The Need for Speed…. REDUCTI0N: Chicago Traffic Safety Analysis

Author: Chris Kucewicz

Photo by Adrian Williams on Unsplash

Introduction and Contents

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Exploratory Data Analysis
5. Modeling and Evaluation
6. Limitations
7. Recommendations
8. Next Steps

Business Understanding

Background | Goals | Success Criteria

Vision Zero Chicago

Goal: ***0*** traffic deaths by 2026

Reality:

- 2024 (to-date): ***115*** traffic deaths
- 2023: ***136*** traffic deaths

Business Understanding

Background → **Goals** → Success Criteria

- **Develop an *inferential* model**
- **Assist Chicago DOT (CDOT) in resource allocation**

"*Serious crash*"
=
Fatal or incapacitating

To learn more about the City of Chicago's definition of an 'incapacitating injury', visit this link: https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if/about_data

Business Understanding

Background → Goals → **Success Criteria**

- **Identify top 3 contributing factors**
- **Precision-Recall Area Under the Curve (PR AUC)**

Business Understanding

Background → Goals → **Success Criteria**

- **Precision-Recall Area Under the Curve (PR AUC)**
  - **Focuses on minority case (serious crashes)**

Data Understanding

Source: Chicago Data Portal (updated: Dec 2024)

Datasets:

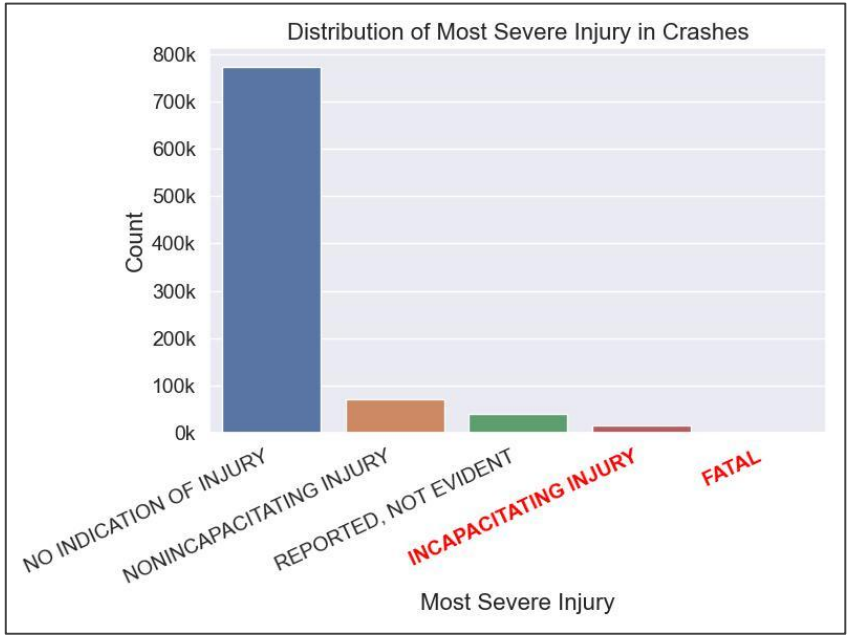
- Crashes: ***900k*** records, ***48*** features
- People: ***~2M*** records, ***29*** features:
- Vehicles: ***1.8M*** records, ***71*** features

Linked by: `crash_record_id`

Data Preparation: Phase 1

1. Processed ***3*** datasets
    a. Removed irrelevant features
    b. Processed features reducing cardinality
    c. Handled Null Values
    d. Reclassified the Target Variable to Binary

Data Preparation: Target Reclassification



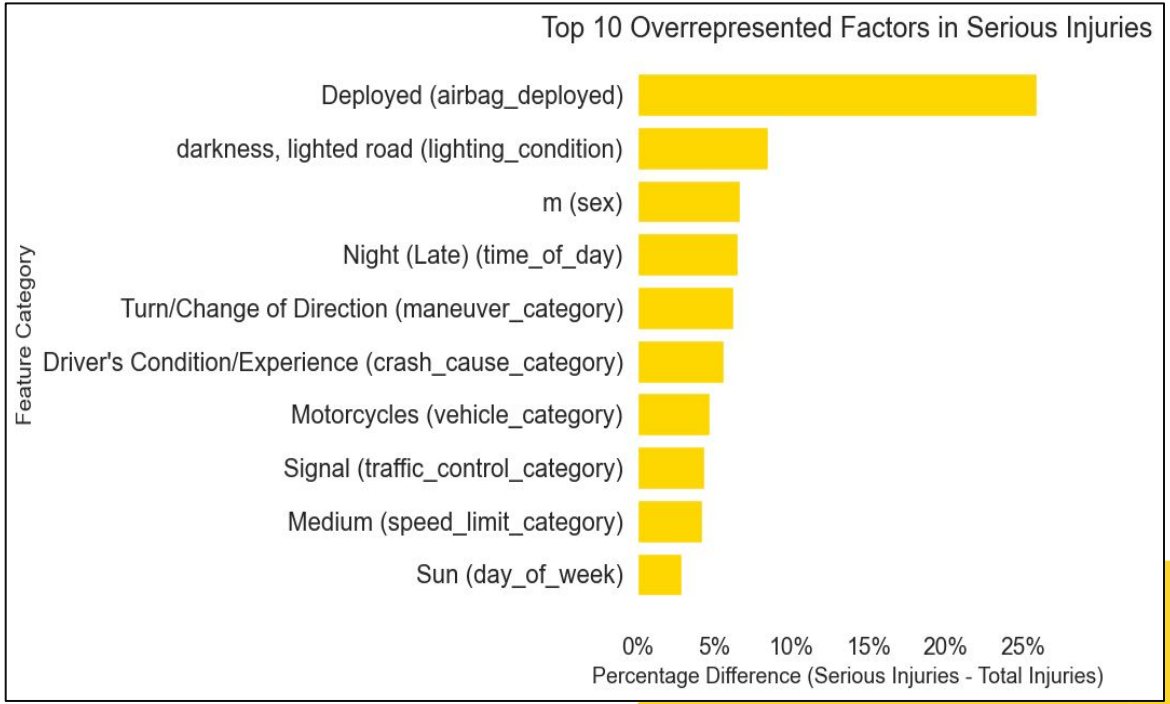
Data Preparation: Phase 2

3. Stratified Sample
    a. Subset: ***15***% of merged_df
    b. Kept target classes proportional
    c. Final dataframe: ***~96k*** records, ***16*** features, ***0*** null values

Exploratory Data Analysis: Key Finding

Overrepresented Categories in Serious Crashes:

- **Sex**: Males
- **Lighting**: Darkness with lighted roads
- **Airbag Deployment**



Modeling: Key Components

Standard Modeling Process:

- Train-test-split = Avoids data leakage
- Baseline models for comparison
- Hyperparameter tuning

Modeling: Addressing Issues

- **Developing interpretable model**
  - White box models (opposite of black box models)

- **Addressing class imbalance**
  - Oversampling (SMOTE) + class weights

Reminder: Minority class = Serious crashes

Evaluation

Metric (PR AUC): Best Decision Tree = ***0.0956***

Class Imbalance: Likely contributing to low performance

Focus: Model inference over predictive accuracy

Perfect Performance:

PR AUC = **1.0**

Evaluation

Decision Trees: Simple, highlight feature importance

Key Features:

- Airbag Deployment (***Not Deployed***)
- Sex (***Male***)
- Seasonality (***Summer, Winter, Spring***)

	Original_Feature	Category	Feature_Importance
0	airbag	Not Deployed	0.0625
1	crash	Unknown/Other	0.0499
2	sex	male	0.0437
3	season	Summer	0.0414
4	season	Winter	0.0397
5	season	Spring	0.0391

Limitations

1. Data Quality: Police-Reported Data

2. Decision Tree Interpretation

3. Computing and Time Constraints

Recommendations

Recommendation 1:
Male Trends Analysis

- Research trend among male *drivers*
- Targeted safety campaign

Recommendations

Recommendation 1:
Male Trends Analysis

- Research trend among male *drivers*
- Targeted safety campaign

Recommendation 2:
Size Regulation

- Implement personal vehicle weight fee
- Money raised → infrastructure projects

Recommendations

Recommendation 1: **Male Trends Analysis**

- Research trend among male *drivers*
- Targeted safety campaign

Recommendation 2: **Size Regulation**

- Implement personal vehicle weight fee
- Money raised → infrastructure projects

Recommendation 3: **Airbag Inspection**

- Annual airbag safety inspections

Next Steps

1. Investigate male injury trends in crashes

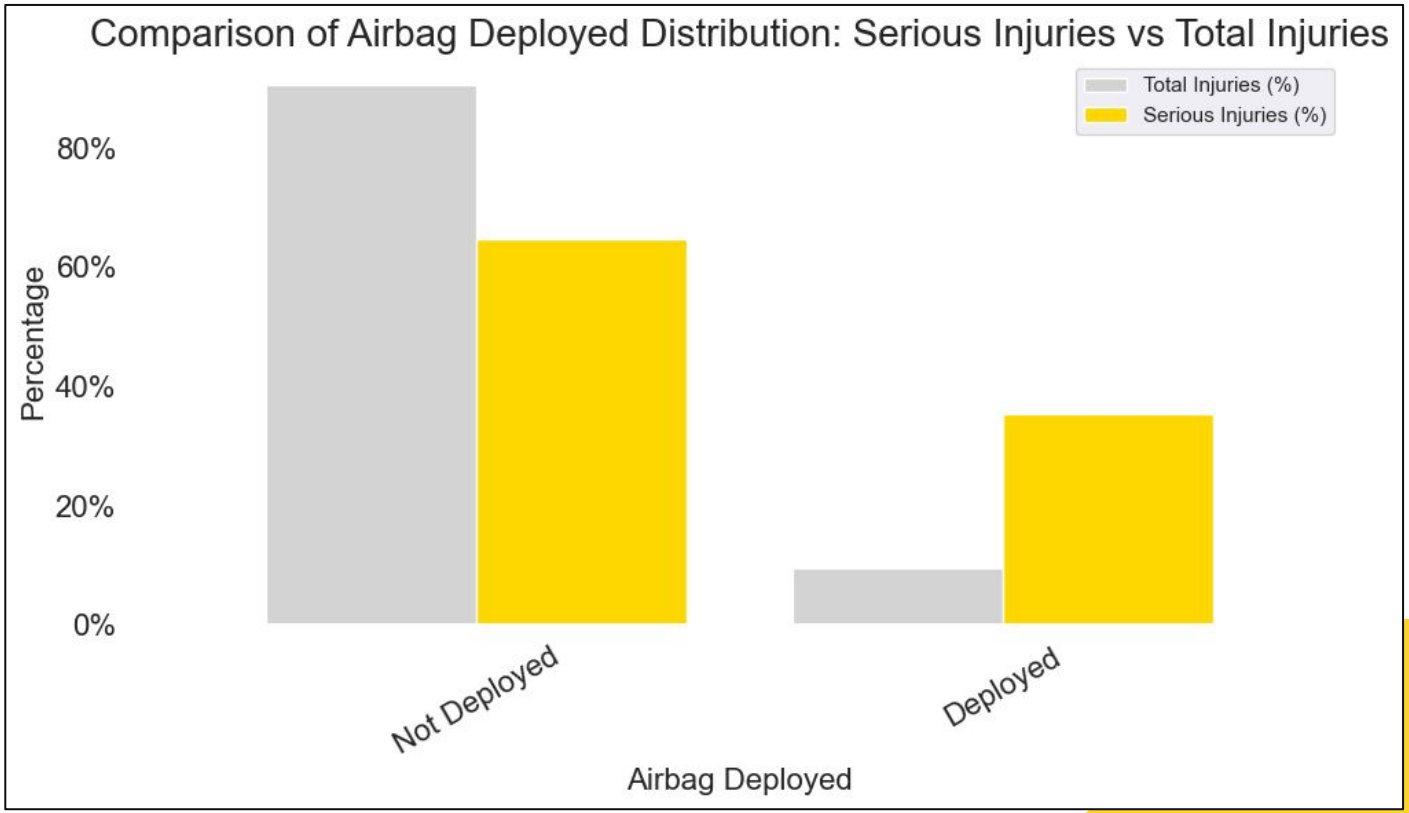
2. Explore key factors contributing to speeding

Thank you!

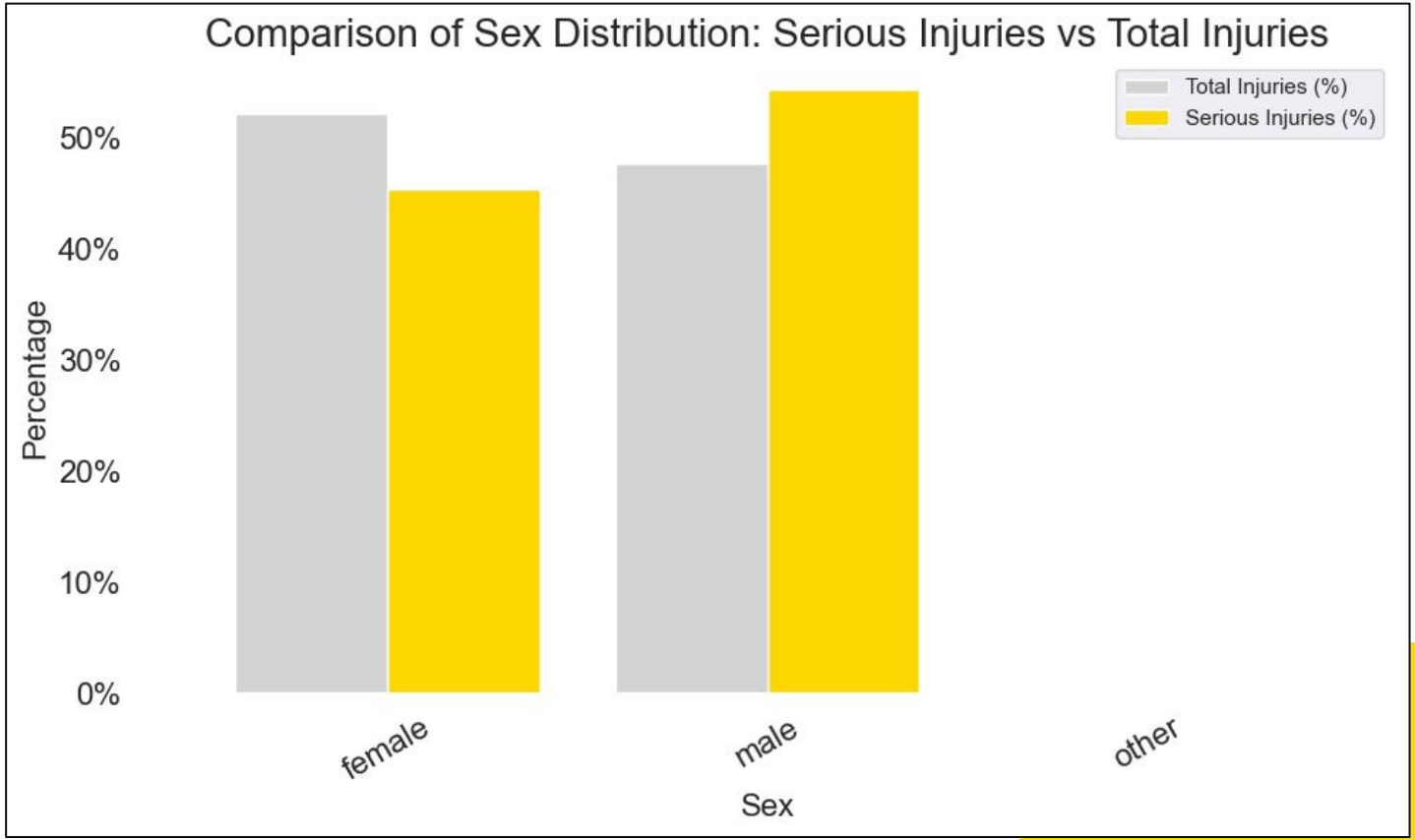
Github Repository:
https://github.com/ckucewicz/traffic_crash_prediction

Contact Chris Kucewicz at
cfkucewicz@gmail.com with additional questions

Additional Visualizations: Airbag Overrepresentation



Additional Visualizations: Male Overrepresentation



Additional Visualizations: Darkness Overrepresentation