



CHICAGO TRAFFIC CRASHES



Team 12 – Jiayi, Jiechen,
Dhaval, Divya

MOTIVATION



ALLEVIATING
HUMAN IMPACT



IMPROVING
PUBLIC SAFETY



MITIGATING
FINANCIAL TOLL



PROACTIVE
PROTECTION



LIFESAVING
INTERVENTIONS





GATHERING DATA

- **Sources:**
 - Traffic Crashes data from the City of Chicago Data Portal
 - Weather Data from Visual Crossing
- **Contents:** Detailed incident-level records on all reported motor vehicle crashes in the City of Chicago from 2015 to present
- **Volume:** Over 700,000 crash observations, spanning 49 attributes
- **Key Strengths:**
 - Large in-scope duration - 7 years of data
 - Granular details and insights per incident
 - Tailored to city infrastructure and traffic conditions



DATA OVERVIEW


 **Location:** *Latitude and Longitude, Distance to Downtown*

 **Date and Time:** *Hour of day/ day of week/day of month*

 **Injuries:** *Number and type of injuries that occurred*

 **Damages:** *Estimated damage cost of crash*

 **Crash Type:** *Type of crash – no right of way, intersection related etc.*

 **Weather Conditions:** *Weather, Lighting, traffic, traffic control devices, roads etc.*



RESEARCH QUESTIONS



How **long** does it take for the **police to be notified** upon accident?

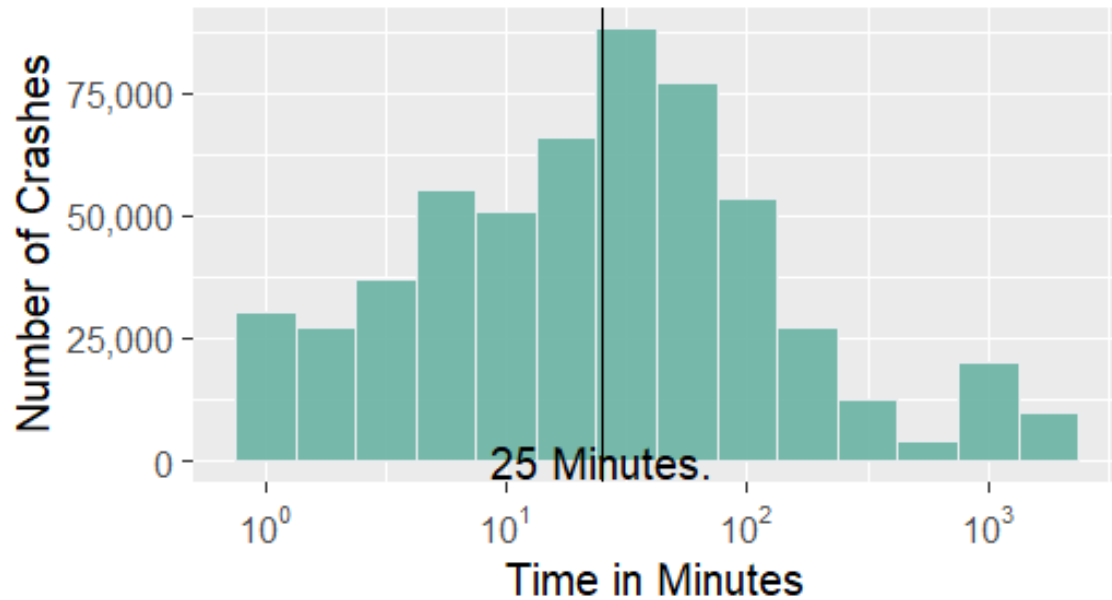


Can we identify **high-impact accidents** based on crash characteristics?

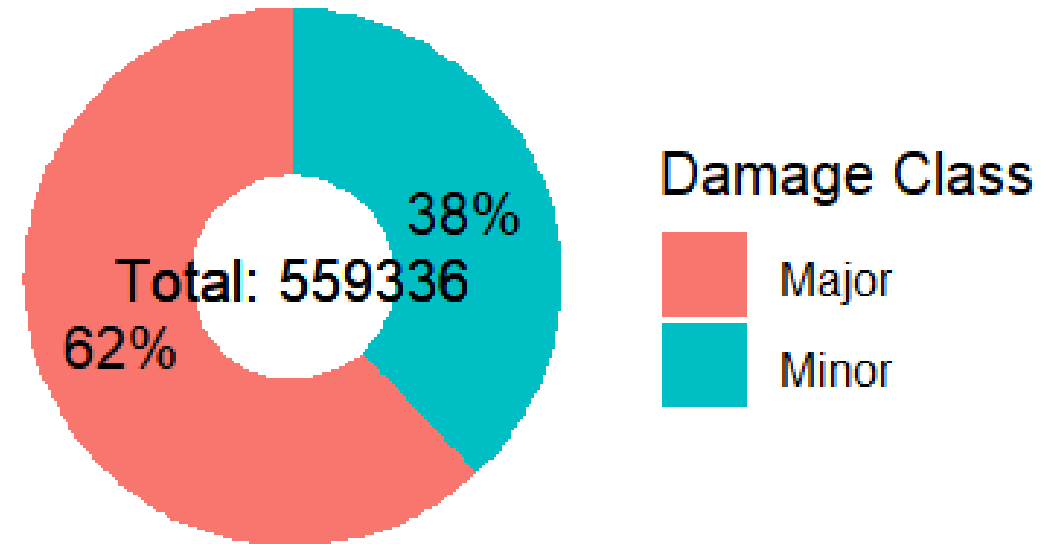


EXPLORATORY DATA ANALYSIS

Police Notification Time

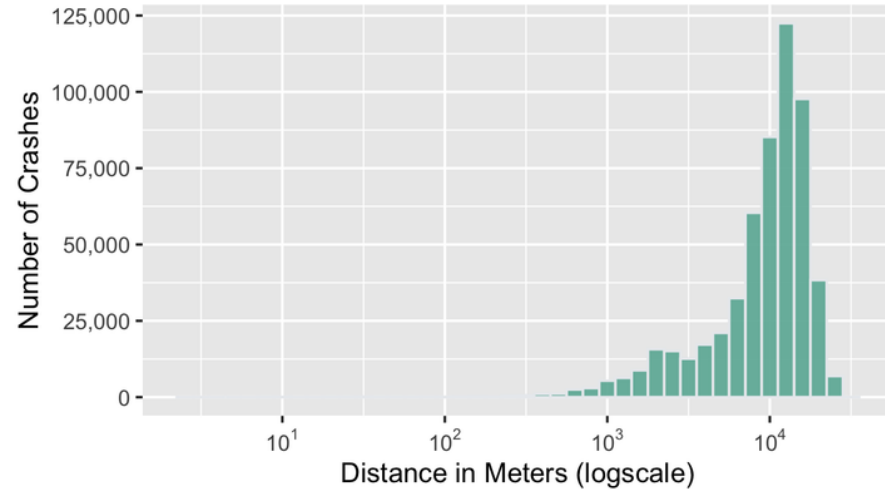


Distribution of Damage Type

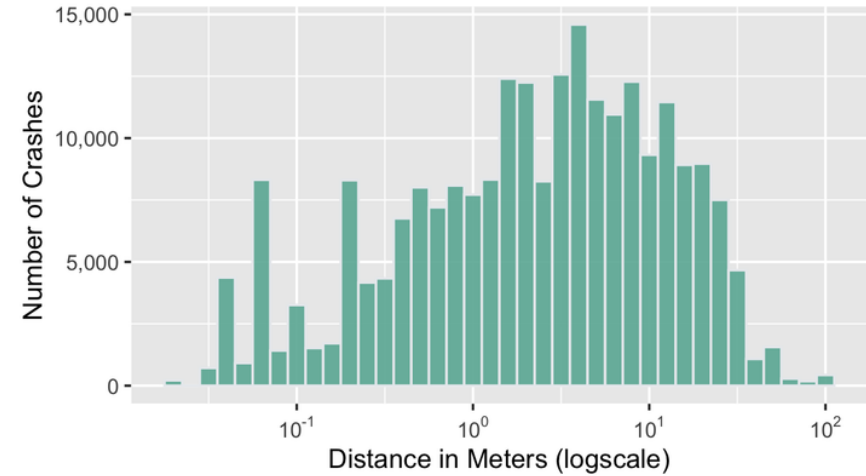


EXPLORATORY DATA ANALYSIS

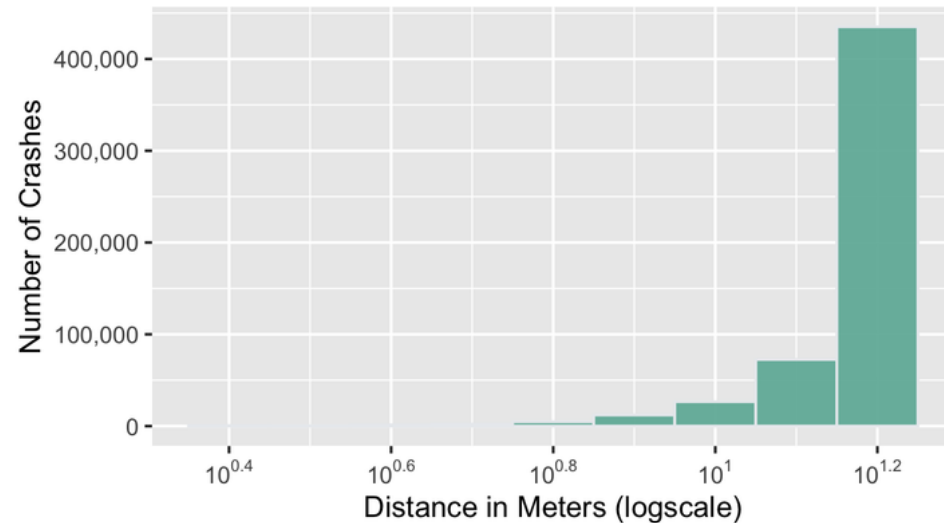
Distribution of Distance to Downtown



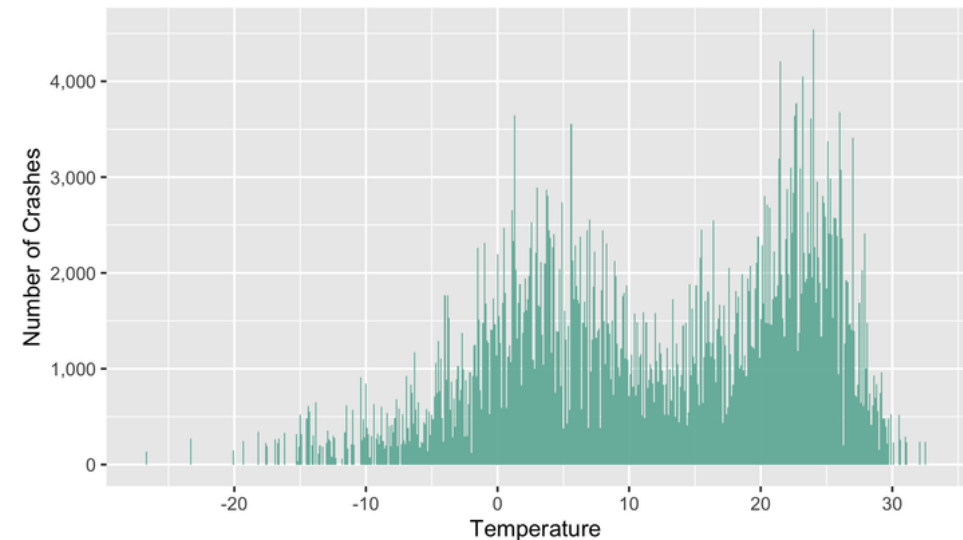
Distribution of Precipitation



Distribution of Visibility



Distribution of Temperature





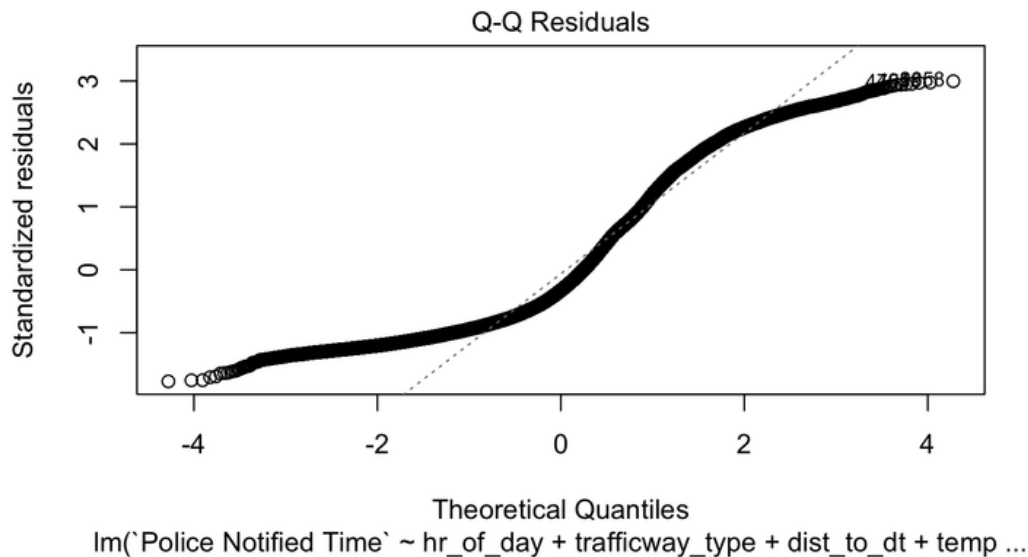
MODEL 1: TIME TO NOTIFY POLICE

Question: How long, in minutes, does it take for the police to be notified upon accident?



Method: Estimate Time to Notify Police

- Time Taken: Continuous (Minutes)
- Method: Linear Regression

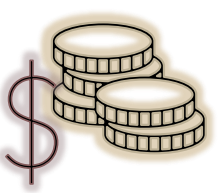


Results



- The Q-Q plot on the left shows significant tail-behavior and indicates possible non-linearity in the data
- Additional weather/location variables didn't improve model fit, with a 4.5% R2 score





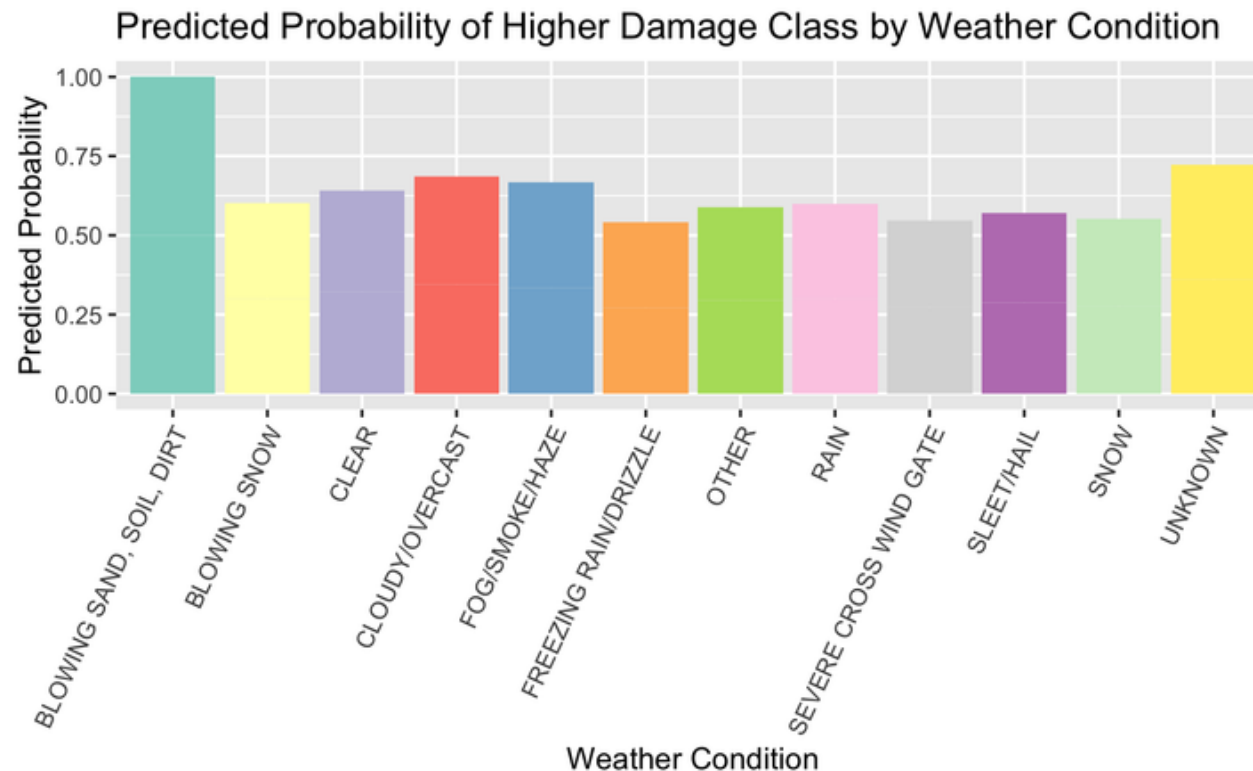
MODEL 2: COST IMPACT CLASSIFICATION

Question: Can we identify high-impact accidents based on crash characteristics?



Method: Cost Impact Classification

- Damage Size: $< \$1.5K$ or $> \$1.5K$
- Method: Logistic Regression



Results

- Even Distribution in Prediction Probability
- Stability Post-Outlier Exclusion
- Mixed Levels of Multicollinearity
- Moderate Accuracy in Severe Accident Prediction (True Positives)



LIMITATIONS AND FUTURE WORK



Limitations

- Challenge: Removed missing values because no information on missing
- Risk: Added external data that may misrepresent the dataset

Potential Improvement

- Address Missing Values: Identify reasons for missing data and perform imputation
- Enhance Dataset: Include additional variables for a comprehensive view
 - Driver-related Variables: Age, gender, years of driving, drinking alcohol
 - Behavioral Variables: Speeding, distracted driving (phone use, etc.), seatbelt usage
- Introduce new models: Decision Trees, Random Forest





THANK YOU

