Mass Shootings: Understanding Underlying Causations By Association Rule Analysis

Joseph Cruz
josephbcruz@lewisu.edu
DATA-51000-
Data Mining and Analytics
Lewis University

I. INTRODUCTION

Gun violence is the act in which a gun is utilized to perform some sort of crime or harm to either one or more people or themselves. Throughout the years, there have been many different forms of gun violence such as suicide, homicide, armed robbery, or even accidental death (i.e., hunting accidents or criminal negligence). In 2017, 60% of deaths attributed to guns in the United States were declared suicides while only 37% of gun deaths were from homicide [1]. Although homicide is responsible for fewer deaths compared to suicide, homicide seems to be discussed much more frequently than suicide. More specifically, mass shootings receive a large amount of media coverage when they occur [2]. Mass shootings are under the umbrella of homicide, but they are defined as multiple homicide incidents where four or more victims are killed with a firearm within one or more locations in relatively proximity within one event in public places [3]. These mass shooting events happen much less frequently than normal homicides, where in 2019 they only attributed to 0.2% of the total United States firearm deaths [4]. Even though they occur less frequently, mass shootings have devasting effects such as multiple fatalities, mass-trauma infliction to those involved, and wide-spread panic at the locale that may contribute to more injury or death, just to name a few.

Because mass shootings have such devastating effects, it is important to understand potential key factors that may be consistent with mass shooting events. On way to analyze these potential key factors or relationships, is through analyzing different instances of mass shootings for similar details, or general trends that occur. The determination of relationships or similarities between instances of mass shootings can potentially allow for a deeper understanding of what may cause these occurrences or may give some insight on what may need to be investigated to reduce the likelihood of these instances occurring. To start identifying these potential relationships and similarities between different mass shooting events, one can perform an association analysis. In performing an association rule analysis, it is possible to identify potential association rules and relationships between instance data, which can then be analyzed to determine the validity of some of these association made between mass shooting events.

In this paper, an association rule analysis is performed to determine potential association rules or relationships of mass shooting events that occur in the United States. This analysis will be performed on a database containing mass shootings in public places that resulted in either four or more casualties from the year 1982 to the current year (2021) [5]. Throughout the course of this report, a description of the dataset will be explored as well as an explanation of the methodology used for analysis. Furthermore, a results and discussion section will be covered with a conclusion from the data to follow. In Section II, the dataset will be described in detail. Section III will be an overview of the methodology used for the association rule analysis. In Section IV, results of the analysis will be reported, and a discussion will ensue to make sense of the results. Finally, in Section V, the conclusions of the findings will be presented.

II. DATA DESCRIPTION

The dataset that will be used for this association rule analysis is a database originally from *Mother Jones* that contains records of mass shootings in public places that resulted in either four or more casualties from the year 1982 to the current 2021 [5]. There are a total of 124 rows, or instances, in the dataset with a total of 24 features. The features of the dataset are displayed in Table 1. The first feature of the dataset is "case", and this is the name of the shooting event, such as the "Las Vegas Strip Massacre". The next feature of the dataset is the "location" which is in reference to the location the event occurred. "Date" is the date in which the shooting event occurred. The "summary" feature is a descriptive and detailed summary of what occurred during the event. The "fatalities" feature is used to describe the quantity of fatal casualties that occurred during the event. The "injured" feature describes the quantity of injuries accrued during the event. The "total_victims" feature is the quantity of both injuries and fatalities attributed to the event. The second "location" feature is meant to describe the public place that the event occurred as a category. For instance, there were labels like "Workplace", "Military", "Religious", etc. to describe the place that the event occurred. The "age_of_shooter" feature is meant to contain the age of the shooter for that event. The feature "prior_signs_mental_health_issues" contains either a positive, negative, or neutral response (yes, no, or unknown) to whether the shooter had evidence of mental health issues prior to the event. The feature

TABLE I. ATTRIBUTES OF MASS SHOOTING DATASET FOR 1982-2021

Attribute	Type	Example Value	Description
case	Nominal (string)	"Las Vegas Strip massacre"	Description of mass shooting event by name it is commonly associated with.
location	Nominal (string)	"San Jose, California"	Location of the event.
Date	Nominal (string)	"5/26/2021"	Date event occurred.
summary	Nominal (string)	"Ahmad Al Aliwi Alissa, 21, carried out a mass shooting…"	Descriptive detailing of the event.
fatalities	Numeric (integer)	10	Quantity of fatalities that occurred in event.
injured	Numeric (integer)	1	Quantity of injured that occurred in event.
total_victims	Numeric (integer)	15	Quantity of total victims in event.
location	Nominal (string)	"Workplace"	Category of public place event occurred.
age_of_shooter	Numeric (integer)	57	Age of the shooter in the event.
prior_signs_mental_health_issues	Nominal (string)	"Unclear"	Presence of mental health issues before the event.
mental_health_details	Nominal (string)	"Brother described him as paranoid; multiple account…"	Description of mental health prior to event.
weapons_obtained_legally	Binary (string)	"Yes"	Weapons used in event purchased legally or not.
where_obtained	Nominal (string)	"A gun store in Simi Valley"	Location the weapon(s) were obtained legally.
weapon_type	Nominal (string)	"semiautomatic handguns"	Type of weapon(s) used in the event.
weapon_details	Nominal (string)	"Ruger AR-556; weapon was purchased six days before the attack"	Descriptive detail about the weapon(s) used in the event.
race	Nominal (string)	"Asian"	Race of the shooter in the event.
gender	Binary (string)	"M"	Gender of shooter in event.
sources	Nominal (string)	"https://www.denverpost.com/2021/03/22/polic…."	News source data originated from for event.
mental_health_sources	Nominal (string)	"https://www.nytimes.com/2021/05/26/u"	News source data originated from for event for mental health status.
sources_additional_age	Nominal (string)	"https://www.chron.com/news"	News source data originated from for age of shooter.
latitude	Numeric (decimal)	39.68663	Latitude coordinate of the event.
longitude	Numeric (decimal)	-86.3231	Longitude coordinate of the event.
type	Nominal (string)	"Mass"	Detail as to whether event was a spree or a mass shooting.
year	Numeric (integer)	2021	Year that event took place.

"mental_health_details" contained specific details of the evidence of mental health issues the shooter may have had prior to the event. The feature "weapons_obtained_legally" is meant to hold either a positive, negative, or neutral response dependent upon if the shooter had obtained their weapon(s), used in the shooting event, legally, illegally, or unclear. The "where_obtained" feature describes the location that the shooter obtained the weapon(s) used in the shooting event. The "weapon_type" is meant to describe the kind of weapon(s) used in the event by the shooter. The "weapon_details" attribute provides a specific detail of the weapon utilized for the event. The "race" and "gender" features are meant describe the race and gender of the shooter, respectively. The "sources" feature contains the sources used to gather the information about the event. The "mental_health_sources" and "sources_additional_age" features contain the sources that were used to gather the mental health and age data for the event. The "latitude" and "longitude" features describe the latitude and longitude coordinates in which the event occurred. The "type" feature is meant to further describe these events as either "mass" for mass shooting events, or "spree" for spree event. The difference between a mass shooting and a spree is that a mass shooting occurs in one location over a continuous period of time with at least four fatalities, whereas a spree can occur in multiple locations with at least two fatalities [6]. Finally, the "year" feature is the year in which the shooting event occurred.

Of the dataset, the following attributes were used for the association analysis: "fatalities", "injured", "total_victims", "location (2)" (2nd location with public area labels), "age_of_shooter", "prior_signs_mental_health_issues", "race", "gender", "type", and "weapons_obtained_legally". The frequency distributions of the attributes used can be found in Figures 1-10. The average, standard deviation, and percent coefficient of variation (%CV) of the "fatalities", "injured", "total_victims", and "age_of_shooter" attributes

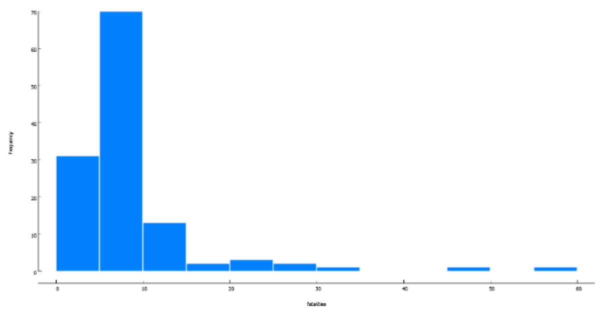

Fig. 1. Frequency distribution of fatalities attribute.

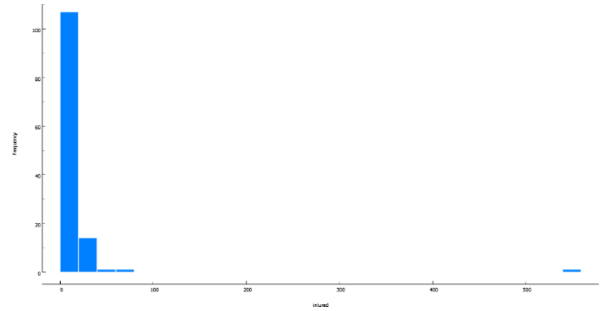

Fig. 2. Frequency distribution of injured attribute.

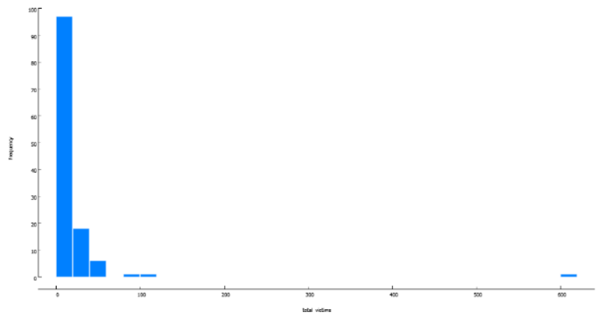

Fig. 3. Frequency distribution of total_victims attribute.

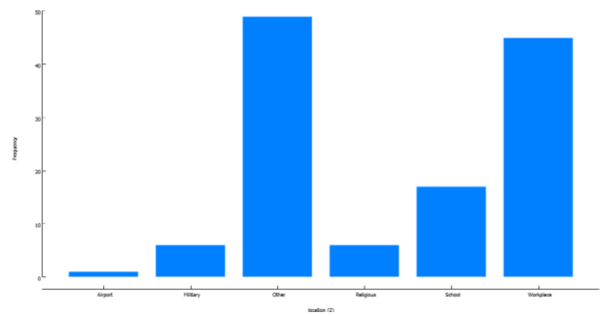

Fig. 4. Frequency distribution of location(2) attribute.

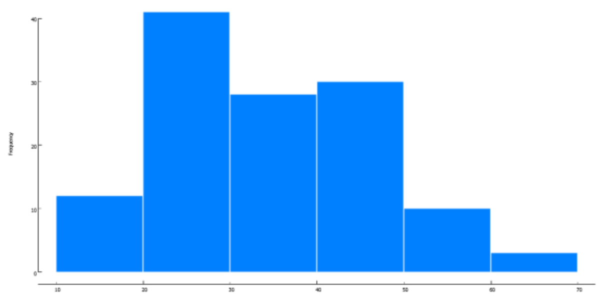

Fig. 5. Frequency distribution of age_of_shooter attribute.

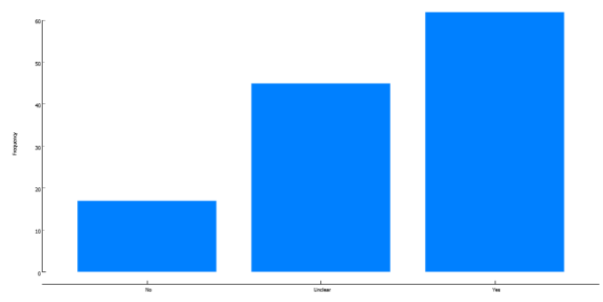

Fig. 6. Frequency distribution of prior_signs_mental_health_issues attribute.

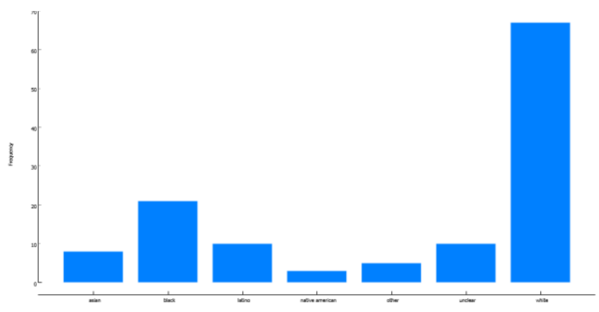

Fig. 7. Frequency distribution of race attribute.

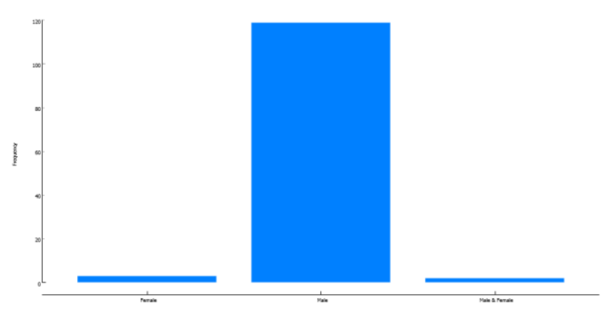

Fig. 8. Frequency distribution of gender attribute.

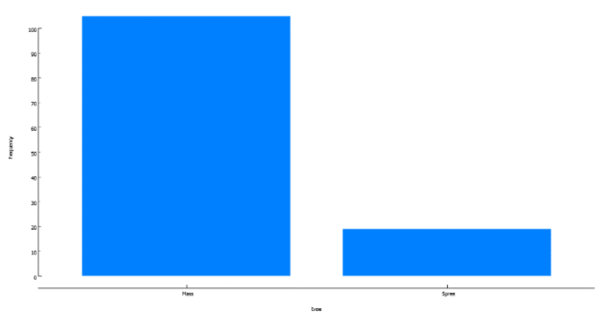

Fig. 9. Frequency distribution of type attribute.

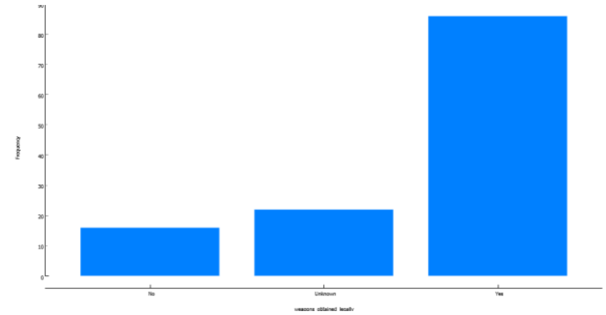

Fig. 10. Frequency distribution of "weapons_obtained_legally" attribute.

TABLE II. STATISTICAL ANALYSIS OF NUMERIC ATTRIBUTES FOR THE MASS SHOOTING DATASET FOR 1982-2021

	fatalities	injured	total_victims	age_of_shooter
Average (AVG)	8.03	11.71	19.74	34.16
Standard Deviation (SD)	7.77	49.48	54.93	12.16
Percent CV (%CV)	96.70	422.59	278.25	35.60

are shown in Table II. The "location" attribute is split into six categories of public places (Airport, Military, Other, Religious, School, or Workplace). The "prior_signs_mental_health_issues" attribute is split into three categories being either "Yes", "No", or "Unclear". The "race" attribute is split into seven categories of races ("asian", "black", "latino", "native american", "other", "unclear", and "white"). The "gender" attribute is split into three categories ("Female", "Male", and "Male & Female"). The "type" attribute consisted of two categories of either "mass" or "spree". Finally, the "weapons_obtained_legally" attribute was split into the three categories of "Yes", "No", or "Unknown".

The "fatalities" attribute had an average of 8.03 deaths per shooting event with a standard deviation of 7.77 and a %CV of 96.7%. The "injured" attribute had an average of 11.71 injuries per shooting event with a standard deviation of 49.48 and the %CV value of 422.59%. The "total_victims" attribute retained an average of 19.74 victims per shooting event with a standard deviation of 54.93 and a %CV of 278.25%. The "age_of_shooter" attribute showed that on average the shooter in a shooting event was approximately 34 (34.16) years old. Furthermore, the standard deviation of this attribute was 12.16 and there was a %CV of 35.6%.

III. METHODOLOGY

To perform the association rule analysis, the data mining toolkit Orange (version 3.26) was used. The flowchart shown in Figure 11 provides a summary of the steps taken to perform the analysis. First, the data was obtained and preprocessed using Excel to remove null values and unify the data. Seven of the attributes required preprocessing before the analysis. In the attribute "location", the null values were replaced with the category "Other" since they were unidentified by the public place. The "-" values for the "age_of_shooter" attribute was replaced with the average age of the data, which was 34.16. For the "prior_signs_mental_health_issues" attribute, the "- ", "TBD", and "Unknown" values were replaced with the value "Unclear" and the values of "yes" were replaced with the value "Yes". The "weapons_obtained_legally" attribute had the null values, "- ", and "TBD" values replaced with the value "Unknown" and all "yes" values replaced with "Yes". The "race" attribute had the "-" values replaced with "unclear". The "gender" attribute was unified by replacing values of "M" or "F" with "Male" or "Female", respectively. Then, the next step was to designate metadata and features of the dataset in Orange. The following attributes were marked as features: "fatalities", "injured", "total_victims", "location (2)" (2^{nd} location with public area labels), "age_of_shooter",

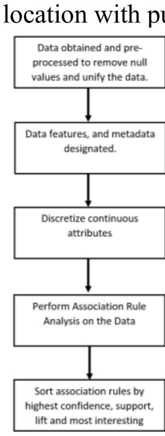

Fig. 11. Flowchart of association rule methodology.

“prior_signs_mental_health_issues”, “race”, “gender”, “type”, and “weapons_obtained_legally”. The rest of the attributes of the dataset were designated as metadata. Then, continuous attributes, such as “fatalities”, “injured”, “total_victims”, and “age_of_shooter” were discretized utilizing the discretize widget in Orange. Discretization was achieved utilizing an equal-frequency discretization with a total of 3 intervals established. The following are the splits for each attribute: “fatalities” categories (<4.5, 4.5-6.5 and ≥ 6.5), “injured” (<1.5, 1.5-6.5 and ≥ 6.5), “total_victims” (<7.5, 7.5-15.5 and ≥ 15.5), “age_of_shooter” (<26.5, 26.5-39.5 and ≥ 39.5). Once the discretization was performed, the preprocessing of the data was officially complete.

To perform the actual association rule analysis, the preprocessed and discretized data was fed into the association rules widget in Orange. The search criteria used to determine association rules were a minimum support of 20% and a minimum confidence of 40%. The antecedent “type =Mass” was used to filter the rules so that all the association rules would be generated for analysis had to deal with the mass shooting type specifically. Soon after, the accumulated association rules were fed into multiple data tables sorted by the maximum levels of confidence, support, and lift to analyze the generated rules. Furthermore, rules were examined for interesting finds not necessarily caught in these data tables and were captured in another data table.

IV. Results and Discussion

A. Results

The association rule analysis yielded a total of 534 rules determined from the dataset. After filtering the rules for the antecedent “type =Mass”, a total of 141 of the 534 rules were retrieved. In Tables III-VI, the association rules were sorted by either support, confidence, lift, or how interesting they were during analysis and the top 10 rules of each sort criteria are shown. While discussing these rules, it is important to note that they will be described as “x → y”, which means x implies y.

Table III is a list of the top ten rules based off the highest support values in the association rules determined with the antecedent filter. The highest support value of the rules was 0.823 while the lowest is 0.419. Furthermore, rules in the table have confidence values ranging from 0.495 to 0.983, coverage scores ranging from 0.468 to 0.847, strength scores ranging from 0.59 to 2.052, and lift scores ranging from 0.99 to 1.034. The rule with the most support in this table is “type=Mass → gender=Male”, where the support score is 0.823, the confidence score is 0.971, the coverage score is 0.847, the strength score is 1.133, and the lift is 1.012. The rule with the least support in the table is “type=Mass → prior_signs_mental_health_issues=Yes” with a support score of 0.419, confidence of 0.495, coverage of 0.847, strength of 0.59, and lift of 0.99.

Table IV is a list of the top ten rules based off the highest confidence values in the association rules determined with the antecedent filter. The highest confidence value of the rules was 1 and the lowest was 0.976. The support of the rules ranged from 0.202 to 0.46, coverage scores ranged from 0.202 to 0.468, strength scores ranged from 2.052 to 4.76, and lift scores ranged from 1.017 to 1.042. Additionally, all the consequents of the table are “gender = Male”. The first six rules of the Table had the highest confidence values, where the first rule was “location (2) =Other, type=Mass → gender=Male” with a support score of 0.315, confidence of 1, coverage of 0.315, strength of 3.051, and lift of 1.042. The rule with the lowest confidence value was “prior_signs_mental_health_issues=Yes, weapons_obtained_legally=Yes, type=Mass→ gender=Male” with a support score of 0.323, confidence of 0.976, coverage of 0.331, strength of 2.902, and lift of 1.017.

TABLE III. Association rules determined by Association Rule analysis, sorted by highest Support values found.

Antecedent	Consequent	Support	Confidence	Coverage	Strength	Lift
type=Mass	gender=Male	0.823	0.971	0.847	1.133	1.012
type=Mass	weapons_obtained_legally=Yes	0.597	0.705	0.847	0.819	1.016
gender=Male, type=Mass	weapons_obtained_legally=Yes	0.573	0.696	0.823	0.843	1.004
type=Mass	weapons_obtained_legally=Yes, gender=Male	0.573	0.676	0.847	0.79	1.01
weapons_obtained_legally=Yes, type=Mass	gender=Male	0.573	0.959	0.597	1.608	1
type=Mass	race=white	0.468	0.552	0.847	0.638	1.022
gender=Male, type=Mass	race=white	0.46	0.559	0.823	0.657	1.034
race=white, type=Mass	gender=Male	0.46	0.983	0.468	2.052	1.024
type=Mass	race=white, gender=Male	0.46	0.543	0.847	0.629	1.02
type=Mass	prior_signs_mental_health_issues=Yes	0.419	0.495	0.847	0.59	0.99

TABLE IV. Association rules determined by Association Rule analysis, sorted by highest Confidence values found.

Antecedent	Consequent	Support	Confidence	Coverage	Strength	Lift
location (2)=Other, type=Mass	gender=Male	0.315	1	0.315	3.051	1.042
fatalities=4.5 - 6.5, type=Mass	gender=Male	0.274	1	0.274	3.5	1.042
location (2)=Other, weapons_obtained_legally=Yes, type=Mass	gender=Male	0.242	1	0.242	3.967	1.042
injured=< 1.5, total_victims=< 7.5, type=Mass	gender=Male	0.226	1	0.226	4.25	1.042
fatalities=4.5 - 6.5, weapons_obtained_legally=Yes, type=Mass	gender=Male	0.202	1	0.202	4.76	1.042
injured=≥ 6.5, race=white, type=Mass	gender=Male	0.202	1	0.202	4.76	1.042
race=white, type=Mass	gender=Male	0.46	0.983	0.468	2.052	1.024
prior_signs_mental_health_issues=Yes, type=Mass	gender=Male	0.411	0.981	0.419	2.288	1.022
weapons_obtained_legally=Yes, race=white, type=Mass	gender=Male	0.355	0.978	0.363	2.644	1.019
prior_signs_mental_health_issues=Yes, weapons_obtained_legally=Yes, type=Mass	gender=Male	0.323	0.976	0.331	2.902	1.017

TABLE V. ASSOCIATION RULES DETERMINED BY ASSOCIATION RULE ANALYSIS, SORTED BY HIGHEST LIFT VALUES FOUND.

Antecedent	Consequent	Support	Confidence	Coverage	Strength	Lift
injured=≥ 6.5, type=Mass	total_victims=≥ 15.5	0.21	0.684	0.306	0.921	2.424
total_victims=< 7.5, gender=Male, type=Mass	injured=< 1.5	0.226	0.8	0.282	1.171	2.42
total_victims=≥ 15.5, type=Mass	injured=≥ 6.5	0.21	0.839	0.25	1.387	2.419
total_victims=< 7.5, type=Mass	injured=< 1.5, gender=Male	0.226	0.778	0.29	1.111	2.411
total_victims=≥ 15.5, gender=Male, type=Mass	injured=≥ 6.5	0.202	0.833	0.242	1.433	2.403
injured=≥ 6.5, type=Mass	total_victims=≥ 15.5, gender=Male	0.202	0.658	0.306	0.895	2.399
injured=≥ 6.5, gender=Male, type=Mass	total_victims=≥ 15.5	0.202	0.676	0.298	0.946	2.394
total_victims=≥ 15.5, type=Mass	injured=≥ 6.5, gender=Male	0.202	0.806	0.25	1.355	2.381
injured=< 1.5, type=Mass	total_victims=< 7.5, gender=Male	0.226	0.824	0.274	1.265	2.375
total_victims=< 7.5, type=Mass	injured=< 1.5	0.226	0.778	0.29	1.139	2.352

TABLE VI. ASSOCIATION RULES DETERMINED BY ASSOCIATION RULE ANALYSIS, MOST INTERESTING RULES FOUND SORTED BY LIFT.

Antecedent	Consequent	Support	Confidence	Coverage	Strength	Lift
prior_signs_mental_health_issues=Yes, type=Mass	weapons_obtained_legally=Yes, race=white, gender=Male	0.226	0.538	0.419	0.942	1.363
total_victims=≥ 15.5, gender=Male, type=Mass	weapons_obtained_legally=Yes	0.21	0.867	0.242	2.867	1.25
prior_signs_mental_health_issues=Yes, race=white, gender=Male, type=Mass	weapons_obtained_legally=Yes	0.226	0.848	0.266	2.606	1.223
fatalities=≥ 6.5, gender=Male, type=Mass	prior_signs_mental_health_issues=Yes	0.202	0.595	0.339	1.476	1.19
prior_signs_mental_health_issues=Yes, type=Mass	weapons_obtained_legally=Yes	0.331	0.788	0.419	1.654	1.137
prior_signs_mental_health_issues=Yes, type=Mass	gender=Male	0.411	0.981	0.419	2.288	1.022
type=Mass	weapons_obtained_legally=Yes	0.597	0.705	0.847	0.819	1.016
type=Mass	gender=Male	0.823	0.971	0.847	1.133	1.012
type=Mass	prior_signs_mental_health_issues=Yes	0.419	0.495	0.847	0.59	0.99
prior_signs_mental_health_issues=Unclear, type=Mass	gender=Male	0.29	0.947	0.306	3.132	0.987

Table V is a list of the top ten rules based upon the highest lift scores in the association rules determined with the antecedent filter. The highest lift score of the rules was 2.424 and the lowest was 2.352. The support of the rules ranged from 0.202 to 0.226, the confidence ranged from 0.658 to 0.839, the coverage ranged from 0.242 to 0.306, and the strength ranged from 0.895 to 1.433. The rule with the most lift was "injured=≥ 6.5, type=Mass → total_victims=≥ 15.5" with a support of 0.21, confidence of 0.684, coverage of 0.306, strength of 0.921, and lift of 2.424. The rule with the least lift was "total_victims=< 7.5, type=Mass→ injured=< 1.5", with a support of 0.226, confidence of 0.778, coverage of 0.29, strength of 1.139, and lift of 2.352.

Finally, Table VI is a list of the top ten rules from the association rules with the antecedent filter that were the most interesting. The rules were selected by analyzing the rules generated and looking for interesting trends and associations between mass shooting events as well as determining appropriate support, confidence, and lift scores. Table VI is sorted by lift since lift is a good measure of correlation. The highest lift score of these rules was 1.363 and the lowest was 0.987. The supports ranged from 0.202 to 0.823, the confidence ranged from 0.495 to 0.981, the coverage ranged from 0.242 to 0.847, and the strength ranged from 0.59 to 3.132.

B. *Discussion*

For the most part, the association rules generated have been insightful for the relationships between mass shooting events. In the highest support rules shown in Table III, many of them deal with the sole antecedent of "type=Mass" which was used to imply "gender =Male", "weapons_obtained_legally =Yes", "race=white", and "prior_signs_mental_health_issues =Yes", as well as combinations of these. These rules tended to have higher coverage values likely because there is only one antecedent that is rather prevalent in the dataset, as "type=Mass" has a much larger distribution than "type=Spree". The lift values for all but one of the rules in Table III is also above 1 indicating that the majority of rules are relatively useful. However, the most interesting rules in Table III were "type=Mass →weapons_obtained_legally=Yes", ""type=Mass → gender=Male", and "type=Mass →prior_signs_mental_health_issues=Yes" and were added to Table VI accordingly.

In Table IV, all the data retained a consequent of "gender=Male" with confidence values of 0.976 or above. Given the fact that the distribution of the gender of the shooters in the dataset primarily favored "Male" these are not necessarily surprising rules to come across. Furthermore, all these rules had lift values above 1, indicating that they are relatively true per the consequent. Of these rules, "prior_signs_mental_health_issues=Yes, type=Mass→ gender=Male" was selected as the most interesting. This is not to say that the other rules are not necessarily interesting, but for the purpose of this paper, these rules were not determined to be the most interesting.

In Table V, the association rules were not able provide much insight on the mass shooting events. Unfortunately, it seems as though most of the rules with highest lift are those about the various combinations of quantity of victims or injured antecedent implying the total victims or injured consequent. Furthermore, these rules are based upon those relationships like total victims implying that there are more than 6.5 injured; however, that is not necessarily a proper set of association rules that can be easily generalized to unseen data. The data set only contains 124 instances, thus associating total victims only based upon injuries is not exactly a meaningful association. This is especially the case because quantities of injured and victims range so drastically pending on numerous different variables that are not necessarily monitored in this dataset (i.e., venue capacity requirements, intent of the shooter, etc.). As a result, none of these rules were selected as the interesting rules of the dataset.

Regarding the association rules in Table VI, six of these association rules were found in the association rules that were not shown in these tables. Each association rule was picked based upon the evaluation of support, confidence, lift, and generally interesting associations to see about the dataset. The ensuing discussion is meant to evaluate these association rules and their implications in no particular order.

The association rule “type=Mass→ gender=Male” is a very interesting one, as it has the highest support (0.823), second highest confidence (0.971) of those found in Table III, and a lift value of 1.012. Further, all the data in Table IV have a consequent of “gender=Male”, with all of them having confidence values 0.976 or above. Given the fact that the distribution of the gender of the shooters in the dataset primarily favored “Male” these are not necessarily surprising rules to come across. Furthermore, given the fact that it has been reported that 98% of mass shootings were committed by men, the association rule is most definitely not surprising [7]. This association rule is a good demonstration of the overwhelming majority of male mass shooters shown in the dataset. Hence, given these factors, it would suffice to say that this determined association rule (“type=Mass→ gender=Male”) could generalize well to unseen data.

The next especially interesting association rule was “type=Mass→ weapons_obtained_legally=Yes”. Although the support is only 0.597, the confidence is still at 0.705, which indicates that the association is frequently true and is generally applicable in the dataset. Additionally, the lift is 1.016 indicating that it is a relatively true per the consequent. This rule would imply that those that perform mass shootings obtained their weapons legally and then committed the crime. This could be explained by a combination of factors such as lax gun control, improper background checks, or (in the rare cases) corrupt “legal” gun selling. Corrupt “legal” gun selling would be like if gun permits were given out in return for bribes [8]. This rule may not necessarily be followed from intuition, as much of the time mass shooters are either portrayed as those with mental disorders that stole a gun from their parents or someone they know (like the Columbine mass shooting). However, in general, the association rule may be able to generalize to unseen data.

The “type=Mass → prior_signs_mental_health_issues=Yes” association rule is another that was particularly interesting. The support for this association rule is 0.419 with a confidence of 0.495, indicating that it is often applicable and frequently true. The lift is also 0.99, indicating that it may not be as frequently true per the consequent as the other rules, but still relatively true. Although this association seems to have grounds to be a relatively strong association rule, it has been one that has caused much misinformation and misunderstanding of the mental health community. Many media frenzies that circulated after these mass shooting events, especially in those before 2010, seemed to demonize those with mental health issues. Citing them as worrisome that they may have a mental breakdown and do the same thing. Consequentially and unfortunately, it seems that this rule would be intuitive given the exposure to these media frenzies. The implications of this rule would insinuate that mass shootings would have a shooter that had previous signs of mental health issues, which does not mean that all those with mental health issues are to-be mass shooters. Hence, if the mass shooter had mental health issues that may have led to their ultimate decision to perform said atrocity because they may have not been in the right mindset, they may have been experiencing some sort of rage episode, or they had decided that if they were going to kill themselves, they may as well take some others down with them. All these examples were from instances within the dataset (specifically in the “mental_health_details” feature). Overall, this is a relatively strong association rule, but, pending upon further data and understanding of a shooter’s mental health backgrounds, it may not necessarily generalize to unseen data well. This is mainly since, realistically, not every mass shooter has mental health issues.

The association rule “prior_signs_mental_health_issues=Yes, type=Mass → gender=Male” has a support of 0.411 with a confidence of 0.981. Thus, this rule is applicable relatively often, but it is generally true almost all the time. The lift is 1.022 indicates it is generally true per the consequent. This rule implies that if the shooter has mental health issues, then they are likely to be male. Also, the association rule “prior_signs_mental_health_issues=Unclear, type=Mass → gender=Male” has a support of 0.29, with a confidence score of 0.947, which indicates that the association is generally true, but the rule is applicable slightly less often. The lift for this rule is 0.987, indicating that it may not be as frequently true per the consequent as the other rules in this list. The implication in this association rule is that if mental health issues are unclear prior to the event, then the shooter is likely to be male. As stated, these associations with the consequent being male may be highly attributed to the large distribution of male shooters. However, it is also worthy to note that mental health issues encompass more than just depression and suicide. Anger, or difficulty controlling anger, can also be considered a mental health issue. Furthermore, intuitively, men tend to be more physically aggressive, are more likely than women to impulsively deal with their anger, and have more motive for revenge [9]. Hence this association could imply that the males with these mental health issues may have been acting either impulsively, with the intent of revenge, or even with the intent to feel better. Regarding the “prior_signs_mental_health_issues=Yes, type=Mass → gender=Male” rule, it does seem to be a strong association rule and may generalize well to unseen data due to the aforementioned rationales. However, for the “prior_signs_mental_health_issues=Unclear, type=Mass → gender=Male” rule, it may be an indicator that men may act more impulsively or with the intent of revenge that these mass shooting occur. In this case, since mental health may still play a role because it is unknown, this association rule may not generalize well to unseen data, unless more data can be supplied to prove the association.

The association rule “fatalities=≥ 6.5, gender=Male, type=Mass → prior_signs_mental_health_issues=Yes” has a support of 0.202 with a confidence of 0.595, which means that the rule is not necessarily applicable often but when it does it is true a good portion of the time. Furthermore, the lift is 1.19 indicating that there is positive correlation between the antecedent and the consequent as well which would mean it is frequently true per the consequent. This rule was interesting because it suggests that given that if a mass shooting event occurs and there are more than or equal to 6.5 fatalities (~7 fatalities due to rounding), then the male shooter may have had prior mental health issues. This could be suggesting that shooters with mental health issues may lead to higher fatality numbers for the mass shooting event, especially given that the interval of discretization made ≥ 6.5 fatalities the highest fatality range of the dataset. This rule may also suggest that those with the mental health issues had intent to kill rather than just to injure, which seems obvious given that they are performing a mass shooting, but this may also be an indication that those with the mental health issues may have had prior homicidal thoughts before the event that fueled the desire to perform the event. The association rule seems that it may generalize well to unseen data, given that there are cases in which those with the right circumstance of mental health problems may be performing the act to kill as many people as possible.

The association rule "total_victims=≥ 15.5, gender=Male, type=Mass →weapons_obtained_legally=Yes" has a support of 0.21 with confidence of 0.867, which indicates that although the association is not always applicable to the dataset, the rule is true for most of the time that it does appear. This rule also has a lift of 1.25, which indicates it is frequently true per the consequent. Furthermore, the rule "prior_signs_mental_health_issues=Yes, type=Mass →weapons_obtained_legally=Yes" can also be related to this rule. This second rule has a support of 0.331 and confidence of 0.788, which indicates that it is more applicable to the dataset than the other rule and is frequently true the times it does appear. This rule also has a lift of 1.137, indicates it is frequently true per the consequent. The combination of these two rules could imply that these mass shooting events may be planned rather than impulsive. To explain, the first rule is an indication that given the number of victims (both injured and killed) being higher than 15.5 and the shooter is male, then the weapons were obtained legally. The weapons may have been bought ahead of time in planning for the event and done so to increase their number of victims. Furthermore, the mental health issue rule could imply that there may have been some form of mindset they had in planning for the event, whether it is one of lack of anger management, tendency to have homicidal thoughts, or even antisocial personality disorder. Regardless in both cases, the guns were obtained legally which could imply that they did not do this out of impulse and there could have been planning for it. Overall, this is speculation, but based upon theses association rules it may be possible that the data could generalize well for unseen future data.

The association rule "prior_signs_mental_health_issues=Yes, type=Mass → weapons_obtained_legally=Yes, race=white, gender=Male" had a support of 0.226 with confidence of 0.538. The association may not be applicable as often, but it does seem to be true a good portion of the time. Further the lift value for this rule is 1.363, which is the highest lift value of the interesting rules selected, thus indicating it is very frequently true per the consequent. This rule implies that given prior signs of mental health issues for a mass shooter, the shooter obtained their guns legally, and that they were a white male. This rule is interesting because it basically profiles the culprit of a mass shooting based upon if there were prior mental health issues with the shooter. Given the fact that the distributions of shooters that are white and shooters that are males are as high as they are, this rule would seem to generalize very well to other unseen data, given that the trend continues.

Finally, the association rule "prior_signs_mental_health_issues=Yes, race=white, gender=Male, type=Mass →weapons_obtained_legally=Yes" has a support of 0.226 with a confidence of 0.848, which insinuates that although this association may not be applicable often, the rule has high accuracy when it does show up. Further, the lift score is 1.223 which indicates it is frequently true per the consequent. The rule implies that given a shooter being a white male with mental health issues, they obtained their guns legally. This association is a somewhat surprising rule. This rule implies that even though individuals had prior signs of mental health issues, they were able to pass the background check to obtain guns legally given that they were white males. This consequently begs the question of the restrictions of background checks to obtaining a weapon regarding mental health. The federal law for mental health illness and gun ownership is based upon whether an individual has been deemed a danger to themselves, others, or cannot manage their own affairs by a court or if they have been involuntarily hospitalized or committed to a mental health facility by a lawful authority [10]. It is possible that the race association is also related to how background checks are done. Perhaps it is possible that the disproportionate number of other races that are imprisoned compared to white people may cause the background checks to have a sort of inherent racist bias primarily based on the data that it has to go off, which may lead to other races not passing the background check and not being able to legally purchase a gun as easy as white people can [11]. Overall, this rule seems to strong association rule and could likely be applied to unseen data with little to no problems.

V. Conclusions

Throughout the course of this paper, an association rule analysis was used to determine potential association rules between instances of mass shooting events. It was determined that of the 534 rules determined, ten of them were selected for overall interest to discuss and explore the implications of the associations. Of these ten, only two would not likely be generalized to unseen data well, however eight of them likely could and were deemed the most useful rules determined by the association rule analysis. These associations provided a unique insight on mass shooting events and further helped to identify some of the correlations between underlying causes of the events themselves. To fully elucidate the associations and their legitimacy/applicability, it would be ideal to perform further data analysis and gathering of information for each of these associations. Furthermore, if these associations are deemed legitimate and applicable for future data, it may lead to the development of social programs and or machine learning algorithms to prevent or reduce the frequency of such heinous atrocities occurring.

REFERENCES

[1] J. GRAMLICH, “What the data says about gun deaths in the U.S.”, Pew Research Center, August 16, 2019. [Online]. Available: https://www.pewresearch.org/fact-tank/2019/08/16/what-the-data-says-about-gun-deaths-in-the-u-s/. [Accessed: June 7, 2021].

[2] D. M. Studdert, Y.Zhang, J. A. Rodden, R. J. Hyndman, “Handgun Acquisitions in California After Two Mass Shootings”, Annals of Internal Medicine, vol. 166, no.10, May 16, 2017. [Online]. Availble: https://www.acpjournals.org/doi/10.7326/M16-1574. [Accessed: June 7, 2021].

[3] D. J. Richardson, “Mass Murder with Firearms: Incidents and Victims, 1999-2013”, Congressional Research Service, July 30, 2015. [Online]. Available: https://fas.org/sgp/crs/misc/R44126.pdf. [Accessed: June 7, 2021].

[4] UC Davis Health, “Facts and Figures”, Ucdavis.edu, 2019. [Online]. Available: https://health.ucdavis.edu/what-you-can-do/facts.html. [Accessed: June 7, 2021].

[5] Data.world, “US Mass Shootings - dataset by awram”, data.world. [Dataset]. Available: https://data.world/awram/us-mass-shootings. [Accessed: June 7,2021].

[6] C. Montaldo, “Mass Murderers, Spree and Serial Killers”, ThoughtCo, June23, 2019. [Online]. Avaliable: https://www.thoughtco.com/defining-mass-spree-and-serial-killers-973123. [Accessed: June 7, 2021].

[7] TCR Staff, “Why Are Almost All Mass Shooters Men?”, The Crime Report, Mar. 29, 2021. [Online]. Available: https://thecrimereport.org/2021/03/29/why-are-almost-all-mass-shooters-men/. [Accessed: June 7, 2021].

[8] K. Whitehouse, “Ex-cop: NYPD gun license division was a bribery machine”, New York Post, April 17, 2018. [Online]. Available: https://nypost.com/2018/04/17/ex-cop-nypd-gun-license-division-was-a-bribery-machine/. [Accessed: June 9, 2021].

[9] M. Dittmann, “Anger across the gender divide”, Monitor on psychology, vol. 34, no. 3, March 2003. [Online] Available: https://www.apa.org/monitor/mar03/angeracross. [Accessed: June 9, 2021].

[10] Giffords Law Center. “Mental Health Record Reporting”, Giffords Law Center, Sept. 13, 2018. [Online] Available: https://giffords.org/lawcenter/gun-laws/policy-areas/background-checks/mental-health-reporting/. [Accessed: June 12, 2021].

[11] UC Berkeley School of Information, “Racial Bias in the National Instant Criminal Background Check System”, UC Berkeley School of Information, February 27, 2018. [Online]. Available: https://blogs.ischool.berkeley.edu/w231/2018/02/27/racial-bias-in-the-national-instant-criminal-background-check-system/. [Accessed: June 12, 2021].