Predicting Fentanyl Presence In Accidental Drug Overdose Deaths In Connecticut, USA

DSI-23: Capstone Project

Ray Tan

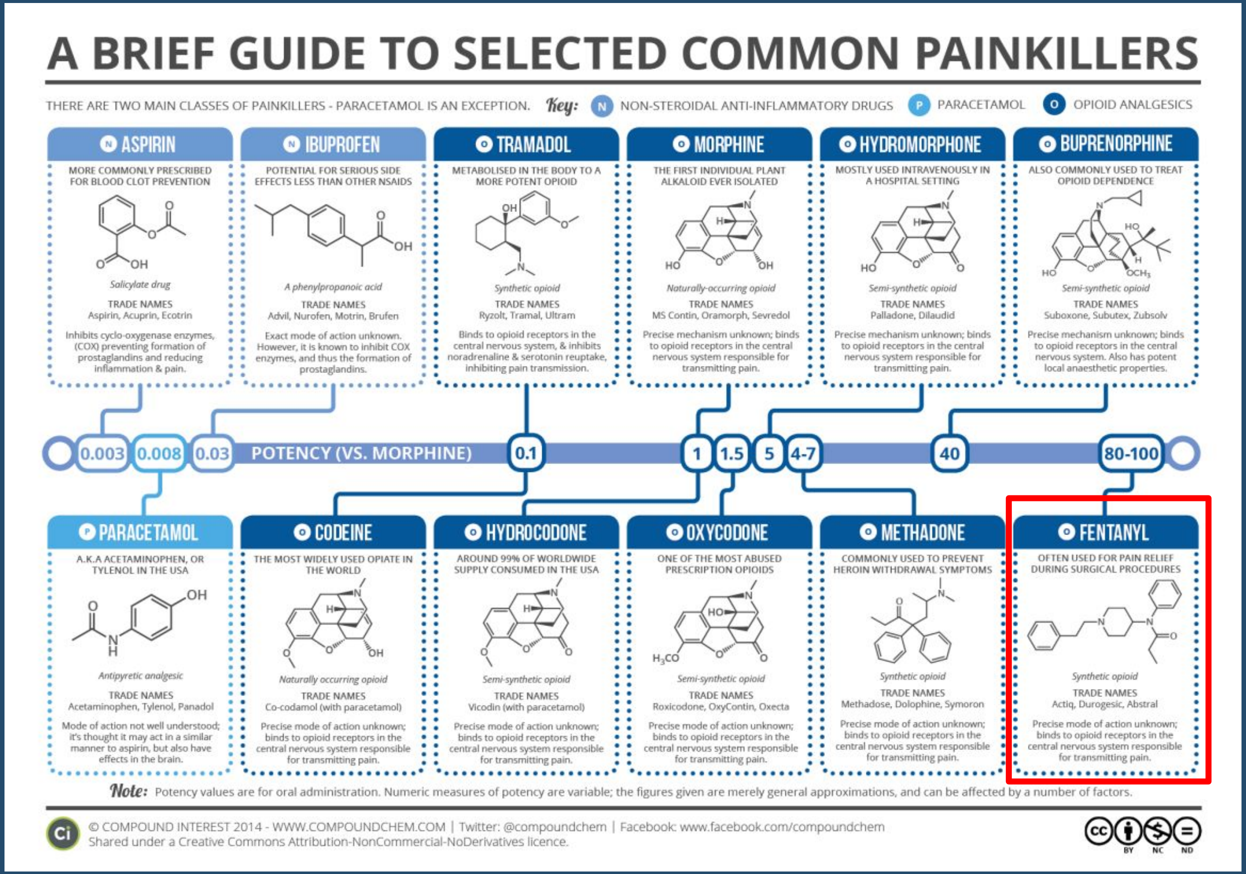
Outline

1. Introduction
2. Data Source
3. Exploration & Visualisation
4. Pre-Processing & Modelling
5. Conclusion

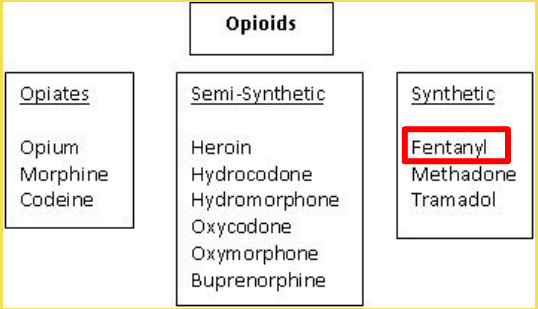
Outline

1. Introduction
2. Data Source
3. Exploration & Visualisation
4. Pre-Processing & Modelling
5. Conclusion

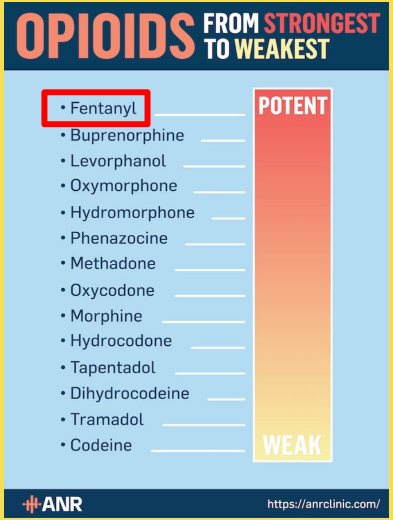
Background



A BRIEF GUIDE TO SELECTED COMMON PAINKILLERS
THERE ARE TWO MAIN CLASSES OF PAINKILLERS - PARACETAMOL IS AN EXCEPTION.
Key: N NON-STEROIDAL ANTI-INFLAMMATORY DRUGS P PARACETAMOL O OPIOID ANALGESICS
N ASPIRIN
MORE COMMONLY PRESCRIBED FOR BLOOD CLOT PREVENTION
Salicylate drug
TRADE NAMES
Aspirin, Acuprin, Ecotrin
Inhibits cyclo-oxygenase enzymes, (COX) preventing formation of prostaglandins and reducing inflammation & pain.
N IBUPROFEN
POTENTIAL FOR SERIOUS SIDE EFFECTS LESS THAN OTHER NSAIDS
A phenylpropanoic acid
TRADE NAMES
Advil, Nurofen, Motrin, Brufen
Exact mode of action unknown. However, it is known to inhibit COX enzymes, and thus the formation of prostaglandins.
O TRAMADOL
METABOLISED IN THE BODY TO A MORE POTENT OPIOID
Synthetic opioid
TRADE NAMES
Ryzolt, Tramal, Ultram
Binds to opioid receptors in the central nervous system, & inhibits noradrenaline & serotonin reuptake, inhibiting pain transmission.
O MORPHINE
THE FIRST INDIVIDUAL PLANT ALKALOID EVER ISOLATED
Naturally-occurring opioid
TRADE NAMES
MS Contin, Oramorph, Sevredol
Precise mechanism unknown; binds to opioid receptors in the central nervous system responsible for transmitting pain.
O HYDROMORPHONE
MOSTLY USED INTRAVENOUSLY IN A HOSPITAL SETTING
Semi-synthetic opioid
TRADE NAMES
Palladone, Dilaudid
Precise mechanism unknown; binds to opioid receptors in the central nervous system responsible for transmitting pain.
O BUPRENORPHINE
ALSO COMMONLY USED TO TREAT OPIOID DEPENDENCE
Semi-synthetic opioid
TRADE NAMES
Suboxone, Subutex, Zubsolv
Precise mechanism unknown; binds to opioid receptors in the central nervous system. Also has potent local anaesthetic properties.
0.003 0.008 0.03 POTENCY (VS. MORPHINE) 0.1 1 1.5 5 4-7 40 80-100
P PARACETAMOL
A.K.A ACETAMINOPHEN, OR TYLENOL IN THE USA
Antipyretic analgesic
TRADE NAMES
Acetaminophen, Tylenol, Panadol
Mode of action not well understood; it's thought it may act in a similar manner to aspirin, but also have effects in the brain.
O CODEINE
THE MOST WIDELY USED OPIATE IN THE WORLD
Naturally occurring opioid
TRADE NAMES
Co-codamol (with paracetamol)
Precise mode of action unknown; binds to opioid receptors in the central nervous system responsible for transmitting pain.
O HYDROCODONE
AROUND 99% OF WORLDWIDE SUPPLY CONSUMED IN THE USA
Semi-synthetic opioid
TRADE NAMES
Vicodin (with paracetamol)
Precise mode of action unknown; binds to opioid receptors in the central nervous system responsible for transmitting pain.
O OXYCODONE
ONE OF THE MOST ABUSED PRESCRIPTION OPIOIDS
Semi-synthetic opioid
TRADE NAMES
Roxicodone, OxyContin, Oxecta
Precise mode of action unknown; binds to opioid receptors in the central nervous system responsible for transmitting pain.
O METHADONE
COMMONLY USED TO PREVENT HEROIN WITHDRAWAL SYMPTOMS
Synthetic opioid
TRADE NAMES
Methadose, Dolophine, Symoron
Precise mode of action unknown; binds to opioid receptors in the central nervous system responsible for transmitting pain.
O FENTANYL
OFTEN USED FOR PAIN RELIEF DURING SURGICAL PROCEDURES
Synthetic opioid
TRADE NAMES
Actiq, Durogesic, Abstral
Precise mode of action unknown; binds to opioid receptors in the central nervous system responsible for transmitting pain.
Note: Potency values are for oral administration. Numeric measures of potency are variable; the figures given are merely general approximations, and can be affected by a number of factors.
© COMPOUND INTEREST 2014 - WWW.COMPOUNDCHEM.COM | Twitter: @compoundchem | Facebook: www.facebook.com/compoundchem
Shared under a Creative Commons Attribution-NonCommercial-NoDerivatives licence.



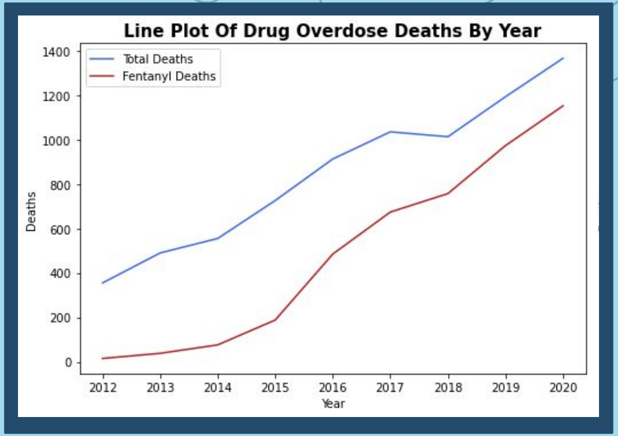
Opioids
Opiates
Opium
Morphine
Codeine
Semi-Synthetic
Heroin
Hydrocodone
Hydromorphone
Oxycodone
Oxymorphone
Buprenorphine
Synthetic
Fentanyl
Methadone
Tramadol

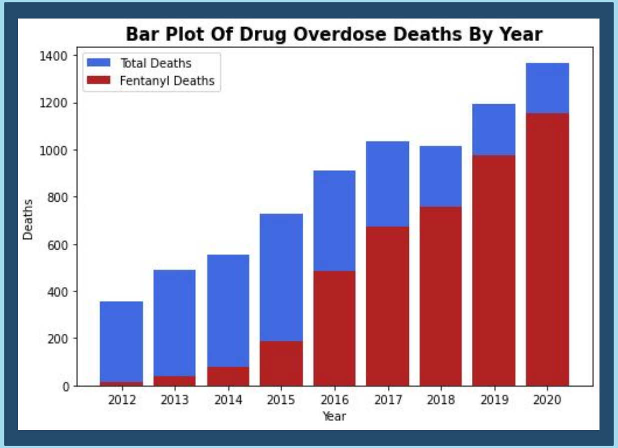


OPIOIDS FROM STRONGEST TO WEAKEST
• Fentanyl
• Buprenorphine
• Levorphanol
• Oxymorphone
• Hydromorphone
• Phenazocine
• Methadone
• Oxycodone
• Morphine
• Hydrocodone
• Tapentadol
• Dihydrocodeine
• Tramadol
• Codeine
POTENT
WEAK
ANR
https://anrclinic.com/

Background

- 2013 marks the **year of emergence** of fentanyl in the illicit drug market
- Fentanyl-related overdose deaths have **skyrocketed in recent years**
- Connecticut has **surpassed** the national death rate for fentanyl-related overdoses since 2013
- 2020 alone has been a **record-breaking year** for fentanyl-related overdose deaths in Connecticut



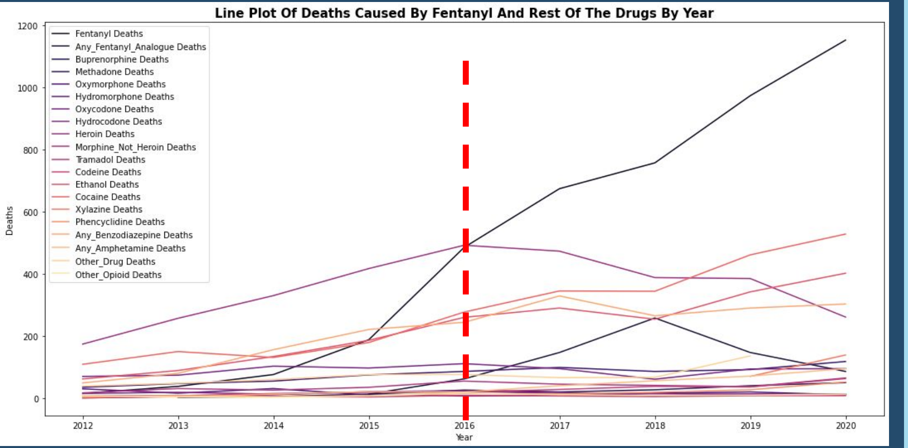


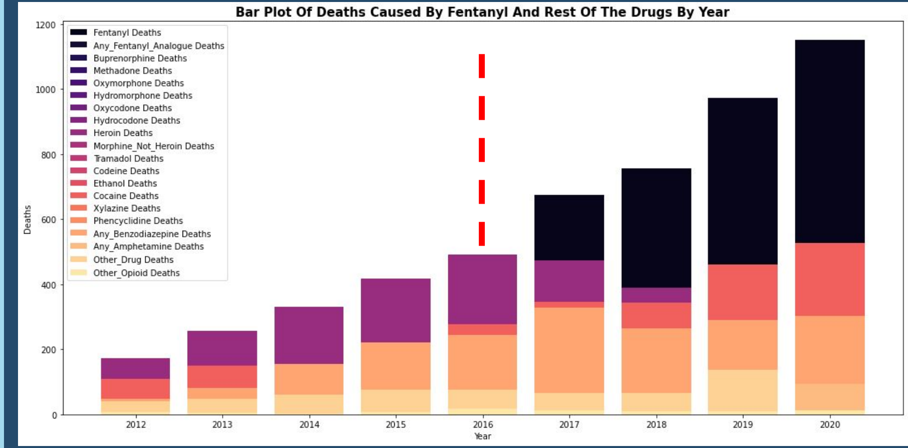
The number of fentanyl deaths has been increasing much faster than the number of total deaths.

Background

- Fentanyl has been implicated in **more overdose deaths** in Connecticut at present as compared to other drugs

- Fentanyl is now the **number one drug** found in overdose deaths in Connecticut





Pre-2016, heroin was responsible for the most number of deaths each year. Post-2016, fentanyl surpassed heroin to be the number one cause of deaths each year.

Problem Statement

- Accidental drug overdose death problem is expected to **continue to worsen** in the coming year

- Fentanyl is the **dominant drug** seen in overdose deaths today

- There is an urgent need for **effective harm-reduction strategies** to tackle the escalating fentanyl crisis and stop deaths from occurring

- As a start, it would be beneficial to know what are some of the **main factors or drivers** causing the surge in fentanyl-related overdose deaths in recent years

Target Audience

Connecticut State Officials

Better Understand

Circumstances surrounding fentanyl-related overdose deaths

Better Recognise

Top features that are most important in explaining fentanyl presence in overdose deaths

Objectives

- Develop several **classification models**
- Evaluate their individual performance based on two metrics: **Accuracy** and **Area Under Curve**
- Choose the **best-performing hyperparameter-tuned model** that achieves an accuracy closest to 1 and an area under curve closest to 1
- Identify the **top features** that are **most important** in explaining fentanyl presence in overdose deaths
- Guide Connecticut state officials on the allocation of **scarce resources** to **key areas** that influence fentanyl presence in overdose deaths

Outline

1 Introduction

2 Data Source

3 Exploration & Visualisation

4 Pre-Processing & Modelling

5 Conclusion

Data Source

About this Dataset

Source: https://data.ct.gov/

Locale: Connecticut

Period: 2012-2020

What is in this Dataset?

Rows	Columns	Each row is a
7,679	42	Death

Columns in this Dataset

- Age
- Sex
- Race
- Location of Death
- Cause of Death
- Other Significant Conditions
- Overdosed Drugs (e.g. Fentanyl)

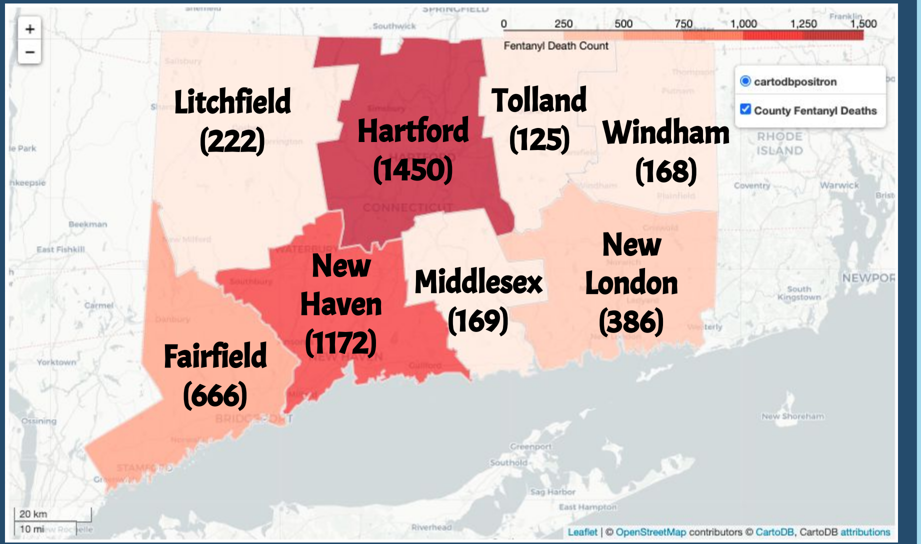
Engineered Columns

- Covid Pandemic
- Number of Drugs
- Number of Conditions

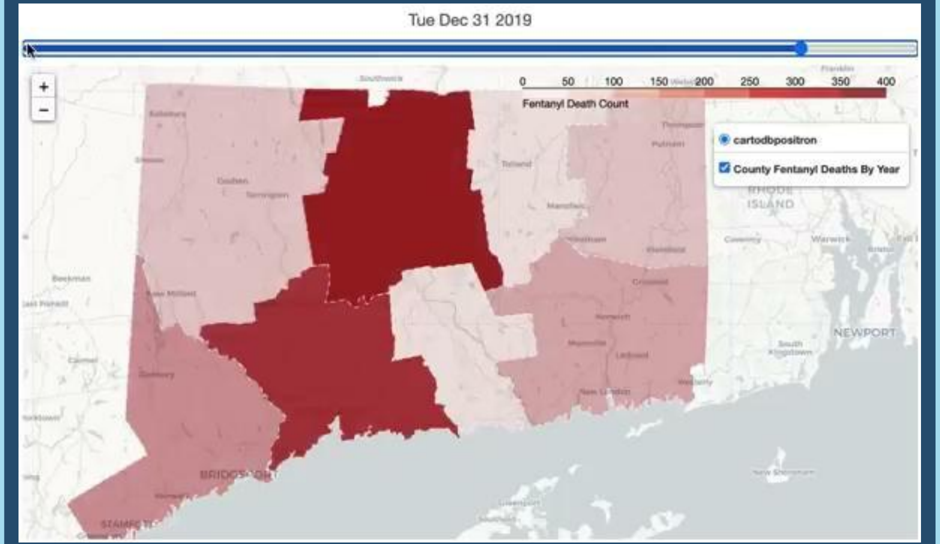
Outline

1. Introduction
2. Data Source
3. Exploration & Visualisation
4. Pre-Processing & Modelling
5. Conclusion

Choropleth Maps Of Fentanyl Deaths



From 2012 to 2020, Hartford County recorded the highest number of fentanyl deaths at 1450 whereas Tolland County recorded the lowest at 125.

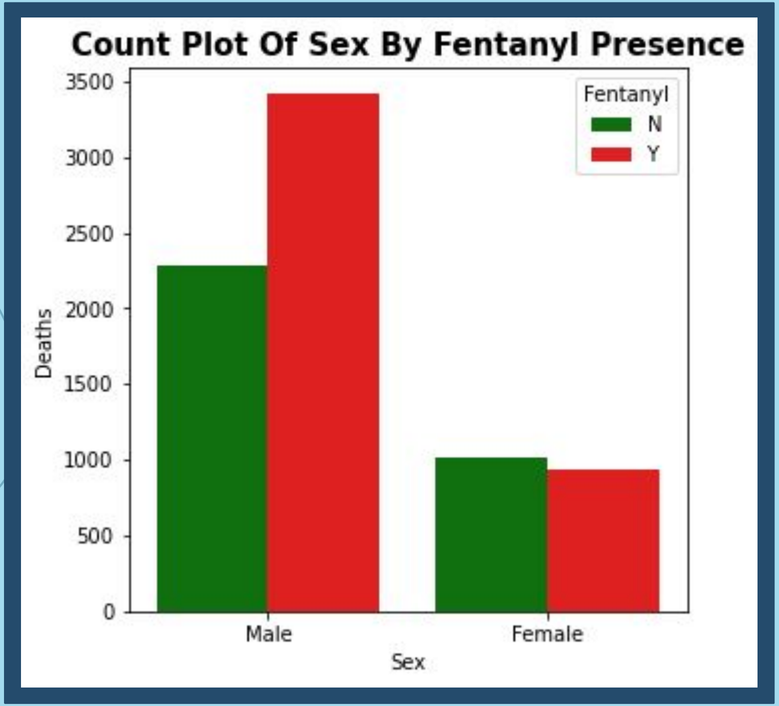


From 2012 to 2020, all counties have been experiencing a rising trend in the number of fentanyl deaths.

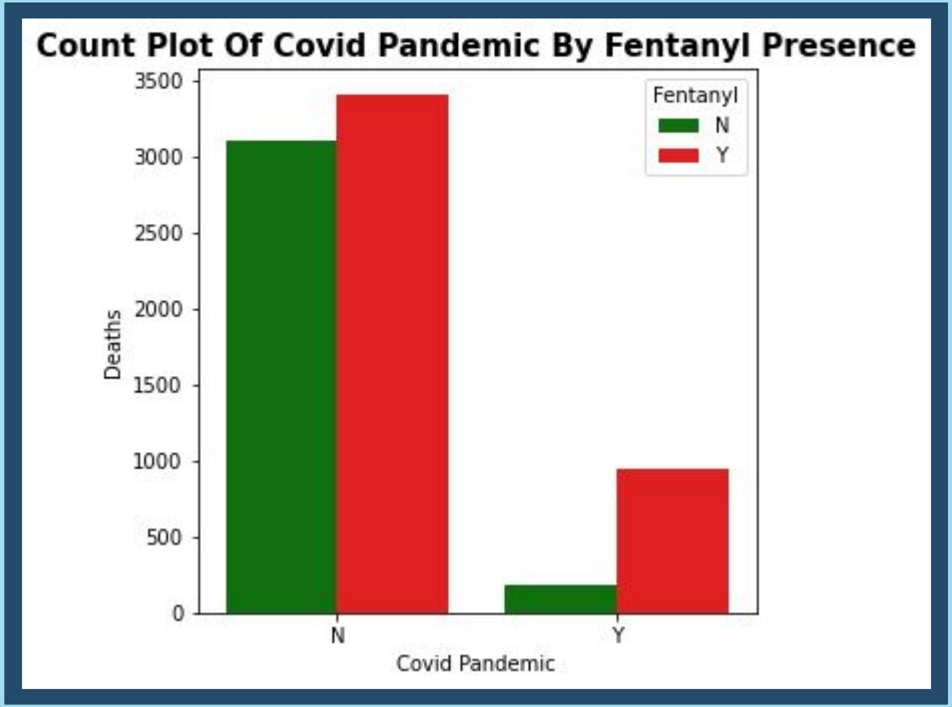
Point Map Of Fentanyl Deaths

From 2012 to 2020, the number of cities recording at least 1 fentanyl death each year has been on the rise.

Plots Of Sex & Covid Pandemic

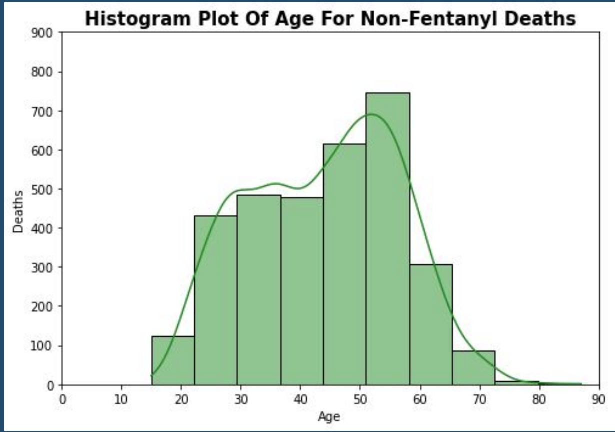


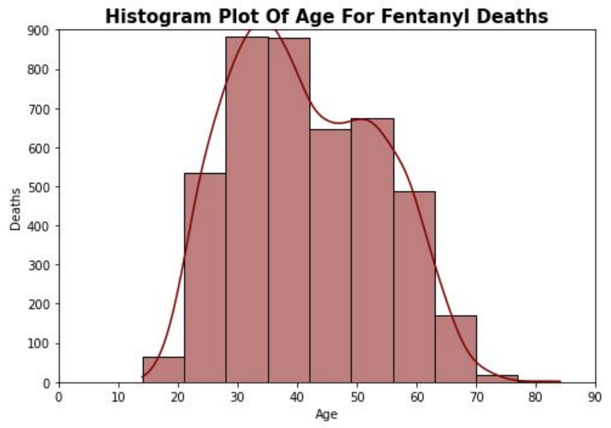
In females, the number of fentanyl deaths and non-fentanyl deaths are comparable. In males, the number of fentanyl deaths far outnumbers that of non-fentanyl deaths.

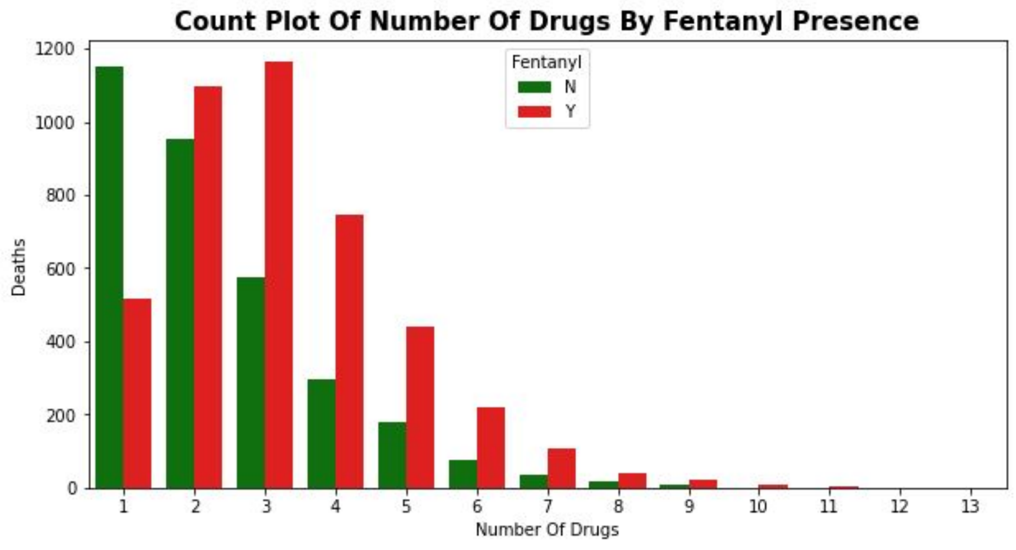


Before Covid-19 was declared a pandemic, there were slightly more fentanyl deaths than non-fentanyl deaths. After Covid-19 was declared a pandemic, there were significantly more fentanyl deaths than non-fentanyl deaths.

Plots Of Age & Number Of Drugs



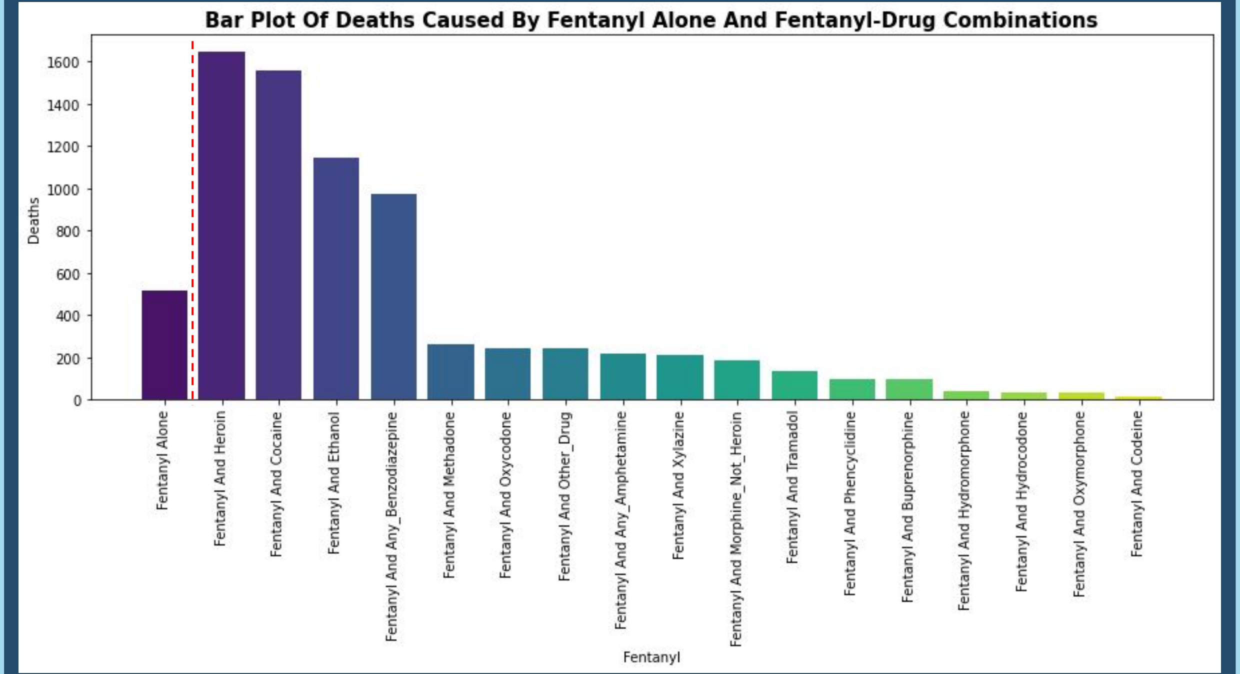




In non-fentanyl deaths, the highest number of deaths is seen among single drug users. In fentanyl deaths, the highest number of deaths is seen among triple drug users, followed by double drug users, and then single drug users.

In non-fentanyl deaths, the distribution of age has a peak number of deaths at 50-60 years. In fentanyl deaths, the distribution of age has a peak number of deaths at 30-40 years.

Plot Of Fentanyl Alone Vs Fentanyl-Drug Combinations



Fentanyl alone has been responsible for 515 deaths. The following drugs with fentanyl - heroin, cocaine, ethanol, or any benzodiazepine - have each resulted in more deaths than fentanyl alone. The most lethal duo is a fentanyl-heroin combination which has been responsible for 1644 deaths.

Outline

1. Introduction
2. Data Source
3. Exploration & Visualisation
4. Pre-Processing & Modelling
5. Conclusion

Workflow

4

Pre-Processing & Modelling

A Set Up Environment

B Compare Models

C Tune Models

D Evaluate Models

Workflow

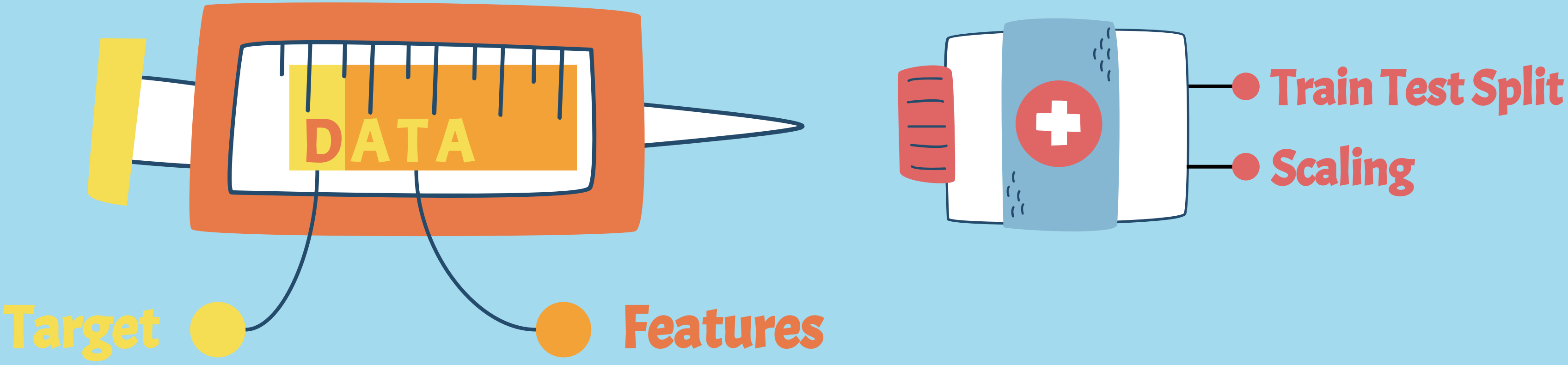
4

Pre-Processing & Modelling

- A — Set Up Environment
- B — Compare Models
- C — Tune Models
- D — Evaluate Models

Set Up Environment

Pre-Processing Pipeline



Workflow

4

Pre-Processing & Modelling

A — Set Up Environment

B — Compare Models

C — Tune Models

D — Evaluate Models

Compare Models

14

Classification Models Compared With Default Hyperparameters

3

Models Shortlisted For Hyperparameter Tuning

- Light Gradient Boosting Machine Classifier
- Gradient Boosting Classifier
- Extreme Gradient Boosting Classifier

Cross Validation on Train Dataset

TOP 3	Model	Accuracy	AUC
lightgbm	Light Gradient Boosting Machine	0.7820	0.8479
gbc	Gradient Boosting Classifier	0.7786	0.8429
xgboost	Extreme Gradient Boosting	0.7745	0.8402
rf	Random Forest Classifier	0.7521	0.8109
et	Extra Trees Classifier	0.7311	0.7745
lr	Logistic Regression	0.7274	0.7818
ridge	Ridge Classifier	0.7246	0.0000
lda	Linear Discriminant Analysis	0.7237	0.7788
ada	Ada Boost Classifier	0.7186	0.7855
svm	SVM - Linear Kernel	0.7156	0.0000
knn	K Neighbors Classifier	0.7037	0.7490
dt	Decision Tree Classifier	0.7005	0.7012
nb	Naive Bayes	0.6026	0.6807
qda	Quadratic Discriminant Analysis	0.5293	0.4944

Test Dataset

TOP 3	Model	Accuracy	AUC
lightgbm	Light Gradient Boosting Machine	0.7899	0.8538
gbc	Gradient Boosting Classifier	0.7847	0.8456
xgboost	Extreme Gradient Boosting	0.7781	0.8391
rf	Random Forest Classifier	0.7585	0.8176
lr	Logistic Regression	0.7328	0.7817
lda	Linear Discriminant Analysis	0.7328	0.7785
et	Extra Trees Classifier	0.7328	0.7782
ridge	Ridge Classifier	0.7323	0.7229
ada	Ada Boost Classifier	0.7280	0.7926
dt	Decision Tree Classifier	0.7110	0.7114
knn	K Neighbors Classifier	0.6922	0.7382
svm	SVM - Linear Kernel	0.6539	0.6763
nb	Naive Bayes	0.6046	0.6794
qda	Quadratic Discriminant Analysis	0.4651	0.4979

Workflow

4

Pre-Processing & Modelling

- A — Set Up Environment
- B — Compare Models
- C — Tune Models
- D — Evaluate Models

Tune Models

Shortlisted Model

- Light Gradient Boosting Machine Classifier
- Gradient Boosting Classifier
- Extreme Gradient Boosting Classifier

1. **Default Model**

- **No tuning** done
- Just the model with its **default hyperparameters**

2. **Tuned Model**

- Tuning done via a **randomised grid search** on a **pre-defined search space**
- Returns the set of hyperparameters that gives the **best accuracy**

3. **Custom Model**

- Tuning done via a **randomised grid search** on a **custom-specified search space**
- Returns the set of hyperparameters that gives the **best accuracy**

Workflow

4

Pre-Processing & Modelling

- **A** Set Up Environment
- **B** Compare Models
- **C** Tune Models
- **D** Evaluate Models

Evaluate Models:

Light Gradient Boosting Machine Classifier

	Train CV Accuracy	Test Accuracy	Train CV AUC	Test AUC
Default LGBMC	0.7820	0.7899	0.8479	0.8538
Tuned LGBMC	0.7724	0.7746	0.8360	0.8411
Custom LGBMC	0.7928	0.7908	0.8533	0.8563

Evaluate Models:

Gradient Boosting Classifier

	Train CV Accuracy	Test Accuracy	Train CV AUC	Test AUC
Default GBC	0.7786	0.7847	0.8429	0.8456
Tuned GBC	0.7713	0.7781	0.8325	0.8420
Custom GBC	0.7799	0.7794	0.8422	0.8453

Evaluate Models:
Extreme Gradient Boosting Classifier

	Train CV Accuracy	Test Accuracy	Train CV AUC	Test AUC
Default XGBC	0.7745	0.7781	0.8402	0.8391
Tuned XGBC	0.7569	0.7677	0.8309	0.8363
Custom XGBC	0.7930	0.7895	0.8538	0.8555

Evaluate Models:

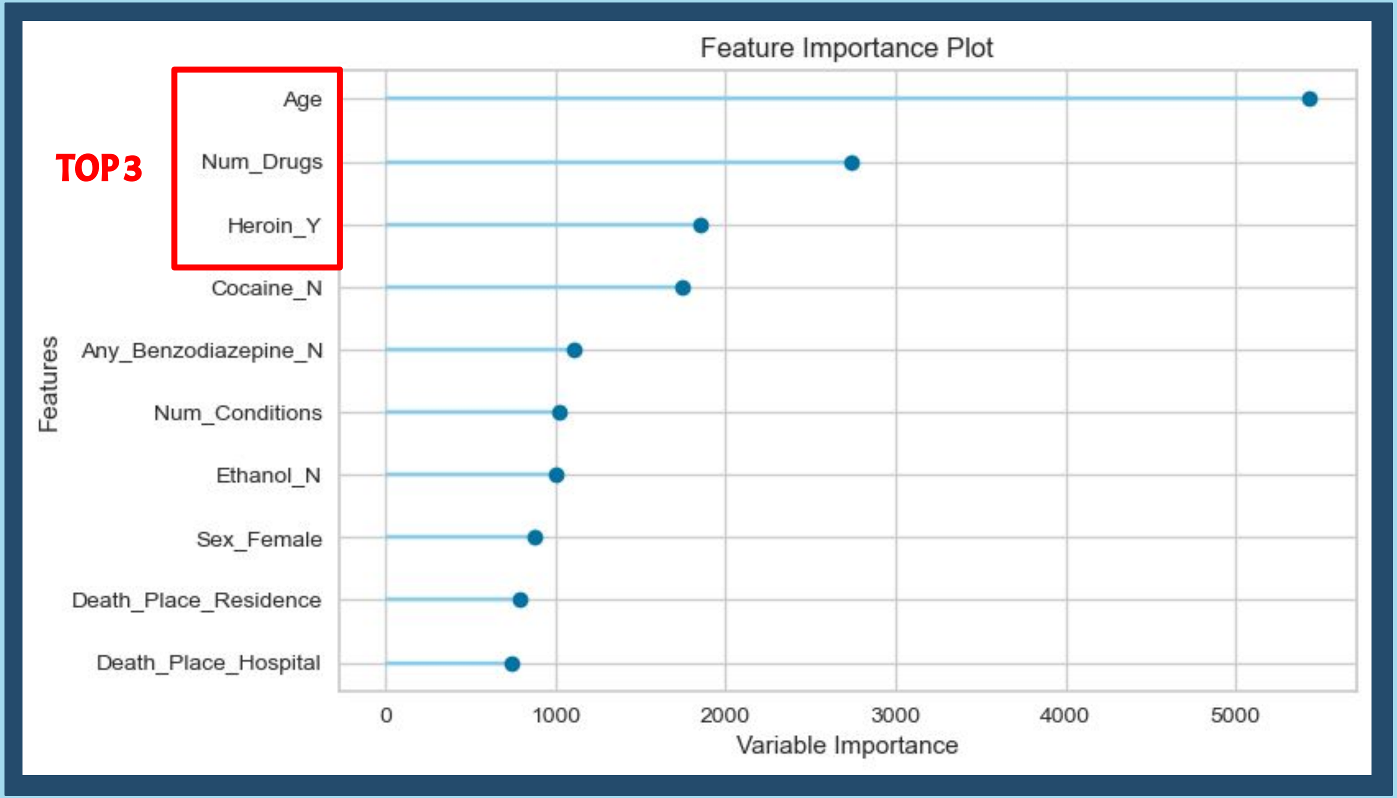
Best Of Each

	Train CV Accuracy	Test Accuracy	Train CV AUC	Test AUC
Custom LGBMC	0.7928	0.7908	0.8533	0.8563
Default GBC	0.7786	0.7847	0.8429	0.8456
Custom XGBC	0.7930	0.7895	0.8538	0.8555

Chosen Model Vs Null Model

Custom LGBMC		Null Model
0.7908	⟺	0.5699
Test Accuracy		Baseline Accuracy
0.8563	⟺	0.5000
Test AUC		Baseline AUC

Feature Importances



Recommendations

- Initiate **education campaigns** in schools and colleges to teach young kids and adults about the **dangers** of using drugs on the illicit drug market, especially on the **overdose risks** and **widespread availability of illegally manufactured fentanyl**

- Pass **stricter laws** that deal out **harsher penalties** to individuals involved in the **purchase and supply of several illicit drugs**, given the high likelihood of finding **illegally manufactured fentanyl laced into the supplies of other illicit drugs**

- Work with doctors to promote the **appropriate use of opioids for pain** to reduce irrational or inappropriate opioid prescribing and dispensing, with added measures for **higher-risk opioids such as fentanyl** that would involve **statewide monitoring of all prescribing and dispensing activities**

Outline

1. Introduction
2. Data Source
3. Exploration & Visualisation
4. Pre-Processing & Modelling
5. Conclusion

Conclusion

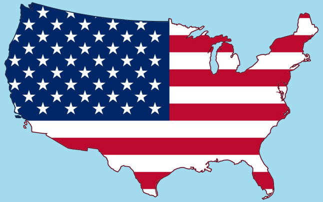
- Fentanyl crisis is a **complex perennial problem** that presents a major challenge for **public health policy**

- Requires a **multi-pronged approach** and a **concerted effort** to overcome successfully

- **Final chosen model**, a Custom Light Gradient Boosting Machine Classifier, was able to **accurately predict 79.08%** of accidental drug overdose deaths to have either no fentanyl present or fentanyl present

- Found the three **most important** features to be **age**, **number of drugs**, and **heroin presence** → To lead and direct the implementation of **harm-reduction strategies** aimed at **reducing fentanyl-related overdose deaths**

Conclusion



- Next steps would involve extending it to **other states of the USA** or investigating the **presence of other drugs** commonly implicated in overdose deaths

- In the context of **Southeast Asia,** while **Singapore** is not embroiled in a drug crisis today, **Myanmar** is now facing a public health disaster as more and more of its young children are getting addicted to **methamphetamine**

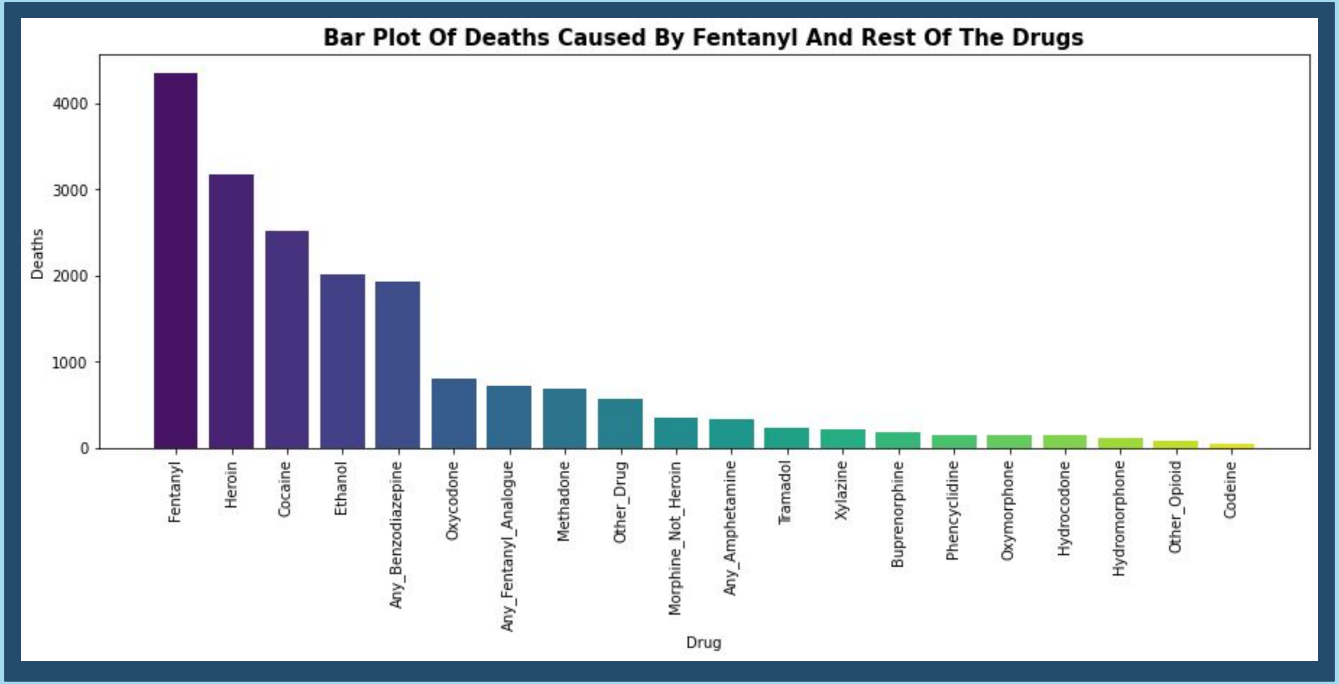
- Future plans could involve applying it to Myanmar as a **starting point** for understanding what are some of the **main factors or drivers** causing the surge in addiction to methamphetamine in recent years

Thank You

And Congratulations To Everyone 🎉

Appendix

Plot Of Fentanyl Vs Rest Of The Drugs



Evaluate Models

No	Model	Train CV Accuracy	Test Accuracy	Accuracy Difference	Train CV AUC	Test AUC	AUC Difference
1	Default LGBMC	0.7820	0.7899	+0.0079	0.8479	0.8538	+0.0059
	Tuned LGBMC	0.7724	0.7746	+0.0022	0.8360	0.8411	+0.0051
	Custom LGBMC	0.7928	0.7908	-0.0020	0.8533	0.8563	+0.0030
2	Default GBC	0.7786	0.7847	+0.0061	0.8429	0.8456	+0.0027
	Tuned GBC	0.7713	0.7781	+0.0068	0.8325	0.8420	+0.0095
	Custom GBC	0.7799	0.7794	-0.0005	0.8422	0.8453	+0.0031
3	Default XGBC	0.7745	0.7781	+0.0036	0.8402	0.8391	-0.0011
	Tuned XGBC	0.7569	0.7677	+0.0108	0.8309	0.8363	+0.0054
	Custom XGBC	0.7930	0.7895	-0.0035	0.8538	0.8555	+0.0017

Acronyms

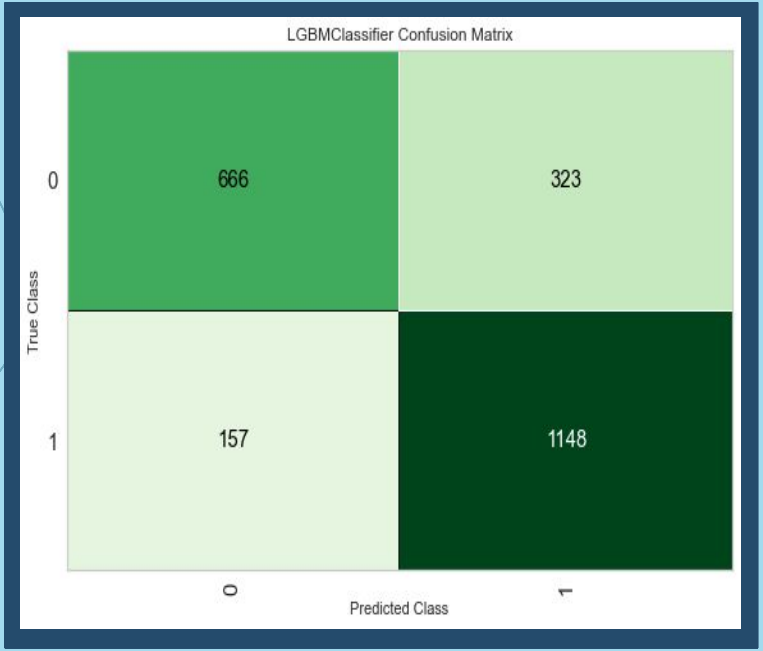
CV: Cross Validation
AUC: Area Under Curve
LGBMC: Light Gradient Boosting Machine Classifier
GBC: Gradient Boosting Classifier
XGBC: Extreme Gradient Boosting Classifier

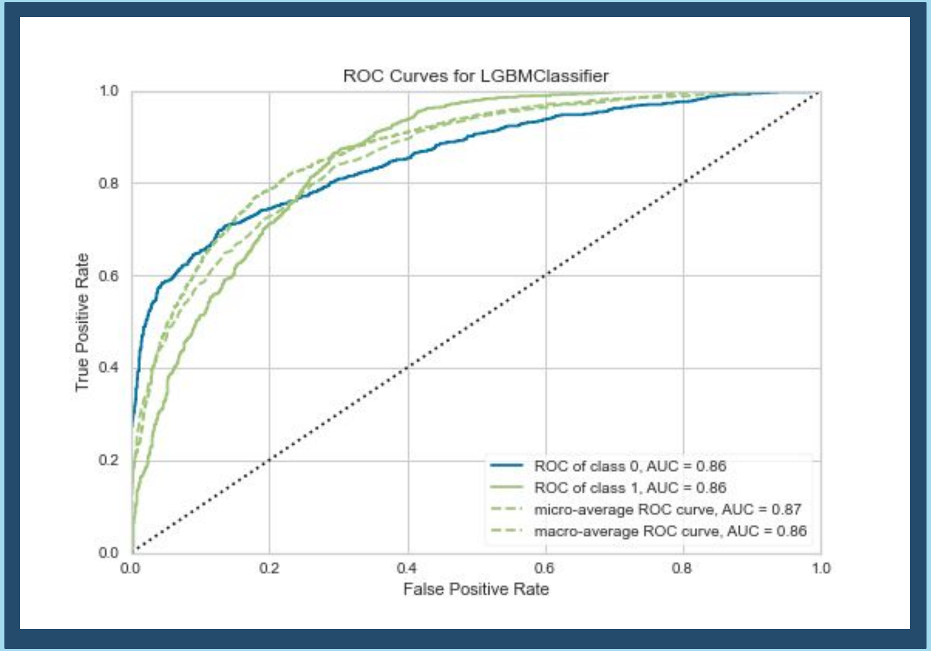
Metrics

Accuracy is a metric that ranges from 0 to 1. The closer it is to 1, the better. It measures how many correct predictions the model made out of all the data points.

AUC is a metric that ranges from 0.5 to 1. The closer it is to 1, the better. It quantifies how well separated the underlying prediction distributions made by the model are.

Confusion Matrix & ROC Curve





Model Inputs & All Feature Importances

	Data Type
Age	Numeric
Sex	Categorical
Race	Categorical
Death_County	Categorical
Death_Place	Categorical
Heroin	Categorical
Ethanol	Categorical
Cocaine	Categorical
Any_Benzodiazepine	Categorical
Covid_Pandemic	Categorical
Num_Drugs	Numeric
Num_Conditions	Numeric
Fentanyl	Label