County Overdose Rates and Information Access/Education
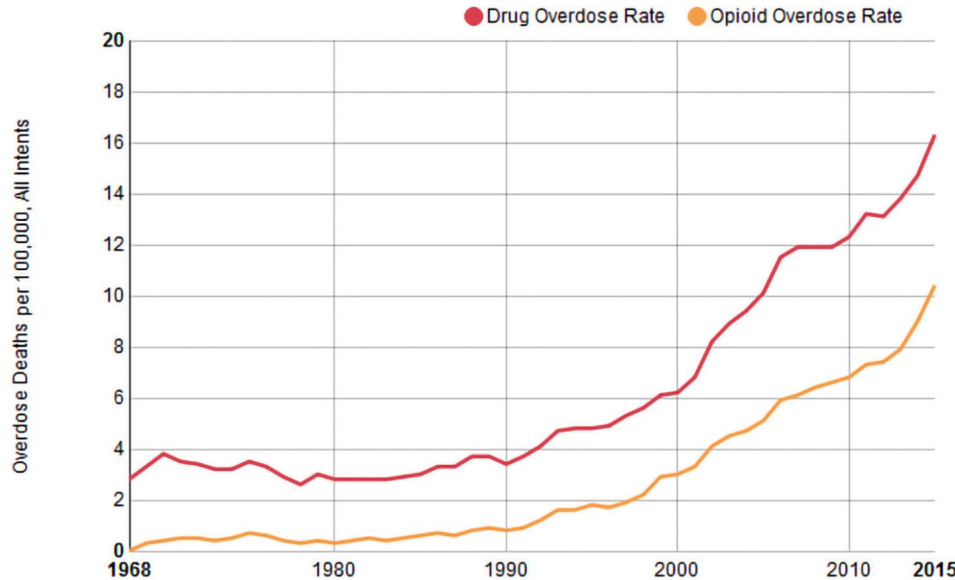
Alex Shropshire, Data Science Consultant (https://www.linkedin.com/in/alexshropshire/)
Mia Iseman, Data Science Consultant (https://www.linkedin.com/in/miaiseman/)

Project Context

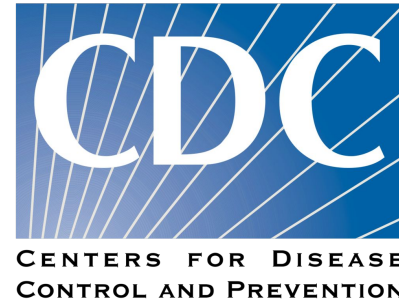Figure 1. Age-Adjusted Overdose Death Rate (All Intents), 1968-2015

- Opioid Epidemic
- Overdose Prevention: A potential initiative at the US Dept. of Health & Human Services
- Prevention vs. Treatment

Data Sources (2013-2017)

US Census API using cenpy
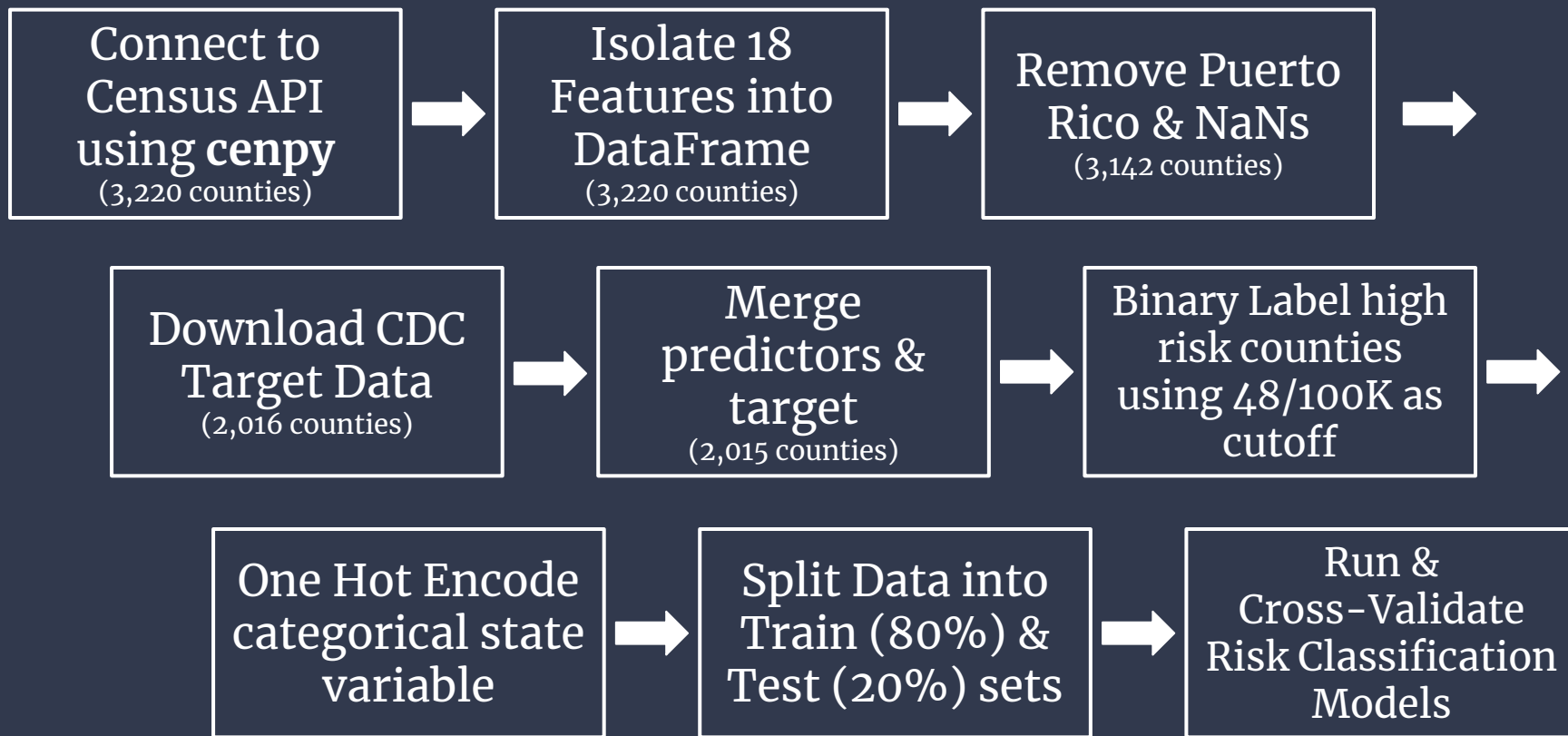**(Education & Connectivity Data)**

CDC Wonder Database
**(Overdose Rate by US County)**

- **18 Predictor Variables:** Various statistics related to educational attainment and internet connectivity by county

- **1 Target Variable:** Drug and/or Alcohol related overdose fatalities per 100,000 people by county

Project Steps

Connect to Census API using cenpy (3,220 counties) → **Isolate 18 Features into DataFrame** (3,220 counties) → **Remove Puerto Rico & NaNs** (3,142 counties) →

Download CDC Target Data (2,016 counties) → **Merge predictors & target** (2,015 counties) → **Binary Label high risk counties using 48/100K as cutoff** →

One Hot Encode categorical state variable → **Split Data into Train (80%) & Test (20%) sets** → **Run & Cross-Validate Risk Classification Models**
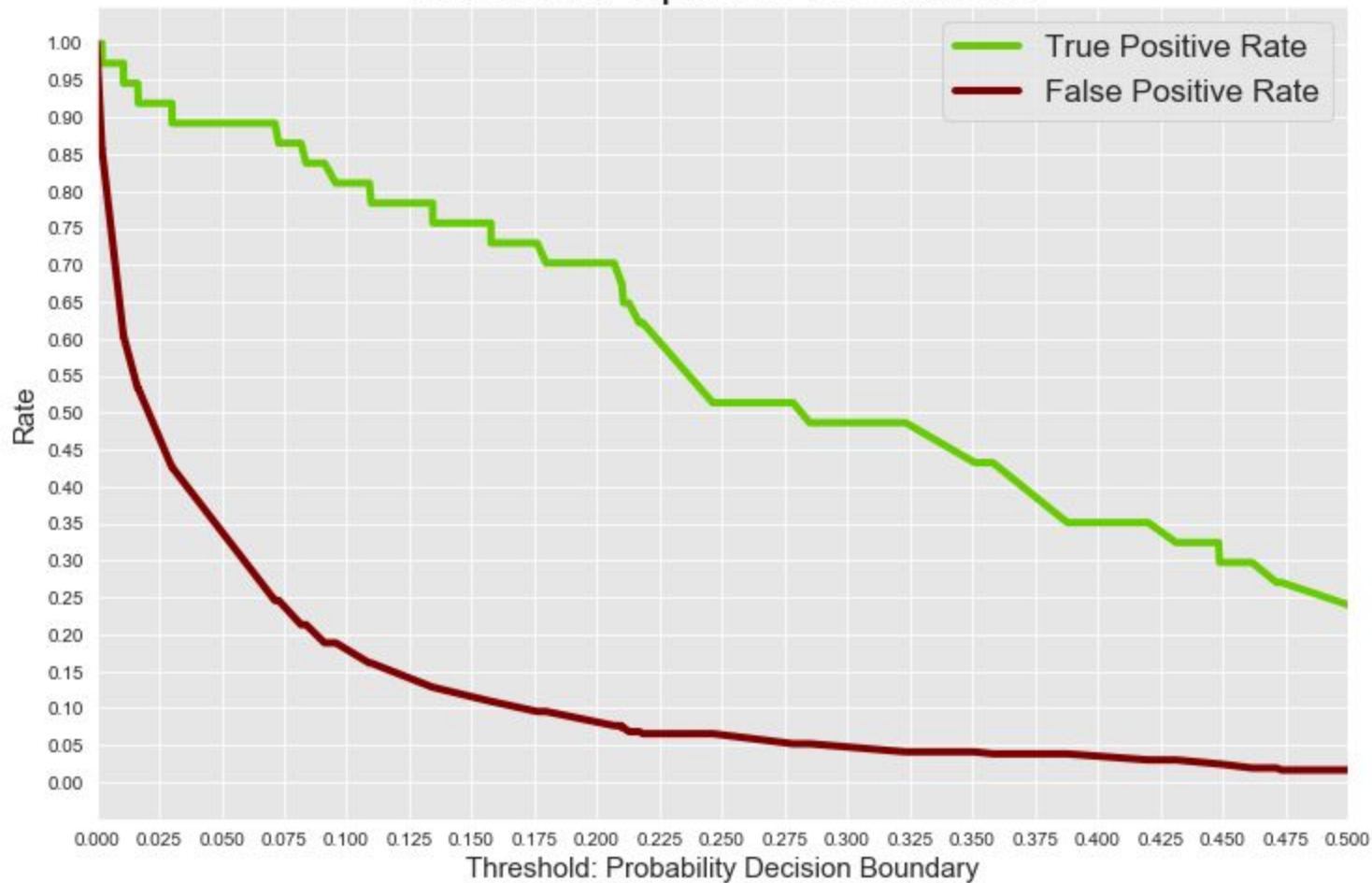
Model Evaluation

- **Area Under Curve (AUC)**: 84%

- **Point A** (Recommended):
  - TPR = 89%
  - FPR = 25%
  - Threshold for Predicted Probability of 'High Risk' = 7%

- **Point B**:
  - TPR = 81%
  - FPR = 16%
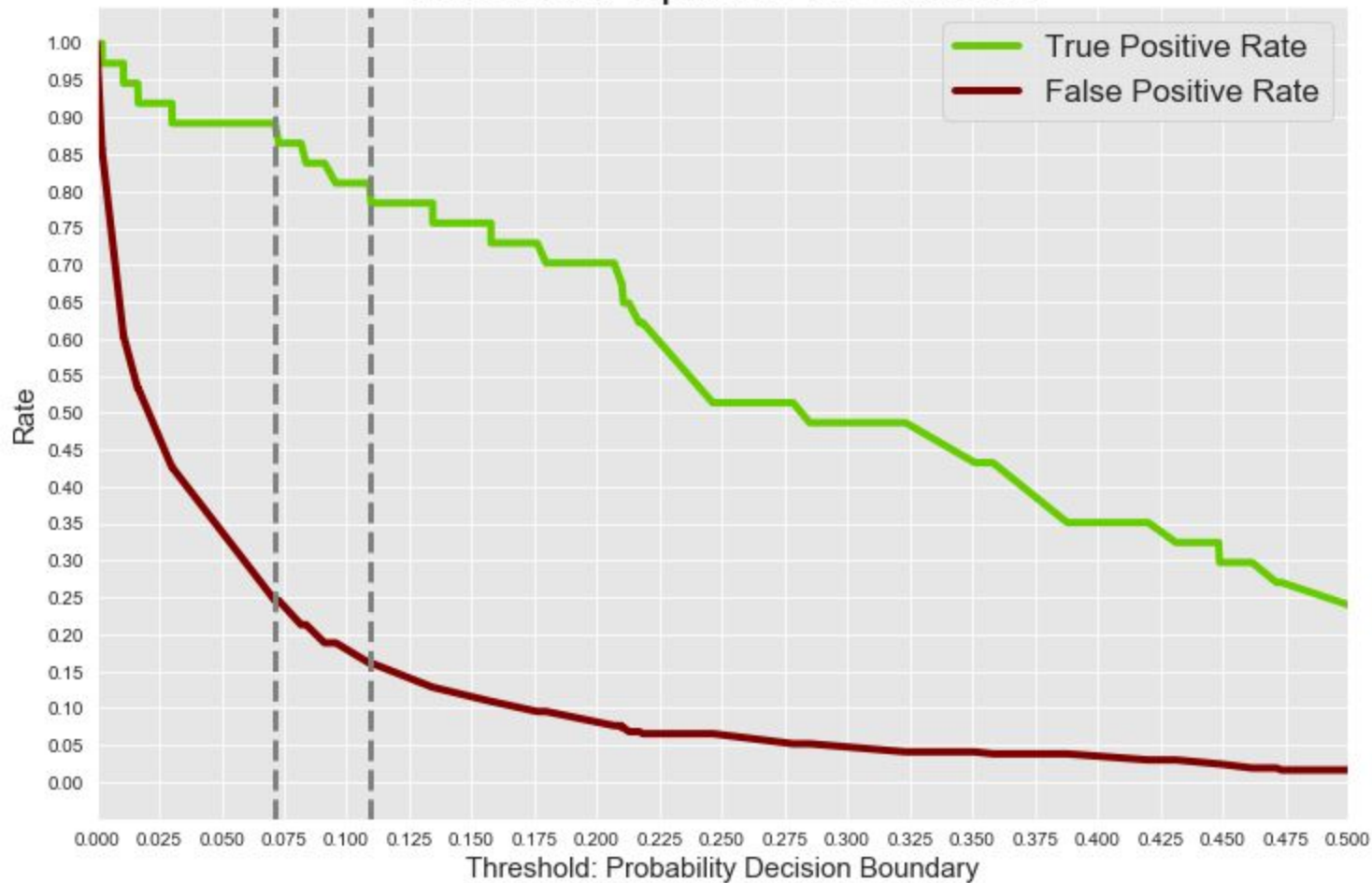 - Threshold for Predicted Probability of 'High Risk' = 11%
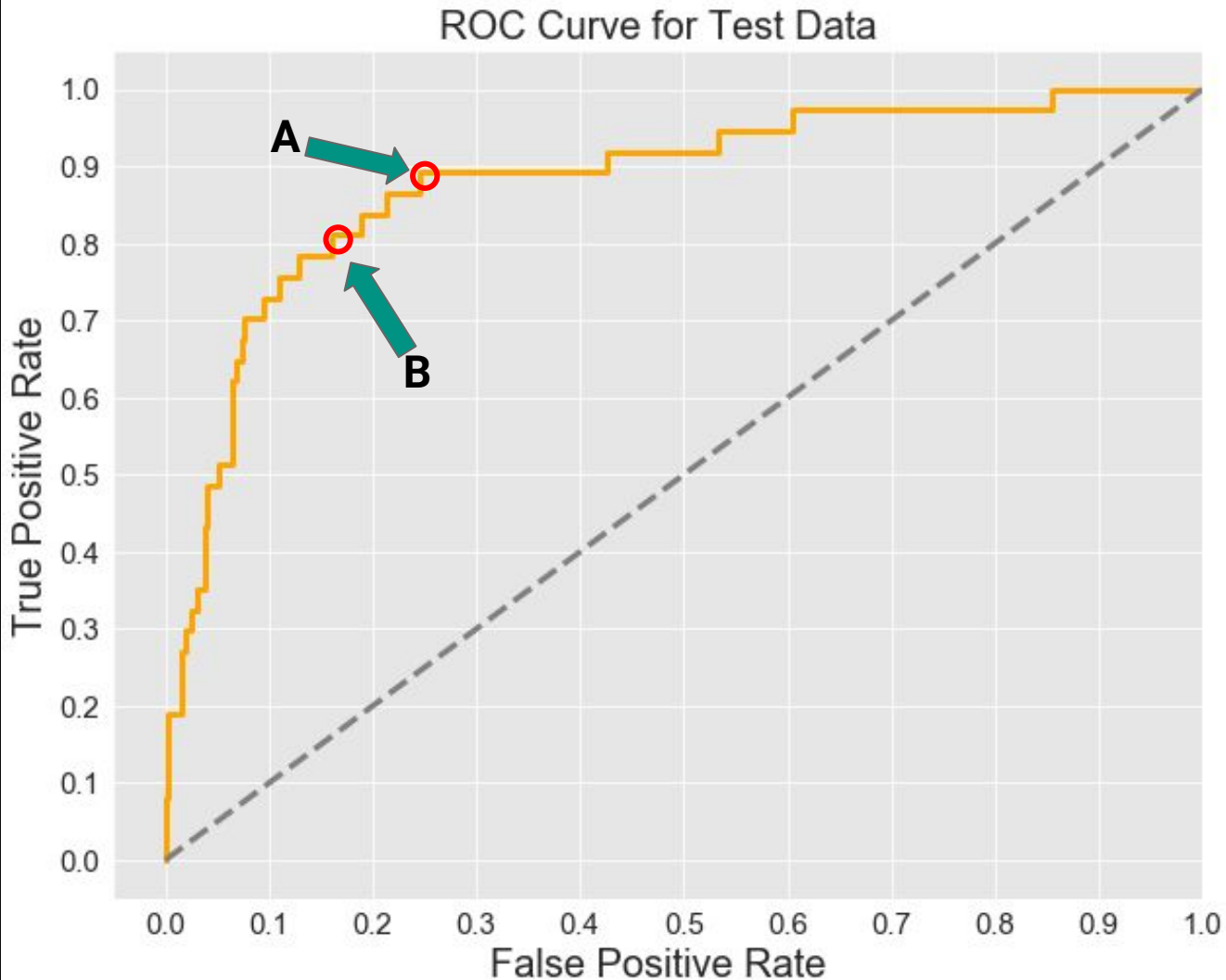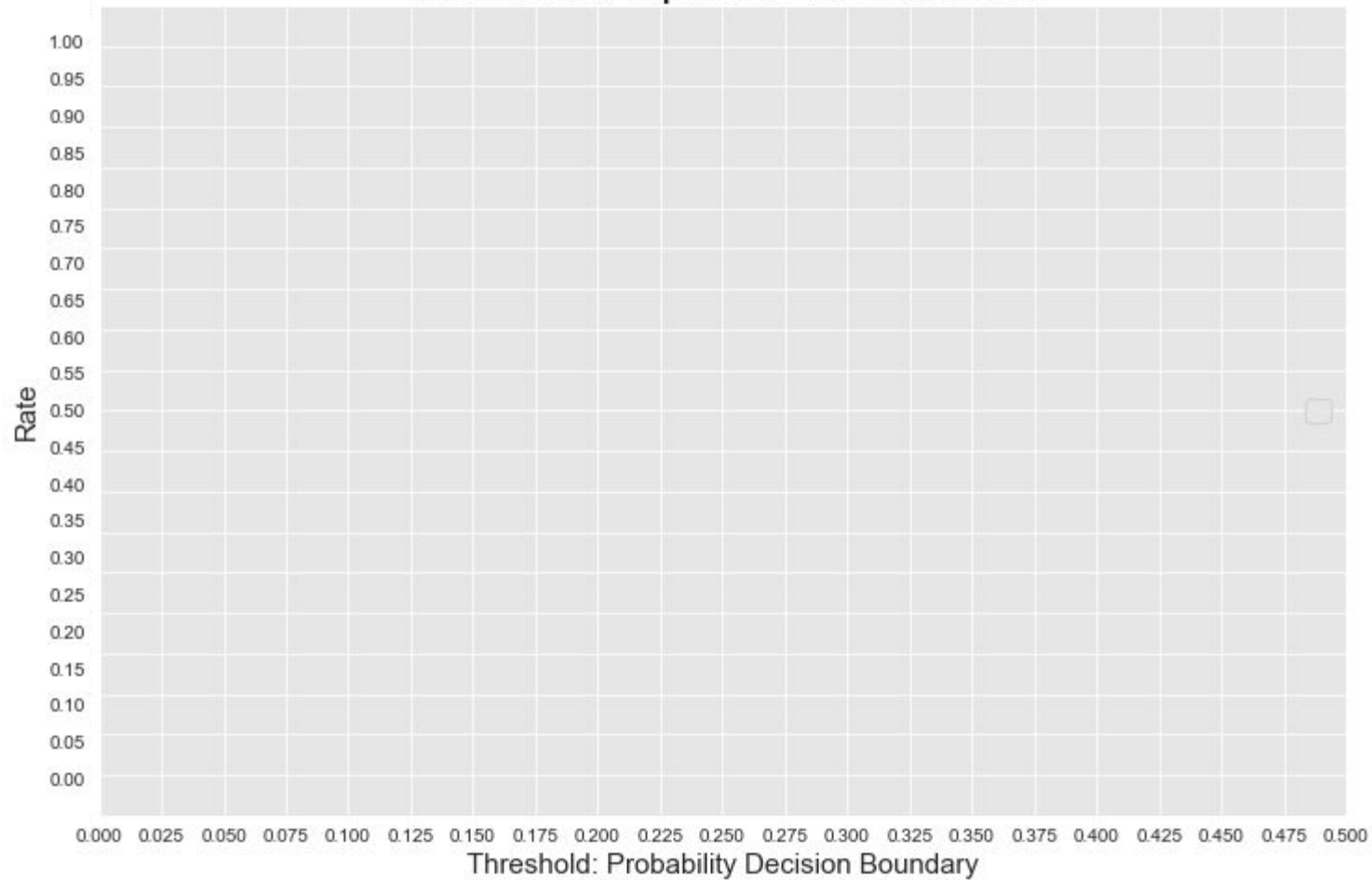
ROC Curve for Test Data

Threshold Options to Consider
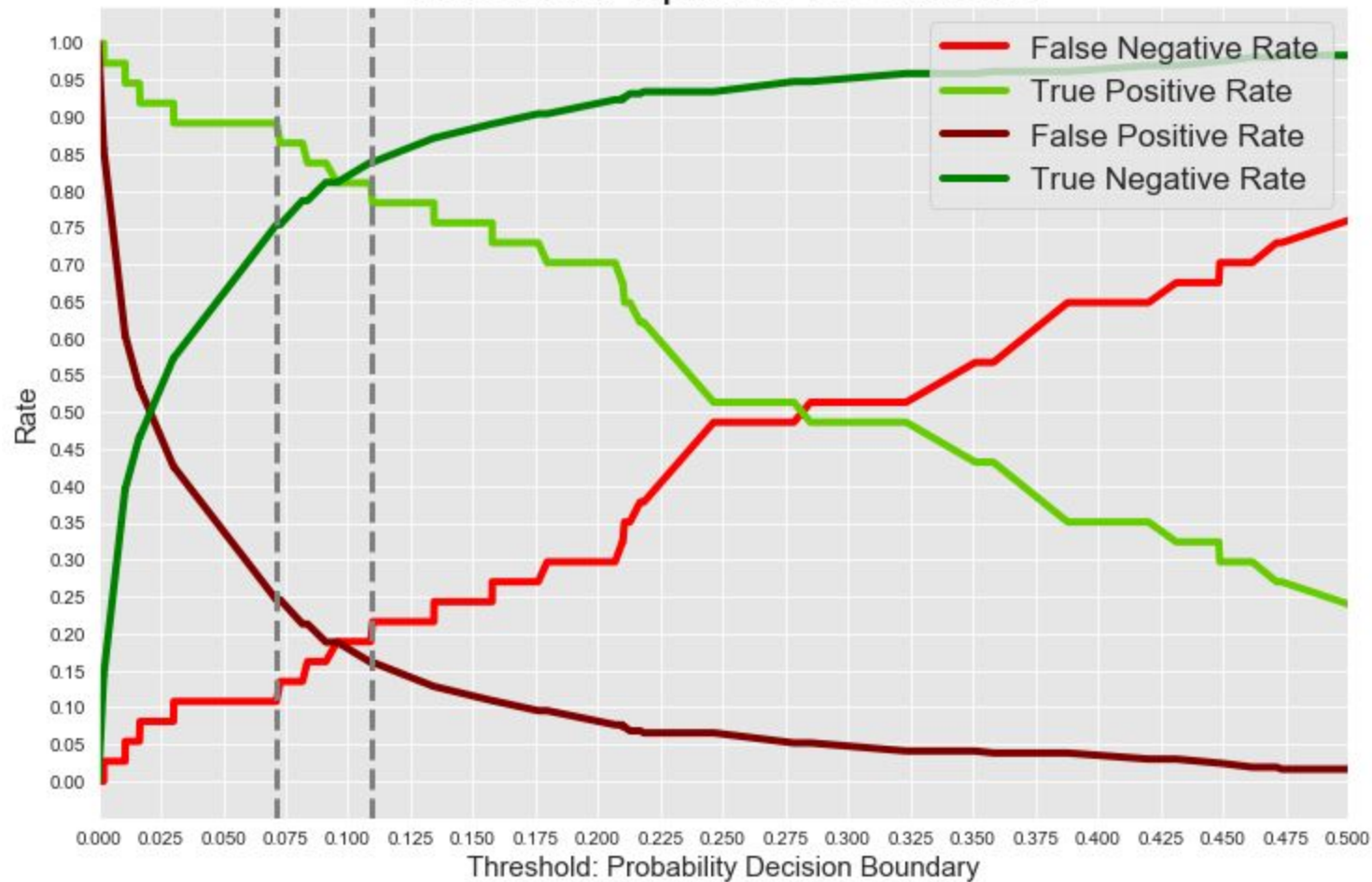
Rate

Threshold: Probability Decision Boundary

Threshold Options to Consider

Threshold Options to Consider

Threshold Options to Consider

Model Evaluation

County	Overdose Crude Rate (per 100K people)	Model Probability of High Risk Classification	Correct Result?
Rio Arriba County, NM	156.2	0.98	True Positive
Roosevelt County, MT	115.3	0.40	True Positive
Menominee County, WI	115.2	0.17	True Positive
Summit County, OH	50.6	0.14	True Positive
Pamlico County, NC	48.2	0.09	False Negative
Tuolumne County, CA	47.9	0.55	False Positive
Fort Bend County, TX	8.37	0.000	True Negative
Walker County, TX	7.93	0.001	True Negative
Lauderdale County, MS	7.90	0.002	True Negative

Most Impactful Predictors

Feature	Log Odds (Logistic Regression coefficient)	Impact on Probability that it is a High-Risk county
% of Total Households with Broadband connectivity	-0.834369	⬇
% of Adults 25+ with some high school (didn't graduate)	0.750014	⬆
% of Adults 25+ with some college (didn't graduate)	0.733234	⬆
% of Adults 25+ who didn't attend high school	-0.544092	⬇
% of Total Households that have a computer	-0.416011	⬇

Next Steps

- Run additional models utilizing upsampling and boosting
- Collect additional predictive factors such as infrastructure data, employment & economic data, community development data
- Understand additional target variables beyond overdose rate
- Actionable recommendations on controllable features

Questions?