Data Mining report

Funaioli Francesca
Karoui Hamza
Vezzuto Samuele

October 16, 2023

Contents

Chapter 1

Data Understanding

In Table 1.1 we show

Feature Name	Initial Type	Cast Type	Description
date	object	Datetime64	date of incident occurrence
state	object	String	state where incident took place
city_or_county	object	String	city or county where incident took place
address	object	String	address where incident took place
latitude	float64	float64	latitude of the incident
longitude	float64	float64	longitude of the incident
congressional_district	int64	Int64	congressional district where the incident took place
state_house_district	int64	Int64	state house district
state_senate_district	float64	Int64	state senate district where the incident took place
participant_age1	float64	Int64	exact age of one (randomly chosen) participant in the incident
participant_age_group1	object	String	exact age group of one (randomly chosen) participant in the incident
participant_gender1	object	String	exact gender of one (randomly chosen) participant in the incident
min_age_participants	object	Int64	minimum age of the participants in the incident
avg_age_participants	object	float64	average age of the participants in the incident
max_age_participants	object	Int64	maximum age of the participants in the incident
n_participants_child	object	Int64	number of child participants 0-11
n_participants_teen	object	Int64	number of teen participants 12-17
n_participants_adult	object	Int64	number of adult participants (18 +)
n_males	float64	Int64	number of males participants
n_females	float64	Int64	number of females participants
n_killed	int64	Int64	number of people killed
n_injured	int64	Int64	number of people injured
n_arrested	float64	Int64	number of arrested participants
n_unharmed	float64	Int64	number of unharmed participants
n_participants	float64	Int64	number of participants in the incident
notes	object	String	additional notes about the incident
incident_characteristics1	object	String	incident characteristics
incident_characteristics2	object	String	incident characteristics

Table 1.1: Features of the Incidents

Feature Name	Initial Type	Cast Type	Description
state	object	String	
year	int64	Int64	
povertyPercentage	float64	float64	poverty percentage for the corresponding state and year

Table 1.2: Features of the Poverty by State

Feature Name	Initial Type	Cast Type	Description
year	int64	Int64	
state	object	String	
congressional_district	int64	Int64	
party	object	String	winning party for the corresponding congressional_district in the state, in the corresponding year
candidateVotes	int64	Int64	number of votes obtained by the winning party in the corresponding election
totalVotes	int64	Int64	number total votes for the corresponding election

Table 1.3: Features of the Winner of Congressional Elections

Chapter 2

Data Preparation

Chapter 3

Clustering