An Overdose of Overdoses: *A County-level Risk Analysis of Fatal Drug Overdoses in America and Relevant Solutions*

The Unit Vector: Team # 19336

April 6, 2025

The Modeling the Future Challenge competition (Competition) is a research competition for high school students sponsored by The Actuarial Foundation and the Institute of Competition Sciences (collectively, the Sponsors) solely for educational purposes.

The mathematical models, conclusions, concepts, ideas, proposals, recommendations, presentations, methods, practices, sources, videos, graphs, tables, charts, assumptions, conclusions, and other information presented by the students (including the winning students) in connection with the Competition (collectively, the "information") are created solely by the students for use in connection with the competition and as a learning tool. The information has not been validated, tested or otherwise confirmed. The information may not be accurate and may not be (i) used or relied upon for any reason, (ii) cited or quoted in a way that would imply accuracy, or (iii) presented as fact. The Sponsors have not validated, and do not "approve" or "endorse" the information or the competitors (including the winners) and the information may not be cited in any way that would imply such approval or endorsement. The students competing in the Competition are not authorized to speak on behalf of the Sponsors or the Competition. Any articles, appearances, interviews or quotes issued by competitors are done in their individual capacity and not on behalf of, or in connection with, the Sponsors or the Competition (unless specially authorized or published by the Sponsors).

By viewing or otherwise accessing the information, you expressly assume all risk of loss, harm or injury resulting from the use or misuse of such information. The Sponsors do not make any guarantee, representation, or warranty, express or implied, at law or in equity, and the Sponsors expressly disclaim all such guarantees, representations or warranties whatsoever, as to the validity, accuracy or sufficiency of any of the information. The Sponsors (including, without limitation, their respective directors, officers, volunteers, employees, agents, attorneys and members) are not responsible for any injuries, claims, losses or damages to persons or property that a viewer (or any third party) may incur arising, directly or indirectly, out of or as a result of any actual or alleged libelous statements; infringement of intellectual property or privacy rights; product liability, whether resulting from negligence or otherwise; or from any use or reliance on any of the information.

Executive Summary:

Since 1999, drug overdoses have claimed the lives of nearly 1 million Americans cumulatively [1]. In 2021, the Centers for Disease Control (CDC) reported that fatal overdoses cost $550 billion that year alone [4]. Safe to say, fatal drug overdoses are an urgent societal malfunction in dire need of amelioration. Thankfully, overdose deaths have begun to drop since 2022 due to medical breakthroughs and government policies and reached a 4-year low of just over 84,000 deaths [22,29]. However, there is still opportunity to reduce risk by accelerating this decline of overdose mortalities, which we aim to analyze in this report.

The risks of fatal overdoses are deaths themselves and the costs resulting not only from drug overdose treatment, but also lost productivity costs that victims would have been able to contribute to the economy had they survived [34]. At a broader level, these costs also impact the government, insurance firms, and society as a whole from the lost productivity the workforce suffers from [35].

To analyze risk and generate mitigation strategies, we construct two different types of mathematical models: The first set of models is classification-based models focused on identifying socio-economic, health and geographic county adjacency to predict counties at a high or low risk of overdose mortality. The second set of models are projection models centered on projecting risk of overdose deaths and their induced monetary loss into the future. The main data we utilize to create and interpret these models consists of the following:

- County level socio-economic and health data obtained from University of Wisconsin's County Health Rankings to create the classifier models [37].
- National overdose mortality data by year to use for the models which project risk of death and expenses [22].
- County Shape Files from the government census to generate a map of all counties for various tasks [24,25].
- 2023 Population Estimates by county from the government census to obtain timely data [36].

For the classification part of the mathematical analysis, we classify counties as "high" or "low" based on their overdose mortality rates. Then we fit a baseline Random Forest and then a Graph Neural Network (GNN), a new type of neural network, to predict the high/low category based on not only county features, but also county adjacency. We find the GNN is the superior model because of its AUC-ROC score of .81, and it allows for wide-scale geographic feature importance interpretation unlike the Random Forest.

For the forecasting portion, we build an ARIMA and Monte Carlo extension of the ARIMA to predict mortality until 2030, and find from the ARIMA that fatal overdoses in America are expected to drop by 9.954% from now to 2030. We use Value of a Statistical Life (VSL) estimates ranging from a low of 5.98 million to a high of $12.78 million to quantify the monetary cost for each life lost. In our Risk Analysis, we use the GNN to identify socioeconomic factors which contribute to and characterize risk of overdose mortality in counties. Then, we use the ARIMA's mean estimate for the number of overdose mortalities to find that in 2030, the expected value of loss for fatal drug overdoses is more than $458 billion (using the low VSL of $5.98 million).

We make recommendations related to social and private insurance which may make them more effective. Based on our models, we also recommend that counties implement wide-scale Take Home Naloxone (THN) programs. We found that in Limestone County, AL (with the lowest overdose mortality rate in the nation), THN programs can save $15.60 million dollars in 2030 even with conservative assumptions such as a low reversal success rate and the low VSL. We also recommend neighboring counties implement smoking deterrence campaigns, as smoking was found to be the 2nd highest predictor of a high overdose mortality from the GNN. We find that if 4 counties in Ohio (Adams, Brown, Pike, Scioto) reduce their smoking by just 5%, they will be saving $27.67 million collectively in 2030, even accounting for the cost of implementing smoking deterrence in the first place.

Table of Contents:

1. Introduction and Background 4
1.1 Problem Statement & Debrief 5
2. Data Methodology 5
2.1 Data Collection 5
2.2 Data Cleaning and Preprocessing 7
2.3 Classifier Feature Selection 7
2.4 Exploratory Data Analysis 8
2.5 Software & Opensource Libraries Used 10
3. Mathematics Methodology 10
3.1 Classification of High-risk Counties 10
3.1.1 Classifier Model Assumptions 11
3.1.2 Random Forest 12
3.1.3 Random Forest Evaluation 12
3.1.4 Graph Neural Network 13
3.1.5 Classifier Selection & Results 14
3.2 Projection of Future Risk of Loss 14
3.2.1 Assumptions 15
3.2.2 ARIMA Forecasting 15
3.2.3 Monte Carlo Simulation 16
4. Risk Analysis 17
4.1 Risk Characterization 17
4.2 Risk Projection 18
4.3 Risk Mitigation Strategy Analysis 20
5. Recommendations 21
5.1 Insurance 21
5.2 Modifying Outcomes (THN) 22
THN Cost-Benefit Analysis 22
5.3 Behavior Change (Smoking Deterrence) 23
Cost-Benefit Analysis 24
6. Conclusion 25
7. Appendix 26
Appendix 1: Brown, Adams, Pike, Scioto counties in Ohio: 26
Appendix 2: Limestone County, AL Map: 26
Appendix 3: Top of a Random Forest Decision Tree: 27
Appendix 4: GitHub Link to all Code 27
8. Acknowledgements 27
9. References 29

1. Introduction and Background

America is one of the most economically developed countries in the world. However, one pervasive societal issue has remained underlying in everyday American life for decades: drug abuse and its induced deaths. The U.S. Department of Health and Human Services defines these drug overdose deaths to be "mortalities resulting from unintentional or intentional overdose of a drug, being given the wrong drug, taking a drug in error, or taking a drug inadvertently" [2].

The size and scope of fatal drug overdoses is extreme, to say the least. Since 1999, drug overdoses have claimed the lives of nearly 1 million Americans cumulatively [1]. Moreover, the CDC reports that in 2022 alone, over 105,000 American deaths were results of drug overdoses, with an average rate of 32.6 deaths per 100,000 people [22, 26]. This presents an upward trend in American overdose mortality for almost all types of drugs in recent years up to 2022. Fentanyl overdoses specifically, have been most responsible for this substantial uptick, with methamphetamine taking second place [23]. Figure 1 illustrates this trend in clear detail.

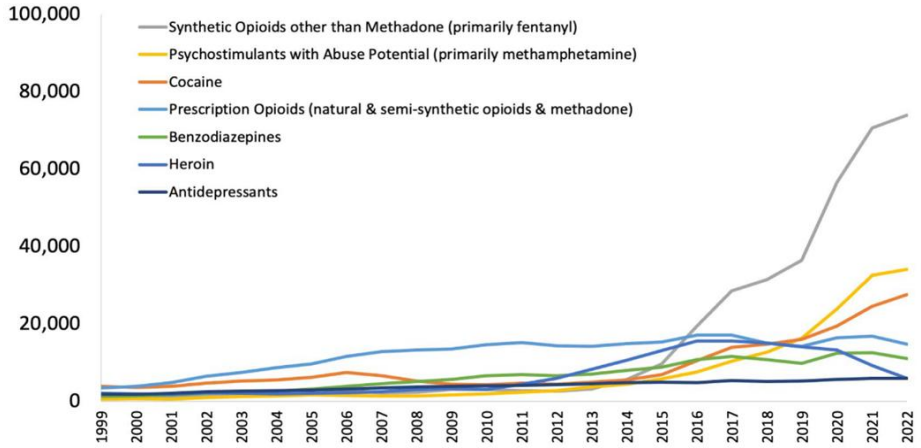


Figure 1: Historic overdose mortality counts by drug category (1999-2022) [23].

To make matters worse, the COVID-19 pandemic caused a surge in overdose mortality rates nationwide [3, 27, 28]. Following the stay-at-home order issued by the government on March 20, 2020, overdose mortality skyrocketed by a staggering 23.7% from 2019 to 2020 in Cook County, IL [3]. Many other counties in America also suffered similar fates, as 13% of Americans surveyed during the pandemic reported increasing substance use as a way to cope with stress [27]. Safe to say, drug overdoses are an urgent societal malfunction in dire need of amelioration due to their pervasiveness in American society that commonly goes overlooked.

However, recent solutions have ameliorated this issue to some extent, leading to a decline in overdose mortalities since 2022. Medical breakthroughs such as naloxone and stricter border policies enacted by the government have significantly reduced illegal opioid supply from foreign territories and reduced their risk of overdoses being fatal [30,31]. The government in 2022 also began pushing for readily available access to naloxone and Narcan, another drug antagonist [32]. Moreover, in 2023, the FDA approved the sale of Narcan as an over-the-counter drug in pharmacies, providing access to lifesaving overdose prevention medication to millions [33].

Despite this current downward trajectory, there is still room for improvement in reducing risk. Tens of thousands of American lives are still lost annually, and implementation of fatal overdose mitigation strategies discussed in this report is intended to solve for exactly this problem.

1.1 Problem Statement & Debrief

In our report, we only focus on risk of loss from fatal overdose deaths, and consider non-fatal overdoses to be out of the scope of this report.

Risk & Who is at Risk:

At the straightforward level, healthcare facilities both local and governmental must allocate expenses into treatment of overdoses and their mortalities. And like many other malfunctions in society, drug abuse has more than one at-risk group.

Employers and companies which the victims of overdose work for also find themselves facing immense risk due to lost productivity from deceased victims. In 2021, the CDC reported from a study by Florence et al. from 2017 that non-fatal drug overdoses cost $471 billion, and fatal overdoses cost an additional $550 billion for a total of over $1 trillion dollars that year [4]. Moreover, there is emotional trauma suffered by affected families and communities. However, we consider these emotional risks outside the scope of our report due to absence of relevant data.

Causes of Risk:

The risk of these fatal overdoses is caused through systemic issues in society, such as over-distribution of painkillers to impoverished areas and laws that shield drug companies from litigation, allowing them to freely market their products. Moreover, illicit supply of drugs from Mexico and China has exacerbated the issue [50].

Effective Mitigation Strategies:

Modifying Outcomes of drug overdoses holds promise in reducing risk, as pre-distribution of medical supplies such as naloxone to hospitals could prevent deaths, but still costs a significant amount.

However, the most promising mitigation strategy elaborated on in this report is Behavior Change. In this scenario, due to the fact that if individuals in counties at a high risk of drug abuse were provided a deterrence from drugs through some form of anti-drug incentivization, they would never overdose in the first place.

The mitigation strategies discussed in this study are intended to be implemented by local governmental or non-profit institutions that have the monetary resources and jurisdiction to implement the strategies on a wide enough scale within counties to have noticeable reductions in overall risk of overdose mortality.

2. Data Methodology

The mathematical modeling portion of this report encompasses two different types of mathematical models: the classification-based models focused on identifying factors of high overdosing counties, and the projection models centered on projecting risk of monetary loss in the future. Therefore, sufficient data, shown below, was collected in order to ensure both models produce well-founded results for relevant interpretation of drug overdose mortality. We also perform relevant cleaning and Exploratory Data Analysis (EDA) to ensure data is consistent with model requirements.

2.1 Data Collection

Table #1: American Counties' Socioeconomic & Health-related Features 2024 Release [37]

- **Source & Reliability:** This table comes from the University of Wisconsin's Population Health Institute, which has been publishing annual county reports since 2010. All county features are aggregated from more than twenty reputable sources such as Stanford University, the National Center for Health Statistics (NCHS), the USDA,

US Census, Bureau of Labor Statistics, and drug overdose mortality rate is obtained from the NCHS as well [20].

- **Purpose:** This table encompasses all the data that the random forest and GNN classifier models will be trained on to predict counties at a high risk of overdose mortality.
- **Features:** This table contains 39 socioeconomic and demographic features for each county as well as the overdose mortality rate per 100,000 people. A description of the 13 features utilized in the classifier model will be included in the Mathematical Modeling section.

Table #2: Annual American Overdose Mortalities (2017-2024) [22]

- **Source & Reliability:** This table comes from the Center for Disease Control, ensuring its credibility.
- **Purpose:** This table contains data that the ARIMA and Monte Carlo models will generate future scenarios of overdose mortality count up to and including 2030.
- **Features:** This table contains time series data for the number of opioid overdose deaths from the years 2017-2024. Originally, it only had data up to 2023, but we impute the most recent data of the number of overdose mortality deaths from September 2023-September 2024 and use this value as 2024's overdose mortality deaths value. This 2024 data point is also from the CDC.

Table #3: Geographic County Shape Files [24,25]

- **Source & Reliability:** This mapping dataset is obtained from the US Census Bureau, a government branch which is highly reputable.
- **Purpose:** This dataset will create the structure for the Graph Neural Network by creating a geographic representation of all counties in America that captures county adjacency.
- **Features:** The data contains relevant mapping data so that an accurate geographic representation of counties in America using Table #1 can be created.

Table #4: 2023 Population by County [36]

- **Source & Reliability:** This population data also comes from the Census Bureau of the government.
- **Purpose:** We use this table to obtain a more recent estimate of population data by county (because *Table #1*'s county population column is from 2020)
- **Features:** This table contains an estimate of the 2023 population by county in America.

Data	Data that helps define historical frequency	Data that helps categorize risks and potential outcomes	Data that helps define the historical range of severity of potential losses
Table #1	**Contains**	**Contains**	**Does not contain**
Table #2	**Contains**	**Does not contain**	**Contains**
Table #3	**Does not contain**	**Contains**	**Does not contain**
Table #4	**Contains**	**Contains**	**Does not contain**

By obtaining data from all categories defined by the Actuarial Process Guide, we ensure that the models we create from this data provide a wide range of insights into causes and future trends of fatal drug overdoses.

Indeed, *Table #1* allows us to identify socioeconomic contributors towards risk of overdose mortality, *while Table #2* permits the projection of overdose mortality deaths into the future, both forming an ironclad analysis of the two major aspects of fatal overdoses worth analyzing in an actuarial report. *Table #3* and *Table #4* provide supplemental data for cleaning and model building purposes, and are still important as well.

2.2 Data Cleaning and Preprocessing

Any counties with missing data are not included in classifier model training, as it would be imprudent to try and synthesize or impute data for these counties. Creating incorrect data could create a highly biased model even if synthesized values were only marginally different from the ground truth, which we have no means of obtaining. In the end, this leaves us with 1,825 counties out of 3,144 total counties.

While this represents only 58% of counties, we note that these counties make up more than 90% of the total American population, which is more than enough to provide meaningful analysis with broad geographic implications in reducing risk from fatal overdoses.

Despite *Table #1* being created in 2024, many attributes, including the county population, and overdose mortality rate per 100,000 people are from years 2019-2021 [20]. Thus, we replace the existing population by county with the most recent county population data from *Table #4*. We only use this population by county data in Cost-Benefit analyses in our Recommendations section, but we include this step here for clarity.

After replacing the population data, we want to ensure that the overdose mortality rates still yield a reasonable estimate for the total number of nationwide overdose deaths using the following equation:

$$Total = \sum_{k=1}^{n} population(overdoseRate/100,000)$$

- ***Where:***
- ***n****= Total number of counties*
- ***Total*** *is the total number of nationwide overdose deaths*
- ***population*** *is the 2023 imputed population data by county from Table #4*
- ***overdoseRate*** *is the 2020 reported overdose mortality rate per 100,00 people by county*

From this, we obtained a total of 87,960 overdose deaths, which is quite close to September 2024's ground truth value of 84,334 deaths [22]. Thus, we can confidently say that despite the overdose mortality rates being collected in 2019-2021, they reflect much similarity with today's values when we factor in the most recent population data available (2023).

Thus, we have ensured our data on overdose mortality both in rate and number is as up-to-date as was possible to make it, because we have utilized the most recent population data while also keeping the total number of drug overdose mortalities as reported by the dataset close to the ground truth value from September 2024.

2.3 Classifier Feature Selection

Before creating classifier models to predict counties being at either a high risk or low risk of overdose mortality, we must first choose certain socioeconomic features available from *Table #1* identified in the Data Sources. All 39 attributes are too many to input to a classifier model for overdose mortality rate like we aim to do in this report. Therefore, we decided to select one third of all features using the following process:

- Create a linear regression model with all 50 initial features being the X values, and the overdose mortality rate per 100,000 people being the Y value the linear regression aims to predict using the initial 50 features.
- After fitting the model, analyze the p-values of each feature in its correlation with overdose mortality rate
- Select the 13 features with the lowest p-values, thereby ensuring we select the most meaningful features for predicting drug overdose mortality and not just random features.

After employing the process above, we select the following 12 features, each with p-values less than .05 (a common decision threshold in statistical analysis):

Full Feature Name	**Initial Form of Measurement**
Adult Smoking	Percentage
Census Participation	Percentage
Children in Single Parent Households	Percentage
Firearm Fatalities	Rate per 100,000
Food Insecurity	Percentage
High School Completion	Percentage
Insufficient Sleep	Percentage
Other Primary Care Providers	Rate per 100,000
Poor or Fair Health	Percentage
Preventable Hospital Stays	Rate per 100,000
Severe Housing Cost Burden	Percentage
% Rural	Percentage

We observe all features are either a percentage, or rate per 100,000 people. We now have ensured that our features are few enough to allow for straightforward classifier model interpretation in our Risk Analysis through the selection of meaningful features. Although this linear regression is technically a mathematical model, we only use it to select meaningful data that will be used as training for the classifier model, so we include the analysis of this model here for clarity.

2.4 Exploratory Data Analysis

Before creating and training classifier models, we must be sure to visually represent the data being used to ensure we as actuaries can understand important attributes of our data when conducting risk analysis. More importantly, exploratory data analysis is commonly done in order to ensure no faulty data skews or hinders model performance [39].

An important first step to take in feature selection of classifier models is to ensure none of the features are too highly correlated [40]. Doing so prevents the model from overfitting to only account for the highly-correlated features, and also facilitates model interpretation when analyzing the most important features. We set the threshold for pairwise feature correlation to be .85. Such a high threshold is permissible because the classifiers we utilize are resistant to overfitting when handling intra-feature correlation. We observe the correlation heatmap in Figure 2:

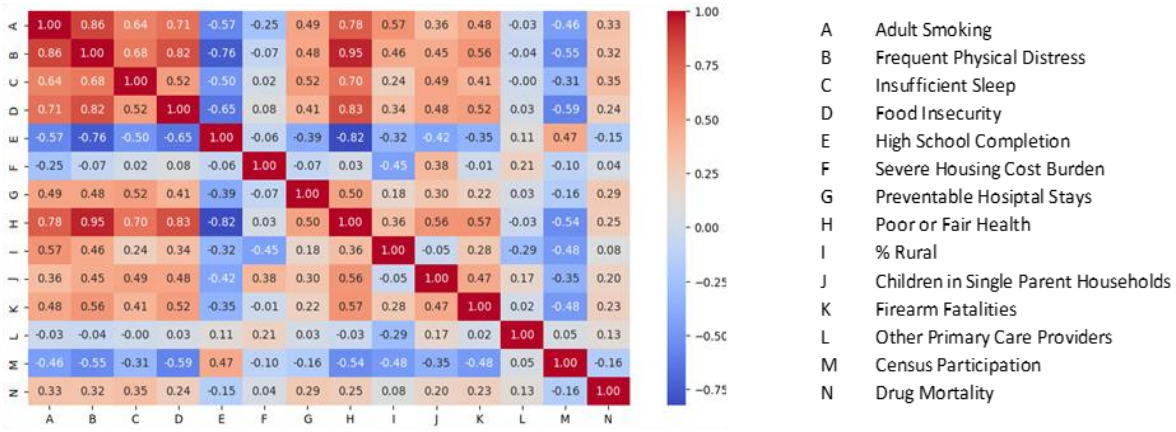


Figure 2: Pairwise correlation heatmap of selected features

We see that the Frequent Physical Distress feature is extremely highly correlated with other features. Thus, we drop it from the classifier models. Also, we see that letter N, our target variable, makes logical sense with its relationship with variables like smoking and insufficient sleep (which are positively correlated). Census participation and high school completion are negatively correlated, which also makes logical sense.

Additionally, our target variable, drug overdose mortality rate per 100,000 people, is represented on a continuous scale with a minimum of 4.86 (Limestone, AL) and a maximum of 148.51 (Logan, WV).

This presents an issue, because in order to create robust binary classifier models, we must convert the drug overdose mortality rate into a binary class. The method we chose was to apply a median split. This well-established statistical practice bins the lower 50% of all values in a continuous scale into one class, and the remaining upper 50% into another class. The important thing to note is that this method guarantees that each class has an equal count of data points.

The particular reason we chose this approach was because of classifier models' necessity for balanced classes in training data. The more unbalanced classes are, the more the classifier will overfit to the overrepresented class, and will suffer from poor performance at predicting the minority class [41].

Applying this on our target variable, drug overdose mortality rate per 100,000 people, we transform it from a continuous scale into binary classes: a low-risk class and a high-risk class.

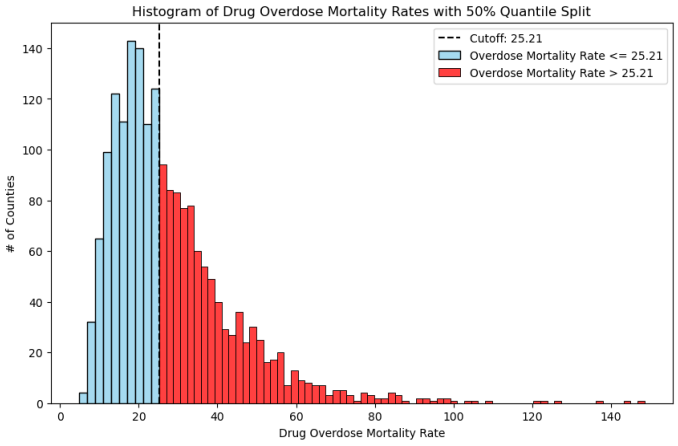


Figure 3: Binarized classes of overdose mortality rate by county. The median is an overdose mortality rate of 25.21 per 100,000 people.

From the figure above, multiple pertinent aspects are revealed about the distribution of overdose mortality rates. The distribution is clearly right skewed, illustrating there are few outlier counties with extremely high drug overdose mortality rates. These select few counties are high-severity (high risk of being susceptible to drug overdoses), but very low frequency.

On the other hand, 50% of counties represented have an overdose mortality of 25.21 per 100,000 people or less. These counties are low-risk, high frequency, due to the significant number of counties at a comparatively lesser risk of overdose mortality packed into this small overdose mortality rate section.

2.5 Software & Opensource Libraries Used

The software packages used are completely opensource, free and non-proprietary and consist of the following:

- Python, Pandas, NumPy, matplotlib, seaborn and Jupyter Notebooks for all the basic coding and charting
- SciPy, sklearn, statsmodels, shap, pytorch, geopandas, networkx, pytorch-geometric to build the models

All of our code is in GitHub, the link for which is included here and in Appendix 4: GitHub Link. Since the code and its outputs were over 50 pages long, we do not include a direct pdf in this report.

3. Mathematics Methodology

As referenced in the Executive Summary, this section will contain models oriented around 2 main goals: classification of high-risk counties of drug overdose mortality and projection of risk of loss from drug overdose mortality into the next 5 years up to and including 2030.

3.1 Classification of High-risk Counties

For recommendations to be effective in reducing overdose mortality in high-risk counties, we must be able to identify socioeconomic and other demographic attributes of counties nationwide that have a high tendency to correlate with significantly large overdose mortality rates. By creating classifier models and then utilizing black box explainer libraries to interpret them, we are able to identify prominent county features that predict a high risk of overdose mortality rates. These feature importances will be utilized in generating risk mitigation strategies in the Risk Analysis section.

The two classifier models we create and evaluate are a baseline Random Forest and a Graph Neural Network. The mathematical foundations of each of these models and the rationale of their choosing will be thoroughly explained in their respective subsections.

The error metric we utilize for both models will be the area under the curve of the receiver-operating characteristic (AUC-ROC) score to evaluate the performance of the model. The AUC-ROC ranges from 0 to 1, with a score of 0 meaning classifying nothing right, .5 being coinflip guesswork, and a score of 1 being the “perfect” model for binary classes.

The score is generated through changing model decision thresholds for classes from 0 to 1, and then plotting a curve of the True Positive Rate (TPR) against the False Positive rate (FPR) as the model navigates through all thresholds from 0 to 1. Then, the AUC-ROC is calculated as follows:

$$\int_0^1 TPR(FPR)d(TPR)$$

We choose this metric because it measures performance across all possible classification thresholds other than the default .5, ensuring robust performance assessment independent of a specific cutoff [42].

Each of the features have their own unique ranges, so we standardize all input data to the classifier models using sklearn's standardScaler library. This prevents overly large values throwing off the model's interpretability in the Risk Analysis section. The standardScaler standardizes the data using the following formula:

$$z = (x - \mu)/\sigma$$

- Where:
- z is the updated feature value,
- x is the original feature value,
- μ is the mean of the feature,
- σ is the standard deviation of the feature.

3.1.1 Classifier Model Assumptions

Assumption: Table #1's data maintains sufficient similarity to the current ground truth data

- **Justification**: The table contains accurate county features aggregated in 2024. Since then, it is unlikely that socioeconomic factors change drastically enough to cause significant alterations in our actuarial process.
- **Necessity for this assumption:** This assumption is necessary if the data we use for the classification models is not accurate to the true data, we will create flawed models and there is nothing we can base our risk analysis off of.

Assumption: Table #1's data contains sufficient spatial representation of overdose mortality for counties in America

- **Justification:** The 1,825 counties represented will be enough for both classifiers to capture relevant feature importance because they capture over 90% of the population
- **Necessity for this assumption:** This assumption is necessary, because if our data was not representative of certain geographic regions, our model and therefore risk analysis and mitigation strategies would not generalize well to all American counties and would overfit to the regions that were fully represented.

Assumption: Data within counties are relatively homogeneous, with no extreme internal variation that would affect the effectiveness of intra-county risk mitigation strategies

- **Justification:** Counties are assumed relatively homogeneous because they typically share similar socioeconomic, geographic, and resource access characteristics within their boundaries.
- **Necessity for this assumption:** This assumption is necessary because if there is a very wide range or outliers of certain features within each county, our predictions would not account for these. Rather they would only solve for the average of the features.

Assumption: Other features that are not included in the classification models are not primary predictors of overdose mortality.

- **Justification:** The feature dataset I will use will have multiple known variables closely related to drug overdose deaths according to scholarly research.

- **Necessity for this assumption:** This assumption is necessary, because if we are not including one or more key predictors of high risk of overdose mortality, then our model is flawed in that it is overlooking very important socioeconomic and health-related features.

3.1.2 Random Forest

The random forest is one of the most powerful tree-based classifier models utilized in data analysis. This reputation is partially attributable to the non-linear feature relationships it can identify, which ultimately allows the model to capture complex decision boundaries other classifier models cannot [43].

Within each decision tree among the forest, the decision split of features that results in the most homogeneous child nodes is selected. This process is repeated recursively, which ensures the tree chooses the feature and threshold that best separates the data until all training data has been classified or the tree reaches a maximum depth specified as a hyperparameter [13].

We avoid underfitting the model by fine-tuning the two major hyperparameters of the random forest model (maximum depth and number of estimator trees) in order to ensure the model finds the best possible parameters. We created random forests with every combination of max depth and number of estimator trees from the following lists:

Max Depth: [1, 3, 5]

Number of Estimator Trees: [1, 3, 5, 7, 9, 11]

3.1.3 Random Forest Evaluation

Running all combinations of the hyperparameters, we observe the following results, where each cell represents a random forest's AUC-ROC with its corresponding hyperparameters.

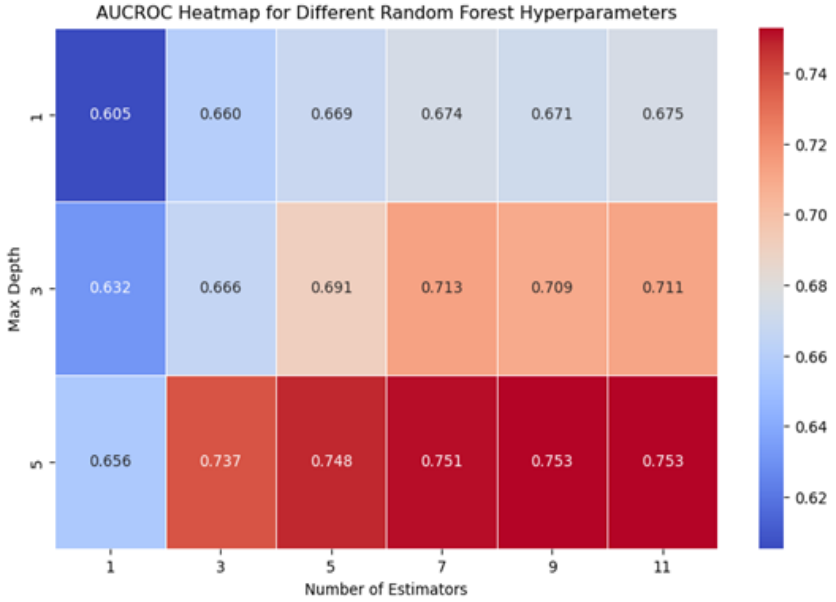


Figure 4: AUC-ROC heatmap of random forests with different max_depth and n_estimator hyperparameters.

We see that the model with 9 estimator trees and a max depth of 5 nodes has the highest AUC-ROC score of .75. In a field as variable as drug overdose mortality, this is an extremely respectable outcome for a Random Forest and will serve as the baseline score for the Graph Neural Network to achieve.

We also avoid overfitting by not searching for trees with n>11 estimators and n>5 max depth, as this creates overly complex trees which tend to overfit to predicting only training data and not unseen data. Moreover, the AUC-ROC score starts to stabilize and decrease after more than 9 estimator trees. The very top of a sample decision tree for the selected random forest is located in Appendix 3.

3.1.4 Graph Neural Network

A **neural network** is a machine learning model inspired by the human brain, consisting of layers of interconnected nodes (neurons) that process input data through weighted connections and activation functions. Each layer extracts progressively complex features, enabling tasks like classification, regression, and pattern recognition (in this case, classification) while trying to minimize a loss function, like many other models [46].

However, basic neural networks such as Multilayer Perceptron networks along with nearly every other classifier (random forests, logistic regression), treat each data point equally and assign no inter-dependency to certain data points. This could overlook the importance of county adjacency in model creation, because neighboring counties tend to share similar attributes.

We therefore utilize a novel type of neural network, the Graph Neural Network, which sets its structure as a graph based on training data rather than a predefined structure like MLP neural networks have [45]. A graph has two components: nodes and edges. In our case, the GNN, sets its nodes as all 1,825 included counties and has edges when two counties are adjacent.

We observe all counties in America colored as either missing data (black), low-risk of overdose mortality (blue), and high risk of overdose mortality (red) to confirm whether the GNN will be effective (if neighboring counties are similar in overdose mortality rate, the target variable we are trying to predict):

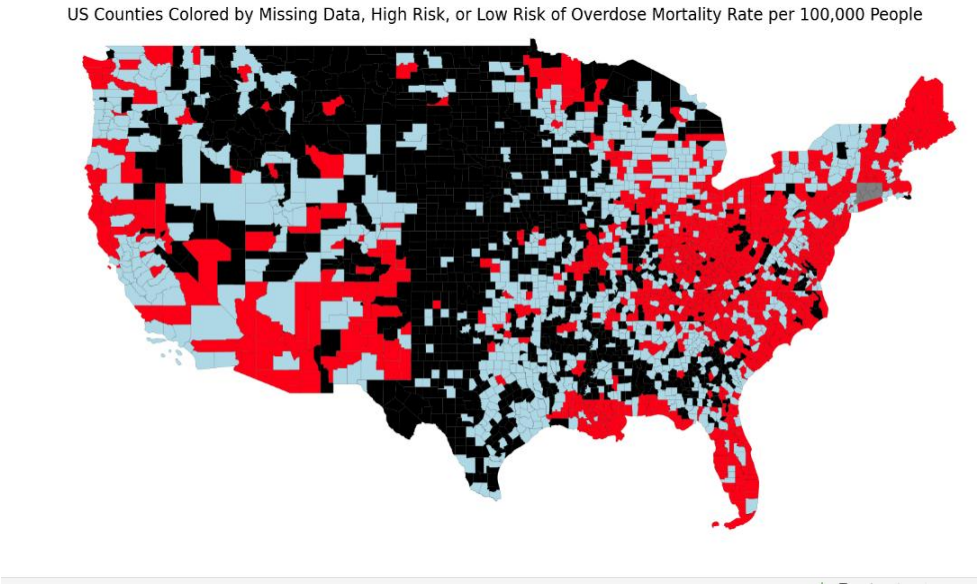


Figure 5: US counties colored by missing data (black), low risk (blue), or high risk (red) of overdose deaths by rate per 100,000 people.

From the figure above, we confirm that there are regions where low and high-risk counties are grouped together, meaning the GNN should perform better than the random forest and produce insights that can be applied to a wide region, not just specific counties. Unfortunately, we observe a paucity of reported data for midwestern counties, but we searched extensively for timely and accurate data here but could not find anything. We still have over 1800 counties with data representing over 90% of the American population, which is plenty for the GNN to train on.

To learn the importance of features at each node, the neural network attempts to minimize a function known as binary CrossEntropy loss, which is a metric that quantifies the difference between the true labels and predicted probabilities. The binary CrossEntropy loss has the following equation [44]:

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log\left(\hat{y}_i\right) + (1-y_i)\log\left(1-\hat{y}_i\right)\right]$$

- ***N*** = number of nodes in the graph being classified (in our case, 1,825 for all counties)
- y_i = true binary label for node/county
- y_i ***hat*** = predicted binary label for node/county

This loss function is minimized through a default algorithm for neural networks, backpropagation, which involves a series of matrix multiplications back and forth across the graph structure of the GNN until the ideal weights for each node/county features are learned. After training the model with 200 iterations through training data, we observe an AUC-ROC score of **.81**.

3.1.5 Classifier Selection & Results

We recommend the GNN for two main reasons. The obvious is that it has a higher AUC-ROC score, meaning that it was able to correctly classify counties at high or low risk of overdose mortality with more accuracy than the baseline random forest model. Examination of the AUC-ROC graph confirms the model's superiority in predictive power over the random forest for all decision thresholds:

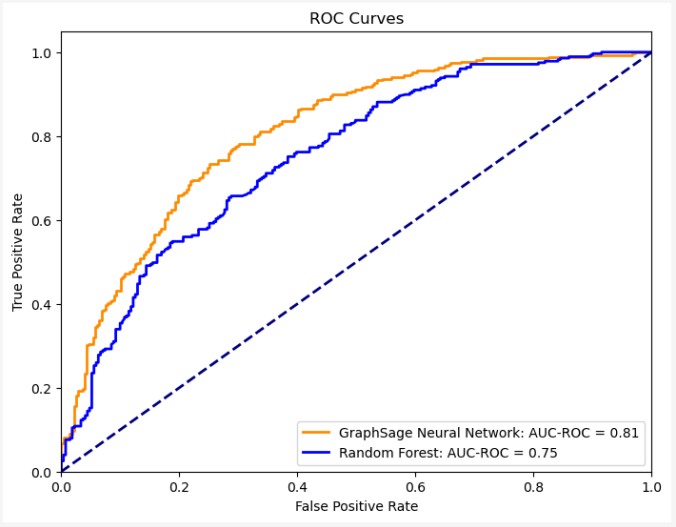


Figure 6: ROC Curves of the baseline random forest model and the Graph Neural Network. Evidently, the GNN performs better with an AUC-ROC of .81, which is .06 more than the random forest.

Additionally, a selection of the GNN will allow for feature importance interpretation that is more applicable to our goals of implementing risk mitigation strategies on a broader geographic area [45]. The GNN will allow for this because it will display prominent features that predict high risk of overdose mortality across neighboring counties unlike the random forest and almost any other classifier model would.

3.2 Projection of Future Risk of Loss

In this section, we build models that project risk of loss (overdose mortality and its associated costs) into the future and will be used in the Risk Analysis to generate an expected value of loss at any given time.

We build an AutoRegressive Integrated Moving Average (ARIMA) simulation to visualize a statistically-founded mean estimate of the number of overdose deaths in years from 2025-2030, and a Monte Carlo simulation that builds off of the ARIMA's mean annual forecast to generate both statistically-founded and noise-incorporated estimates of the number of overdose deaths we expect to see in the same time period. This allows for efficient visualization of a distribution of potential future risks in best-case and worst-case scenarios.

3.2.1 Assumptions

Assumption: No major external shocks (e.g., pandemics or new synthetic drugs being created) occur during the projection period.

- **Justification:** External shocks are unpredictable and can drastically change overdose trends or costs.
- **Necessity for this assumption:** This is necessary, as the projections will be based on existing trends and cannot account for unforeseen events. The output projected costs will contain the expected value of loss assuming that the future trends hold similarity to pre-existing trends.

Assumption: **No technological breakthroughs will develop in the near future that drastically reduce the costs of handling drug overdoses.**

- **Justification**: Even though healthcare evolves at a steady pace, significant breakthroughs (such as Naloxone) are rare and difficult to predict in the short-term future.
- **Necessity for this assumption:** The model depends on a stable relationship between overdose deaths and treatment costs.

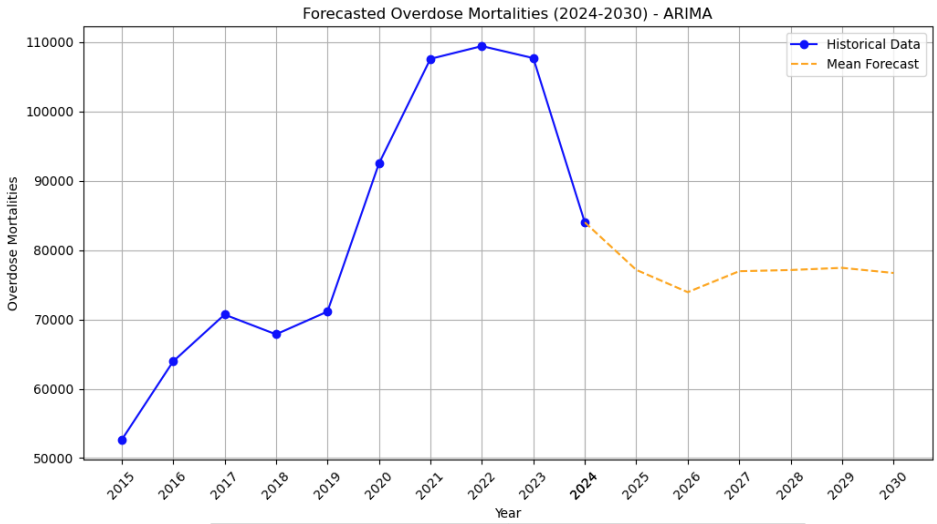
3.2.2 ARIMA Forecasting

We utilize an ARIMA model to project overdose mortality deaths into the next 5 years, as the model's moving average feature captures yearly fluctuations in death counts [12]. It is easy to over-exaggerate the risk of drug overdose mortality into the future because of the extreme spike in years 2019-2021. However, the ARIMA model will take into account the recent decrease of drug overdose deaths, and unlike a simple linear regression, will generate a very specific forecast with priority given to the most recent years. Thus, an ARIMA presented itself as the best choice and we avoid over-exaggeration of risk.

The following parameters were specified to adhere to priority in projection given to recent years:

- AutoRegressive terms: **n=4 →** Captures past 4 years where overdoses flatlined and decreased.
- Differencing Order: **n=1 →** Small differencing order focuses on recent fluctuations (i.e. the decrease since the end of 2022) rather than long-term trends such as the sharp increase from 2015-2021.
- Moving Average terms: **n=2 →** Noise correction for past 2 years, further biasing recent data and therefore generating realistic estimates.

After running the ARIMA with ***Table #2***'s data of annual overdose mortalities from 2015-2024 as past precedence, we observe the following projection:



Year	Forecasted Overdose Mortalities
2025	77,145
2026	73,934
2027	76,952
2028	77,125
2029	77,438
2030	76,699

Figure 7: ARIMA forecast of nationwide overdose mortalities up to 2030

Consistent with news and scholarly literature, our ARIMA predicts that overdose deaths are expected to drop noticeably within the next few years [47]. In 2030, we observe a mean estimate of 76,699 mortalities, which is a 9.954% decrease from September 2024's ground truth value. We use the ARIMA's mean forecast from each year up to 2030 to identify monetary expenses of these overdose mortalities in the Risk Analysis section.

3.2.3 Monte Carlo Simulation

We also implement a Monte Carlo simulation as an extension to the ARIMA model that enhances the forecasting process by incorporating random variations that reflect real-world uncertainty. While the ARIMA provides a structured, deterministic forecast based on past trends and statistical relationships, a Monte Carlo simulation will extend this by generating multiple possible future scenarios, allowing us to estimate a range of outcomes rather than a single projection and view a distribution of future risk [48].

This is particularly useful when forecasting drug overdose mortality, where external factors such as policy changes, economic shifts, and social behaviors introduce uncertainty that a purely statistical model like ARIMA may not fully capture.

We set the mean overdose deaths from 2025-2030 from the ARIMA to be equal to that of the Monte Carlo for each year. We then find 5% and 95% confidence intervals of the projected number of overdose mortality deaths from the Monte Carlo simulation that will be used in best-case and worst-case scenario analysis for 2030. The distribution obtained from the Monte Carlo we ran with 10,000 trials assuming a normal distribution is shown below as an extension to the original ARIMA graph for the years 2025-2030:

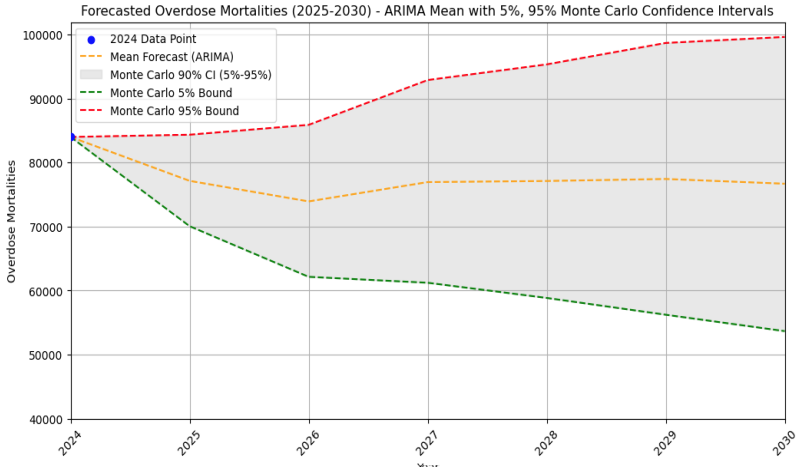
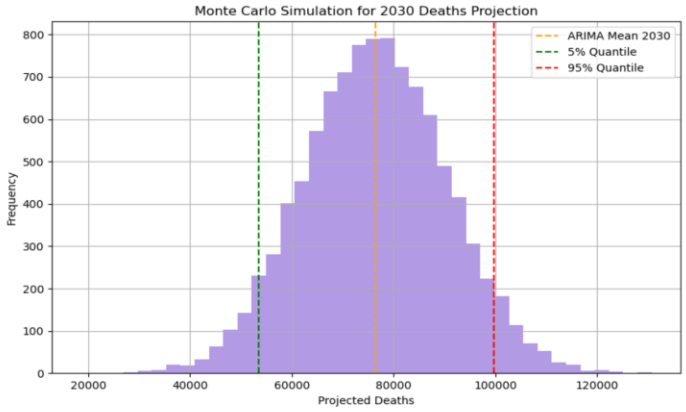


Figure 8: Monte Carlo extension of 5% and 95% confidence intervals onto the ARIMA model.

Indeed, we see that the mean forecasted overdose mortality deaths up to 2030 of the Monte Carlo estimate is equivalent to the mean forecasted overdose mortality deaths of the ARIMA. Moreover, for the 5% estimate all the way up to the mean, the Monte Carlo simulations predict that overdose mortalities will drop noticeably from 2025-2030. The 95% confidence interval, representing a worst-case scenario, forecasts that drug overdoses will spike again, but not to the level that they were at the height of Covid in 2021.

Examination of the Monte Carlo distribution of overdose mortalities specifically for the year 2030 allows for a more in-depth analysis of the distribution of risk of loss that we expect to see:



5%	μ - σ	μ	μ + σ	95%
53,458	62,410	76,699	90,511	99,653

Figure 9: Monte Carlo distribution for overdose mortality deaths in 2030 assuming a standard gaussian distribution.

From the Monte Carlo distribution of risk of deaths in 2030, we will ideally see overdose deaths fall to around 50,000 in 2030, which was the best-case scenario at the 5% confidence interval predicted by simulation above in Figure 9.

4. Risk Analysis

4.1 Risk Characterization

We begin the Risk Analysis with the characterization of what causes risk by identifying strong predictors of overdose mortality through utilizing a black box interpreter (SHapley Additive exPlanations) on our chosen classification model, the Graph Neural Network.

The SHAP module operates by iterating through test instances that the model was evaluated on, and reruns the classifier on slightly altered features of these test instances to record which specific features have the highest impact on the model's prediction when changed slightly [51]. We set the number of test instances to be explained at 900 instances out of the total 1,825 counties. Feature importances as identified by the SHAP run on the GNN are illustrated below:

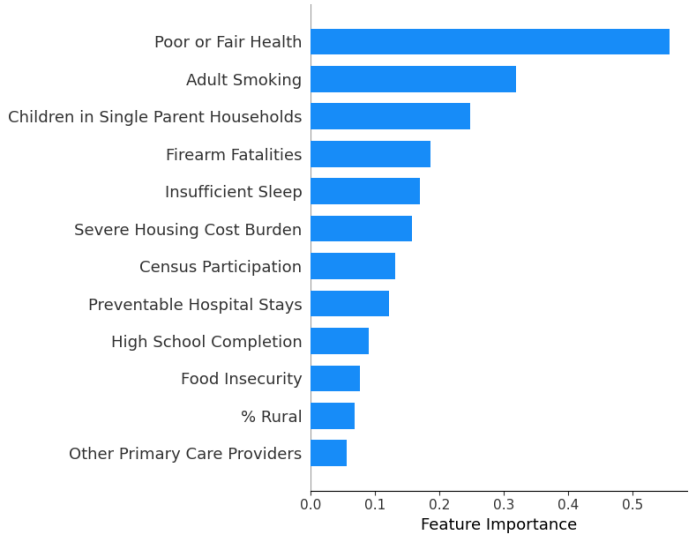


Figure 10: Normalized socioeconomic feature importances of predicting high risk of overdose mortality using the GNN.

In alignment with published literature, the SHAP identifies that many of the health-related features in counties from *Table #1* are strong predictors of overdose mortality [49]. Different features shown in Figure 10 have varying degrees of influence on whether a county is at a high or low risk of overdose mortality, which the SHAP module identifies through feature importances, shown as the X-axis in the figure.

Moreover, different features have both positive and negative correlations for risk of overdose mortality. Referencing Figure 2, the pairwise correlation heatmap of all socioeconomic features, we identify that some features such as High School Completion and Census Participation are negatively correlated with overdose mortality (i.e. more census participation in a county tends to predict lower overdose mortality rates). Conversely, other features such as Adult Smoking and Firearm Fatalities are positively correlated with risk of overdose mortality, as one might expect.

Certain features such as citizens with poor health and single parent households, while very important in predicting fatal overdoses, are systemic issues stemming from poverty. Therefore, effective short-term mitigation strategies designed to ameliorate these issues might be harder to implement.

However, there are features such as smoking, insufficient sleep, and firearm fatalities which mitigation strategies do exist for. We utilize these feature importances to identify behavior change mitigation strategies that will improve county features which tend to predict high risk of overdose mortality.

4.2 Risk Projection

We now quantify in terms of monetary loss the expected value of risk up to and including 2030. To achieve this, we utilize a range of Value of a Statistical Life (VSL) estimates recommended by the U.S. Department of Health and Human Services (A low estimate of $4.40 million, a central estimate of $9.40 million, and a high estimate of $13.10 million in 2015 dollars) [5].

We recalculate these low, central, and high values in today's 2025 dollars and account for inflation, causing 35.93% higher values from 2015 using an inflation index. The modified low estimate becomes $5.98 million, the central estimate becomes $12.78 million, and the high becomes $18.21 million [8].

At first, these values sound extremely high and over-exaggerated, even the low VSL. However, a VSL estimates the lifelong monetary value of one human life, not just the expenses related to the incident of their mortality. Additionally, empirics confirm the validity of the estimates. Referring back to the CDC's remark that the 70,699 fatal drug overdoses in 2017 cost $550 billion dollars, this comes down to an average of roughly $7.78 million per person in 2017 dollars [4]. In today's money after accounting for inflation, this becomes $10.18 million in a VSL estimate per the CDC's own findings from credible journals [8,14].

Therefore, we can safely say that our Low and Central VSL estimates are not over-exaggerated and quite reasonable given historical data. We still include a High VSL to simply provide a worst-case estimate, but we do not actually utilize it in any Cost-Benefit Analysis for recommendation strategies we ultimately analyze.

Utilizing the inflation-updated VSL estimates (Low: $5.98 million - Central: $12.78 million - High: $18.21 million), we revisit the yearly mean of forecasted overdose mortality deaths from the ARIMA model to generate annual estimates from 2024-2030 of the cost of drug overdose mortality deaths using the following equation:

Expenses for year (n) = VSL * # of overdose deaths in year (n)

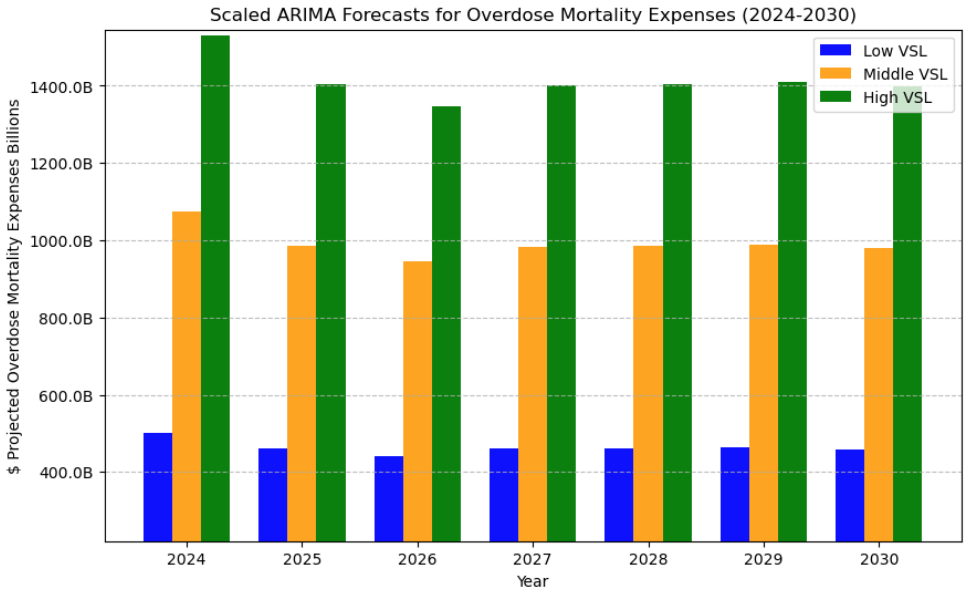


Figure 11: ARIMA-based monetary risk of loss from overdose mortalities using Low, Central, and High VSL estimates

Forecasted Overdose Mortality Expenses	Low VSL ($5.98 million per life lost)	Central VSL ($12.78 million per life lost)	High VSL ($18.21 million per life lost)
2025	$502 billion	$1.07 trillion	$1.53 trillion
2026	$461 billion	$986 billion	$1.41 trillion
2027	$442 billion	$945 billion	$1.37 trillion
2028	$460 billion	$983 billion	$1.40 trillion
2029	$463 billion	$989 billion	$1.41 trillion
2030	**$458 billion**	**$980 billion**	**$1.40 trillion**

We see that like overdose mortalities, their incurred expenses are also predicted to drop noticeably by 2030. Looking at the low VSL for each year, expenses are forecasted to drop by roughly $44 billion from 2025 to 2030, confirming that the ARIMA generated a reasonable future forecast that biases the most recent decline in overdose deaths.

In our Cost-Benefit Analyses in our recommendations, we utilize the ARIMA's 9.954% reduction in overdose mortalities and their expenses in 2030 using the low VSL to quantify risk with no intervention or mitigation strategy implemented. We choose the low VSL in these analyses to ensure we do not over-exaggerate expenses but also maintain realistic estimates for future risk.

4.3 Risk Mitigation Strategy Analysis

Insurance Consideration

Certain insurance programs do have the capability to mitigate risk of death and financial loss. Despite this, poverty in general bars at-risk individuals from having most types of private insurance in the first place [38]. Therefore, the only type of insurance which has viability in reducing overall risk of death and money is social insurance. Social insurance refers to a type of insurance where the government pays expenses for individuals at risk for health or other problems that they might not be able to pay for themselves. This fits perfectly with what a typical at-risk individual of a drug overdose could drastically benefit from.

We therefore do fully support the expansion of current social insurance programs such as Medicaid to at-risk individuals of a fatal drug overdose. However, a lack of published data on expenses and outcomes for beneficiaries as well as the cost of the insurance itself bars us from conducting a cost-benefit analysis. Nonetheless, we elaborate further on this implementation in our Recommendations section.

Modifying Outcomes Consideration

Modifying the outcomes of drug overdoses that do occur holds significant promise in reducing risk of loss of both lives and expenses related to mortalities. We consider numerous historically effective community-based programs intended to reduce drug overdose mortalities and improve on them by simulating what would occur on a county level rather than community level scale.

The program we identify and elaborate on is Take Home Naloxone kit distribution to counties at a high risk of overdose mortality. The justification behind this choosing is that the cost-benefit analysis of over-distribution of these kits to a county population still yielded a far lesser expected value of loss less than there would be in 2030 without intervention. The details of this cost-benefit analysis as well as the THN distribution are included in the Recommendations section for clarity and better flow.

Behavior Change Consideration

Behavior change holds the most promise in reducing risk of monetary loss and life because it can prevent overdoses from occurring in the first place (unlike Modifying Outcomes) thereby saving remarkably more money for all at-risk parties mentioned in the Risk Projection section. To determine effective Behavior Change strategies, we utilize the following workflow:

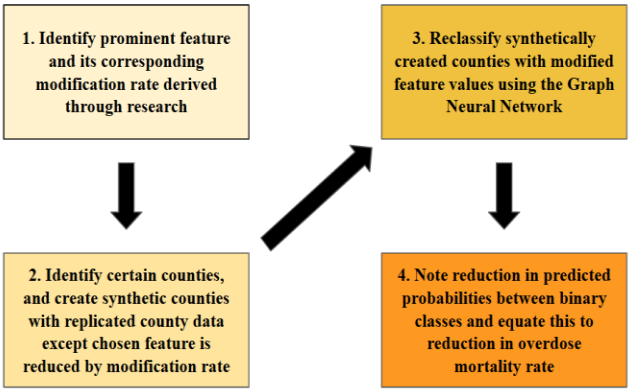


Figure 12: Behavior Change mitigation strategy identification and simulation workflow utilizing the Graph Neural Network.

Through the workflow above, we are able to generate behavior change recommendations based on important features of predicting high risk of overdose mortality that the GNN predicts, and then find new expected values of loss based entirely off of our mathematical models, rather than just choose random behavior change incentives and use only the projection models to quantify loss.

The main socioeconomic feature that we choose to mitigate is Adult Smoking through the implementation of smoking deterrence programs targeted towards smokers in counties at a high risk of overdose mortality. We find that the more counties can reduce their smoking rate, the lesser the expected value of loss becomes. Like with our Modifying Outcomes risk mitigation strategy, the cost benefit analysis of smoking deterrence implementation is discussed in the Recommendations section.

5. Recommendations

5.1 Insurance

We recommend that the government invest into further developing responsible social insurance programs designed to provide financial aid for medical treatment and overall well-being of at-risk individuals. Typically, uninsured or underinsured individuals experience treatment delays, reduced care, and higher chances of relapses, potentially resulting in overdose mortalities.

The government can achieve this by making educational programs available to individuals, so that they are aware of all the insurance options and can take advantage of them in needed situations. Additionally, the insurance programs themselves can be enhanced by further data gathering and analysis which can make the insurance options more targeted. Having a centralized database across the country not only for fatalities but also for all overdose incidents will not only help caregivers, but also government and private insurance providers in analyzing and better pricing the risk.

Finally, policies and regulations also play a key role in the insurance market. Different policies cause inefficiencies for both insurance providers and insurance seekers. Having a more uniform set of policies (driven by better data as discussed above) that can streamline the market especially for drug overdose would benefit all people, including all types of insurance providers.

Despite extensively searching for public data, we were unable to find data noting specific costs and effects of social insurance on drug overdose mortality rates specifically. This makes a rigorous cost-benefit analysis

mathematically impossible, but the benefits of social insurance still likely stand to outweigh the costs of implementation if offered towards high-risk individuals correctly.

5.2 Modifying Outcomes (THN)

We recommend county and state governments to approve and regulate the distribution of Take-Home-Naloxone (THN) kits to citizens in counties of a high risk of overdose mortality, which can be used at will to reverse an overdose before a critical threshold is reached and recovery becomes impossible. These agencies should collectively begin distributing these THN kits with utmost priority so that the risk of a fatal overdose occurring is minimized as soon as possible due to the fact that naloxone will be more available to susceptible individuals. We cannot predict when an overdose will happen next, and therefore immediacy if this distribution would be the utmost priority. Since the government can purchase naloxone in bulk quantities, they should negotiate discounts with manufacturers and potentially find ways to address patent issues under national emergency criterion.

First proposed in the 1990s, THN programs were meant to educate families and prepare opioid users for a worst-case scenario in which an overdose occurs [17]. They have been implemented in countries globally to date, and various studies confirm the effectiveness of these American THN programs in reducing overdose mortality [7].

Indeed, a meta-analysis from King's College London in 2016 of over 20 THN programs in the past revealed that when the naloxone from a THN kit was used to reverse an overdose, it was successful more than 90% of the time in all but one of these programs (which had an 83% reversal success) [22].

THN Cost-Benefit Analysis

Looking at a paper in December 2024 assessing effectiveness of THN allocation, distribution of naloxone kits cost $76 in 2017 [7]. In 2025 dollars, this becomes $99.42 accounting for inflation [8]. We verified the reasonability of this estimate in today's world by finding that current naloxone kits sell from around $50 to $100 dollars. To remain conservative, we use our higher-end estimate of $99.42.

We then look at the county in America with the lowest overdose mortality rate from *Table #1*, which is Limestone County, Alabama, with a rate of 4.86 overdose mortalities per 100,000 people. We evaluate the effectiveness of THN programs across this county because this choice will confirm the effectiveness of THN in worst case scenarios, (i.e. where overdose mortalities are already low to begin with). The geographic location of Limestone County is shown in Appendix 2.

Limestone's population is expected to grow by approximately 3,786 people annually up to 2030. Using this annual increase in Limestone's population, Limestone's population is estimated to be 141,156 in 2030 [19].

Assuming the ARIMA's prediction of a 9.954% nationwide reduction in overdose mortalities from 2024 to 2030 will occur in Limestone, they will have an overdose mortality rate of 4.40 in 2030, or 6.21 total overdose deaths using the 2030 projected population of 141,156. We utilize the low VSL estimate of $5.98 million to find a baseline expected value of loss with no intervention, which becomes approximately $37.11 million dollars.

Then, we conduct sensitivity analysis in which this distributed naloxone to the entire 2030 Limestone population of 141,156 has an 80%, 95%, and 99% effectiveness percentage in reversing an overdose when one occurs. We view the new Cost Benefits through the following table, where each column is calculated as follows:

- ***THN Distribution Expenses*** *= (Cost per THN Kit) * (2030 Limestone Population)*
- ***Fatal Overdoses Prevented*** *= (THN Reversal success %) * (# of projected OD deaths in Limestone in 2030)*
- ***New Expected Value of VSL Loss*** *= $37.11 million - (Low VSL * Fatal ODs prevented)*

- ***Cost Benefit in 2030*** *= $37.11 million - THN Distribution Expenses - New Expected Value of VSL Loss for each THN effectiveness %*

THN Overdose Reversal Success %	THN Distribution Expenses	Fatal Overdoses Prevented	New Expected Value of VSL Loss	Cost-Benefit in $ in 2030
80%	$14.03 million	4.96	$7.48 million	**+$15.60 million**
90%	$14.03 million	5.58	$3.74 million	**+$19.34 million**
95%	$14.03 million	5.90	$1.87 million	**+$21.21 million**

Even with an extremely conservative 80% THN reversal success estimate and the lowest VSL, lowest risk of overdose mortality, county-wide THN distribution has the potential to save over $15 million in Limestone Alabama in 2030, where overdose mortality rate is the lowest in the nation. Indeed, even with over-distribution to the entire county population, THN remains extremely cost effective in 2030 when overdose deaths are projected to be 9.954% lower than the status quo.

It may seem obvious to implement THN on a wide scale due to clear benefits, but there are a few limitations and reasons why a cost-benefit analysis could be misleading. Primarily, struggles with distribution of naloxone in rural areas where overdose mortality rates tend to be highest might undermine the entire goal of THN, as people who need it most may find it difficult to obtain naloxone vials [6]. Also, an existing shortage of naloxone globally may make distribution on a large scale quite challenging [18]. A targeted strategy at distributing THN kits to only higher-risk individuals would be more effective at solving this problem if such a method to identify these specific individuals existed.

Nonetheless, we strongly recommend development of THN programs in regions or counties where it is feasible and cost-effective to do so, because we have expanded on well-established THN attempts by amplifying the scale of distribution to a whole county instead of a community alone and showing clear opportunity for reduced risk of loss of both life and expenses.

5.3 Behavior Change (Smoking Deterrence)

We recommend government and non-profit organizations in neighboring counties at a high risk of drug overdose mortality collectively implement smoking awareness and deterrence campaigns, as a high smoking rate was one of the strongest predictors of counties being at a high risk of overdose mortality generated from the feature interpretation of the GNN. Organizations that have the capability and jurisdiction to employ smoking deterrence experts within counties with high smoking rates should immediately begin this strategy we recommend, as it will likely reduce county smoking rates, thereby reducing the risk of fatal overdoses in tandem.

However, to make this conclusion, we ensure that smoking is not just correlated to overdose mortality, but is also a cause of overdose mortality. A study conducted in 2008 by Shenghan Lai et. al. from Johns Hopkins University confirms this. Utilizing logistic regression and other models to identify the relationship between drug use and smoking, Lai et al. found that 90% of study participants who ever used heroin could have been because of smoking [10]. Moreover, 78% of people who ever tried cocaine could be attributable to smoking [10]. Thus, smoking is indeed a causal feature of fatal drug overdoses and reducing it will also reduce overdose mortalities.

Cost-Benefit Analysis

Referring back to the workflow in the Risk Analysis section in Figure 12, we identify a smoking deterrence program implemented by Mundt et al. from the University of Wisconsin 2023, in which smoking cessation experts were dispersed throughout hospitals to deter smoking use from at-risk patients. Ultimately, gathering patient data from 2017-2020, the study found that for just over 10,000 patients, smoking cessation programs cost an average of $6.44 per patient each month, which is $8.13 per patient per month in today's money [8, 11].

To conduct specific cost-benefit analysis, we identify four neighboring counties in Ohio at a high risk of overdose mortality as defined by the binary split we made in Figure 3: Pike County, Brown County, Adams county, and Scioto county. A map of these counties is shown in Appendix 1.

First, we find the expected value of monetary loss for all of these counties using the ARIMA's forecasted 9.954% decline in overdose mortalities and assume that it will happen to each of these four counties in 2030 as well. We again use the low VSL of $5.98 million per overdose mortality and find the expected value of loss the same way we found the expected loss for Limestone County from Modifying Outcomes in section 5.1.

After calculating this, we determine an expected value of loss of $767.80 million across all four counties without any smoking deterrence program implementation.

In order to generate a cost-benefit analysis for the year 2030, we then find the number of adult smokers in all four counties using the Adult Smoking percentage feature and the county population data. Then, we find the cost of smoking deterrence programs for all smokers in these counties using the $8.13 per person cost derived just above, and obtain the following results:

County	2023 Population	% Adults Reporting Smoking	Monthly cost of Smoking Deterrence for all smokers	Implementation cost from 2029-2030 (12 months)
Pike	27,001	27.2%	$60,000	$720,000
Brown	43,777	24.3%	$86,000	$1.03 million
Adams	27,521	24.8%	$55,000	$660,000
Scioto	71,969	27.4%	$160,000	$1.92 million

In total, cost of implementation is approximately $4.33 million. Now, we conduct sensitivity analysis in which these smoking deterrence campaigns across these four counties reduce their smoking percentages by 5%, 10%, and 25% from their current values. After reducing the smoking feature, we adhere to the workflow in Figure 12 and observe the percentage change in these modified counties' predictions by the GNN before and after modifying the smoking feature.

We recall that a prediction of .5 or higher refers to the GNN predicting the county at a high risk of overdose mortality, and below .5 means a prediction of a low-risk county. We observe the following results after performing the sensitivity analysis for different smoking percentage reductions:

County	Original GNN Risk prediction	GNN Risk Prediction after 5% reduced smoking	GNN Risk Prediction after 10% reduced smoking	GNN Risk Prediction after 25% reduced smoking
Pike	.88	.85 (-3.96%)	.80 (-9.15%)	.62 (-30.18%)
Brown	.84	.79 (-6.8%)	.71(-15.61%)	.42 (-50.54%)
Adams	.94	.92 (-2.9%)	.88 (-6.99%)	.66 (-30.19%)
Scioto	.93	.89 (-3.99%)	.83 (-9.98%)	.56 (-39.87%)

Understandably, we see that the original GNN prediction of these 4 counties is above .5, confirming the model recognized these counties as high risk of overdose mortality and assigning a reasonable prediction probability to each of them. Moreover, the more we decrease the smoking feature, the smaller the prediction probability becomes, elucidating the GNN understands that smoking has a directly proportional relationship with overdose mortality rate. For Brown County, we do see extremely large changes in risk prediction after changing smoking compared to the other three counties. However, the GNN captures complex non-linear relationships which are likely a cause for this occurrence.

We equate this percentage change of the GNN's risk prediction of these counties to be the percentage change of the counties' overdose mortality rate, because this is essentially what the GNN aims to predict in the first place.

We can now find expected values of loss for each smoking deterrence effectiveness by noting the percentage reduction in overdose mortality for each county after reducing smoking for each deterrence reduction %.

Smoking deterrence effectiveness %	Smoking program cost from 2029-2030	New Expected value of loss	Original expected value of loss	Cost-Benefit in $ in 2030
5%	$4.33 million	$735.8 million	$767.8 million	**+$27.67 million**
10%	$4.33 million	$690 million	$767.8 million	**+$73.47 million**
25%	$4.33 million	$471.1 million	$767.8 million	**+$292.37 million**

We see that for all reductions, we see monetary gain from this program in these counties and therefore encourage neighboring counties at a high risk to work towards reducing smoking.

6. Conclusion

For our Modifying Outcomes strategy of Take-Home Naloxone (THN) kit distribution, we still used the ARIMA model as a base case for the cost-benefit analysis in Limestone, AL to underscore that THN will be extremely effective up until 2030, and likely even later.

For our Behavior Change strategy of neighboring counties working together to implement smoking deterrence, we created simulated scenarios where smoking deterrence programs reduced smoking and observed the change in overdose mortality rate using the Graph Neural Network.

This risk mitigation strategy was not only evaluated using mathematical modeling, but identified in the first place by it too, making it by far the most robust strategy discussed in our report.

The only significant limitation of our study was that midwestern counties lacked overdose mortality rate data and therefore were not included in the classifier models, as seen in Figure 5. Again, we recall that over 90% of the entire American population was included in the 1,825 counties with data, thus ensuring that the Graph Neural Network's feature importances and interpretation was still extremely valuable.

Our Cost-Benefit analyses for both recommendation strategies were well-reasoned, as we actually quantify a new expected value of loss using extremely rigorous mathematical justification with different sensitivities of success of THN programs and smoking deterrence instead of just listing possible strategies.

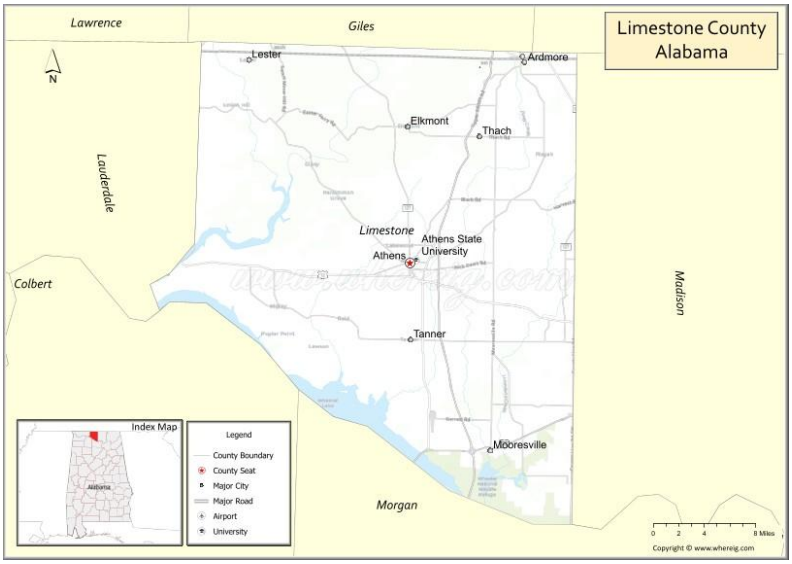
Next steps: We encourage actuaries interested in mitigating risk of fatal drug overdoses to simulate cost-effectiveness of similar mitigation strategies we discussed but in different counties than we did, in order to gain a broader geographic understanding of combating fatal drug overdoses.

The future is in our hands. Not as counties, but as a people united against drug abuse. The time to take action against it is now.

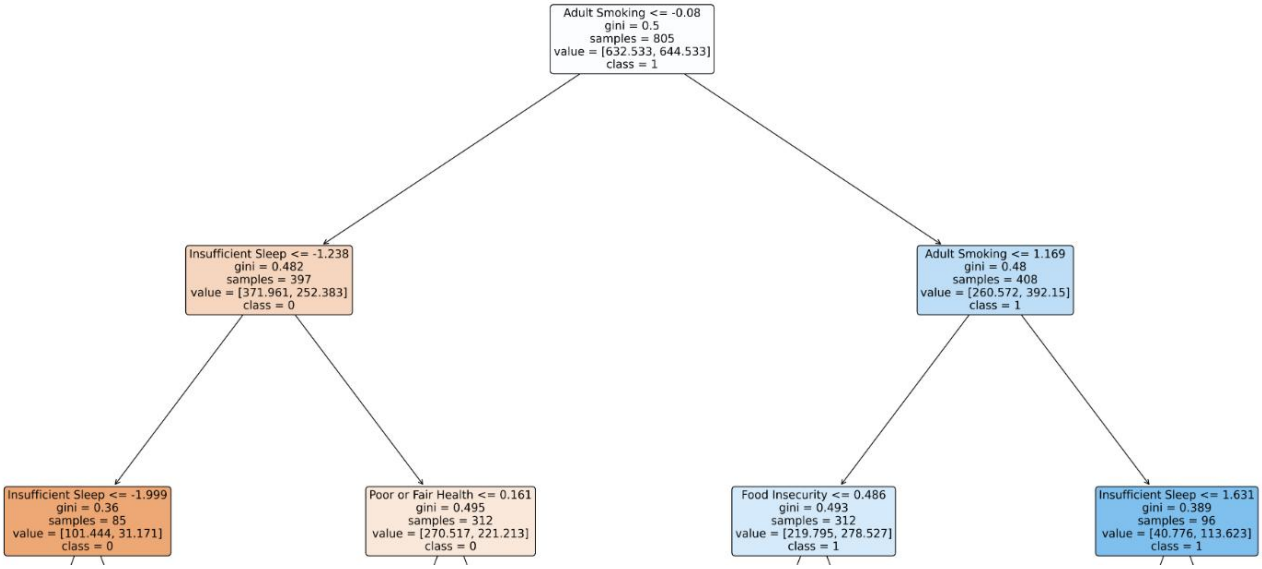
7. Appendix

Appendix 1: Brown, Adams, Pike, Scioto counties in Ohio:

Appendix 2: Limestone County, AL Map:



Appendix 3: Top of a Random Forest Decision Tree:



Appendix 4: GitHub Link to all Code

Link to public code:
https://github.com/optimusprime0077/public_code/blob/main/Python_machineLearning/drug_od_analysis/python/analyze_drug_od.ipynb

8. Acknowledgements

We are very grateful to our coach Ken Heard and mentor Laura Mitchell for their insightful guidance and unwavering support. We would also like to thank Nicole Semprit for helping organize the mentor meetings. We

would like to thank the Actuarial Foundation for creating the Modeling the Future Challenge and giving us the resources and opportunity to produce this research. Finally, we also extend sincere gratitude to all sponsors of the MTFC for their unrelenting contributions in seeing the competition flourish and be a beacon of data analysis competitions amongst high schoolers.

9. References

[1]:https://drugabusestatistics.org/drug-overdose-deaths/#:~:text=Opioids%20are%20the%20deadliest%20drug,terms%20of%20highest%20death%20count.
[2]:https://www.cdc.gov/nchs/products/databriefs/db491.htm#:~:text=The%20age-adjusted%20rate%20of%20drug%20overdose%20deaths%20increased,and%20increased%20for%20those%20age%2035%20and%20older.
[3]: https://link.springer.com/article/10.1007/s11524-022-00675-x?fromPaywallRec=true
[4]: https://www.cdc.gov/mmwr/volumes/70/wr/mm7015a1.htm
[5]:https://trumpwhitehouse.archives.gov/sites/whitehouse.gov/files/images/The%20Underestimated%20Cost%20of%20the%20Opioid%20Crisis.pdf
[6]: https://harmreductionjournal.biomedcentral.com/articles/10.1186/s12954-020-00375-2
[7]: https://jamanetwork.com/journals/jamanetworkopen/fullarticle/2827627?utm_source=chatgpt.com
[8]: https://www.bls.gov/data/inflation_calculator.htm
[9]:https://pmc.ncbi.nlm.nih.gov/articles/PMC5975355/pdf/nihms966245.pdf
[10]:https://pubmed.ncbi.nlm.nih.gov/11110061/
[11]:https://pubmed.ncbi.nlm.nih.gov/37844710/
[12]:https://ademos.people.uic.edu/Chapter23.html
[13]:https://library.fiveable.me/introduction-to-advanced-programming-in-r/unit-8/decision-trees-random-forests/study-guide/S0xNewxyyJsrdqpj
[14]: https://www.sciencedirect.com/science/article/pii/S0376871620305159
[15]: https://www.whereig.com/usa/states/alabama/counties/limestone-county-map-al.html
[16]:https://www.mappr.co/counties/new-york-counties-map/
[17]: https://pmc.ncbi.nlm.nih.gov/articles/PMC5071734/
[18]:https://www.health.gov.au/our-work/take-home-naloxone-program/about-the-take-home-naloxone-program#:~:text=The%20Take%20Home%20Naloxone%20
[19]: https://worldpopulationreview.com/us-counties/alabama/limestone-county
[20]:https://www.countyhealthrankings.org/sites/default/files/media/document/2024%20CHRR%20Technical%20Document_1.pdf
[21]: https://pmc.ncbi.nlm.nih.gov/articles/PMC5071734/
[22]: https://www.cdc.gov/nchs/nvss/vsrr/drug-overdose-data.htm
[23]: https://nida.nih.gov/research-topics/trends-statistics/overdose-death-rates#Fig2
[24]: https://www2.census.gov/geo/tiger/TIGER2024/COUNTY/
[25]: https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html
[26]: https://www.cdc.gov/nchs/products/databriefs/db522.htm#:~:text=The%20age%2Dadjusted%20rate%20of,adults%20age%2055%20and%20older
[27]: https://pmc.ncbi.nlm.nih.gov/articles/PMC8896880/#:~:text=According%20to%20the%20Centers%20for,nationwide%20in%20opiate%20overdoses2.
[28]: https://harmreductionjournal.biomedcentral.com/articles/10.1186/s12954-022-00653-1

[29]: https://www.cdc.gov/media/releases/2025/2025-cdc-reports-decline-in-us-drug-overdose-deaths.html#:~:text=CDC%20Reports%20Nearly%2024%25%20Decline%20in%20U.S.%20Drug%20Overdose%20Deaths,-Release&text=Provisional%20data%20shows%20about%2087%2C000,month%20period%20since%20June%202020.
[30]: https://pmc.ncbi.nlm.nih.gov/articles/PMC9838196/#:~:text=Overdose%20education%20and%20naloxone%20distribution%20(OEND)%20is%20an%20evidence%2D,2019%3B%20Pitt%20et%20al.%2C
[31]: https://www.cbp.gov/border-security/frontline-against-fentanyl
[32]: https://www.presidency.ucsb.edu/documents/fact-sheet-biden-harris-administration-announces-over-250-organizations-made-voluntary
[33]: https://www.npr.org/2023/03/29/1166750095/narcan-fda-approval-naloxone-over-the-counter-otc
[34]: https://pmc.ncbi.nlm.nih.gov/articles/PMC6565486/
[35]: https://slack.com/blog/collaboration/effects-low-productivity-business-growth
[36]: https://www.census.gov/data/tables/time-series/demo/popest/2020s-national-detail.html
[37]: https://www.countyhealthrankings.org/health-data/methodology-and-sources/data-documentation
[38]: https://pmc.ncbi.nlm.nih.gov/articles/PMC8247994/
[39]: https://www.ibm.com/think/topics/exploratory-data-analysis#:~:text=The%20main%20purpose%20of%20EDA,interesting%20relations%20among%20the%20variables.
[40]: https://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/
[41]: https://encord.com/blog/an-introduction-to-balanced-and-imbalanced-datasets-in-machine-learning/]
[42]: https://www.evidentlyai.com/classification-metrics/explain-roc-curve#:~:text=The%20ROC%20AUC%20score%20is%20the%20area%20under%20the%20ROC,and%201%20indicates%20perfect%20performance.
[43]: https://www.deepchecks.com/glossary/decision-boundary/#:~:text=A%20non%2Dlinear%20decision%20border,vector%20machines%2C%20and%20neural%20networks.
[44]: https://www.datacamp.com/tutorial/the-cross-entropy-loss-function-in-machine-learning
[45]: https://medium.com/@hassaanidrees7/graph-neural-networks-gnns-unlocking-the-power-of-relational-data-e324004ab2d4
[46]: https://www.ibm.com/think/topics/neural-networks
[47]: https://sph.emory.edu/news/news-release/2024/10/drug-overdose-death-rate-insight.html
[48]: https://www.ibm.com/think/topics/monte-carlo-simulation#:~:text=When%20a%20Monte%20Carlo%20Simulation,36%20combinations%20of%20dice%20rolls.
[49]: https://onlinelibrary.wiley.com/doi/abs/10.1002/9781119788409.ch13#:~:text=This%20chapter%20presents%20a%20systemic,experiences%2C%20parenting%2C%20and%20peers.
[50]: https://www.reuters.com/investigates/section/fentanyl-express/
[51]: https://shap.readthedocs.io/en/latest/