Examining the Applicability of Waze Crash Alert as a Real-Time Crash Detection Tool

Presenter:
Branislav Dimitrijevic

Authors:
Md Tufajjal Hossain, Joyoung Lee, Dejan Besenski, Lazar Spasovic

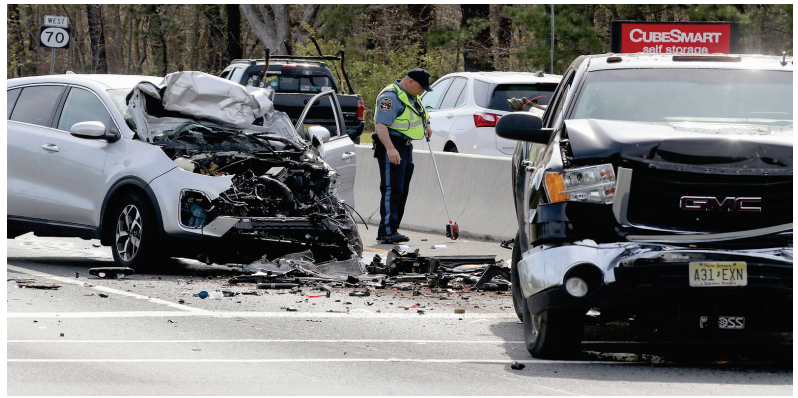
New Jersey Institute of Technology
Department of Civil and Environmental Engineering
NJIT ITS Resource Center

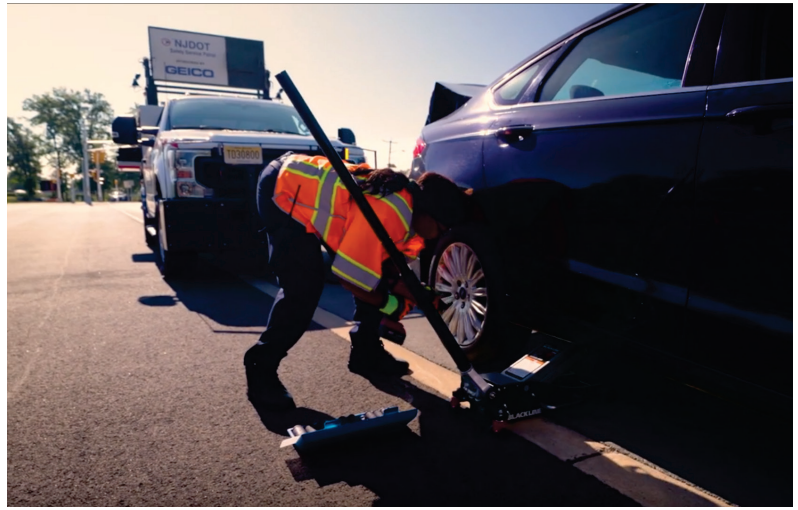
26th Annual NJDOT Research Showcase
October 23, 2024

Mercer County Community College – Conference Center at Mercer, West Windsor, NJ 08550

Incident Detection and Response

- Timely and reliable detection of traffic incidents is critical to effective emergency response and congestion management.
- Traditionally, the traffic operations personnel rely on stationary data collection technologies for incident detection, such as loop detectors, cameras, and microwave-based systems, as well as verified incident reports from the scene.
- The challenge: limited coverage (temporal and spatial) and due to limited resources.
- In recent years, there has been an increased focus on crowdsourced data for incident detection due to its cost-effectiveness and extensive coverage.
- Despite its potential, crowdsourced data is largely “unsanctioned” and unverified, as the contributors are not traditionally trained and lack responsibility for report accuracy and reliability.

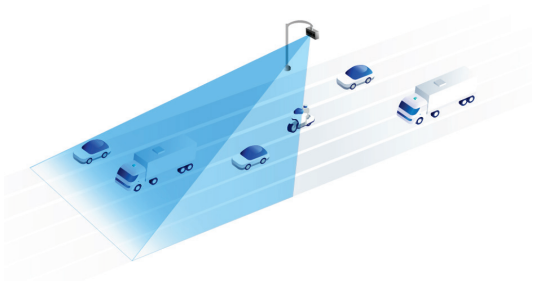


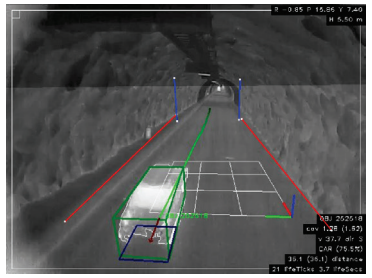


Background

Incident Detection Methods

Traditional Methods
(e.g., loop detectors, cameras)





- High costs (installation and maintenance)
- limited coverage.

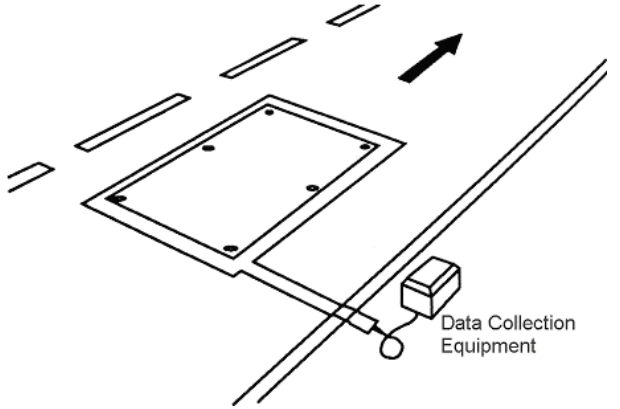
Crowdsourced Data
(e.g., Waze Data)

- Cost-effective, extensive coverage.
- Unverified, questionable reliability and accuracy, contributors lack formal training

Background - Cont'd

- **Importance of Incident Detection**
  - Effective traffic incident and emergency response and traffic incident management rely on quick incident detection, particularly for motor vehicle crashes.
  - Timely detection -> quicker response -> less traffic impact

- **Challenges in Incident Detection**
  - High cost and limited coverage of roadside detection.
  - Limited human resources for roadway monitoring
  - Limited coverage by the SSP, reporting by the law enforcement.



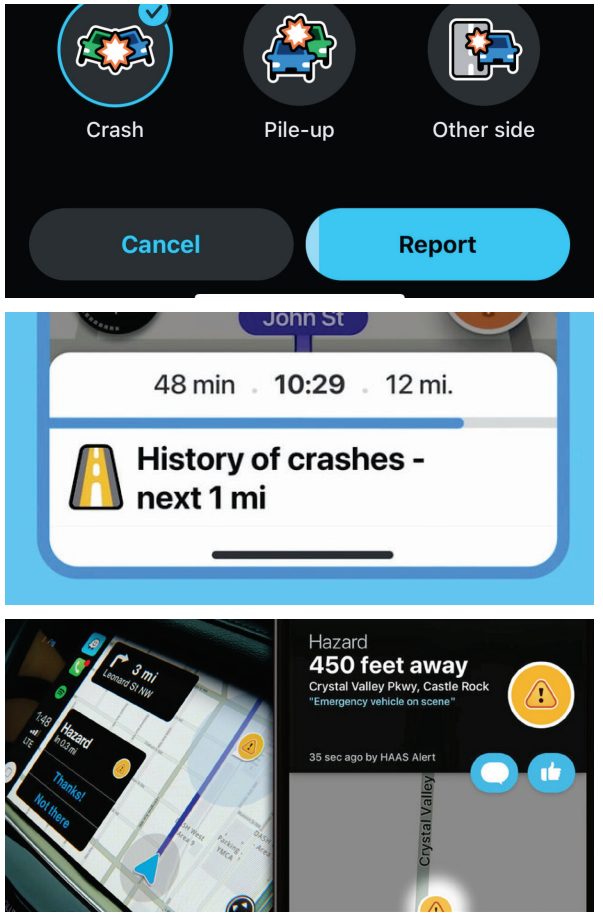
Background - Cont'd

- **Leveraging crowdsourced data for Enhanced Response**
  - To further improve SSP's response times and reduce congestion during incidents, crowdsourced data could be utilized to detect incidents, especially crashes.
  - Among various sources of crowdsourced data, Waze is one of the most popular. Waze user-generated alerts offer significant potential for real-time incident detection.
  - However, ensuring the validity and accuracy of these alerts remains a challenge as false reports can occur.

Background - Cont'd

Introduction and Working Procedure of Waze

- Waze is a crowd-sourced navigation application created in 2009. It provides satellite navigation software on smartphones and other GPS-enabled devices. It efficiently guides users by leveraging information contributed by other users.
- Users report traffic "events" like crashes, congestion, traffic stops, police traps. Waze turns these into alerts and offer alternative routes.
- Users can also update maps and landmarks.
- Waze integrates state agency's data on road construction for enhanced accuracy.



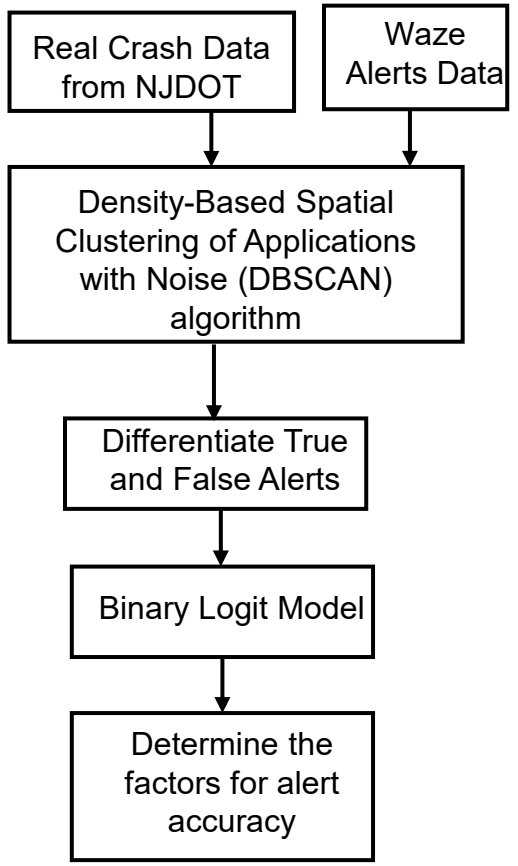
Goals and Methods

1) Evaluate Waze Crash Alert Matching Rate

- Comparison of Waze crash alerts with actual crash records.
- Use historical crash records from New Jersey Crash Report (NJTR-1) and Waze alerts.
- Apply the DBSCAN algorithm to identify true and false alerts.

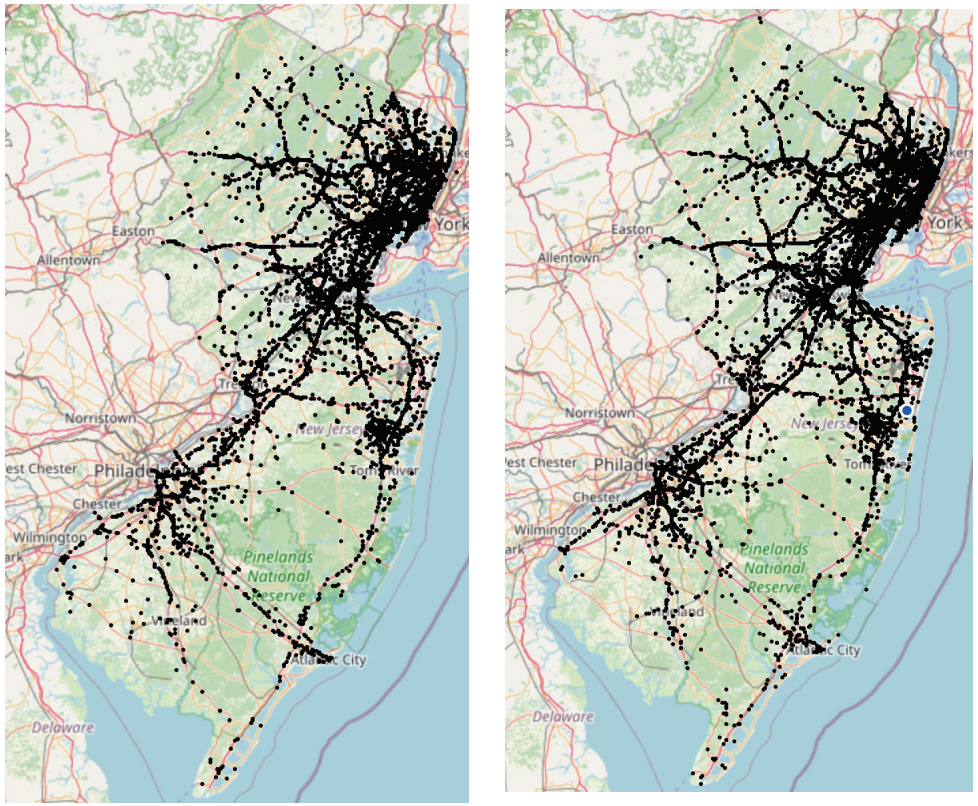
2) Identify Factors Affecting Waze Alert Matching

- Determination of the factors influencing whether Waze alerts match real crashes.
- Conduct statistical analysis to explore key variables.
- Develop a binary logit model to analyze factors that contribute to true alert cases.



Data Overview

- **Waze Crash Alert Data Overview**
  - ***Timeframe of data used in analysis:***
    - 2021: September to December
    - 2022: January to April
    - 2023: June to July
  - ***Data Captured:***
    - Unique system ID
    - Datetime
    - Geographic coordinates
    - Road type
    - Reliability
    - Report description
    - Report rating
    - Confidence
    - User feedback ("thumbs up")



Location of Alerts (Left: October 2021; Right: June 2023)

Data Overview

❑ Waze Crash Alert Data Overview

➢ ***Key Data Attributes:***

I. *Reliability Score (0-10)*

- Based on user reactions ("thumbs up", "not there") and reporter level, reflects alert trustworthiness, starting at 5.
- Scores rarely fall below 5 unless a highly ranked editor submits a "not there" report.

II. *Confidence Score (0-5)*

- Positive feedback level from Waze users.

III. *Road Types*

- Streets
- Primary Streets
- Secondary Streets
- Freeways, and
- Ramps

IV. *Crash Subtypes*

- Minor
- Major, and
- No Subtype

V. *Type of Day*

- Weekend
- Weekday

VI. *Time Periods*

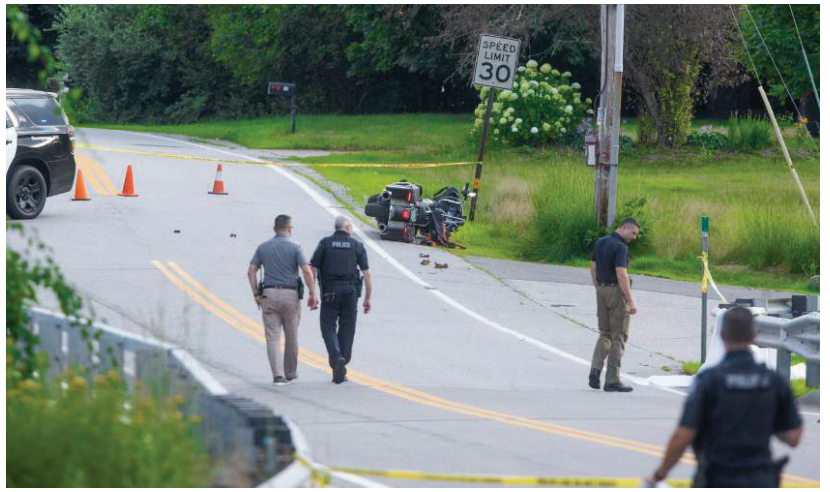
- **AM Peak:** 7 am - 9 am
- **PM Peak:** 4 pm - 7 pm
- **Morning Non-Peak:** 9 am - 12 pm
- **Afternoon Non-Peak:** 12 pm - 4 pm
- **Evening Non-Peak:** 7 pm - 9 pm
- **Nighttime Non-Peak:** 9 pm - 7 am

VII. *Report Rating (User Levels 1-6):*

- **Level 1:** New Users
- **Level 2 to 5:** Editors with map-editing privileges based on experience.
- **Level 6:** Waze Champions with nationwide access

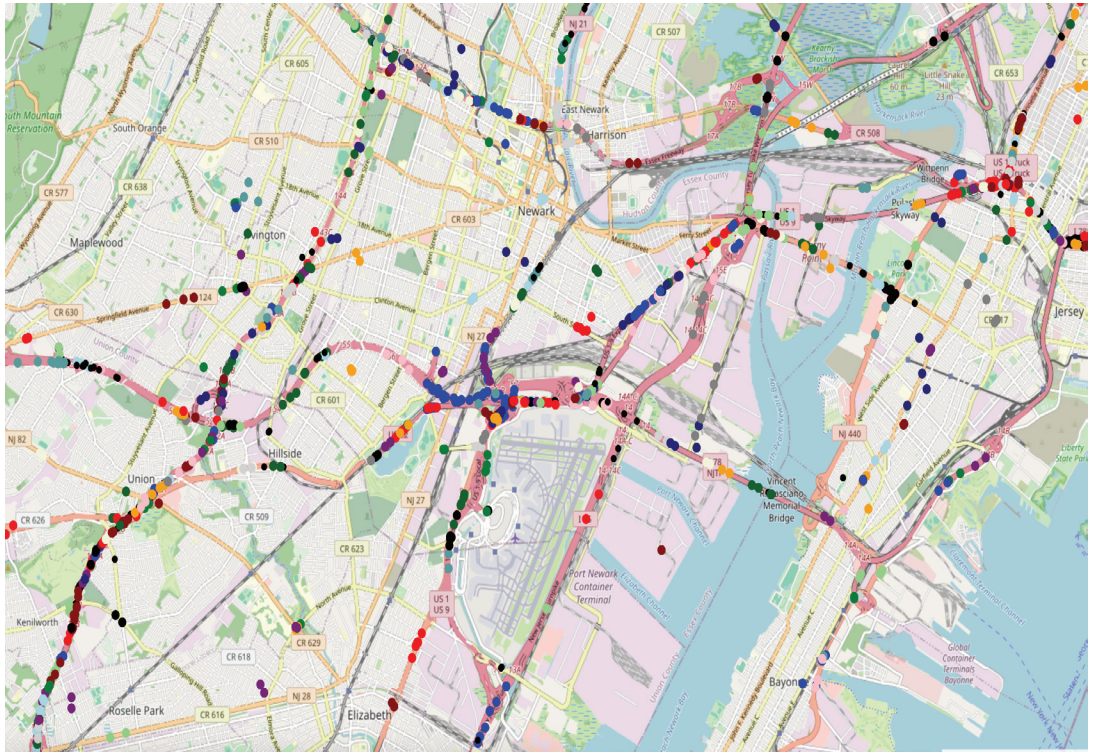
Data Overview – Cont.

- **New Jersey Police Crash Report (NJTR-1) Overview**
  - NJTR-1 collects detailed data on motor vehicle crashes in New Jersey.
  - **Focuses on critical elements:** crash location, date/time, severity, driver/vehicle details, and roadway infrastructure.
  - Utilizes 153 data blocks for comprehensive crash reporting.
  - Pedestrian and cyclist involvement only recorded when a motor vehicle is involved. It also captures commercial vehicle involvement.
  - NJDOT and law enforcement agencies use this data to study crash patterns and improve traffic safety.



Addressing False Waze Alerts

- Waze generates numerous alerts for the same crash, as reported by different Waze users in the vicinity of the crash at different times.
- Waze users may also report false alerts, which can be caused by a variety of reasons, including misuse of the Waze app, mischaracterization of incidents, or even GPS error on their Waze devices.
- To address such issues, a data filtering process is performed based on correlation between the Waze crash alerts and the actual crashes.
- This method employs the Density-Based Spatial-Temporal Clustering of Applications with Noise (DBSCAN) algorithm to filter out false alerts.



Crash alerts near Newark and Liberty Airport (Jan 2022). The alerts of the same color indicate that they are generated to report the same crash within an hour.

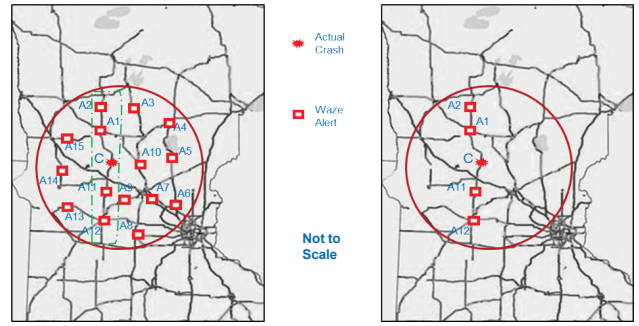
Addressing False Waze Alerts – Cont.

- **True and False Alert Identification**

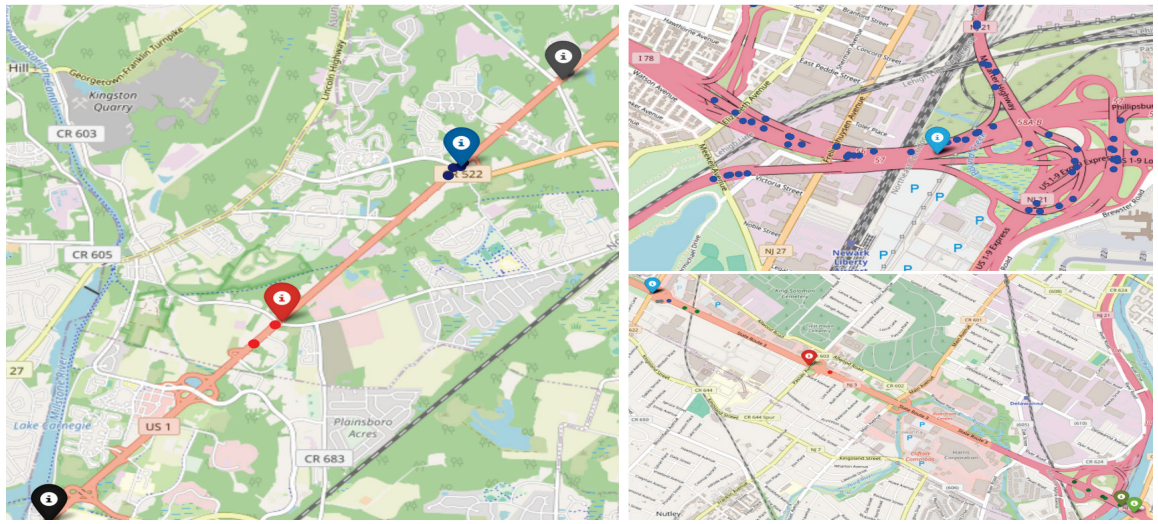
1. DBSCAN applies a 1-hour time window and 0.5-mile radius around true crash locations from NJTR-1.

2. Waze alerts within the radius are checked using road distance data from OpenStreetMap.

3. Only alerts within a 0.25-mile roadway distance are retained as related to the actual crash and identified as true alerts.



Conceptual Representation of Detection of Matched Waze Alerts



Example of Crash-Alerts Match Case. The pinpoints and the dots represent the actual crash locations and alerts matched with the crash, respectively.

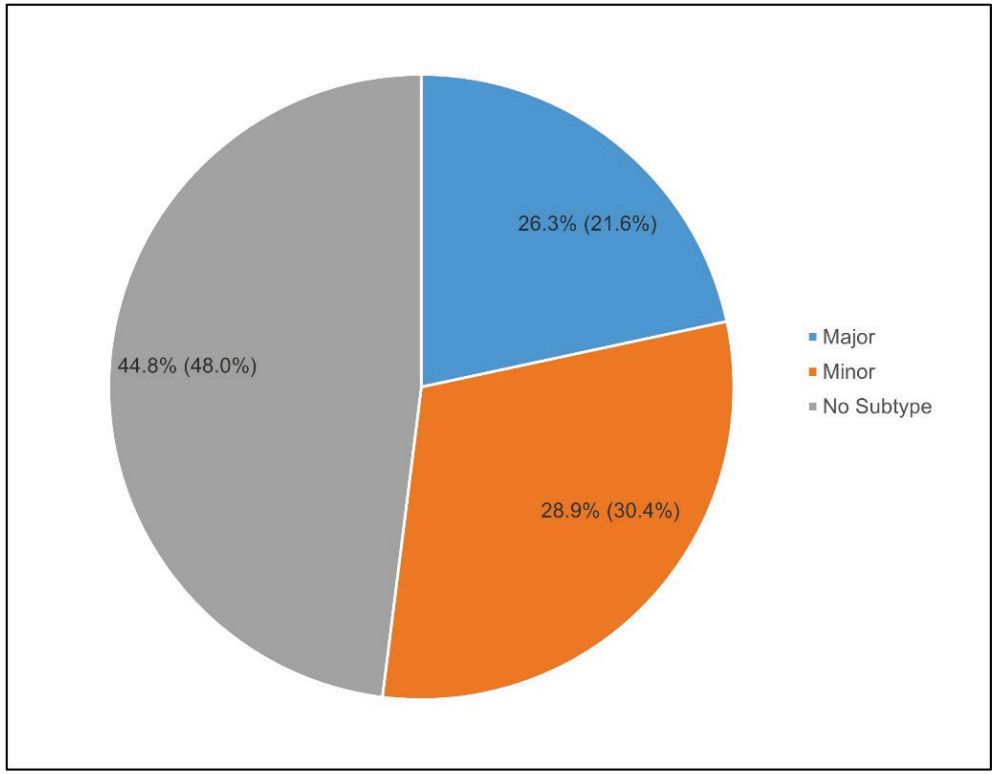
Matching Rate Analysis

- **Total Waze Alerts:** 119,481 records
  - Matched Alerts: 18,667 (15.6%)
  - Unmatched Alerts: 100,814 (84.4%)
- **Yearly Distribution:**
  - **2021:** 42.9% of total alerts (Average Matching Rate: 18.4%)
  - **2022:** 33.6% of total alerts (Average Matching Rate: 13.1%)
  - **2023:** 23.4% of total alerts (Average Matching Rate: 13.9%)
- **Highest matching rate:** 19.4% (Dec 2021), 19.0% (Oct 2021)
- **Lowest matching rate:** 12.4% (Feb 2022), 12.0% (Mar 2022)
- **Overall matching rate:** 15.6%

Year	Month	Matched with Actual Crash	Unmatched with Actual Crash	Matching Rate (%)
2021	September	469	2,333	16.7
	October	3,436	14,644	19.0
	November	2,936	12,976	18.5
	December	2,814	11,702	19.4
2022	January	1,677	10,399	13.9
	February	1,512	10,718	12.4
	March	1,703	12,491	12.0
	April	234	1,432	14.1
2023	June	1,826	10,933	14.3
	July	2,060	13,186	13.5
Overall Matching Rate				15.6

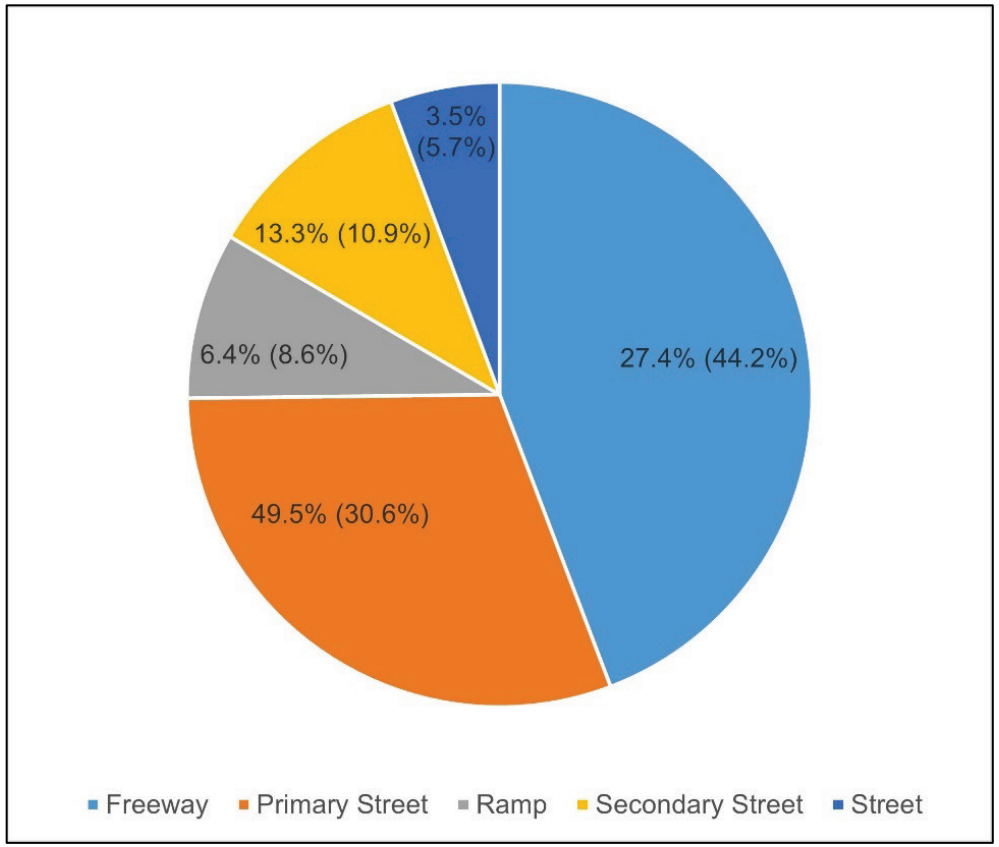
Matching Rate Analysis – Cont.

- **Crash Type Distribution (Matched Cases)**
  - **Major Crashes:**
    - 21.6% of total alerts
    - 26.3% of matched cases
    - Higher likelihood of matching with actual crashes
  - **Minor Crashes:**
    - 30.4% of total alerts
    - 28.9% of matched cases
    - Slightly lower matching rate compared to "Major" crashes
  - **No Subtype:**
    - 48% of total alerts
    - 45% of matched cases



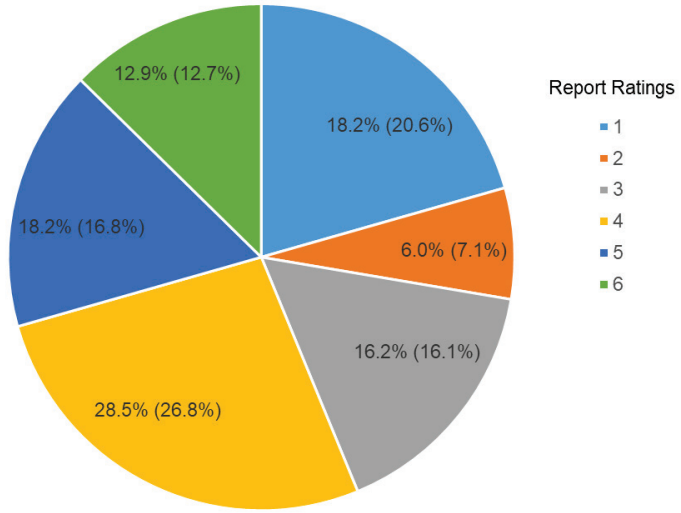
Matching Rate Analysis – Cont.

- **Road Type Distribution (Matched Cases)**
  - **Freeways:**
    - 44.2% of total alerts
    - 27.4% of matched crashes
    - Higher share of alerts, more false positives
  - **Primary Streets:**
    - 30.6% of total alerts
    - 49.5% of matched crashes
    - Highest matching rate
  - **Secondary Streets:**
    - 10.9% of total alerts
    - 13.3% of matched crashes
    - Moderate matching rate
  - **Other Road Types:**
    - ***Ramps and Streets:*** Fewer alerts and lower matching rates



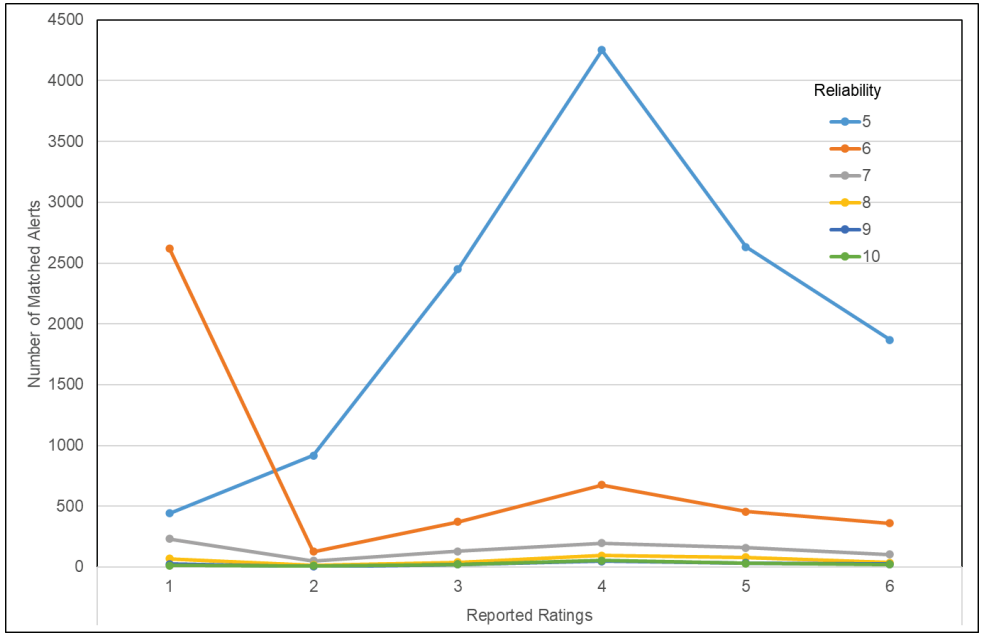
Matching Rate Analysis – Cont.

- **Report Ratings Distribution (Matched Cases)**
  - **Rating 1:**
    - 20.6% of total alerts
    - 18.2% of matched alerts
  - **Rating 2:**
    - Smallest share: 7.1% of total alerts
    - 6.0% of matched alerts
  - **Ratings 3 and 6:**
    - Rating 3: 16.1% of total alerts, 16.2% of matched alerts
    - Rating 6: 12.7% of total alerts, 12.9% of matched alerts
    - Close alignment between these two categories
  - **Rating 4:**
    - Highest share: 26.8% of total alerts
    - 28.5% of matched alerts
  - **Rating 5:**
    - 16.8% of total alerts
    - 18.2% of matched alerts



Matching Rate Analysis – Cont.

- **Report Ratings vs Reliability (Matched Cases)**
  - **Moderate Reliability (5.0):**
    - Most frequent, especially with report ratings between 3 and 5.
  - **Higher Reliability (5.0 to 10.0):**
    - Decreasing number of alerts with the increasing of reliability rating.
    - Moderate report ratings often associated with higher reliability scores.
  - **Advanced Editors (Rating 4):**
    - Most dependable, contributing to more reliable crash reports due to their expertise.
  - Matching rate of crash alerts is more influenced by report ratings than reliability scores.



Factors Affecting the Alert Matches

Logit Regression Model

- For identifying significant factors, a logit model is calibrated to the data.
- The significance of each report attribute (i.e., explanatory variable) is assessed by examining the p-values of the corresponding coefficients.
- Variables with p-values below 0.05 are considered significant, implying that they have a statistically significant impact on the validity of crash report by Waze users.
- Furthermore, the magnitude and direction of the coefficients indicate how much each attribute influences the (expected) validity of the report.

Variables	Predictors	Coefficient	Std. Error	Z-value	P-Value	Odds Ratio
	Intercept	-8.054	43.954	-0.183	0.855	0.000
Day of the week	Weekend	0.000	-	-	-	-
	Weekday	0.105	0.020	5.15	0.000	1.111
Time category	Nighttime non-peak	0.000	-	-	-	-
	AM Peak	0.395	0.033	12.128	0.000	1.484
	Morning non-peak	0.211	0.034	6.265	0.000	1.235
	Afternoon non-peak	0.263	0.028	9.494	0.000	1.300
	PM Peak	0.316	0.028	11.091	0.000	1.371
Road Type Category	Secondary Street	0.000	-	-	-	-
	Ramp	-0.577	0.038	-15.026	0.000	0.562
	Street	-0.733	0.047	-15.505	0.000	0.480
	Primary Street	0.377	0.026	14.746	0.000	1.458
	Freeway	-0.775	0.028	-27.944	0.000	0.461
Reliability	Reliability4	0.000	-	-	-	-
	Reliability5	5.918	43.954	0.135	0.893	371.755
	Reliability6	6.084	43.954	0.138	0.890	438.914
	Reliability7	6.040	43.954	0.137	0.891	419.930
	Reliability8	6.028	43.954	0.137	0.891	414.761
	Reliability9	6.013	43.954	0.137	0.891	408.698
	Reliability10	6.017	43.954	0.137	0.891	410.502
Report rating	Report Rating1	0.000	-	-	-	-
	Report Rating2	0.105	0.041	2.563	0.010	1.110
	Report Rating3	0.276	0.032	8.631	0.000	1.317
	Report Rating4	0.328	0.029	11.21	0.000	1.388
	Report Rating5	0.347	0.032	11.011	0.000	1.415
	Report Rating6	0.295	0.034	8.745	0.000	1.343
Confidence	Confidence0	0.000	-	-	-	-
	Confidence1	0.102	0.069	1.475	0.140	1.107
	Confidence2	0.212	0.173	1.229	0.219	1.237
	Confidence3	0.477	0.246	1.941	0.052	1.612
	Confidence4	0.224	0.351	0.64	0.522	1.252
	Confidence5	0.474	0.489	0.968	0.333	1.606
Crash Type	No Subtype	0.000	-	-	-	-
	Minor	-0.073	0.020	-3.727	0.000	0.929
	Major	0.257	0.021	12.357	0.000	1.293

Factors Affecting the Alert Matches

1. Day of the Week

Predictors:	Weekday
Coefficient:	0.105
Odds Ratio:	1.111
Effect:	Positive

- Alerts are 11.1% more likely to match in Weekday than Weekends.
- **Explanation:** Higher traffic volumes and consistent road usage during weekdays increase alert matching rate.

2. Time of Day

Predictors:	AM peak, Morning and Afternoon non-peak, PM peak
Coefficient:	0.395, 0.211, 0.263, 0.316
Odds Ratio:	1.484, 1.235, 1.300, 1.371
Effect:	Positive

- Alerts are more likely to match as follows: AM Peak: +48.4%, PM Peak: +37.1%, Morning Non-peak: +23.5%, Afternoon Non-peak: +30%.
- **Explanation:** Peak hours lead to higher engagement and more accurate reporting due to increased traffic and driver attentiveness.

3. Road Type

Predictors:	Ramp, street, primary street, freeway
Coefficient:	-0.577, -0.733, 0.377, -0.775
Odds Ratio:	0.562, 0.480, 1.458, 0.462
Effect:	Mix

- Alerts on primary streets are 45.8% more likely to match while on freeways they are 53.9% less likely to match.
- **Explanation:** Higher speeds on freeways make it harder for drivers to generate crash alerts.

4. Report Rating

Predictors:	2,3,4,5,6
Coefficient:	0.105, 0.276, 0.328, 0.347, 0.295
Odds Ratio:	1.110, 1.317, 1.388, 1.415, 1.343
Effect:	Positive

- Higher report ratings increase matching likelihood: Rating 2: +11.0%, Rating 5: +41.5%
- **Explanation:** Higher ratings indicate more reliable alerts from experienced users.

5. Crash Type

Predictors:	Minor and Major crash
Coefficient:	-0.073, 0.257
Odds Ratio:	0.929, 1.293
Effect:	Positive with higher severity level

- Alerts for major crashes are 29.3% more likely to match While for minor crashes, they are 7.1% less likely to match.
- **Explanation:** Major crashes cause more disruption, making drivers easier to detect and generate alerts.

Conclusions

- Higher likelihood of valid reports:
  - Weekdays
  - Peak hours
  - Primary streets,
  - Higher report ratings, and
  - Major crashes
- User engagement is higher during peak hours due to increased traffic and greater impact of crashes..
- Familiarity with street conditions and lower speeds improve alert matching.
- In contrast, higher speeds make real-time reporting more difficult, especially for minor incidents.
- Major crashes attract more attention, leading to higher matching rates.
- Higher report ratings improve reliability of reports

Future Research

- **Explore Seasonality:** Analyze temporal trends and their effect on Waze alert validity.
- **Additional Variables:** Integration of additional variables, such as weather conditions or user demographics, to further refine predictive models and improve “verified” real-time traffic incident reporting.
- **Machine Learning Integration:** Dynamically adjust the weight of user ratings based on historical matching rate could further optimize the reliability of crowdsourced traffic alerts.

Questions & Discussion