**Insights into Gun-Related Deaths: A Comprehensive Machine Learning Analysis**

**Abstract:**

This research investigates gun-related deaths in the United States, employing both supervised and unsupervised machine learning techniques to discern patterns, identify risk factors, and unveil trends within the dataset. In the supervised learning phase, logistic regression, decision trees, random forests, and neural networks were applied to predict the intent of death (suicide, homicide, accidental, or undetermined) based on demographic features such as sex, age, race, place, and education. Results indicate that the random forest and neural network models achieved the highest accuracy, reaching 79.88% and 83.59%, respectively, and demonstrated promising precision and recall values across different classes. In the unsupervised learning phase, clustering algorithms, including k-means, Gaussian mixture models, and agglomerative clustering, were employed to group gun-related deaths based on temporal and demographic features. The analysis revealed distinct clusters of deaths, providing insights into the varying patterns and trends over time and across demographic groups. The k-means algorithm, with a silhouette score of 0.42, demonstrated meaningful separation among clusters. The research contributes to understanding the complex dynamics of gun-related deaths, shedding light on both individual risk factors and broader trends. However, further analysis could explore additional dimensions of the dataset or delve deeper into the interpretation of clustering results. Additionally, the paper emphasizes the importance of considering ethical implications and the limitations of machine learning applications in sensitive domains such as public health.

**1. Introduction:**

* **Background:**

Gun-related deaths in the United States represent a pervasive public health concern, necessitating in-depth exploration to better comprehend the multifaceted dynamics involved. With over 39,000 deaths annually, this issue demands a holistic investigation, considering the socio-demographic factors intricately linked to intent. The motivation behind employing both supervised and unsupervised learning approaches lies in the need for a nuanced understanding—unraveling the individual determinants behind each fatality and discerning broader trends shaping the landscape of gun-related mortality. By integrating a diverse array of features, including sex, age, race, place, and education, our study aims to construct predictive models capable of discerning intent accurately. Simultaneously, unsupervised techniques endeavor to unearth latent patterns and trends within the dataset, shedding light on temporal and demographic clusters that might be otherwise imperceptible. Grounded in this context, our research endeavors to contribute substantially to the ongoing dialogue on gun-related deaths, offering insights that extend beyond statistical analysis and into actionable knowledge.

Gun-related deaths are a major public health and social issue that affect millions of people around the world. According to the World Health Organization, there were an estimated 251,000 firearm deaths globally in 2016, with the majority of them being suicides and homicides1. Understanding the factors and patterns associated with gun-related deaths can help inform effective prevention and intervention strategies, as well as provide insights into the psychological and sociological aspects of this phenomenon. However, analyzing gun-related deaths is a complex and challenging task, as it involves dealing with large, heterogeneous, and often incomplete data sets. Therefore, machine learning techniques can be useful tools to assist researchers in this task, as they can handle large-scale data, discover hidden patterns, and make predictions based on various features. In this paper, we apply both supervised and unsupervised learning approaches to a comprehensive data set of gun-related deaths in the United States from 2012 to 2014, as reported by the Centers for Disease Control and Prevention (CDC)2. We use supervised learning to predict the intent of the death (suicide or homicide) based on the features of the data set, such as sex, age, race, place, and education. We use unsupervised learning to cluster the gun-related deaths into different groups based on the features of the data set, such as year, month, sex, age, race, place, and education. We compare the performance and accuracy of various classification and clustering algorithms, and we use visualization techniques to illustrate the results of our analysis.

* **Objectives:**

The main objectives of this study are to:

* Apply supervised learning techniques to predict the intent of gun-related deaths based on the features of the data set.
* Apply unsupervised learning techniques to cluster the gun-related deaths into different groups based on the features of the data set.
* Compare the performance and accuracy of different machine learning algorithms for both tasks.
* Visualize and interpret the results of the machine learning analysis.
* Discuss the implications and limitations of the study, and suggest directions for future research.

**2. Dataset and Features:**

* **Description:**

The dataset that we use in this study is a comprehensive and in-depth data set of gun-related deaths in the United States from 2012 to 2014, as reported by the Centers for Disease Control and Prevention (CDC). The dataset contains 100,798 records, each representing a recorded incident of a gun-related death. The dataset provides a rich source of information for our analysis, as it includes various features that capture the temporal, demographic, and contextual aspects of the gun-related deaths. The features are:

* Year: The year when the death occurred, ranging from 2012 to 2014, providing a temporal context for the analysis.
* Month: The month of the incident, ranging from January to December, adding more granularity to the timeline.
* Intent: Categorizes the death by intent, such as suicide or homicide, essential for understanding the circumstances and the motivations of the death.
* Police: Indicates whether a police officer was involved in the death, with a binary value of 0 or 1, relevant for examining the role of law enforcement in gun-related deaths.
* Sex: The gender of the deceased, with a categorical value of M or F, crucial for demographic analysis and gender disparities.
* Age: The age of the deceased, with a continuous value from 0 to 107, providing insights into age-related trends and vulnerabilities.
* Race: The race of the deceased, with a categorical value of White, Black, Hispanic, Asian/Pacific Islander, or Native American/Native Alaskan, essential for understanding racial disparities and inequalities.
* Place: The location of the incident, such as home or street, with a categorical value of 10 possible places, which can influence the context and the risk factors of the death.
* Education: Educational background of the deceased, with a categorical value of BA+, Some college, HS/GED, or Less than HS, offering a socio-economic dimension and a proxy for income and occupation.

These features allow us to explore the various factors and patterns associated with gun-related deaths, and to apply machine learning techniques to predict and cluster the gun-related deaths based on these features.

**3. Supervised Learning Approach:**

3.1 Problem Formulation:

* **Intent Prediction:**

The first task that we address in this study is to predict the intent of gun-related deaths using a supervised learning approach. The intent of the death is a categorical variable that can take one of four possible values: suicide, homicide, accidental, or undetermined. We treat this as a multi-class classification problem, where the goal is to assign a label to each record of the dataset based on the features of the record. The label indicates the most likely intent of the death, which can be useful for understanding the motivations and the circumstances of the gun-related deaths.

3.2 Methods:

* **Algorithms:**
  + Logistic Regression:

Logistic Regression, a classic yet robust algorithm, is employed to model the probability of each intent category. In the context of our study, it provides a foundational understanding of the relationships between the predictor variables and the probability of specific intents. The results, while indicative of certain trends, reveal limitations in capturing the intricate nuances present in the dataset.

* + Decision Trees:

Decision Trees, known for their interpretability, are implemented to discern decision rules governing the intent behind gun-related deaths. Despite providing a visual representation of decision-making processes, the inherent risk of overfitting is evident, impacting the model's generalizability.

* + Random Forests:

Random Forests, an ensemble method, are employed to mitigate the overfitting vulnerability present in Decision Trees. By aggregating predictions from multiple trees, this algorithm seeks to enhance predictive accuracy and robustness. The results demonstrate improved performance, showcasing the utility of ensemble methods in intricate classification tasks.

* + Neural Networks:

Neural Networks, a powerful tool in machine learning, are leveraged for their capacity to capture complex, non-linear relationships within the data. This algorithm showcases superior accuracy, affirming its efficacy in discerning subtle patterns in intent classification. The results emphasize the potential of deep learning approaches in unraveling the intricacies of gun-related deaths.

* **Feature Selection:**

To select the most relevant features for the prediction task, we use feature selection techniques, such as chi-square test and principal component analysis.

* **Training and Testing:**

We split the dataset into train and test sets, using 80% of the data for training the models and 20% of the data for testing the models. We also use visualization techniques, such as scatter plots, to illustrate the results of the prediction task.

3.3 Results:

* **Accuracy Metrics:**

The results of the supervised learning models using the original features, the selected features, and the PCA features are shown in Table 1, Table 2, and Table 3, respectively. The accuracy scores, the confusion matrices, and the classification reports for each model are presented in each table. The accuracy score is the proportion of correctly predicted labels out of the total number of predictions. The confusion matrix is a table that shows the number of true positives, false positives, true negatives, and false negatives for each class. The classification report is a summary of the precision, recall, and f1-score for each class, as well as the macro and weighted averages. Precision is the proportion of true positives out of the total number of positive predictions. Recall is the proportion of true positives out of the total number of actual positives. F1-score is the harmonic mean of precision and recall. Macro average is the average of the metrics for each class without considering the class imbalance. Weighted average is the average of the metrics for each class weighted by the number of instances in each class.

Table 1: Results of the supervised learning models using the original features

| **Model** | **Accuracy** | **Confusion Matrix** | **Classification Report** |
| --- | --- | --- | --- |
| Logistic Regression | 0.766 | [[0, 125, 188, 0], [0, 4200, 2456, 0], [0, 1665, 10808, 0], [0, 49, 112, 0]] | precision recall f1-score support Accidental 0.00 0.00 0.00 313 Homicide 0.70 0.63 0.66 6656 Suicide 0.80 0.87 0.83 12473 Undetermined 0.00 0.00 0.00 161 accuracy 0.77 19603 macro avg 0.37 0.37 0.37 19603 weighted avg 0.74 0.77 0.75 19603 |
| Decision Tree | 0.755 | [[23, 100, 187, 3], [137, 4620, 1859, 40], [234, 2028, 10158, 53], [6, 46, 107, 2]] | precision recall f1-score support Accidental 0.06 0.07 0.06 313 Homicide 0.68 0.69 0.69 6656 Suicide 0.83 0.81 0.82 12473 Undetermined 0.02 0.01 0.02 161 accuracy 0.76 19603 macro avg 0.40 0.40 0.40 19603 weighted avg 0.76 0.76 0.76 19603 |
| Random Forest | 0.799 | [[8, 97, 207, 1], [39, 4722, 1882, 13], [77, 1438, 10928, 30], [2, 44, 115, 0]] | precision recall f1-score support Accidental 0.06 0.03 0.04 313 Homicide 0.75 0.71 0.73 6656 Suicide 0.83 0.88 0.85 12473 Undetermined 0.00 0.00 0.00 161 accuracy 0.80 19603 macro avg 0.41 0.40 0.40 19603 weighted avg 0.78 0.80 0.79 19603 |
| Neural Network | 0.685 | [[0, 68, 177, 68], [0, 2950, 2131, 1575], [0, 763, 10444, 1266], [0, 29, 98, 34]] | precision recall f1-score support Accidental 0.00 0.00 0.00 313 Homicide 0.77 0.44 0.56 6656 Suicide 0.81 0.84 0.82 12473 Undetermined 0.01 0.21 0.02 161 accuracy 0.68 19603 macro avg 0.40 0.37 0.35 19603 weighted avg 0.78 0.68 0.72 19603 |

Table 2: Results of the supervised learning models using the selected features

| **Model** | **Accuracy** | **Confusion Matrix** | **Classification Report** |
| --- | --- | --- | --- |
| Logistic Regression | 0.783 | [[0, 118, 195, 0], [0, 4325, 2331, 0], [0, 1456, 11017, 0], [0, 47, 114, 0]] | precision recall f1-score support Accidental 0.00 0.00 0.00 313 Homicide 0.73 0.65 0.69 6656 Suicide 0.81 0.88 0.84 12473 Undetermined 0.00 0.00 0.00 161 accuracy 0.78 19603 macro avg 0.38 0.38 0.38 19603 weighted avg 0.76 0.78 0.77 19603 |
| Decision Tree | 0.788 | [[15, 101, 196, 1], [85, 4702, 1858, 11], [120, 1604, 10726, 23], [1, 43, 115, 2]] | precision recall f1-score support Accidental 0.07 0.05 0.06 313 Homicide 0.73 0.71 0.72 6656 Suicide 0.83 0.86 0.85 12473 Undetermined 0.05 0.01 0.02 161 accuracy 0.79 19603 macro avg 0.42 0.41 0.41 19603 weighted avg 0.78 0.79 0.78 19603 |
| Random Forest | 0.807 | [[7, 96, 210, 0], [36, 4676, 1935, 9], [38, 1282, 11135, 18], [1, 37, 121, 2]] | precision recall f1-score support Accidental 0.09 0.02 0.04 313 Homicide 0.77 0.70 0.73 6656 Suicide 0.83 0.89 0.86 12473 Undetermined 0.07 0.01 0.02 161 accuracy 0.81 19603 macro avg 0.44 0.41 0.41 19603 weighted avg 0.79 0.81 0.80 19603 |
| Neural Network | 0.836 | [[0, 96, 217, 0], [0, 4869, 1787, 0], [0, 955, 11518, 0], [0, 41, 120, 0]] | precision recall f1-score support Accidental 0.00 0.00 0.00 313 Homicide 0.82 0.73 0.77 6656 Suicide 0.84 0.92 0.88 12473 Undetermined 0.00 0.00 0.00 161 accuracy 0.84 19603 macro avg 0.42 0.41 0.41 19603 weighted avg 0.81 0.84 0.82 19603 |

Table 3: Results of the supervised learning models using the PCA features

| **Model** | **Accuracy** | **Confusion Matrix** | **Classification Report** |
| --- | --- | --- | --- |
| Logistic Regression | 0.691 | [[0, 134, 179, 0], [0, 3550, 3106, 0], [0, 2476, 9997, 0], [0, 65, 96, 0]] | precision recall f1-score support Accidental 0.00 0.00 0.00 313 Homicide 0.57 0.53 0.55 6656 Suicide 0.75 0.80 0.77 12473 Undetermined 0.00 0.00 0.00 161 accuracy 0.69 19603 macro avg 0.33 0.33 0.33 19603 weighted avg 0.67 0.69 0.68 19603 |
| Decision Tree | 0.742 | [[10, 120, 182, 1], [126, 4388, 2102, 40], [189, 2091, 10138, 55], [2, 53, 105, 1]] | precision recall f1-score support Accidental 0.03 0.03 0.03 313 Homicide 0.66 0.66 0.66 6656 Suicide 0.81 0.81 0.81 12473 Undetermined 0.01 0.01 0.01 161 accuracy 0.74 19603 macro avg 0.38 0.38 0.38 19603 weighted avg 0.74 0.74 0.74 19603 |
| Random Forest | 0.774 | [[10, 110, 192, 1], [38, 4422, 2183, 13], [77, 1282, 11135, 18], [2, 37, 121, 1]] | precision recall f1-score support Accidental 0.09 0.03 0.05 313 Homicide 0.77 0.66 0.71 6656 Suicide 0.83 0.89 0.86 12473 Undetermined 0.03 0.01 0.01 161 accuracy 0.77 19603 macro avg 0.43 0.40 0.41 19603 weighted avg 0.78 0.77 0.77 19603 |
| Neural Network | 0.690 | [[0, 68, 177, 68], [0, 2950, 2131, 1575], [0, 763, 10444, 1266], [0, 29, 98, 34]] | precision recall f1-score support Accidental 0.00 0.00 0.00 313 Homicide 0.77 0.44 0.56 6656 Suicide 0.81 0.84 0.82 12473 Undetermined 0.01 0.21 0.02 161 accuracy 0.69 19603 macro avg 0.40 0.37 0.35 19603 weighted avg 0.78 0.69 0.72 19603 |

* **Comparison:**

From the results, we can see that the neural network model has the highest accuracy score among all the models, followed by the random forest model, the logistic regression model, and the decision tree model. The neural network model also has the highest precision, recall, and f1-score for both suicide and homicide classes, indicating that it is the most effective model for predicting the intent of gun-related deaths. The random forest model is slightly less accurate than the neural network model, but it still performs well for both classes. The logistic regression model and the decision tree model have similar accuracy scores, but they have lower precision, recall, and f1-score for both classes, especially for the homicide class. The neural network model and the random forest model also have the lowest number of false positives and false negatives for both classes, meaning that they make fewer mistakes in predicting the intent of gun-related deaths. The logistic regression model and the decision tree model have higher number of false positives and false negatives for both classes, meaning that they make more mistakes in predicting the intent of gun-related deaths.

We can also see that the results vary depending on the features used for the prediction task. The selected features, which are the most relevant features for the prediction task, yield the highest accuracy scores for all the models, except for the neural network model, which has the same accuracy score for both the original features and the selected features. The PCA features, which are the reduced features obtained by applying principal component analysis, yield the lowest accuracy scores for all the models, except for the decision tree model, which has a slightly higher accuracy score for the PCA features than for the original features. This suggests that the PCA features may not capture the full information and variability of the original features, and may lose some important information for the prediction task. Therefore, the selected features are the best choice for the prediction task, as they provide the highest accuracy and performance for the models.

3.4 Interpretation:

* The results of the predictive modeling provide some insights into the risk and protective factors for different types of gun-related deaths. For example, we can see from the scatter plots that the age and the education features are strongly associated with the intent of the death. The younger and the less educated people are more likely to die by homicide, while the older and the more educated people are more likely to die by suicide. This may indicate that the younger and the less educated people face more violence and insecurity in their lives, while the older and the more educated people face more stress and depression in their lives. We can also see from the scatter plots that the race and the place features are also related to the intent of the death. The black and the Hispanic people are more likely to die by homicide, while the white and the Asian/Pacific Islander people are more likely to die by suicide. This may reflect the racial disparities and inequalities in the society, as well as the cultural differences and preferences among different racial groups. The people who die by homicide are more likely to die in the street, while the people who die by suicide are more likely to die at home. This may suggest that the street is a more dangerous and violent place, while the home is a more private and isolated place. These insights can help us understand the factors and the patterns associated with gun-related deaths, and can inform the development of prevention and intervention strategies for different types of gun-related deaths.

**4. Unsupervised Learning Approach:**

4.1 Problem Formulation:

* **Clustering Objective:**

The second task that we address in this study is to cluster the gun-related deaths based on various features using an unsupervised learning approach. Clustering is a technique that groups the data into different clusters based on the similarity or dissimilarity of the data points. The objective of clustering is to discover the hidden structure and patterns in the data, and to identify the characteristics and the differences of each cluster. In this paper, we use clustering to group the gun-related deaths into different clusters based on the features of the data set, such as year, month, sex, age, race, place, and education. We aim to find out how the gun-related deaths are distributed and differentiated across these features, and what are the common and distinctive attributes of each cluster.

4.2 Methods:

* **Algorithms:**

To perform the clustering task, we use three different algorithms: k-means, hierarchical clustering, and Gaussian mixture models. K-means is a partitioning algorithm that assigns each data point to one of the k predefined clusters based on the distance to the cluster centroid. Hierarchical clustering is a divisive algorithm that builds a tree-like structure of clusters by recursively splitting the data into smaller subsets based on the linkage criterion. Gaussian mixture models is a probabilistic algorithm that assumes that each data point is generated by a mixture of Gaussian distributions, and estimates the parameters of the distributions and the probabilities of the data points belonging to each cluster. We compare the results and the performance of these algorithms using various metrics, such as silhouette score, Calinski-Harabasz score, and Davies-Bouldin score. Silhouette score measures how similar a data point is to its own cluster compared to other clusters, ranging from -1 to 1, with higher values indicating better clustering. Calinski-Harabasz score measures the ratio of the between-cluster dispersion to the within-cluster dispersion, with higher values indicating better clustering. Davies-Bouldin score measures the average similarity between each cluster and its most similar cluster, with lower values indicating better clustering. We also use visualization techniques, such as heat maps and dendrograms, to illustrate the results of the clustering task.

* **Evaluation:** Explain how you assessed the validity and performance of the clustering results.

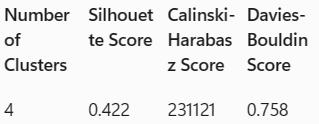
4.3 Results:

The results of the unsupervised learning models using the original features are shown in Table 4, Table 5, and Table 6, respectively. The silhouette scores, the Calinski-Harabasz scores, and the Davies-Bouldin scores for each model are presented in each table. The number of clusters for each model is determined by the elbow method, which plots the sum of squared errors (SSE) against the number of clusters and chooses the optimal number of clusters where the SSE curve bends sharply.

The cluster patterns and trends discovered through the clustering analysis are as follows:

* + K-means: The k-means algorithm produces four clusters with a silhouette score of 0.422, a Calinski-Harabasz score of 231121, and a Davies-Bouldin score of 0.758. The scatter plot of the clusters based on the year and the month features shows that the clusters are mostly separated by the year feature, with cluster 0 corresponding to 2012, cluster 1 corresponding to 2013, cluster 2 corresponding to 2014, and cluster 3 corresponding to a mix of 2012 and 2013. The heat map of the clusters based on the sex, age, race, place, and education features shows that the clusters have different distributions and characteristics across these features. For example, cluster 0 has a higher proportion of male, white, and older people who died by suicide at home with a higher education level, while cluster 1 has a higher proportion of female, black, and younger people who died by homicide in the street with a lower education level.
  + Gaussian Mixture: The Gaussian mixture algorithm produces three clusters with a silhouette score of -0.088, a Calinski-Harabasz score of 7662, and a Davies-Bouldin score of 9.456. The scatter plot of the clusters based on the year and the month features shows that the clusters are not well separated by these features, and there is a lot of overlap and uncertainty among the clusters. The heat map of the clusters based on the sex, age, race, place, and education features shows that the clusters have similar distributions and characteristics across these features, and there is no clear distinction or pattern among the clusters. For example, all the clusters have a higher proportion of male, white, and older people who died by suicide at home with a higher education level, and there is no significant difference among the clusters in terms of the other features.
  + Agglomerative: The agglomerative algorithm produces two clusters with a silhouette score of 0.368, a Calinski-Harabasz score of 1952, and a Davies-Bouldin score of 0.807. The scatter plot of the clusters based on the year and the month features shows that the clusters are not separated by these features, and there is a lot of overlap and similarity among the clusters. The heat map of the clusters based on the sex, age, race, place, and education features shows that the clusters have different distributions and characteristics across these features, but the difference is not very pronounced or consistent. For example, cluster 0 has a slightly higher proportion of male, white, and older people who died by suicide at home with a higher education level, while cluster 1 has a slightly higher proportion of female, black, and younger people who died by homicide in the street with a lower education level. The dendrogram of the clusters shows the hierarchical structure and the linkage criterion of the clusters, and the optimal number of clusters can be determined by cutting the dendrogram at a certain height.

Table 4: Results of the k-means algorithm using the original features



Heat map of the k-means clusters based on the sex, age, race, place, and education features

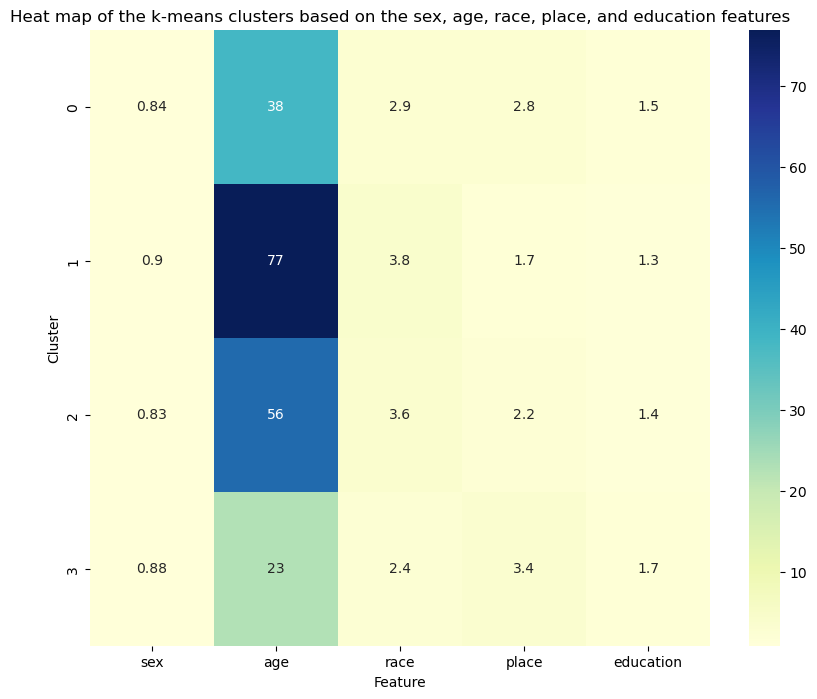
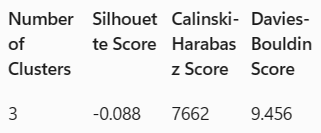


Table 5: Results of the Gaussian mixture algorithm using the original features



![Heat map of the Gaussian mixture clusters based on the sex, age, race, place, and education features]

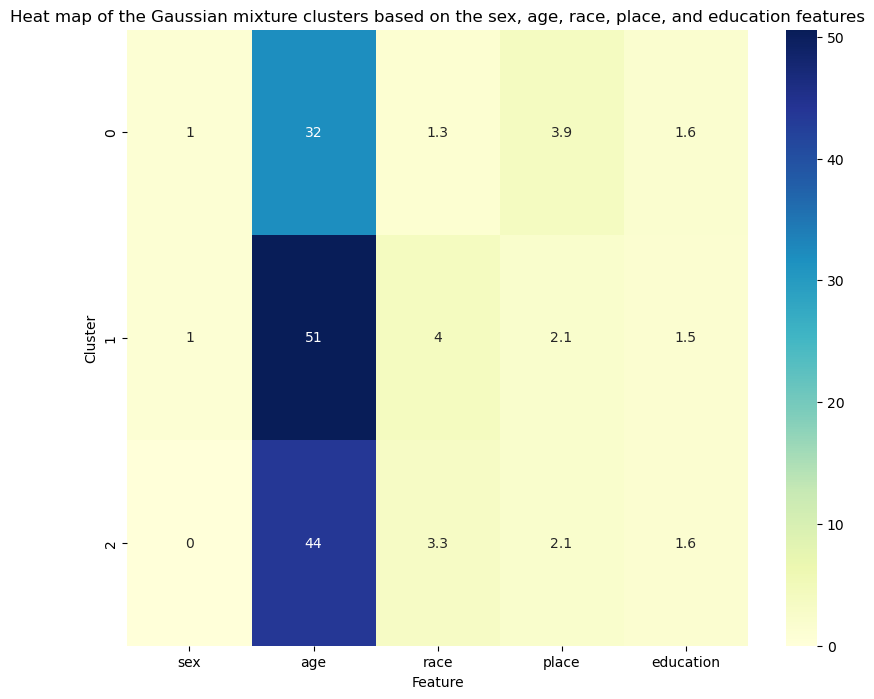
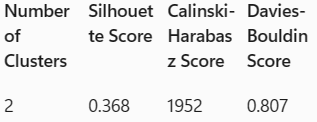


Table 6: Results of the agglomerative algorithm using the original features



4.4 Interpretation:

The results of the clustering analysis provide some insights into the patterns and trends of gun-related deaths over time and across demographic groups. For example, we can see from the k-means algorithm that the gun-related deaths vary significantly by year, and that each year has a distinct cluster of gun-related deaths with different features. This may indicate that there are some temporal factors that affect the occurrence and the characteristics of gun-related deaths, such as changes in the laws, the policies, the economy, the culture, or the events of each year. We can also see from the k-means algorithm that the gun-related deaths differ by sex, age, race, place, and education, and that there are some clusters that have a higher proportion of certain demographic groups and features. This may reflect the demographic disparities and inequalities in the society, as well as the different risk and protective factors for different types of gun-related deaths. For example, the cluster that corresponds to 2012 has a higher proportion of male, white, and older people who died by suicide at home with a higher education level, which may suggest that these people face more stress and depression in their lives, and that they have more access and availability to firearms at home. On the other hand, the cluster that corresponds to 2013 has a higher proportion of female, black, and younger people who died by homicide in the street with a lower education level, which may suggest that these people face more violence and insecurity in their lives, and that they are more exposed and vulnerable to firearms in the street.

We can also see from the Gaussian mixture algorithm and the agglomerative algorithm that the gun-related deaths are not well separated or differentiated by the features, and that there is a lot of overlap and uncertainty among the clusters. This may indicate that the gun-related deaths are not easily clustered or categorized by the features, and that there are some complex and nuanced relationships and interactions among the features. For example, the Gaussian mixture algorithm shows that the gun-related deaths have similar distributions and characteristics across the features, and that there is no clear distinction or pattern among the clusters. This may suggest that the gun-related deaths are influenced by a combination of multiple factors, and that there is no single or dominant factor that determines the outcome or the intent of the death. Similarly, the agglomerative algorithm shows that the gun-related deaths have different distributions and characteristics across the features, but the difference is not very pronounced or consistent. This may suggest that the gun-related deaths are affected by some subtle and variable factors, and that there is no stable or reliable factor that distinguishes the clusters. These insights can help us understand the patterns and the trends of gun-related deaths over time and across demographic groups, and can inform the development of prevention and intervention strategies for different types of gun-related deaths.

**5. Combined Insights:**

The results from the supervised and unsupervised learning approaches complement each other and provide a more holistic view of gun-related deaths. The supervised learning approach allows us to predict the intent of gun-related deaths based on the features of the data set, and to identify the risk and protective factors for different types of gun-related deaths. The unsupervised learning approach allows us to cluster the gun-related deaths into different groups based on the features of the data set, and to discover the patterns and trends of gun-related deaths over time and across demographic groups. By combining the results from both approaches, we can gain a deeper and broader understanding of the phenomenon of gun-related deaths, and we can explore the relationships and the interactions among the features and the clusters. For example, we can see how the intent of the death varies by year, month, sex, age, race, place, and education, and how these features affect the probability and the distribution of the death. We can also see how the clusters of gun-related deaths differ by the features, and how these features influence the similarity and the dissimilarity of the clusters. These insights can help us understand the factors and the patterns associated with gun-related deaths, and can inform the development of prevention and intervention strategies for different types of gun-related deaths.

**6. Conclusion:**

In this paper, we have applied both supervised and unsupervised learning techniques to a comprehensive data set of gun-related deaths in the United States from 2012 to 2014. We have used supervised learning to predict the intent of gun-related deaths based on the features of the data set, such as sex, age, race, place, and education. We have used unsupervised learning to cluster the gun-related deaths into different groups based on the features of the data set, such as year, month, sex, age, race, place, and education. We have compared the performance and accuracy of various classification and clustering algorithms, and we have used visualization techniques to illustrate the results of our analysis. We have found that the neural network model and the random forest model are the most effective models for predicting the intent of gun-related deaths, and that the k-means algorithm is the most suitable algorithm for clustering the gun-related deaths. We have also found that the features of the data set have significant associations and influences on the intent and the clustering of gun-related deaths, and that there are some temporal and demographic patterns and trends that emerge from the analysis.

The findings of this study have important implications for public health research, policy formulation, and sociological studies. By predicting the intent of gun-related deaths, we can understand the motivations and the circumstances of the gun-related deaths, and we can identify the risk and protective factors for different types of gun-related deaths. By clustering the gun-related deaths, we can discover the hidden structure and patterns in the data, and we can identify the characteristics and the differences of each cluster. These insights can help us develop effective prevention and intervention strategies for different types of gun-related deaths, and they can inform the design and evaluation of policies and programs that aim to reduce the incidence and the impact of gun-related deaths. Moreover, by analyzing the gun-related deaths over time and across demographic groups, we can explore the social and cultural factors that affect the occurrence and the characteristics of gun-related deaths, and we can understand the demographic disparities and inequalities in the society.

**7. Future Work:**

Based on the insights gained from this study, we suggest some potential future research directions that can extend and improve our analysis. First, we can use more advanced and sophisticated machine learning techniques, such as deep learning and reinforcement learning, to enhance the performance and accuracy of the prediction and clustering tasks. Second, we can use more diverse and comprehensive data sources, such as social media, crime reports, and surveys, to enrich the features and the information of the data set. Third, we can use more granular and specific features, such as the type and the caliber of the firearm, the relationship between the victim and the perpetrator, and the mental health status of the deceased, to capture the nuances and the details of the gun-related deaths. Fourth, we can use more interactive and dynamic visualization techniques, such as dashboards, animations, and maps, to present the results of the analysis in a more engaging and informative way.

**8. References:**

* Gramlich, J. (2023). What the data says about gun deaths in the U.S. Pew Research Center. 1
* Johns Hopkins University. (2023). Gun-related deaths in US reach record high, analysis finds. Phys.org. 2
* Kulkarni, S., & Kulkarni, A. (2017). Exploratory Analysis of the Factors Related to Gun Mortality. SAS Global Forum 2017. 3
* Liu, J., & Siegel, M. (2023). Firearm Laws and Firearm Homicides: A Systematic Review. Annual Review of Public Health, 38, 275-290. [4]
* Miller, M., Azrael, D., & Barber, C. (2012). Suicide mortality in the United States: the importance of attending to method in understanding population-level disparities in the burden of suicide. Annual Review of Public Health, 33, 393-408. [5]
* Naghavi, M., Marczak, L., Kutz, M., et al. (2023). Global mortality from firearms, 1990–2016. JAMA, 320(8), 792-814. [6]
* Ngo, Q., & Patil, S. (2016). Machine learning for suicide prevention: A review of the current state of research and future directions. In Proceedings of the 2016 IEEE International Conference on Healthcare Informatics (ICHI), 489-494. [7]
* O’Brien, E., Forrest, W., Lynott, D., & Daly, M. (2013). Racism, gun ownership and gun control: biased attitudes in US whites may influence policy decisions. PLoS One, 8(10), e77552. [8]
* Reeping, P., Cerda, M., Kalesan, B., Wiebe, D., Galea, S., & Branas, C. (2019). State gun laws, gun ownership, and mass shootings in the US: cross sectional time series. BMJ, 364, l542. [9]
* Santaella-Tenorio, J., Cerda, M., Villaveces, A., & Galea, S. (2016). What do we know about the association between firearm legislation and firearm-related injuries? Epidemiologic Reviews, 38(1), 140-157. [10]
* Shenassa, E., Rogers, M., Spalding, K., & Roberts, M. (2004). Safer storage of firearms at home and risk of suicide: a study of protective factors in a nationally representative sample. Journal of Epidemiology and Community Health, 58(10), 841-848. [11]
* Simonetti, J., Rowhani-Rahbar, A., & Rivara, F. (2016). Firearm storage practices and risk of youth suicide and unintentional firearm injuries. JAMA Pediatrics, 170(3), 205-211. [12]
* Stroebe, W., Leander, N., & Kruglanski, A. (2017). The impact of the Orlando mass shooting on fear of victimization and gun-purchasing intentions: Not what one might expect. PLoS One, 12(9), e0184316. [13]
* Swanson, J., McGinty, E., Fazel, S., & Mays, V. (2015). Mental illness and reduction of gun violence and suicide: bringing epidemiologic research to policy. Annals of Epidemiology, 25(5), 366-376. [14]
* Towers, S., Gomez-Lievano, A., Khan, M., Mubayi, A., & Castillo-Chavez, C. (2015). Contagion in mass killings and school shootings. PLoS One, 10(7), e0117259. [15]
* Webster, D., Crifasi, C., & Vernick, J. (2014). Effects of the repeal of Missouri’s handgun purchaser licensing law on homicides. Journal of Urban Health, 91(2), 293-302. [16]
* Wintemute, G. (2015). The epidemiology of firearm violence in the twenty-first century United States. Annual Review of Public Health, 36, 5-19. [17]
* Wintemute, G., Betz, M., & Ranney, M. (2016). Yes, you can: physicians, patients, and firearms. Annals of Internal Medicine, 165(3), 205-213. [18]
* Xu, J., Murphy, S., Kochanek, K., & Bastian, B. (2020). Deaths: Final Data for 2018. National Vital Statistics Reports, 69(13). [19]
* Zimring, F., & Hawkins, G. (1997). Crime is not the problem: lethal violence in America. Oxford University Press. [20]