



# Sentanyl


Localized forecasting for fentanyl overdose prevention

**Team:** Lucy Herr, David Daniels, Jose DelValle, Oscar Garcia, Hans Hernandez

Capstone Project

UC Berkeley Masters in Information & Data Science

Spring 2024



# Meet the Sentanyl Team



**Lucy Herr**  
*Data Scientist*



**David Daniels**  
*Portfolio Analytics  
Analyst*



**Jose DelValle**  
*Senior Data Analyst*



**Oscar Garcia**  
*Senior Data Scientist*



**Hans Hernandez**  
*Software Engineer*



# Agenda

**1 The Problem**

**2 Demo (MVP)**

**3 Technical Solution**

**4 Results**

**5 Conclusion**

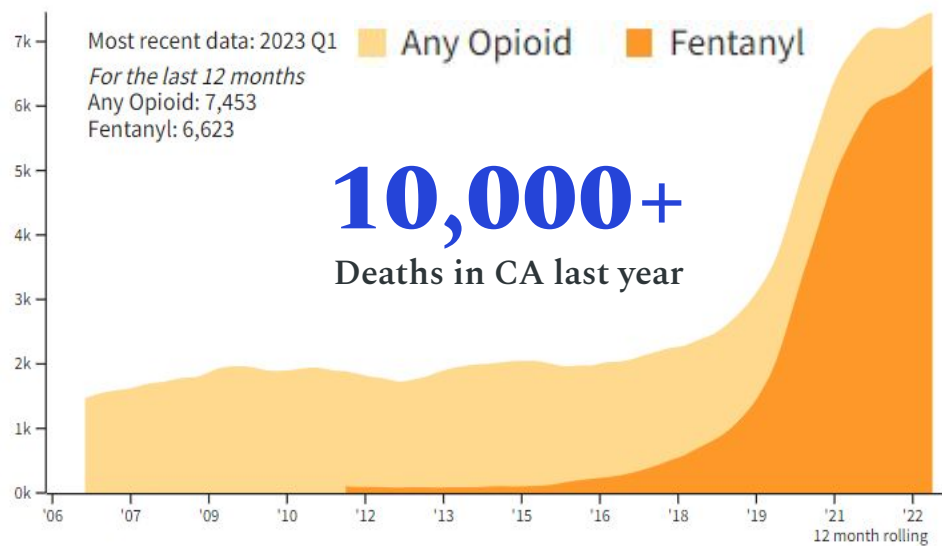


# The Escalating Opioid Crisis

## 1 The Problem

- 2 Demo
- 3 Technical Solution
- 4 Results
- 5 Conclusion

Opioid-related overdose deaths



Over 80% of opioid deaths since 2019 involved Fentanyl

## Why Fentanyl?

What makes Fentanyl so dangerous?

1. 50x more potent than Heroin
2. Frequently mixed with other drugs
3. Cheap & easy to manufacture



# Who Does This Impact?



## Target Users of Our Product

Healthcare systems

Local & state  
government agencies

Non-profit harm  
reduction organizations

Law enforcement

Friends & family of drug  
users



### 1 The Problem

### 2 Demo

### 3 Technical Solution

### 4 Results

### 5 Conclusion

Overdoses could impact anyone:

- Death rates increased across all demographic groups since 2019
- Drop in national life expectancy

Who is most at risk?

- Men Ages 25-45
- Black Men
- Rural and Urban settings

# A Preventable Tragedy

## 1 The Problem

Over 65% of drug overdose deaths in the last 3 years were considered **preventable**

## 2 Demo

## 3 Technical Solution

## 4 Results

## 5 Conclusion

### Naloxone (NARCAN®) ...

can reverse an opioid overdose

must be administered as quickly as possible

is becoming more available to the public

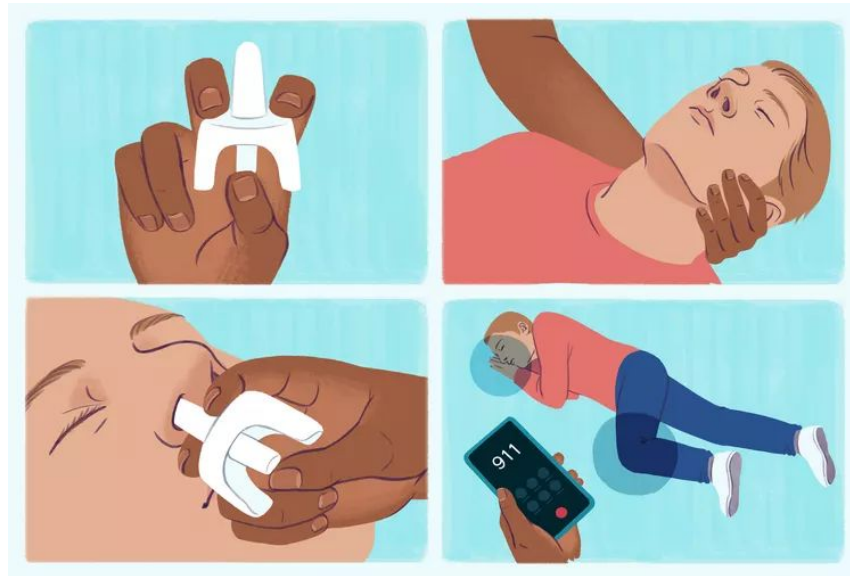


Illustration by Mira Norian for Verywell Health



# Problem Definition

1

The Problem

2

Demo

3

Technical  
Solution

4

Results

5

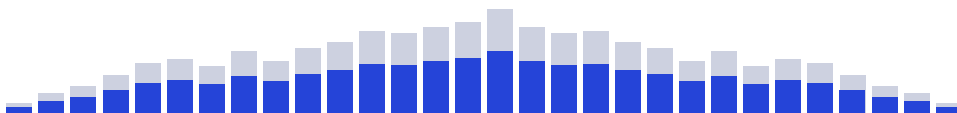
Conclusion

**Problem:** In recent years, California has seen a dramatic rise in fentanyl-related overdoses, which tend to come in waves

**Goal:** Predict monthly fentanyl overdose death rates for each county in California within  $\pm 1$  death per 100K

- Forecast range: 1 year (12 months)

**Purpose:** Enable targeted, life-saving interventions





# Existing Solutions

1

The Problem

2

Demo

3

Technical  
Solution

4

Results

5

Conclusion

Current dashboards and trackers:

- Limited data granularity or visibility
- Restricted user access
- No forecasting



## The Sentanyl Difference

Benefits of our solution:

- 1. Future predictions months in advance**  
Users can plan interventions ahead of potential surges
- 2. Innovative Dataset**  
Compiled data from a wide range of sources, granularity to the month and county level
- 3. User-friendly dashboard available to the public**





# Target User Research



1  
The Problem

2  
Demo

3  
Technical  
Solution

4  
Results

5  
Conclusion

User Sector	Example Interviewee	Product Development Insights
Addiction Medicine & Research	Dr. Leslie Suen (UCSF School of Medicine)	Overdose surge insights High-risk populations Mobile public health solutions
Social Welfare & Public Health	Dr. Jamie Chang (UCB School of Social Welfare)	Demographic subgroups Harm reduction research
Law Enforcement & EMS	Detective Cindy Buechner Firefighter Pete Buechner	Integration of key crime variables linked to drug supply
Healthcare Systems	Kaiser Permanente Research Contracting	Strategies for government health data access



# Sentanyl: Fentanyl Overdose Forecasting

1  
The Problem

2  
Demo

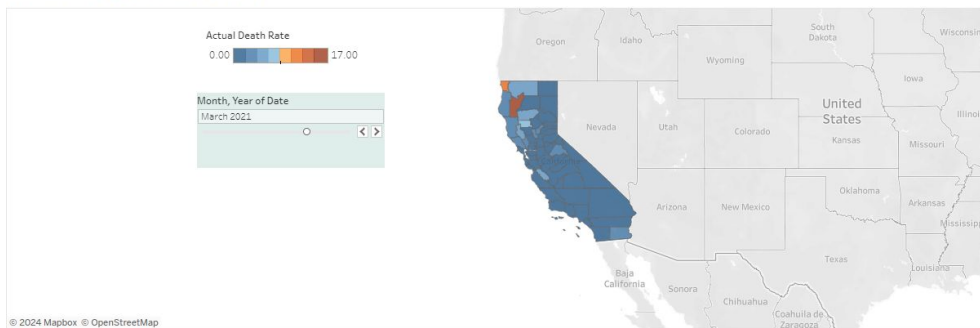
3  
Technical  
Solution

4  
Results

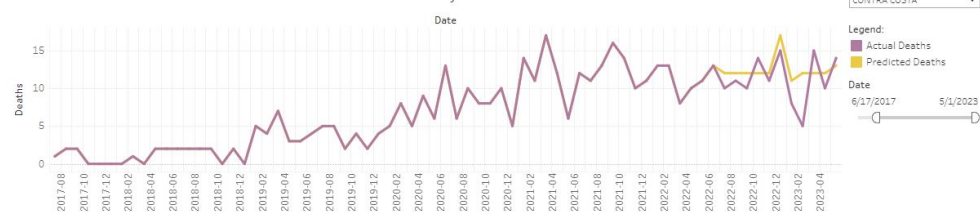
5  
Conclusion

- **Interactive Dashboard**
  - Data by county
  - Projection timeframe
  - Indicator feature of expected spike in overdose deaths
- **Historical Trend Analysis**
  - Explore data from 2016 to 2023
  - Multiple sources
- **Link:**  
<https://groups.ischool.berkeley.edu/Sentanyl/#>

Death Rate by Month and County



Historic Number of Deaths and Past Predictions for county CONTRA COSTA



# Data Sources

1  
The Problem

2  
Demo

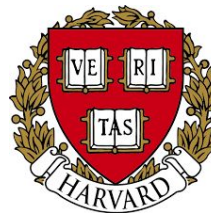
3  
Technical  
Solution

4  
Results

5  
Conclusion

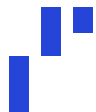
## Outline of Data Sources:

- Monthly Drug Overdose Deaths/Rates
  - CDC
  - CA Dept of Public Health
  - CA Overdose Surveillance Dashboard
- Population Estimates & Demographics
  - US Census Bureau
  - CA Dept of Finance
- Crime Data
  - FBI
  - United Nations
  - US Border Patrol
- SS Data
  - Social Security Administration
- Agency Features
  - US Census Bureau
- Social Media Features
  - Harvard/Twitter



(Used for website but not in AI models)





# Feature Breakdown

- 1 The Problem
- 2 Demo
- 3 **Technical Solution**
- 4 Results
- 5 Conclusion



**Temporal Granularity: Monthly**  
5,569 monthly records  
(2016-23)



**Geographical Features (22)**  
death-related data and rolling  
averages for adjacent counties



**Demographic Features (90)**  
Age ranges & Proportions by  
Gender



**Socioeconomic Features (11)**  
Poverty, Income, Social Security  
recipients by age



**Crime Features (1)**  
Fentanyl seizures (lbs)

**Target Feature (Y)**

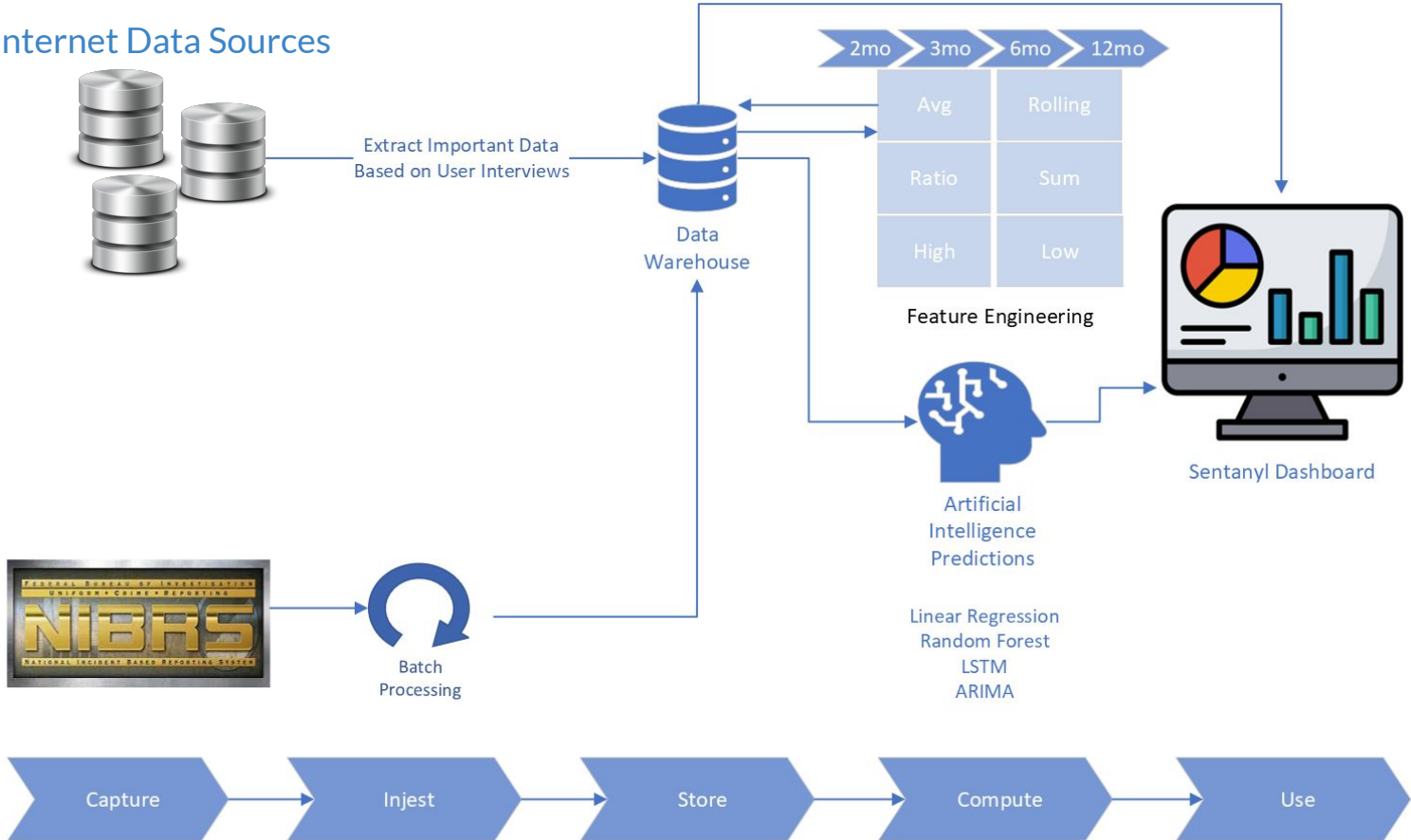
**Monthly  
Fentanyl Death  
Rate (per 100K)**



# Overall Architecture

- 1 The Problem
- 2 Demo
- 3 Technical Solution
- 4 Results
- 5 Conclusion

## Internet Data Sources



# County-Level Differences



1  
The Problem

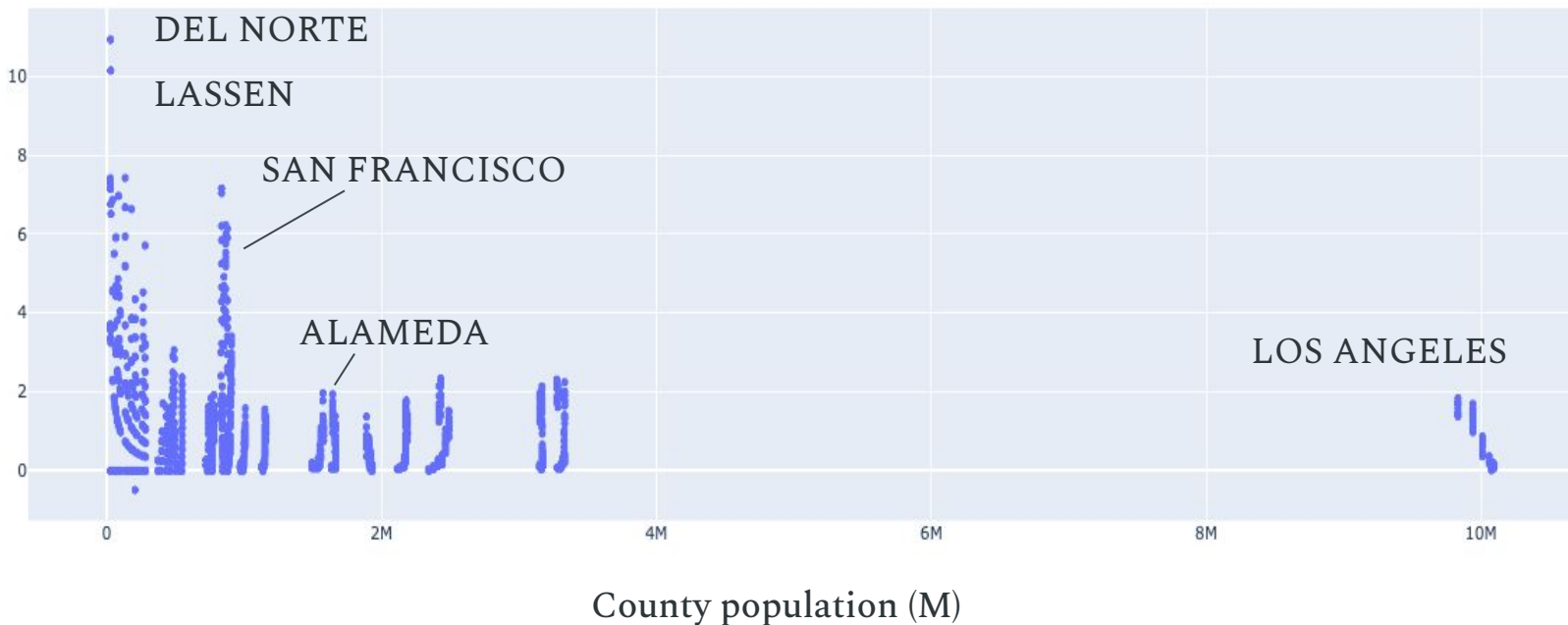
2  
Demo

3  
Technical  
Solution

4  
Results

5  
Conclusion

Fentanyl death rates by county population  
Deaths rate/100K pop.



# Modeling Strategy/Overview



1  
The Problem

2  
Demo

3  
Technical  
Solution

4  
Results

5  
Conclusion

## We test multiple models across counties

- **Experiment with range of models and features for each county**
  - ARIMA, Dynamic Theta, Random Forest, LSTM
  -
- **Select the best model for each county:**
  - High-volatility counties (frequent death surges): use model that best predicts surges
  - Low-volatility counties: use model with lowest weighted RMSE (to penalize underprediction)

$$\text{Weighted RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N w_i (y_i - \hat{y}_i)^2}$$



# Model Description: LSTM

1  
The Problem

2  
Demo

3  
Technical  
Solution

4  
Results

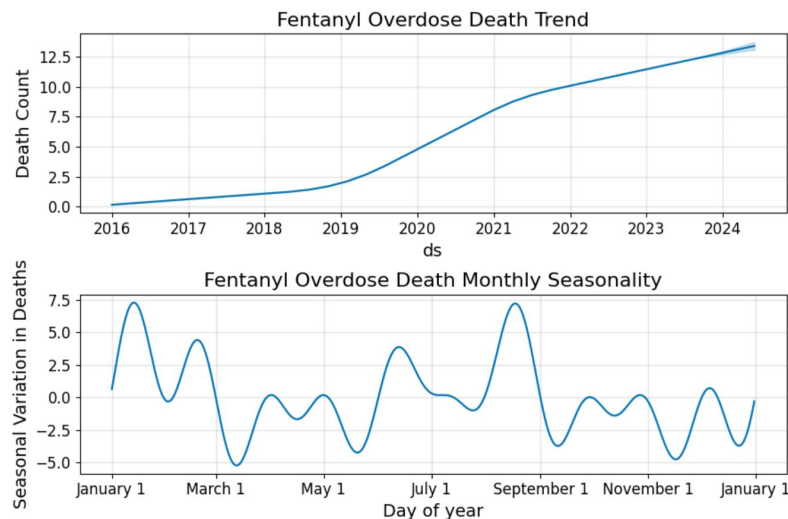
5  
Conclusion

## Why LSTM ?

- Related research on similar public health problems (e.g., Choudhuri et al., 2019)

## Pre-processing decisions

- Address not stationarity
- Cyclical encoding for time period features
- Cross validation with Hyperparameter tuning





# Model Description: LSTM

1  
The Problem

2  
Demo

3  
Technical  
Solution

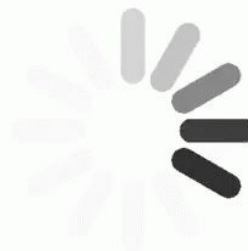
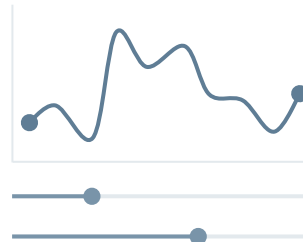
4  
Results

5  
Conclusion

## LSTM Limitations

- Single LSTM across all counties
- Sensitivity to non-stationarity
- Sensitivity to the length of input sequences

## Overfitting



Loading





# We identify “spikes” and select models that perform best at them

1  
The Problem

2  
Demo

3  
Technical  
Solution

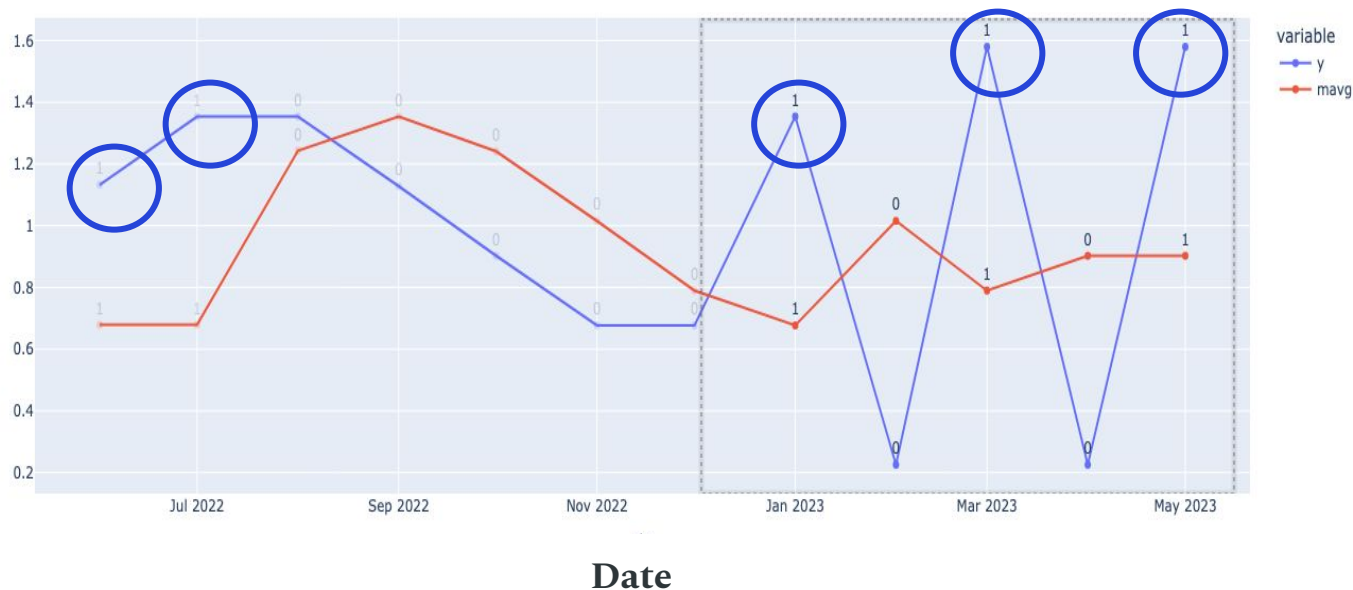
4  
Results

5  
Conclusion

Death  
rate

Santa Barbara county

Spikes and **moving average**: 150% ‘step’



# Spikes are important to predict



1  
The Problem

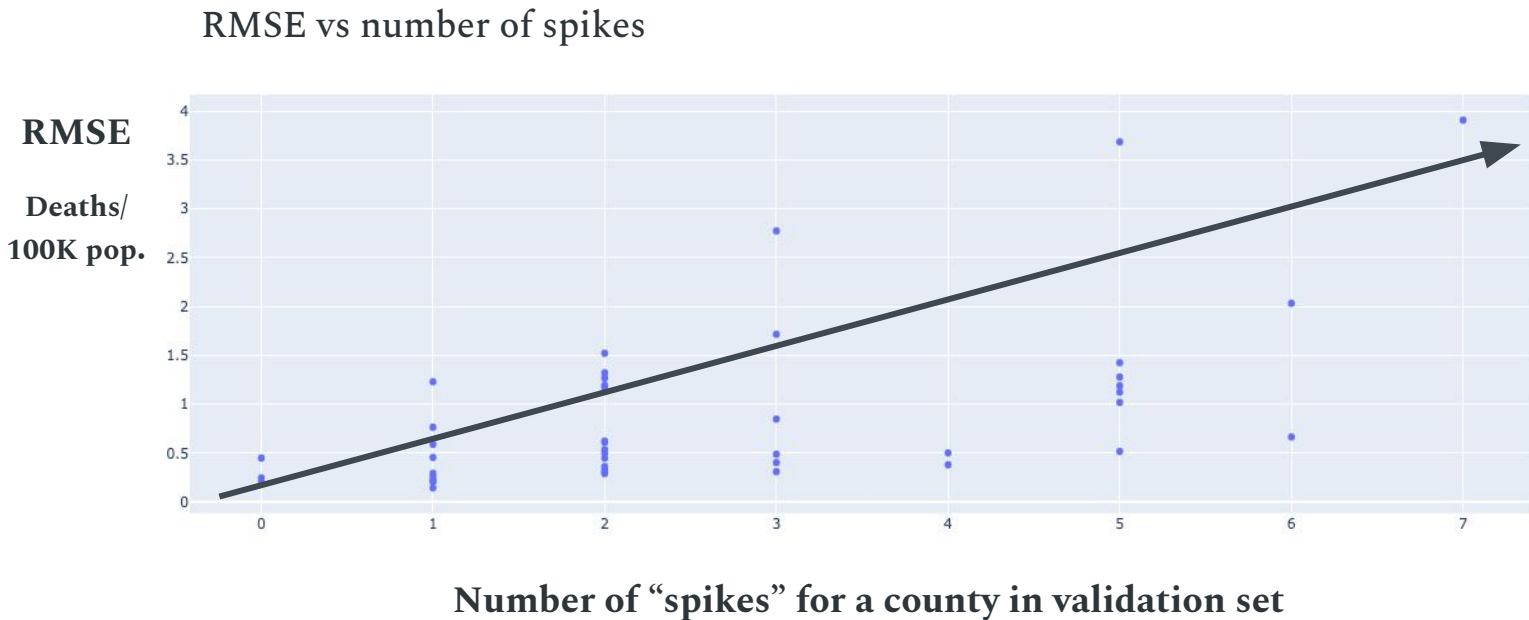
2  
Demo

3  
Technical  
Solution

4  
Results

5  
Conclusion

We observe that in general, county RMSE increases with the number of spikes



# Example: choose best model for spikes

1  
The Problem

2  
Demo

3  
Technical  
Solution

4  
Results

5  
Conclusion

**Best RMSE overall**  
**(Dynamic Theta)**

**0.30**

**Best RMSE spikes**  
**(Random Forest)**

**0.15**

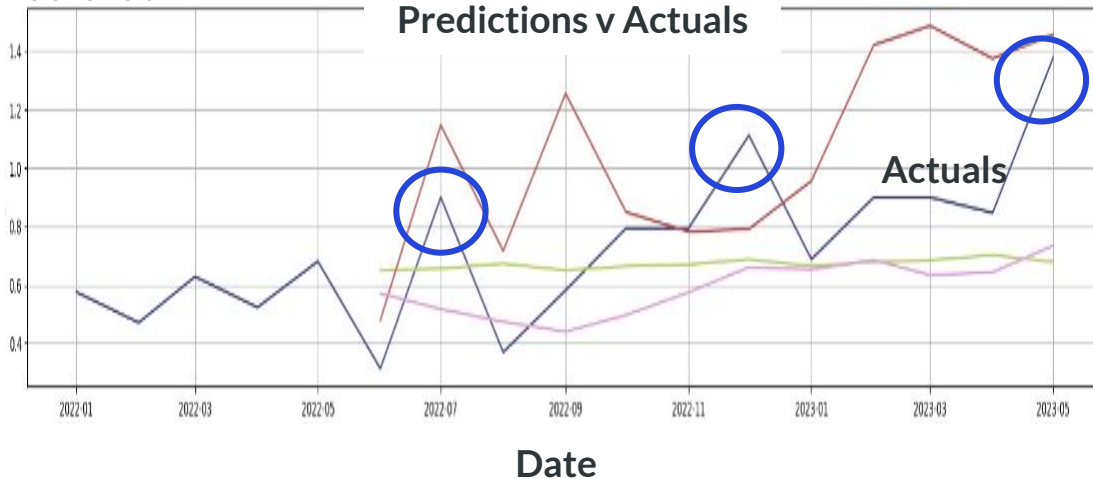


**Chosen model**  
**Random Forest)**

**0.38**

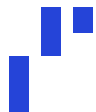
**Deaths rate**

**SANTA CLARA**  
**Predictions v Actuals**



— Random Forest  
— Dynamic Theta





# Model Results

- 1 The Problem
- 2 Demo
- 3 Technical Solution
- 4 Results
- 5 Conclusion

NAIVE (Baseline)

1.16

RMSE  
(deaths/100K)

**Features Used:**  
Previous Death Rate

Dynamic Theta

1.02

RMSE

**Features Used:**  
Historic Death Rates

**Best model in  
15% of  
counties**

Random Forest

0.89

RMSE

**Features Used:**  
All

**Best model in  
33% of  
counties**

LSTM

0.96

RMSE

**Features Used:**  
All

**Best model in  
37% of  
counties**

Best per county

0.84 ★

RMSE

**28%  
improvement  
over baseline**

**Pred. err. +/-1  
59% of months.  
Overpredicts 2+  
deaths 22%**



# Challenges and Future Directions

## Product Development Challenges

- Limited data access
  - Health data privacy restrictions
  - Inherent lag in final (non-provisional) overdose data
- Low granularity in OD data reporting

## Future Goals

- How-to guide for adapting methods
- Set up integration with government & health systems (real-time data)
- Develop dashboard based on user feedback

To help battle the U.S. opioid epidemic, Santa Clara proactively installs a no-charge Narcan vending machine in Benson.

May 9, 2023 | By Tracy Seipel



Community care: A Narcan vending machine is now located in the lower level of Benson Memorial Center. The machine, which contains free Narcan kits, is believed to be the first installed at a college or university campus in Santa Clara County. Photos by Josie Lepe.



# Sentanyl's Mission

Our forecasting tools empower preventative strategies and informed policy decisions. By enabling users to pivot from past data to future insights, we can look forward to saving lives impacted by the opioid epidemic.



1  
The Problem

2  
Demo

3  
Technical  
Solution

4  
Results

5  
Conclusion



# Appendix







# Who does this impact?

Quick stats on key racial / gender / age groups at highest risk

- If possible, highlight disparities in overall users of fentanyl vs those who die from it
- Who would benefit most from potential interventions if resources were allocated where and when they were needed most?



# We choose models that predict spikes better



1  
The Problem

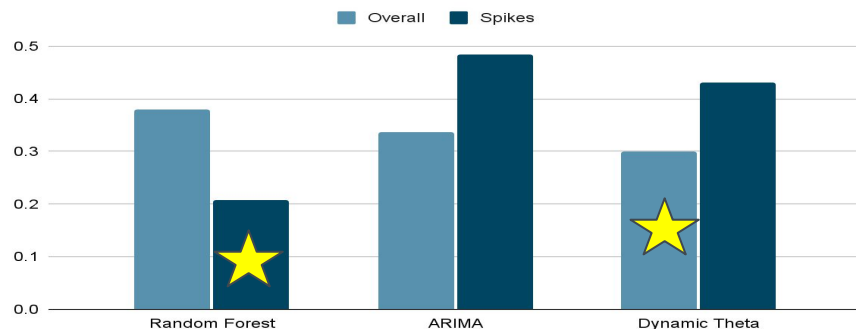
2  
Demo

3  
Technical  
Solution

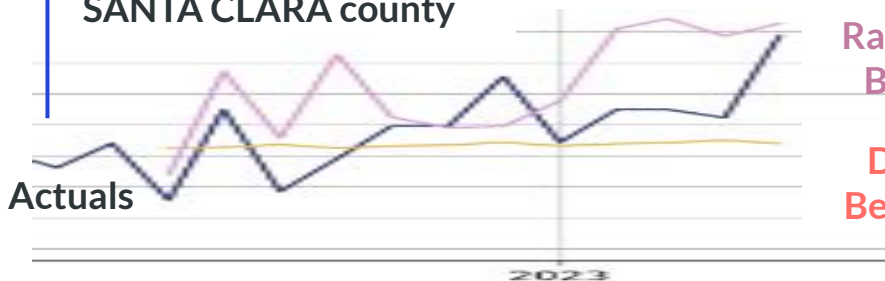
4  
Results

5  
Conclusion

RMSE for various models



SANTA CLARA county



Random Forest.:  
Best at spikes

D. Theta:  
Best overall

## How we use spikes

By default, we choose models that minimize RMSE for a county.

However for counties with high number of spikes, we choose model with best RMSE for spikes

Why?

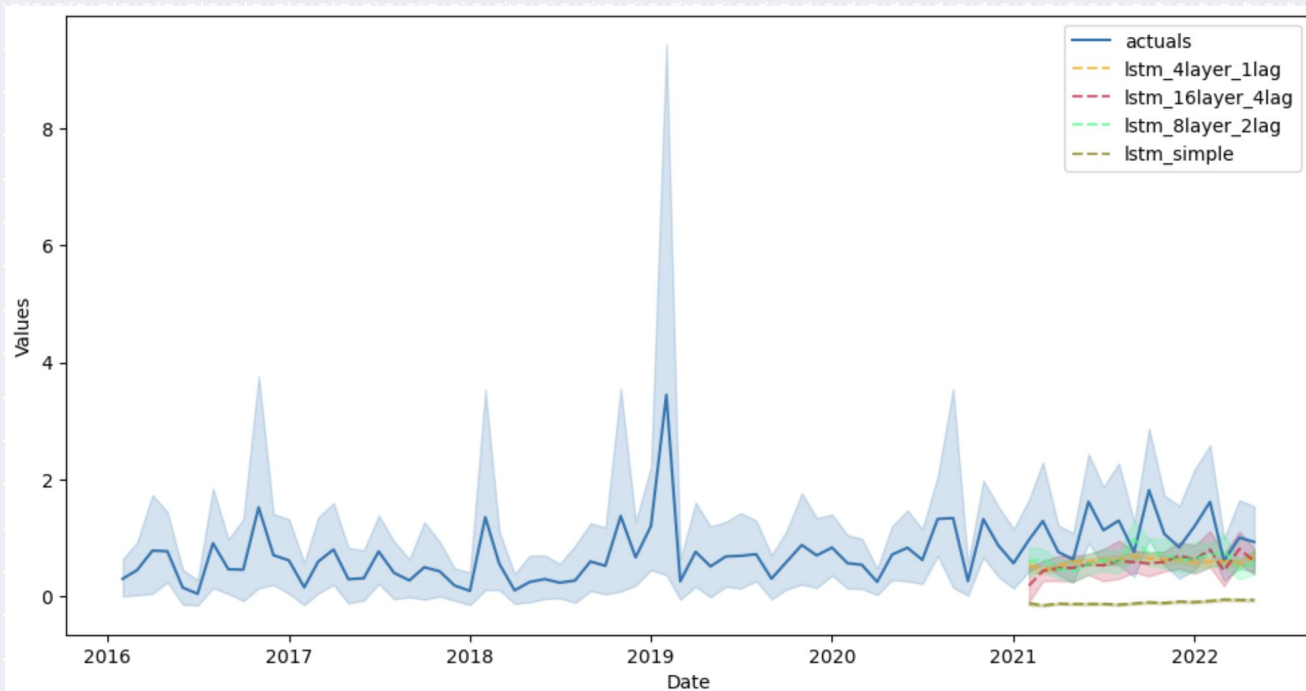
- 1. Saves more lives**  
By predicting sudden increases, we can help plan resources in advance
- 2. Helps predictions**  
With weighted RMSE, we penalize models that tend to underpredict

# MVP Description

- **Interactive Dashboard**
  - Data by county
  - Projection timeframe
  - Indicator feature of expected spike in overdose deaths
- **Historical Trend Analysis**
  - Explore data from 2016 to 2023
  - Multiple sources



# Current LSTM Results

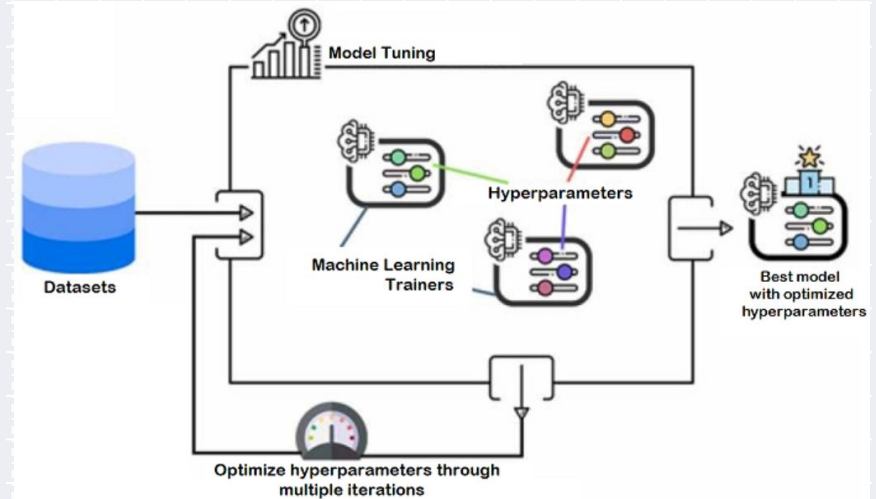


## Current best model:

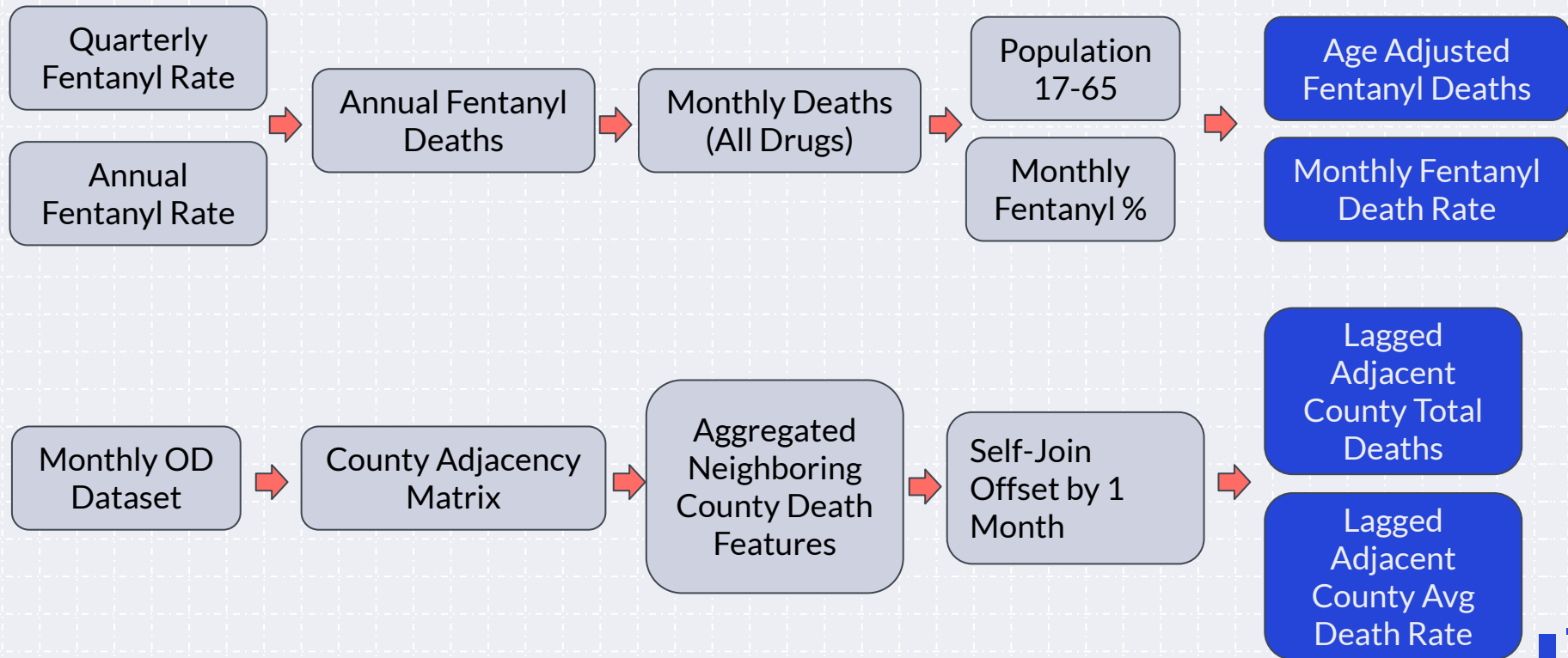
- Layers: 8x32
- Lags: 4
- Optimizer: Adam
- Learning rate: 0.01
- Epochs: 25
- Validation RMSE = 0.019

# Remaining Key Areas

- Ensemble model
- Transformers model for time series data
- Model Tuning
- Additional Feature Engineering



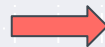
# Feature Engineering



# Feature Engineering

## Original Raw Features

- Monthly Overdose Deaths (All Drugs)
- Quarterly Fentanyl Death Rate
- Yearly County Population Estimate



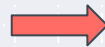
## Final Target Feature

Monthly Fentanyl  
Death Rate



## Additional External Dataset

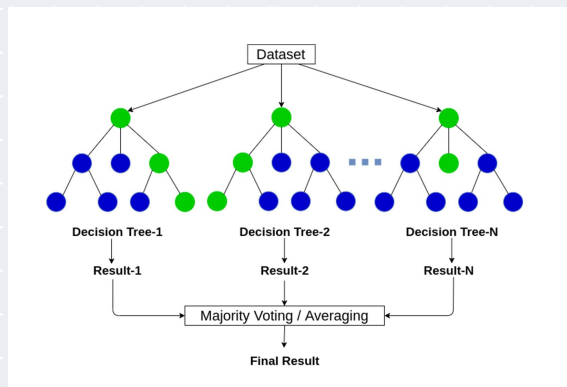
County Adjacency Matrix



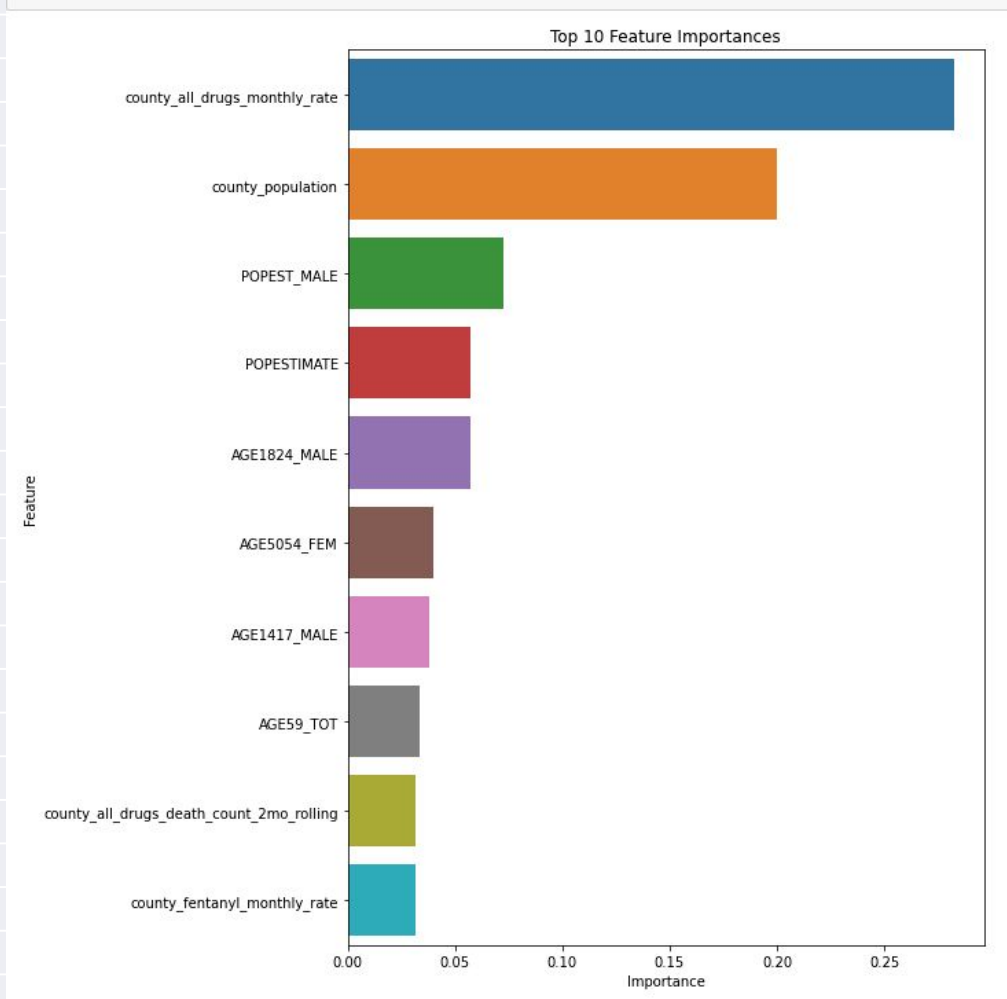
## Key Independent Variable

Lagged  
Adjacent  
County Avg  
Death Rate

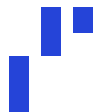
# Random Forest



- All features used for training
- Dropped 2023 and used Nov 2022 - Dec 2022 for validation
- Ran GridSearchCV to find best hyperparameters
- Allows us to find the most important features in our data
- RMSE 0.107 (all counties)







# Feature Importance

- 1 The Problem
- 2 Demo
- 3 Technical Solution**
- 4 Results
- 5 Conclusion

