US TRAFFIC ACCIDENTS 2016-2013

INFRASTRUCTURE AND URBAN-RURAL DISPARITIES

-LAURA ROJAS-

Introduction & Methodology

Data Source: U.S. Accidents (Kaggle), ~7.2M records. 10% sample used for analysis (N = 719,282).

This project analyzes the US Accidents (2016–2023) dataset, a rich source of over 7 million accident records with environmental, temporal, and contextual features.

The primary business goal is to generate actionable insights to support data informed infrastructure planning and to highlight disparities between urban and rural settings in accident severity and outcomes.

- Do accident severity patterns differ significantly between urban and rural areas?
- How do interactions between urbanicity and infrastructure affect accident outcomes? Is the lack of POIs in rural areas contributing to more severe accidents?
- What role does the rush hour play, and how does it interact infrastructure factors?

Cleaning, Imputation and Additional Features

Exploratory Data Analysis:

Visualizations and summary statistics informed hypothesis formulation.

Multivariate Analysis:

- Logistic Regression on high-severity (binary target).
- Feature importance via coefficients and odds ratios.

Statistical Tests:

- **Categorical:** Chi-square, Cramér's V
- **Ordinal/Continuous:** Kruskal-Wallis, Mann-Whitney U, Spearman Correlation
- **Interaction Effects** via contingency tables (rush hour × traffic control)

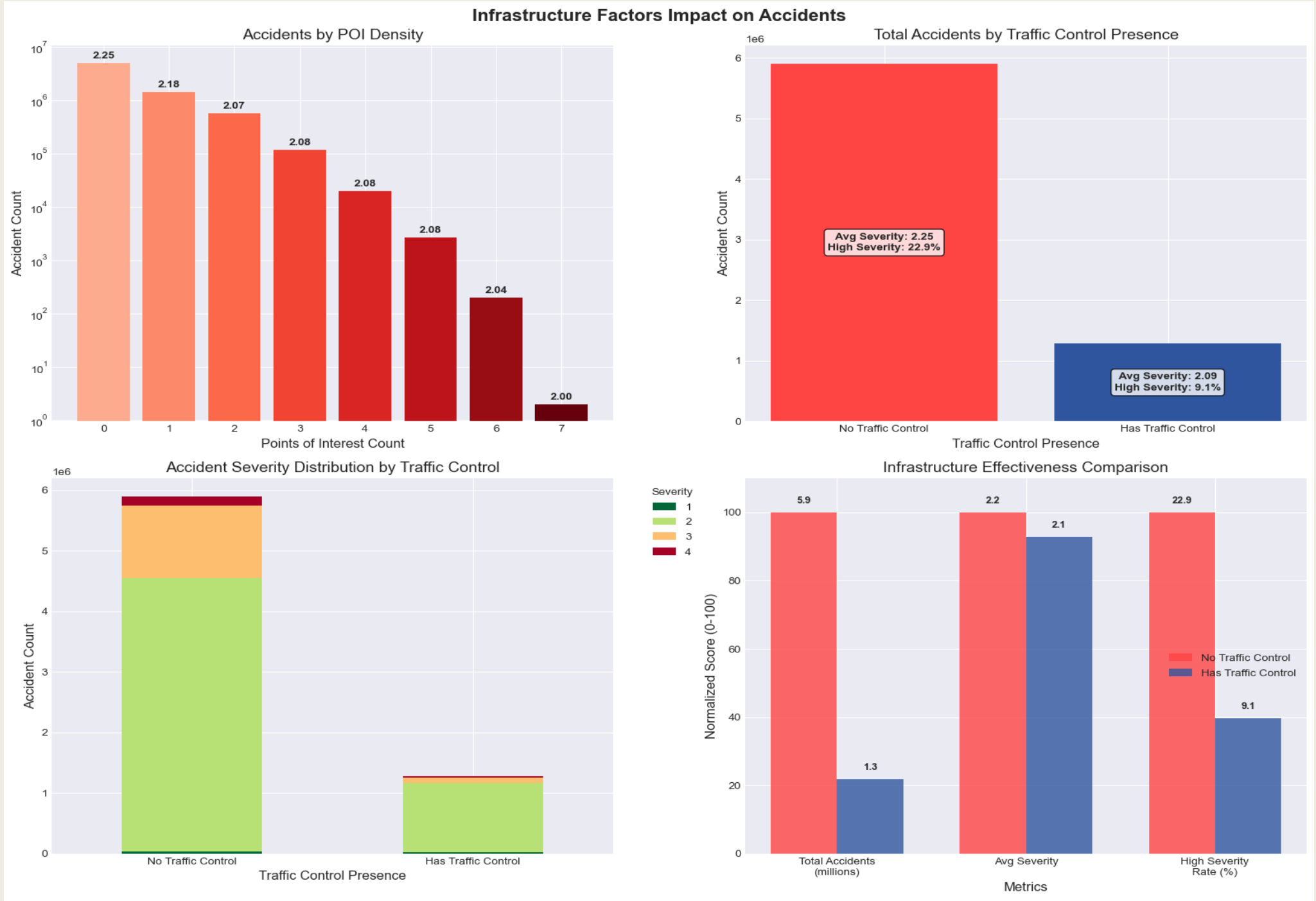
Initial Data Exploration: Key Findings

This analysis specifically examines how Points of Interest (POI) density and traffic control infrastructure correlate with accident severity to identify actionable intervention strategies.

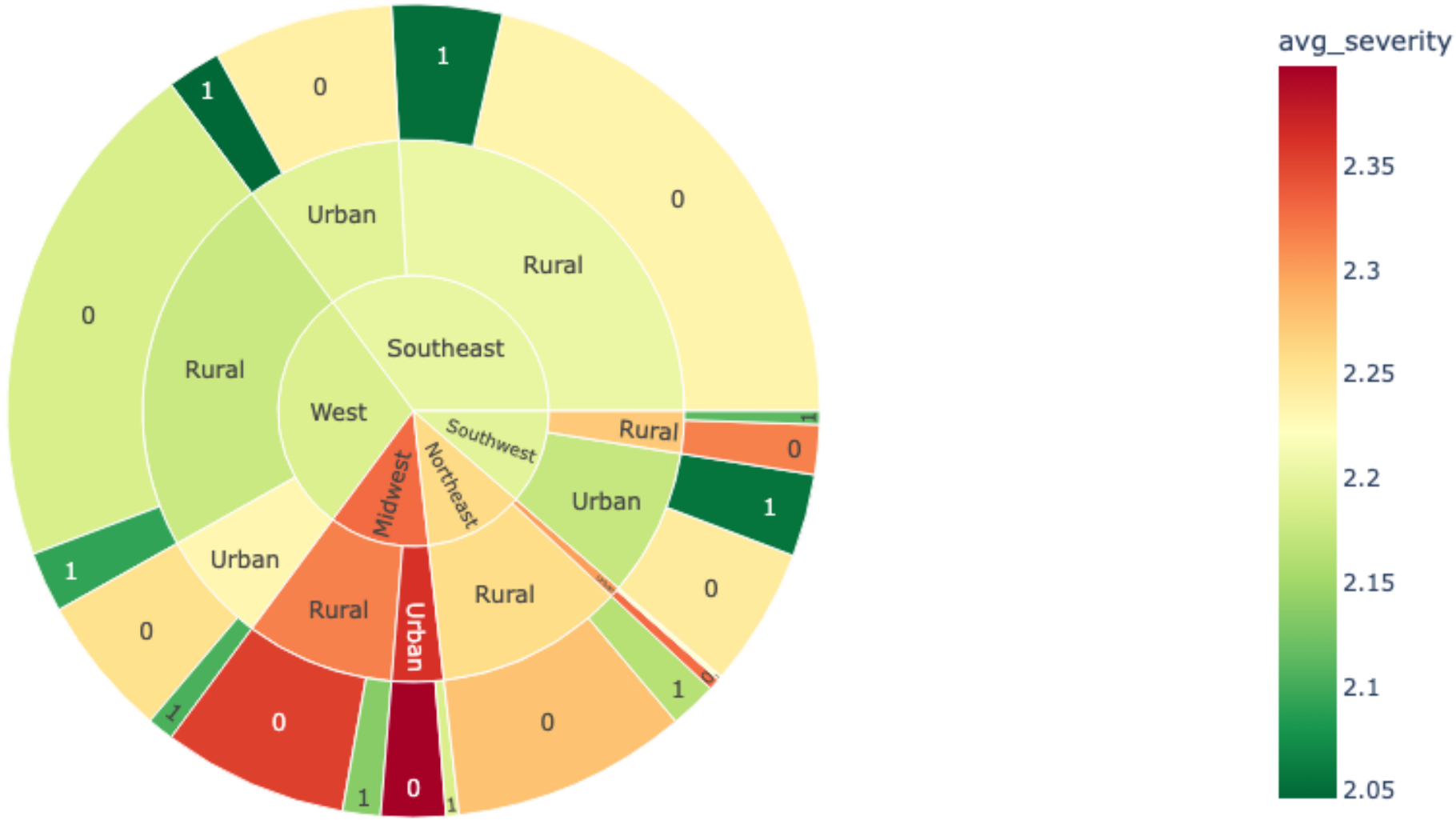
1.POI Impact: Areas with 0 POIs have 2.253 avg severity vs 2.176 for areas with 1 POIs (8.1% reduction)

2. Traffic Control Effectiveness: 7.2% reduction in average severity

3. High Severity Reduction: 60.4% reduction in high-severity accident rate No Control: 22.9% vs With Control: 9.1%


Infrastructure Factors Impact on Accidents
Accidents by POI Density
Accident Count
Points of Interest Count
2.25
2.18
2.07
2.08
2.08
2.08
2.04
2.00
Total Accidents by Traffic Control Presence
1e6
Accident Count
Traffic Control Presence
Avg Severity: 2.25
High Severity: 22.9%
Avg Severity: 2.09
High Severity: 9.1%
No Traffic Control
Has Traffic Control
Accident Severity Distribution by Traffic Control
1e6
Accident Count
Traffic Control Presence
No Traffic Control
Has Traffic Control
Severity
1
2
3
4
Infrastructure Effectiveness Comparison
Normalized Score (0-100)
Metrics
5.9
1.3
2.2
2.1
22.9
9.1
No Traffic Control
Has Traffic Control
Total Accidents (millions)
Avg Severity
High Severity Rate (%)

Key Risk Factors: Region → Urban/Rural → Infrastructure



STATISTICAL ANALYSIS & INSIGHTS

The process began multivariate logistic regression to evaluate the relative importance of predictors such as POI density, traffic control presence, and urbanity. Although the model's predictive power was limited (ROC-AUC = 0.5961), the feature coefficients revealed meaningful associations, the most important being a negative relationship between POI count and severity, and a mitigating effect of traffic control.

- poi_count: Odds Ratio = 0.65 Fewer POIs mean Higher severity
- urban_rural_num: Odds Ratio= 1.38 Urban areas mean Higher severity
- urban_poi_interaction: Odds Ratio = 0.81 Suggests mitigation effect of POIs in urban areas.

- H1: More POIs → Lower Severity

- Chi-Square significant ($p < 0.001$), Cramér's $V = 0.11$ → Small effect

- Kruskal-Wallis also significant ($p < 0.001$)

- Descriptive stats: Slight reduction in mean severity from No_POI to Medium_POI

Mean ranges from 2.25 → 2.07 (subtle but important in the great scheme)

- H2: Traffic Control Presence

- Chi-Square significant ($p < 0.001$), Cramér's $V = 0.13$ Small to moderate effect

- Mann-Whitney U significant ($p < 0.001$), effect size = -0.15

- Descriptive: Mean severity lower when traffic control is present (2.25 → 2.09)

- H3: Urban vs Rural

- Chi-Square significant ($p < 0.001$)

- POI Count differs significantly: Urban Mean = 0.61 vs Rural = 0.34

- H4: Infrastructure & Time Interaction

- Non-rush: Traffic control reduces high severity from 22% to 9.5%

- Rush hour: Traffic control reduces it even further: 23.5% to 8%

- Strong interaction effect, especially important for urban planning and stakeholders

RECOMMENDATIONS

Strategic Traffic Control Infrastructure

Implement a data-driven traffic control installation program targeting high-accident corridors with low infrastructure density, prioritizing locations with POI counts ≤2 and no existing traffic control.

Statistical Justification

- Traffic control presence reduces mean severity from 2.25 to 2.09 (7.1% improvement)
- 86.5% reduction in high-severity accidents during rush hours (23.5% → 8.1%)
- Areas without traffic control show 22.3% high-severity rate vs 8.9% with control
- Logistic regression coefficient: -0.431 for POI count (OR = 0.65)

POI - Based Safety Corridor Development

Create an Integrated safety corridor by providing more POI around high accident areas to reduce speed and increase driver awareness.

Statistical Justification

- Each additional POI reduces severity odds by 35% (OR = 0.65)
- Areas with 3+ POIs show consistently lower severity rates
- Urban areas with higher POI density (0.61 vs 0.34 rural) demonstrate better safety outcomes
- Strong correlation between POI presence and traffic control (r = 0.754)

Rush Hour Dynamic management system

Design and deploy an intelligent transportation system that adjust traffic control and warning during rush hours

Statistical Justification

- Rush hour increases severity differences: 23.5% vs 8.1% high-severity rates without vs with traffic control
- Time-based patterns show 20% higher accident rates during rush hours
- Weather severity correlation (0.035) suggests environmental responsiveness needed
- Poor visibility correlation (-0.109) indicates opportunity for adaptive lighting

Traffic Accidents - Dashboard