_stolen	Status of guns involved in the crime (i.e. Unknown, Stolen, etc...)
gun_type	Typification of guns used in the crime
incident_characteristics	Characteristics of the incidence
latitude	Location of the incident
location_description	Description of the location
longitude	Location of the incident
n_guns_involved	Number of guns involved in incident
notes	Additional information of the crime
participant_age	Age of participant(s) at the time of crime (victicms nad suspects)
participant_age_group	Age group of participant(s) at the time crime
participant_gender	Gender of participant(s)
participant_name	Name of participant(s) involved in crime
participant_relationship	Relationship of participant to other participant(s)
participant_status	Extent of harm done to the participant
participant_type	Type of participant (victim or suspect)
sources	Participants source

Column Name	Column Description
state_house_district	Voting house district
state_senate_district	Territorial district from which a senator to a state legislature is elected.

Data Exploration

Missing Data

Takeaways from Data Exploration Excersize

Data Preparation

Data Imputing

Modeling and Evalutation

Feature Selection

Data Upsampling

Partitioning Clustering Approach

Hierarchical Clustering Approach

Density-based Clustering Methods

Clustering Method Evaluation

Model Deployment

Conclusion

Note from the Authors

This file was generated using *The R Journal* style article template, additional information on how to prepare articles for submission is here - Instructions for Authors. The article itself is an executable R Markdown file that could be downloaded from Github with all the necessary artifacts.

Sumaira Afzal
York University School of Continuing Studies

Viraja Ketkar
York University School of Continuing Studies

Murlidhar Loka
York University School of Continuing Studies

Vadim Spirkov
York University School of Continuing Studies