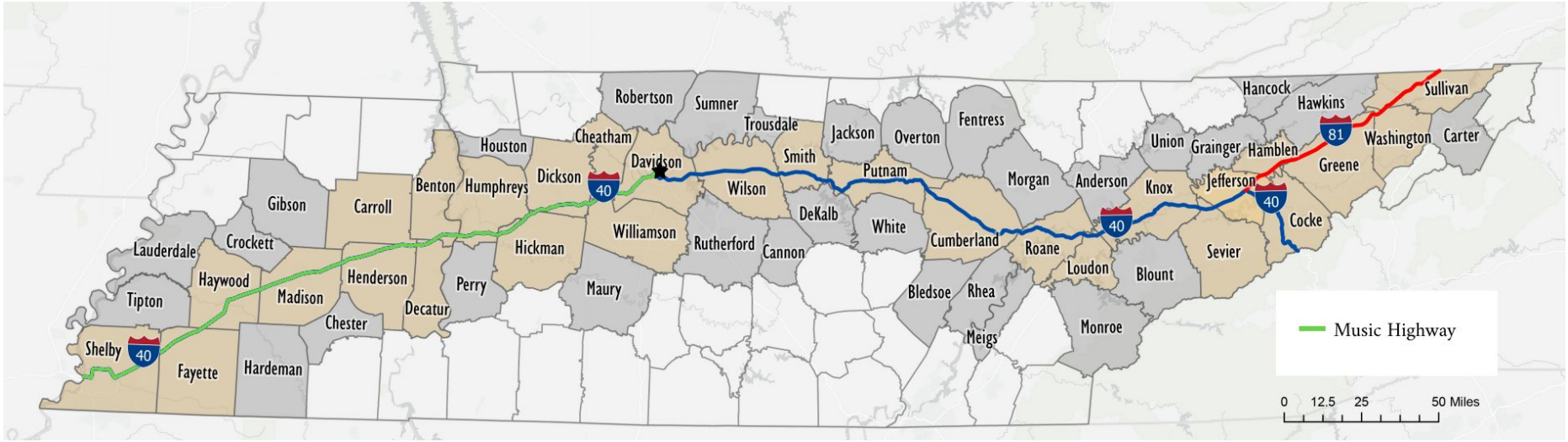
Team Members:

Ruixuan Ding
John Hurtado
Yang Mo
James O'Quinn
Chilambwe Wapamenshi

Music Highway



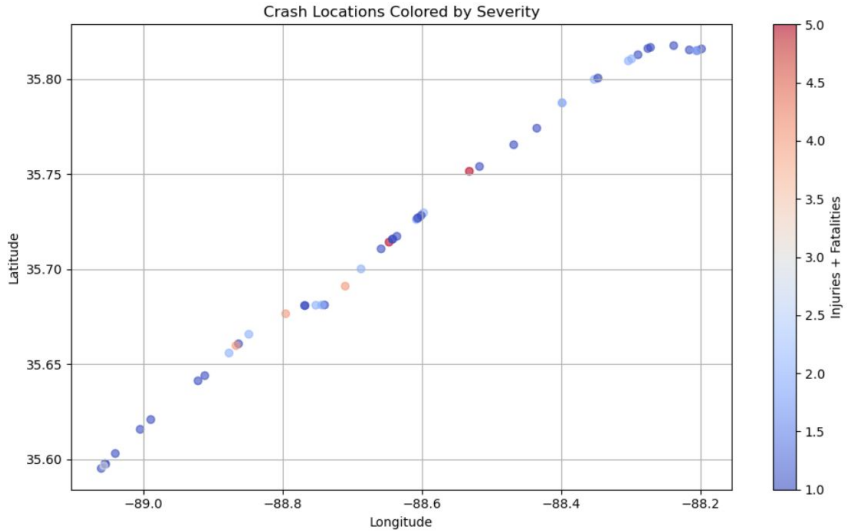
Dataset Generation

- **Crash Data (2023–2025)** from Tennessee DoSHS; focus: I-40 in Madison & Henderson counties
- **Raw Features:** Latitude, longitude, crash severity
- **Enrichment:** Tagged road features (guardrails, lighting, pavement, etc.) via **Google Street View**

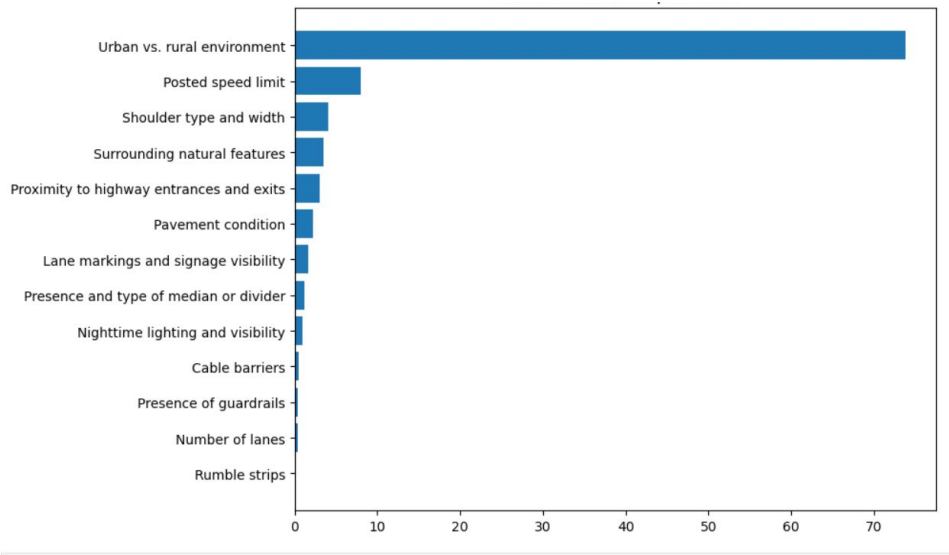
Issue: Class imbalance.

Key Findings from Exploratory Analysis

Crash Severity by Location



Feature Importance



Objective

Evaluate the impact of specific safety interventions on crash severity along I-40 in Madison and Henderson Counties.

Modeling Approach:

- Binary target: injury vs. no injury
- Held-out segment to simulate interventions
- Oversampling to balance data

Models Used

Logistic Regression for easy interpretation of feature effects.

CatBoost Classifier for handling complex categorical data efficiently.

Logistic regression model:

Classification Report:

	precision	recall	f1-score	support
0	0.82	0.56	0.67	173
1	0.24	0.53	0.33	45
accuracy			0.56	218
macro avg	0.53	0.55	0.50	218
weighted avg	0.70	0.56	0.60	218

Confusion Matrix:

```
[[97 76]
 [21 24]]
```

CatBoost Classifier:

Confusion Matrix:

```
[[74 99]
 [12 33]]
```

Classification Report:

	precision	recall	f1-score	support
0	0.86	0.43	0.57	173
1	0.25	0.73	0.37	45
accuracy			0.49	218
macro avg	0.56	0.58	0.47	218
weighted avg	0.73	0.49	0.53	218

Hypothesis Testing

- We simulated each safety intervention by changing the relevant features in the data.
- Model predictions before and after these changes were compared to estimate impact.
- Statistical significance was tested using bootstrap resampling with 10,000 samples.
- The null hypothesis assumed interventions did not change injury rate

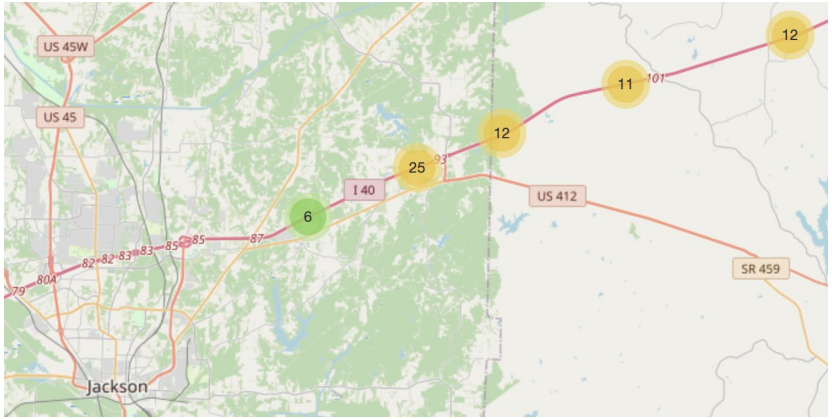
Evaluation Metrics

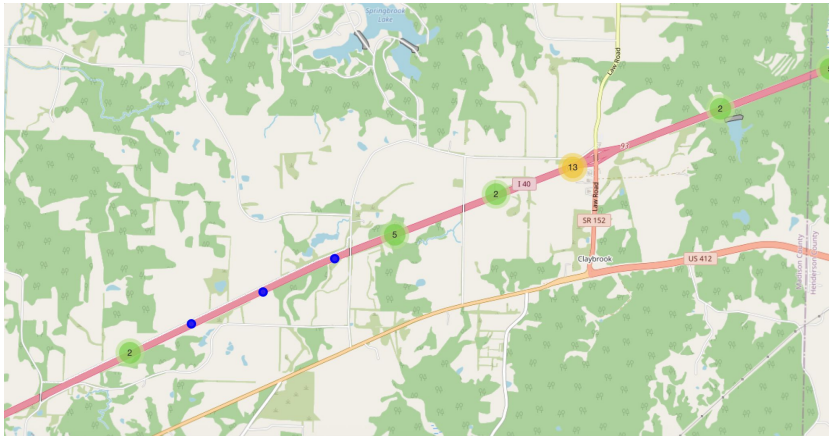
- We focused on recall to catch as many injury cases as possible.
- F1 score was used to find the best prediction threshold balancing precision and recall.
- Significance of intervention effects was confirmed using p-values and confidence intervals.

Results

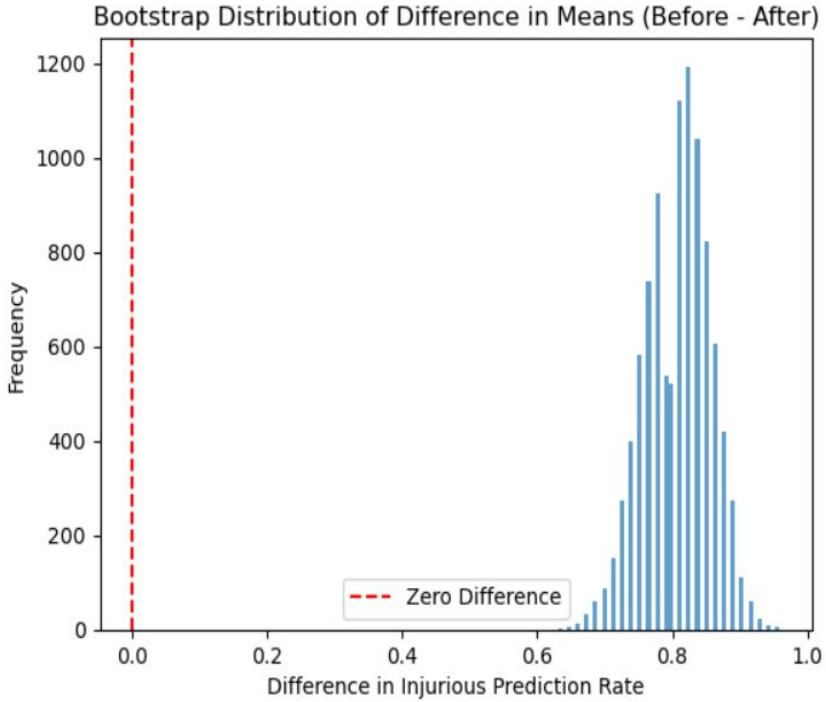
Segment Details

- Segment Length: 3.26 miles
- Longitude range: (-88.663, -88.605]
- Time Period of Observation: 2023–2025
- Observed Injurious Crashes (Actual): 25

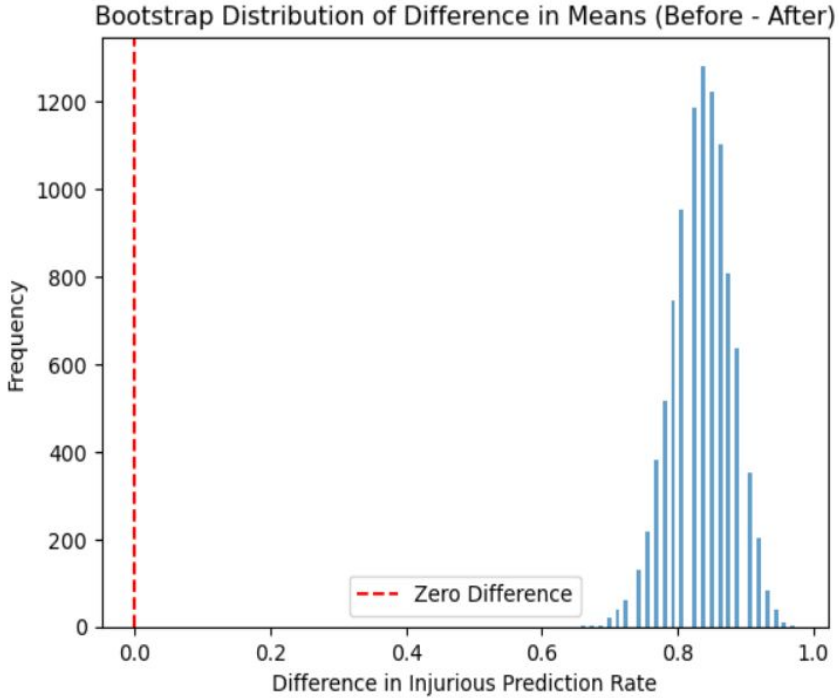




Results



Logistic Regression:
Mean Before = 0.836 (95% CI: 0.740 – 0.918)
Mean After = 0.027 (95% CI: 0.000 – 0.068)
p-value = 0.0000



CatBoost:
Mean Before = 0.836 (95% CI: 0.740 – 0.918)
Mean After = 0.000 (95% CI: 0.000 – 0.000)
p-value = 0.0000

Results

Injury Reduction Predictions			
Model	**Predicted Before**	**Predicted After**	**Reduction**
Logistic Regression	61	2	**59**
CatBoost	61	0	**61**

Societal Cost of Injury (Conservative Assumption)	
Cost per Injurious Crash	$302,600
Prediction threshold with LR	50%
Estimated Societal Benefit	**0.5 x 59 x $302,600 = $8,926,700**

Source: US Department of Transportation Cost-Benefit Analysis 2022 Update)

Intervention Cost Estimation			
Component	**Cost per Mile**	**Segment Length**	**Total**
Guardrails	$ 626,600	3.26 miles	$ 2,042,716
Lane Markings & Signage	$ 60,000	3.26 miles	$ 195,600
Total Estimated Cost			**$ 2,238,316**

Source:FHWA Roadway Design

Net Benefit = $8,926,700 - $2,238,316 = $6,688,384

Future Directions:

- Causal modeling
- Recognizing highway features with deep learning
- Investigating crash prevention


EAST
INTERSTATE
40
WEST
INTERSTATE
40