Opioid Overdose Prediction

Mid-Atlantic Opioid Task Force

Aviv Farag af3228@drexel.edu

Joseph Logan jrl362@drexel.edu

Daniel Turnbow djt96@drexel.edu

Silas McKee sm4543@drexel.edu

Abstract – This paper describes the study of opioid overdose in Pennsylvania in which related datasets were pulled from Census, CDC and PAGOV. Those datasets were read and merged in Python to create new features to explore the overdose epidemic and to develop machine learning models for various applications.

Keywords – ML, Python, CDC, PAGOV, Regression, Random Forest, Cluster Analysis

GitHub Repository – https://github.com/Joeyloganpython/Capstone

Webpage – https://joeyloganpython.github.io/Capstone/

1. Introduction

In 2021 deaths due to drug overdoses in the United States hit an all-time high of 100,000 [1]. Opioid overdoses are a major issue leading to many of those deaths. One of the most sinister elements of the current opioid crisis is the addition of fentanyl. Fentanyl a highly concentrated opioid, which is sometimes laced into other drugs, which can be dangerous for those who ingest them.

Many of the publicly available datasets, such as those from the CDC, or state level datasets, are challenging to work with due to incomplete data, very large datasets, or differences in data collection practices. Our project goal is to bridge those gaps by providing actionable analysis, dashboards, and visualizations that public health officials or volunteer organizations can use to serve their communities.

1.1. Fentanyl

Fentanyl is a highly concentrated, synthetic opioid [2], and is sometimes laced into other drugs, often causing those who take them to ingest a profoundly higher dose than intended. Fentanyl is extremely potent, which means that even small absolute amounts can lead to an overdose, especially for users who have not developed a tolerance to opioids.

Fentanyl is similar to morphine, though 50 – 100 times more potent [2]. In 2022 the DEA has issued a public warning for fentanyl related overdose death, stating that fentanyl poisoning, and deaths are at an all-time high. Fentanyl is cheap to illegally manufacture and can come in pills, eye drops, and paper form. The pills can be made to look like other prescription medications.

1.2. Naloxone

According to drugabuse.gov [3], a site maintained by the National Institute on Drug Abuse: Naloxone is a medication designed to rapidly reverse opioid overdose. It is an opioid antagonist meaning that it binds to opioid receptors and can reverse and block the effects of other opioids. It can very quickly restore normal respiration to a person whose breathing has slowed or stopped because of overdosing with heroin or prescription opioid pain medications.

The brand Narcan is a form of naloxone administered as a nasal spray. It is packaged in a carton containing two doses to allow for repeat dosing if needed. Narcan is an easy-to-administer form of the medication, and the training does not take long.

2. Datasets

This study examines data on both the state and county level. Therefore, data was harvested from the CDC [4] (Centers for Disease Control) for all the states and from PA.gov [5] for counties within Pennsylvania. Furthermore, census data [6] was used to incorporate state and county population data into our features to enable meaningful population-based comparison across geographic areas (e.g., overdose deaths per 10,000 people per state)

2.1. Data Acquisition

Data was pulled from three primary sources: PA.gov, the CDC, and the Census. The PA.gov and CDC datasets were all available via Socrata APIs [7] which provide access to a variety of open data resources from governments as well as other organizations.

The Census data is accessed via the Census microdata API [8]. A guide for calling the API can be found at Census Microdata API User Guide [9].

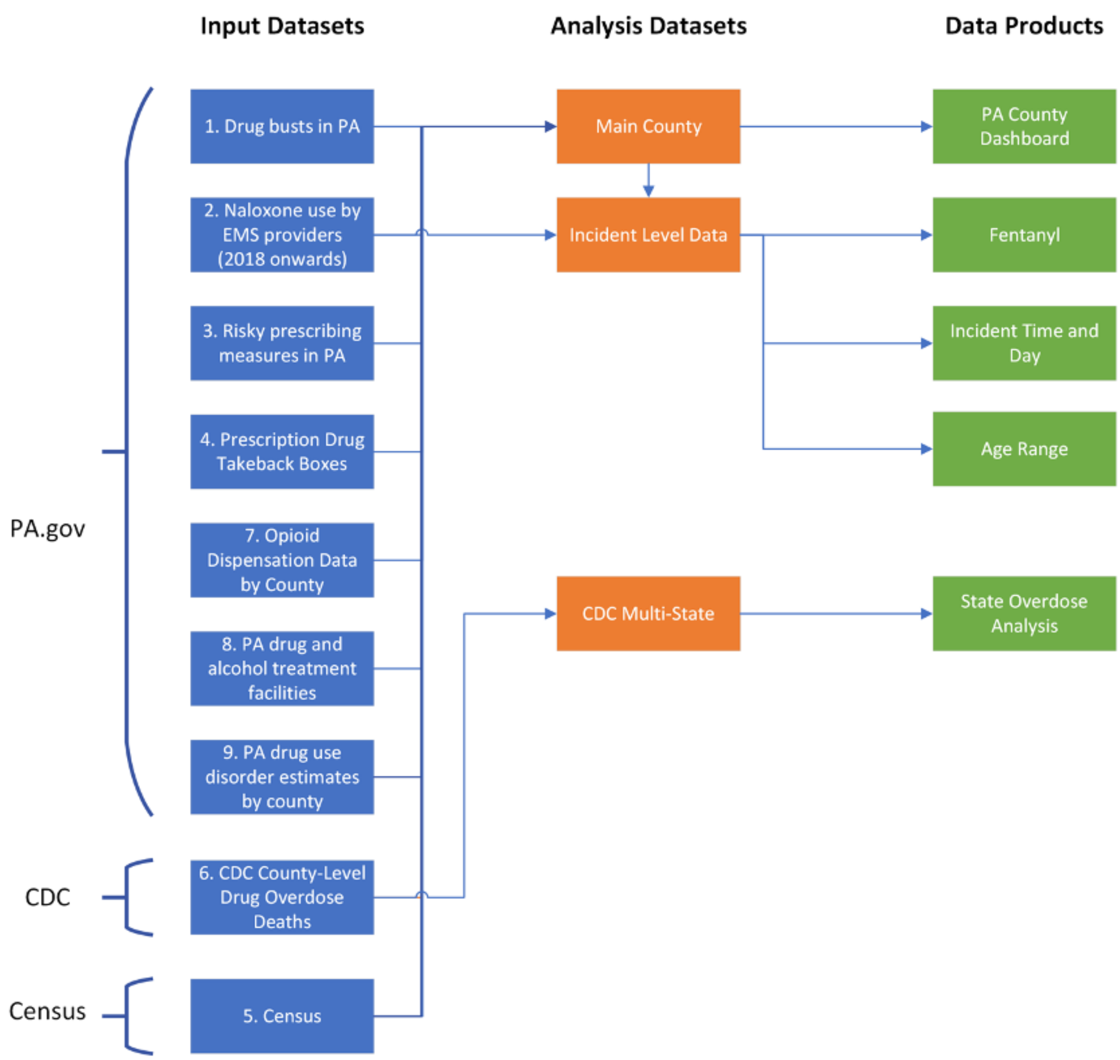

Figure 1- Data Preprocessing Diagram

2.2. Data Preprocessing

Figure 1 shows the preprocessing procedure. The left side (input datasets) contains a list of all datasets that were pulled and used in this study, and they were grouped by their source (PA.GOV, CDC, and Census). The middle column (Analysis Datasets) is where the datasets were combined, cleaned, and analyzed to produce the Data Products in the last column.

3. Exploratory Data Analysis

This study examines several rich datasets from different sources. To efficiently analyze the data in both County and State levels, Dashboards were created using Tableau Public. Those dashboards are shared on a website that was developed using GitHub Pages.

3.1. Basic Metrics

PA.GOV dataset was filtered by the drug that was used, so that the new dataset contains record related to opioids. Then, information regarding the variables was printed as shown in the figure below.

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 19235 entries, 2 to 36256
Data columns (total 10 columns):
 #   Column          Non-Null Count  Dtype
---  ------          --------------  -----
 0   Incident ID     19235 non-null  int64
 1   County          19235 non-null  object
 2   Narcan Admin    19235 non-null  int64
 3   Survive         19235 non-null  int64
 4   Gender Desc     19235 non-null  object
 5   Incident Date   19235 non-null  object
 6   Day             19235 non-null  object
 7   Incident Time   19235 non-null  object
 8   Age Range       19235 non-null  object
 9   Year            19235 non-null  int64
```

Figure 2 - Variables Information

This dataset contains 19,325 rows and there are no missing values. Next, univariate analysis was executed by creating histogram and listing unique values. County unique values are as follows:

```
array(['Delaware', 'Chester', 'Beaver', 'Bucks', 'Philadelphia',
       'Washington', 'Cumberland', 'Northumberland', 'Montgomery', 'Pike',
       'Armstrong', 'Carbon', 'Centre', 'Bradford', 'Dauphin', 'Lehigh',
       'Blair', 'Allegheny', 'Erie', 'York', 'Lebanon', 'Monroe',
       'Franklin', 'Lancaster', 'Berks', 'Mifflin', 'Westmoreland',
       'Crawford', 'Mercer', 'Luzerne', 'Susquehanna', 'Elk', 'Lycoming',
       'Snyder', 'Lackawanna', 'Lawrence', 'Wayne', 'Juniata', 'Perry',
       'Northampton', 'Adams', 'Cambria', 'Schuylkill', 'Clearfield',
       'Tioga', 'Columbia', 'Potter', 'Fulton', 'Wyoming', 'Butler',
       'Somerset', 'Fayette', 'Indiana', 'Jefferson', 'Clarion',
       'Bedford', 'Greene', 'Huntingdon', 'Union', 'McKean', 'Montour',
       'Forest', 'Warren', 'Clinton', 'Venango', 'Sullivan'], dtype=object)
```

Figure 3 - County Unique Values

There are 67 counties in PA, but there are only 66 listed in Figure 3, meaning there is one missing county which is Cameron. Distribution analysis of Gender, Year, Survive and Naloxone Administered is shown in Figure 4 below.

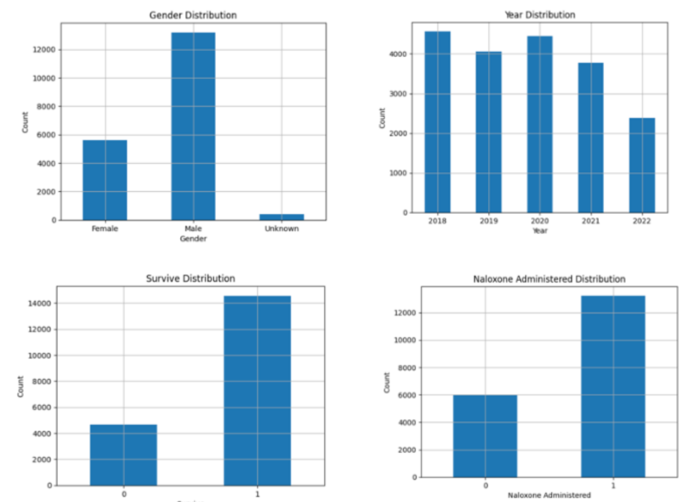

Figure 4 – Histogram of Gender (Top Left), Year (Top Right), Survive (bottom Left), and Naloxone Administered (Bottom Right)

Gender shows that more males are more involved in overdoses, and that the number of incidents is consistent throughout the years. Note that at the time of data preparation, (fall 2022) the data for 2022 was incomplete and therefore much lower than other years. The survival rate was high with ~14,000 overdoses where the victim survived compared to ~4,000 overdose deaths, but this is probably due to Naloxone administration which is also high (~12,000) compared to around 6,000 cases without Naloxone.

4. Machine Learning Models

4.1. Cluster Analysis

Method

Our cluster analysis began with the County Characteristics (Main County) dataset. This dataset combines opioid-related and demographic features for the counties in Pennsylvania. The features are listed in the table below. Each feature in the dataset is numeric and provides either percentages or counts (on a per 10,000 resident basis). The one exception is County Population, which was numeric but not calculated on a per 10,0000 resident basis. Because our clustering seeks to provide meaningful clusters related to opioid survival and Naloxone administration, we excluded % Naloxone Administration and % OD Survival from the clustering to avoid data leakage.

Table 1- Features Description

Feature	Description
% Naloxone Administration	% of overdose incidents where Naloxone is administered
% OD Survival	% of overdose incidents where the victim survived
Opioid Overdoses	# of overdoses per 10,000 residents
3+ Prescribers and 3+ Dispensers'	# of individuals per 10,000 residents with 3 or more prescribers and 3 or more dispensers of drugs
Average Daily MME > 50'	# of individuals per 10,000 residents with greater than 50 morphine milligram equivalents prescribed per day
Overlapping Opioid / Benzodiazepine Prescriptions'	# of individuals per 10,000 residents with overlapping Opioid and Benzodiazepine prescriptions
Total Drug Dispensation'	# of drug dispensations per 10,000 residents
Total Prescriptions'	# of prescriptions per 10,000 residents
Incidents - Fentanyl'	# of incidents per 10,000 residents that involved fentanyl
Incidents - Heroin'	# of incidents per 10,000 residents that involved heroin
Incidents - Opium'	# of incidents per 10,000 residents that involved opium
Arrests - Fentanyl'	# of arrests per 10,000 residents that involved fentanyl
Arrests - Heroin'	# of arrests per 10,000 residents that involved heroin
Arrests - Opium'	# of arrests per 10,000 residents that involved opium
Drug Quantity - Fentanyl'	Total quantity fentanyl seized per 10,000 residents
Drug Quantity - Heroin'	Total quantity heroin seized per 10,000 residents
Drug Quantity - Opium'	Total quantity opium seized per 10,000 residents
'%Incidents Fentanyl'	% of incidents that involved fentanyl
%Arrests Fentanyl'	% of arrests that involved fentanyl
%Quantity Fentanyl'	Fentanyl percentage of overall drugs seized
Drug Take-Back Sites'	# of drug take-back sites in the county
Drug Treatment Locations'	# of drug treatment locations in the county
County Population'	County population
County Percent Change Since 2010'	% change in county population since 2010

Results

After removing % Naloxone Administration and % OD Survival, the dataset was filtered to include only the 2021 data, thereby giving an overall structure of one record per county. The resultant dataset was standardized using a standard scaler and then subjected to three different clustering algorithms: (1) DBSCAN (2) Agglomerative Clustering (3) KMeans. We compared the algorithms by calculating the silhouette scores and Caliniski-Harabsz scores for the resultant clusters. The results are listed in the following table.

Table 2- Results

Clustering Algorithm	**Silhouette Score**	**Calinski-Harabasz Score**	**Comments**
DBScan	-0.069	2.091	Variety of clusters but poor silhouette score
KMeans	0.195	10.333	Scores for optimized K, a good mix of cluster sizes.
Agglomerative Clustering	0.169	8.952	Scores for optimized K

Initial results gave relatively low silhouette scores and Calinski-Harabasz scores, indicating that the clusters were not very distinct. Of the algorithms we examined, KMeans appeared to give the best results and we examined the silhouette scores and inertia values for values of K between 2 and 9. The results of this analysis are shown in the following figures.

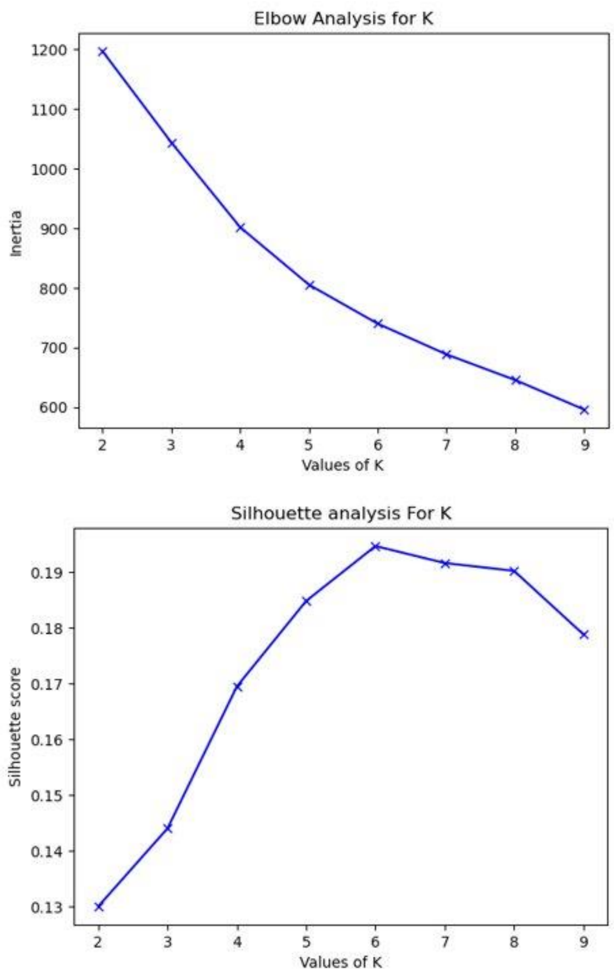

Figure 5 - Elbow analysis and Shilouette analysis as function of k

These analyses agreed on an optimal value of K=6. Using KMeans (K=6) we generated assigned labels to the clusters. We then combined joined the cluster mapping to the incident level dataset and calculated the percent survival for each cluster. The results are shown in the table 3.

There was considerable variation in the percent survival from ~54% in cluster 2 and ~85% in cluster 5. This indicates that clustering is providing some meaningful information about the likelihood of survival. Unfortunately, when we used this clustering as a feature for the classification problem, it did not materially improve our results.

Table 3- Results

Cluster	Cases with Survival	Total Cases	% Survival
0	9665	12667	76.3
1	2429	3499	69.4
3	968	1402	69.0
4	2194	2715	80.8
5	168	197	85.3

4.2. Feature Selection Survival

For survival, we measured correlation of features against the target variable, further we implemented Sklearn Feature Selection library SelectKBest class to complete a Chi-square score on the features verses the target variable. We began by label encoding the strings to determine important features.

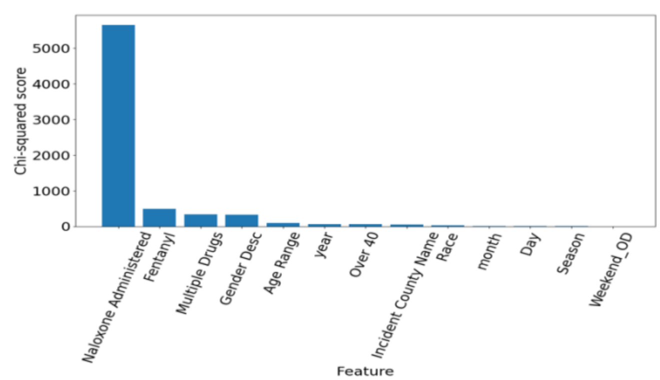

Figure 6 - Feature selection following one-hot encoding.

Features such as Naloxone administration, Fentanyl, Multiple Drugs, Gender, and Age Range were the most important. However, there are 67 distinct counties in PA, and this could likely lead to confusion in the algorithm. Next, We one hot encoded the features and re-ran the test, and the results are shown in figure 7.

Incident county of Philadelphia was an important feature. Much of our early exploratory analysis and public health articles confirm that Philadelphia has both a growing rate of Fentanyl incidents and low survival rate. We determined that we would explore either using Philadelphia as a binary feature, or cluster labels rather than incident county name is a better strategy.

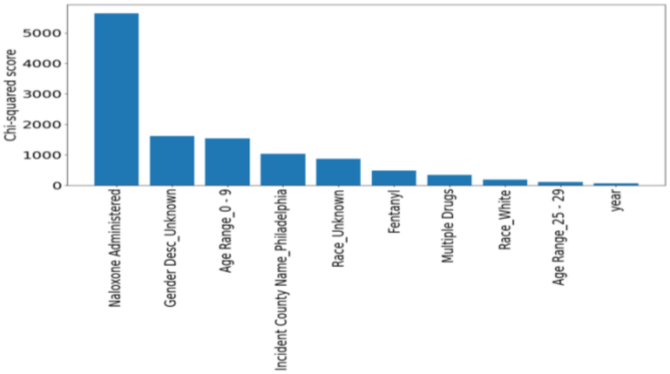

Figure 7 - Feature selection following one-hot encoding.

4.3. Feature Selection Naloxone Administration

For naloxone administration, we also measured correlation of features against the target variable, further we implemented Sklearn Feature Selection library SelectKBest class to complete a Chi-square score on the features verses the target variable. We began by label encoding the strings to determine important features.

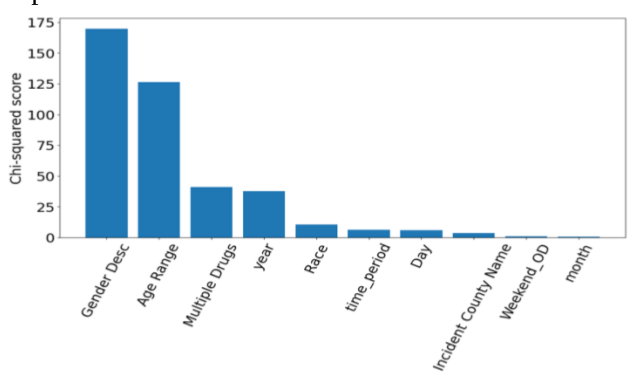

Figure 9 - Feature selection for Naloxone Administration model

However, the incident county feature was surprisingly not important, even though exploratory analysis from our first report indicated that it would be. This is likely due to there being 67 distinct counties, many with low population. Next, we one hot encoded the features and re-ran the test, whose results are shown in figure 9.

After we one-hot encoded the features, much like with Survival, we discovered that some counties were important features. Likewise, the age range of 0-9 was important. This algins with our exploratory analysis from the first session.

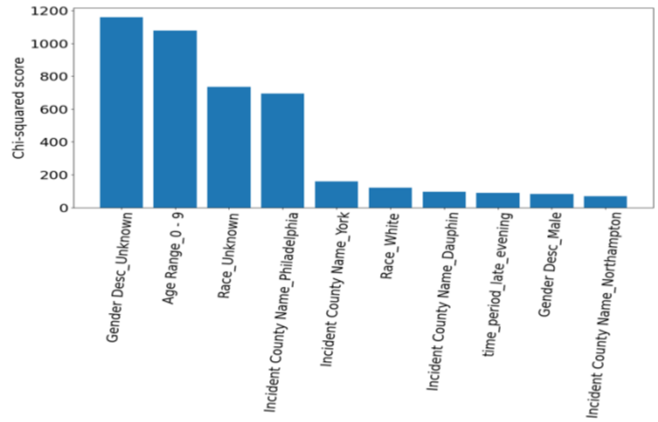

Figure 8 - Feature selection following one-hot encoding.

4.4. Unbalance Class Distribution

For both survival and naloxone administration there is unbalanced class distribution. Out of the samples in the survival analysis: 75.2 % Survived and 24.8% died. Likewise, from the Naloxone administration analysis 68.7% had Naloxone administered and 31.3% did not. We wanted to ensure that the sensitivity and specificity were as close as possible rather than relying on accuracy score as a metric. With that in mind, we applied both class weights and probability threshold to both models.

4.5. Classification Survival

method

We chose three models: Random Forrest, Logistic Regression, and Decision Tree for survival classification. The three models had similar metrics. Due to the class imbalance, we wanted to ensure that the models performed at 70% for sensitivity and specificity. Because of this, we implemented both class weights and thresholding. To pick the optimal threshold, we measured both the Matthews score and AUC at different thresholds.

We used the following features to predict survival:

- If the victim is Over 40
- If the victim was given Naloxone
- If it was a Fentanyl incident
- If it was a Multiple Drug incident
- If the incident occurred in Philadelphia

Results

Random Forrest:

Table 4 - Confusion Matrix

	Predict Survive	Predict Died
Survived	12,023	3,449
Died	1,522	3,575

AUC: 0.74
Matthews Correlation: 0.44

```
Classification Report:
             precision  recall   f1-score   support

        Died      0.51     0.70      0.59      5097
    Survived      0.89     0.78      0.83     15472

    accuracy                         0.76     20569
   macro avg      0.70     0.74      0.71     20569
weighted avg      0.79     0.76      0.77     20569
```

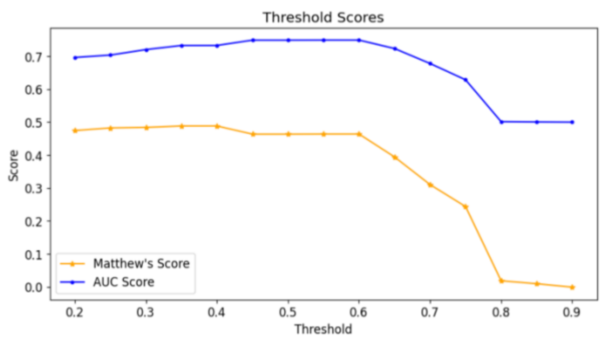

Figure 10 - AUC and Matthew's score as function of Threshold

Logistic Regression:

Table 5 - Confusion Matrix

	Predict Survive	Predict Died
Survived	11,483	3,989
Died	1,405	3,692

AUC: 0.73
Matthews Correlation: 0.41

```
Classification Report:
             precision  recall   f1-score   support

        Died      0.48     0.72      0.57      5097
    Survived      0.89     0.74      0.81     15472

   micro avg      0.74     0.74      0.74     20569
   macro avg      0.68     0.73      0.69     20569
weighted avg      0.79     0.74      0.75     20569
```

Decision Tree:

Table 6 - Confusion Matrix

	Predict Survive	Predict Died
Survived	10,969	4,503
Died	1,350	3,747

AUC: 0.72
Matthews Correlation: 0.39

```
Classification Report:
             precision  recall   f1-score   support

        Died      0.45     0.74      0.56      5097
    Survived      0.89     0.71      0.79     15472

   micro avg      0.72     0.72      0.72     20569
   macro avg      0.67     0.72      0.68     20569
weighted avg      0.78     0.72      0.73     20569
```

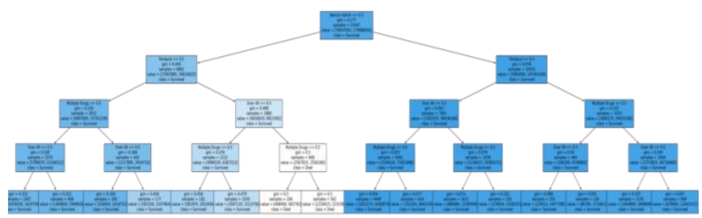

Figure 11 - Decision Tree Result

4.6. Classification Naloxone

Method

A victim getting Naloxone during an overdose encounter is highly predictive. With that in mind, we also wanted to examine if we could predict if Naloxone would be used.

We used the following features:

- Season of event (winter, spring, summer, fall)
- If the event was during the weekend
- If the victim took Fentanyl
- Race of the victim
- Incident County Name
- Time period of the incident

Results

	Predicted Not Given Naloxone	Predicted Given Naloxone
Not Given Naloxone	3670	2763
Given Naloxone	5898	8238

AUC: 0.62

Classification Report:

```
                   precision   recall   f1-score   support

Not Given Naloxone  0.38        0.57      0.46       6433
    Given Naloxone  0.75        0.58      0.66      14136

         micro avg  0.58        0.58      0.58      20569
         macro avg  0.57        0.58      0.56      20569
      weighted avg  0.63        0.58      0.59      20569
```

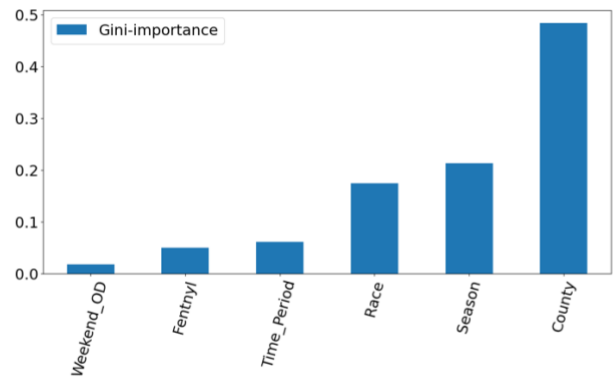

Figure 12- Importance Index

4.7. Regression (State Level)

Method

Linear regression was utilized to predict the number of overdose death per 10,000 in the state level. CDC overdose death data was merged with population estimate data between 2015 and 2021 to achieve overdose death per 10,000 for each state as can be seen below:

	State	2015	2016	2017	2018	2019	2020	2021
0	Alabama	18.172649	17.115158	19.593861	18.818740	17.627318	21.143117	27.968937
1	Alaska	19.959376	21.498241	21.238340	19.615338	21.106015	21.721886	32.975147
2	Arizona	21.362653	23.631796	24.917632	27.497812	30.208621	39.657364	44.879854
3	Arkansas	15.500086	15.094059	15.909534	17.563020	15.703472	17.677921	22.390099
4	California	14.389469	14.592343	14.839489	15.995555	17.952673	23.496865	32.628711

Figure 14 - Data Sample

Fitting occurs while looping over State, predicting a year at a time, till the desired year is reached. Since there are many states, a sample was randomly chosen and is shown in figure 12.

Results

In the graph, we see the results of a model that predicted up to 2025. Pennsylvania is blue, Delaware is red, Texas is green, and the average of all states is in orange.

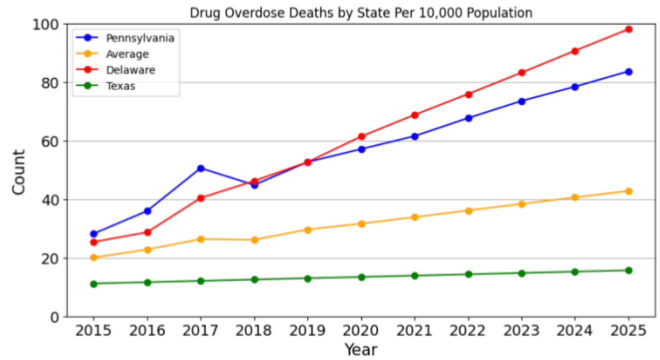

Figure 15- Linear Regression Results

4.8. Time Series (County Level)

Method

Univariate time-series approach was utilized on the county level to forecast the number of Opioid overdose incidents up to a desired date. For that purpose, the incidents data was read, and after filtering for the chosen county, missing months were imputed with 0 counts, and 5 lags were inserted into each row, meaning there are 5 sequential features (previous counts), and 5 sequential targets (future counts) as can be seen in figure 14. Sequential means that they are 1 month apart from one another, for instance, in the example below, the column Count is related to 2018-01-01, column x_1 is related to the following month 2018-02-01, and so on. Column y_0 is following x_4, so in this example, x_4 is related to 2018-05-01, then y_0 is 2018-06-01.

Date	Count	x_1	x_2	x_3	x_4	y_0	y_1	y_2	y_3	y_4
2018-01-01	3.0	4.0	2.0	3.0	1.0	5.0	8.0	8.0	4.0	2.0
2018-02-01	4.0	2.0	3.0	1.0	5.0	8.0	8.0	4.0	2.0	1.0
2018-03-01	2.0	3.0	1.0	5.0	8.0	8.0	4.0	2.0	1.0	0.0
2018-04-01	3.0	1.0	5.0	8.0	8.0	4.0	2.0	1.0	0.0	8.0
2018-05-01	1.0	5.0	8.0	8.0	4.0	2.0	1.0	0.0	8.0	0.0

Figure 13 - Data Sample

To forecast up to a specific date, the algorithm appends a new row with the following month and shift all values to the left. The last column (y_4) is set to NaN, until the forecasted value is placed.

Results

There are 67 counties in Pennsylvania, 66 of them are presented in this dataset, so to efficiently present the data, and for easier hyperparameter tuning, an interactive plot was created using ipywidgets package. This plot runs on Jupyter Notebook and allows filtering results as can be seen in figure 16.

In the upper left corner, there are dropdowns and sliders that enable filtering County, Regressor, Year, Month, Criterion, N_estimators (for Gradient Boosting and Random Forest), and max_features. Once any widget's value is changed, the process of filtering county, filling missing values, inserting lags, fitting, and forecasting starts. Once it is done, a new plot is drawn showing a few graphs:

1. The upper graph which includes: input – blue, test- red, and forecast – orange.
2. The lower graphs are additive and multiplicative decomposition. Those show the different components of the observed time-series which are trend, seasonality, and residual (noise).

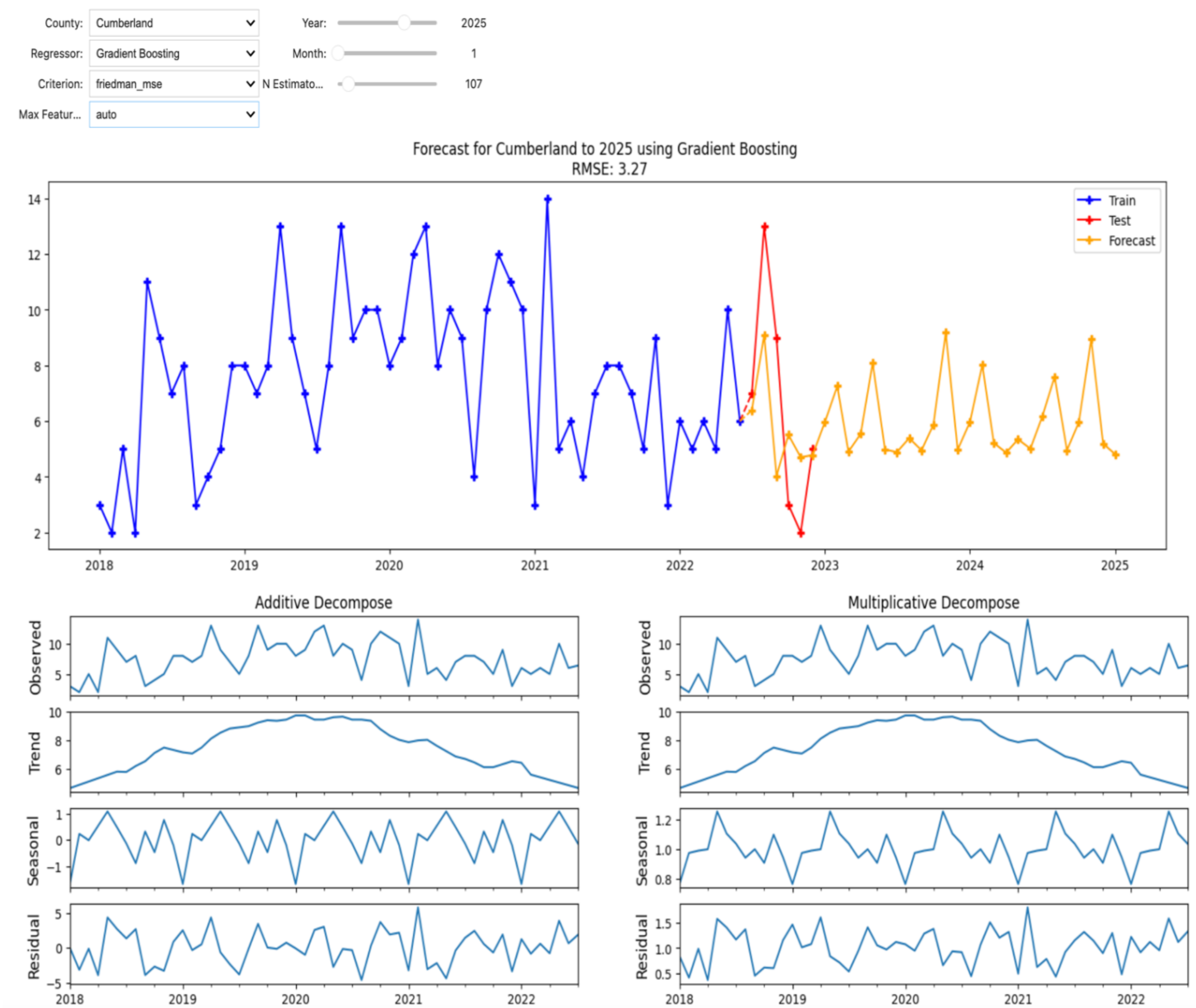

Figure 16 - Time-series interactive plot results

5. Discussion

Several models were developed in this project, each of which focused on a different level: incidents (two classification models), county level (cluster analysis and univariate time-series), and state level (linear regression).

5.1. Findings

Cluster analysis yield promising results grouping together counties with similar characteristics. Further analysis is needed to compare counties both within clusters and cross clusters to find the features that increase or reduce efficiency of dealing with the epidemic.

There are two classification models: one for survival and another for Naloxone administration. Naloxone is highly correlated with survival, and therefore the performance of the naloxone administration is worse than survival. We conclude that Naloxone is an efficient drug, that should be widely distributed, especially in counties who are more affected by the epidemic such as Philadelphia. Some measures have already been put in place in some counties as described in the next section (5.2).

Another finding from the classification model is that geography is an important feature. We used it as a binary feature for Philadelphia, but other counties might be good candidates.

Univariate time series was utilized on the 66 counties (one county was missing in the dataset) and developed an interactive plot for better presentation as well as easier hyperparameter tuning process. An interesting finding is that the model didn't work well for Philadelphia which had a spike of overdose incidents during covid-19 (2021) as can be seen in figure 17 (Appendix A). We arrived at the conclusion that multi-variate time-series might perform better because its ability to adjust to short-term events and trends.

Last model is a linear regression for every state. State level datasets that were found in this project were short of important features, and therefore a richer dataset is required to develop enhanced and more accurate models to fight Opioid overdose nationwide.

5.2. Recent Public Health Policy Changes

Naloxone is a prescription medication, however on January 17th, 2023, a standing order was issued in Pennsylvania that allows pharmacists to dispense naloxone without requiring an individual prescription. Many other states have similar legislation in place, and commercials for Naloxone are run online by the Pennsylvania department of public health. Furthermore, other states such as New Jersey have put similar measures in place. If data on dispensing Naloxone on the Pharmacy level is ever published, that data could also be used for future work. Lastly, it will be worth measuring if these changes in public health policy helps increase the percentage of overdoses treated with Naloxone.

6. Conclusions

The opioid crisis in the United States has been a public health issue for many years. Our project aimed to provide machine learning tools on different levels, so that public health officials or volunteer organizations can use to serve their communities. We utilized several rich datasets from different sources, including PA.GOV, CDC, and County Characteristics dataset.

Our analysis revealed several important features related to survival and naloxone administration, including Naloxone administration, Fentanyl, Multiple Drugs, Gender, and Age Range. Our cluster analysis identified six meaningful clusters if counties which significantly predicted the likelihood of survival, and our classification models achieved high sensitivity and specificity scores.

We also utilized regression to predict the number of overdose deaths per 10,000 at the state level, and time series analysis to forecast the number of opioid overdose incidents up to a desired date at the county level.

Overall, our project provides valuable insights into the opioid crisis in the United States and can be used to inform public health interventions and policies.

References

1. Center for Disease Control and Prevention (CDC), *U.S. Overdose Death In 2021 Increased Half as Much as in 2020 – But Are Still Up 15%*, May 11, 2022, https://www.cdc.gov/nchs/pressroom/nchs_press_releases/2022/202205.htm.

2. United States Drug Enforcement Administration (DEA), *Fentanyl*, April 2020, https://www.dea.gov/factsheets/fentanyl.

3. National Institute on Drug Abuse (NIDA), 2022, January 11. Naloxone DrugFacts. Retrieved from https://nida.nih.gov/publications/drugfacts/naloxone on 2022, November 26.

4. CDC, *Provisional County-Level Drug Overdose Death Counts*, October 26, 2022. https://www.cdc.gov/nchs/nvss/vsrr/prov-county-drug-overdose.htm

5. Pennsylvania State Police, *Overdose Information Network Data CY January 2018 - Current Monthly County State Police*, December 27, 2018, https://data.pa.gov/Opioid-Related/Overdose-Information-Network-Data-CY-January-2018-/hbkk-dwy3

6. Census, State Population Totals 2010 - 2019, December 16, 2021. https://www.census.gov/data/datasets/time-series/demo/popest/2010s-state-total.html#par_textimage_1873399417

7. SODA Developers, Socrata API, https://dev.socrata.com.

8. Census Microdata API, https://api.census.gov/data.html

9. Census, Census Microdata API User Guide, July 22, 2021. https://www.census.gov/data/developers/guidance/microdata-api-user-guide.How_To_Create_API_Queries.html

10. Abraham, Aswini et al. "*An Examination of Seasonal Trends in Delaware Drug Overdoses, 2016-2020.*" Delaware journal of public health vol. 7,5 44-51. 15 Dec. 2021, doi:10.32481/djph.2021.12.014

Appendix A – Philadelphia Univariate Time Series

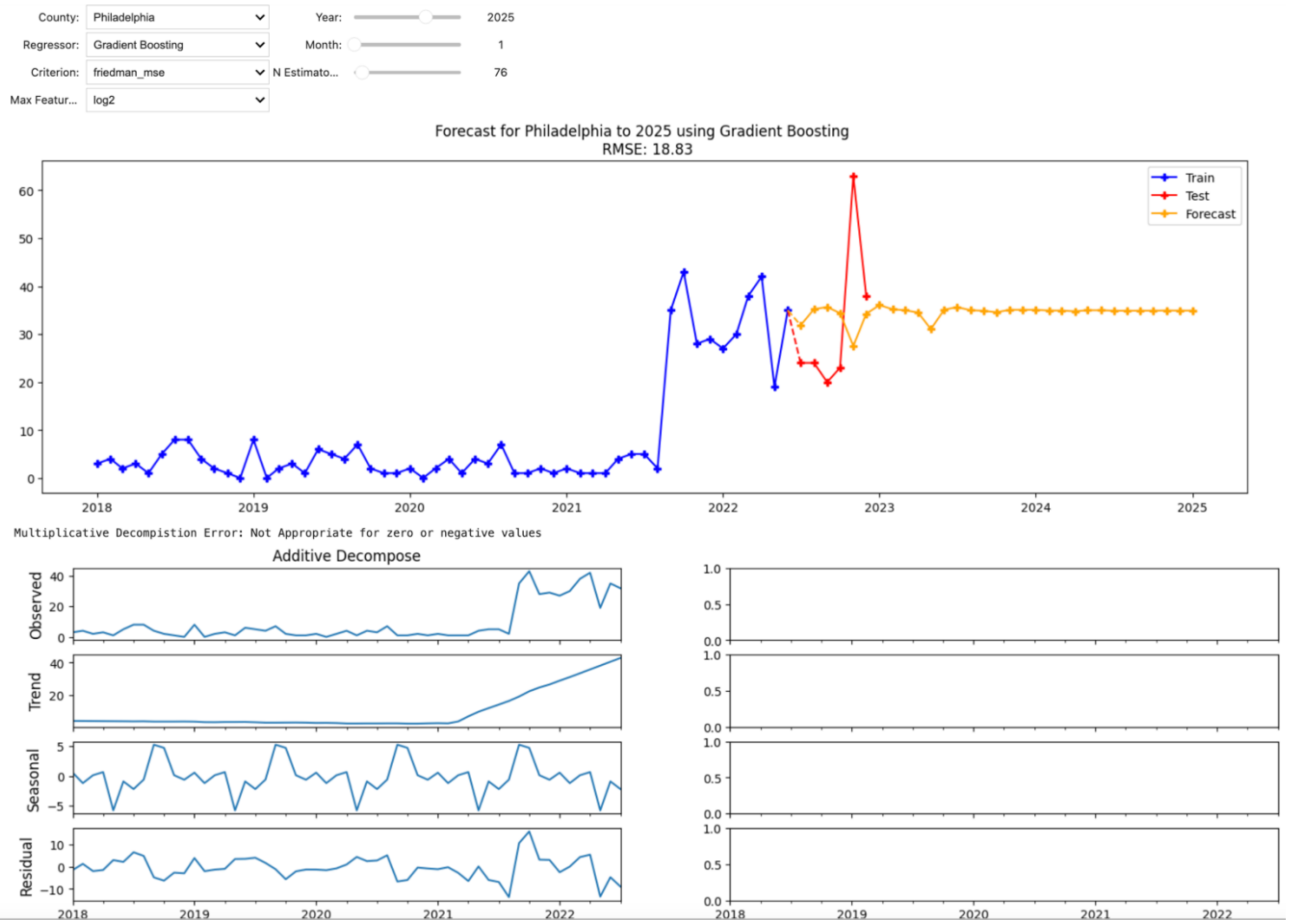

Figure 17 - Philadelphia time-series model performance is not satisfying.

In the figure above, one can see a spike during 2021 (covid-19), and the forecast looks flat, similar to the pattern prior to the spike. Therefore, we conclude that a multi-variate time series might perform better. Note that the right bottom graphs are missing, that is due to imputation of missing months with 0. In other words, months that had 0 incidents, were inserted, and imputed with 0, but multiplicative decomposition raises an error in those cases.