Non-Technical Presentation: Empowering Traffic Management with Predictive Insights

This presentation outlines our project to develop a **predictive tool** for understanding and preventing car crashes in Chicago, designed specifically for business stakeholders in the Traffic Management Department.

Slide 1: Overview - Enhancing Road Safety with Predictive Power

- **Our Mission:** To equip the City of Chicago's Traffic Management Department with a **smart predictive tool** that forecasts *why* car crashes occur.
- **The Vision:** Shift from reacting to accidents to **proactively preventing them**, making Chicago's roads significantly safer for everyone.
- **Core Value:** Deliver **actionable, data-driven insights** to guide strategic decisions and targeted safety initiatives.

Slide 2: Business & Data Understanding - The Need for a Predictive Edge

- **The Challenge:** Car crashes impose substantial human and economic costs. Current methods often address symptoms. How can we pinpoint the *root causes* to prevent future incidents?
- **The Solution:** A **predictive tool** allows us to anticipate and understand the underlying reasons for crashes. This enables more efficient resource allocation and impactful interventions.
- **Our Foundation:** We leverage comprehensive historical crash data from the Chicago Data Portal, detailing:
    - Crash circumstances (time, conditions).
    - People involved (injuries, alcohol tests).
- **What We Predict:** The tool's core function is to predict the **"Primary Contributory Cause"** of each crash.

Slide 3: Modeling - How Our Predictive Tool Learns

- **Smart Data Preparation:** We meticulously prepared the data, combining crash and individual information. We engineered new features like "rush hour" and "weekend" and grouped rare crash causes into an "OTHER" category to make the problem manageable.
- **Automated Preprocessing:** We built an automated system to consistently clean and prepare all data, ensuring the predictive tool operates on reliable information.

- **Choosing the Best Brain:** After exploring several machine learning "brains" (models), our **LightGBM Classifier** emerged as the most powerful and efficient. It's designed to learn complex patterns for accurate predictions.
- **Handling Rare Events:** Since some crash causes are uncommon, we used **SMOTE** to balance our training data. This ensures our predictive tool doesn't ignore rare, but critical, crash scenarios.

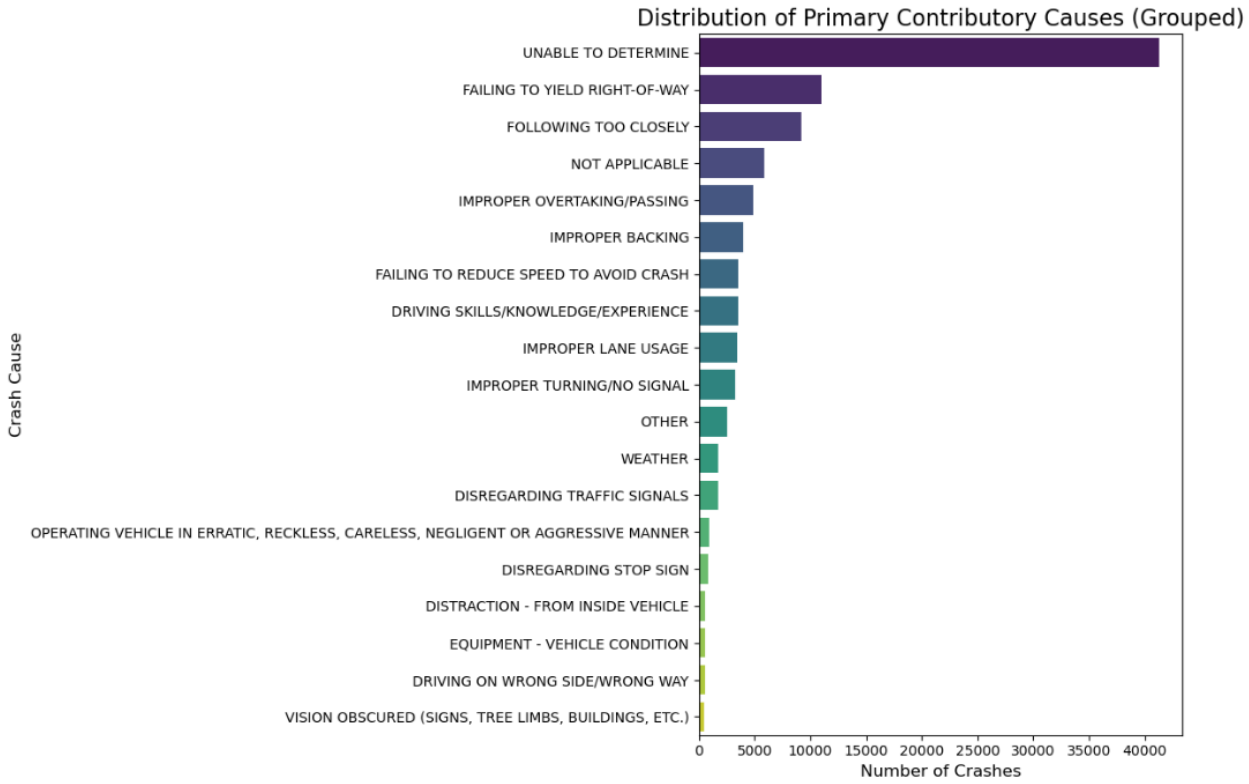
Slide 4: Evaluation - Assessing Our Predictive Tool's Capability

- **Overall Predictive Accuracy:** Our LightGBM model achieved a test accuracy of approximately **41.4%**. While this might seem modest, predicting one out of 19 complex and often overlapping crash causes is a significant step.
- **Why LightGBM Excels:** It consistently provides more accurate predictions and generalizes better to new, unseen crash data compared to simpler models.
- **Understanding Predictions (Confusion Matrix):** While the tool predicts well overall, the Confusion Matrix helps us understand its specific strengths and areas for improvement:
    - It's **effective at predicting common causes** like "UNABLE TO DETERMINE" and "FAILING TO REDUCE SPEED TO AVOID CRASH."
    - It faces challenges with **predicting very rare causes**, sometimes confusing them with more frequent ones. This is a common hurdle in real-world data.
- **Key Drivers of Prediction (Feature Importance):** The tool reveals what information is most critical for its predictions:
    - The **hour of the crash** and the **specific police beat/location** are the strongest predictors.
    - Other significant factors include the **day of the week, month, year, posted speed limit**, and the **initial type of collision** (e.g., rear-end). This tells us *what data points* are most valuable for predicting crash causes.

Slide 5: Visual Insights - Data-Driven Justification for Our Predictive Findings

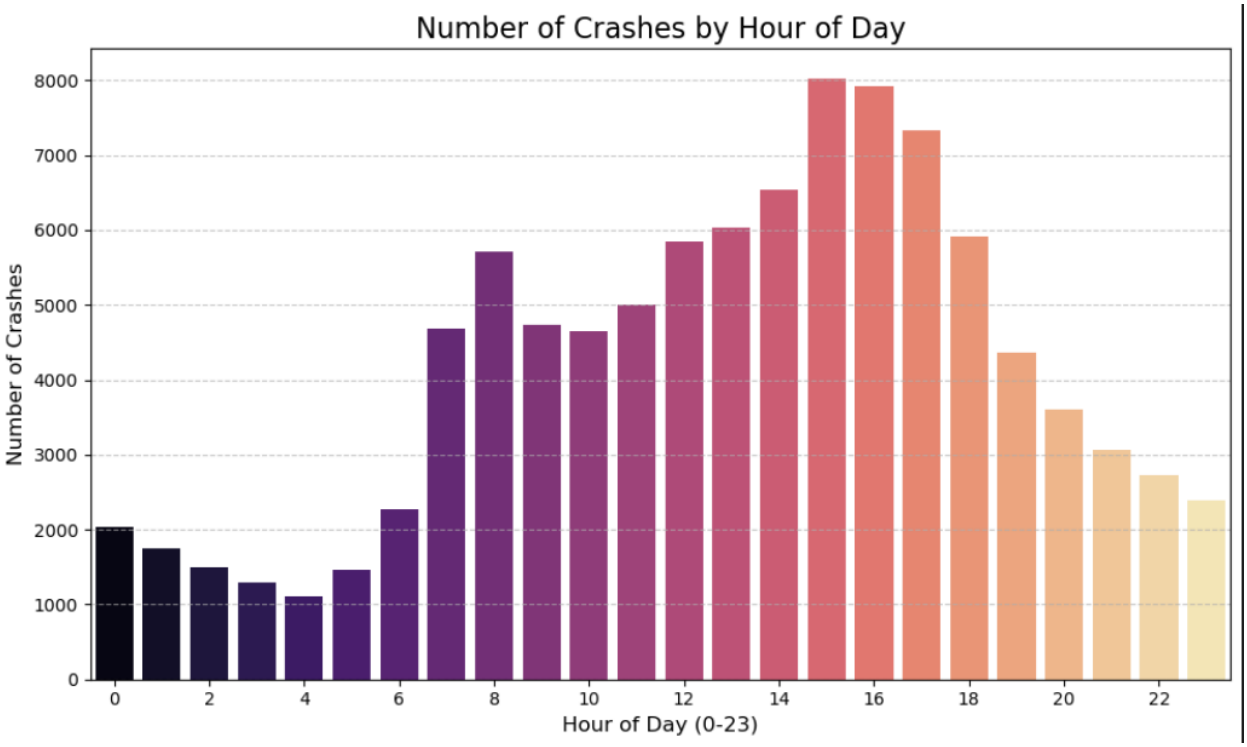
These visuals provide the evidence and context for our predictive tool's capabilities and the insights it generates, directly supporting our findings and recommendations for the Traffic Management Department.

- **Visual 1: Distribution of Primary Contributory Causes (Grouped)**



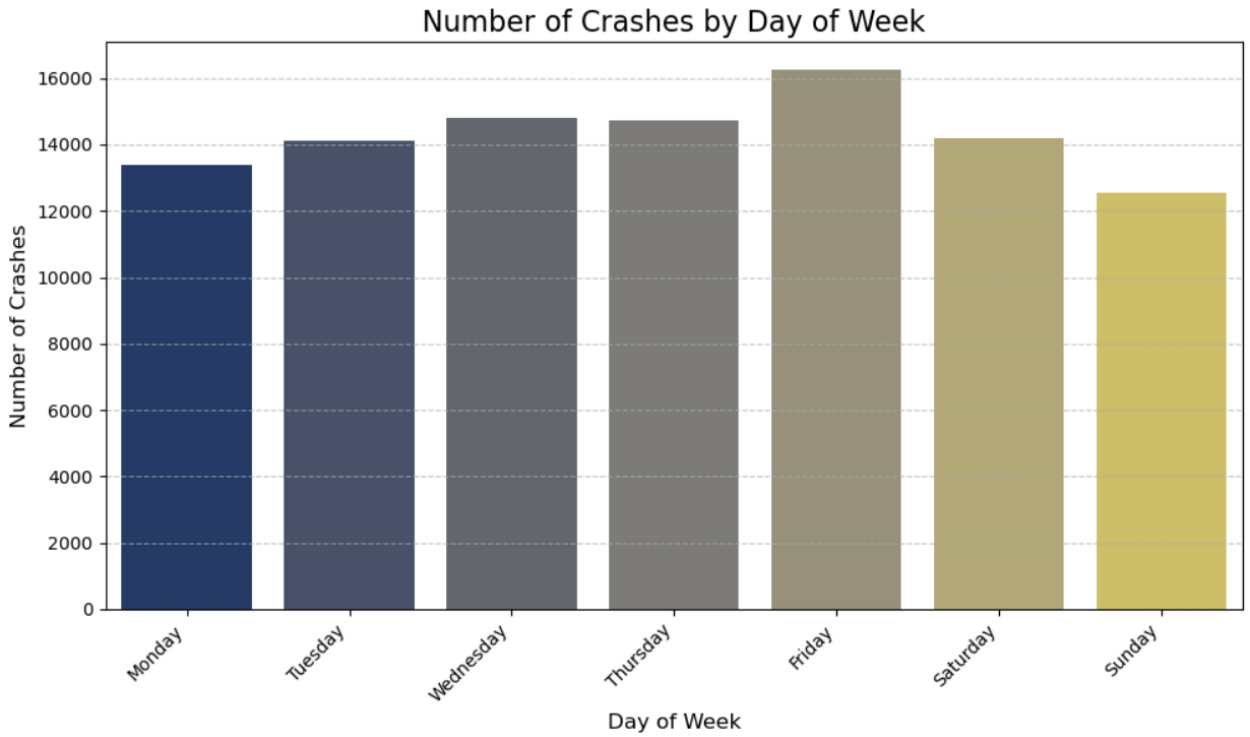
- *Explanation:* This chart shows the most frequent reasons why crashes occur, with less common causes grouped into 'OTHER'. It helps the predictive tool focus on the most impactful areas for safety improvements by highlighting which causes are most prevalent.

- **Visual 2: Number of Crashes by Hour of Day**



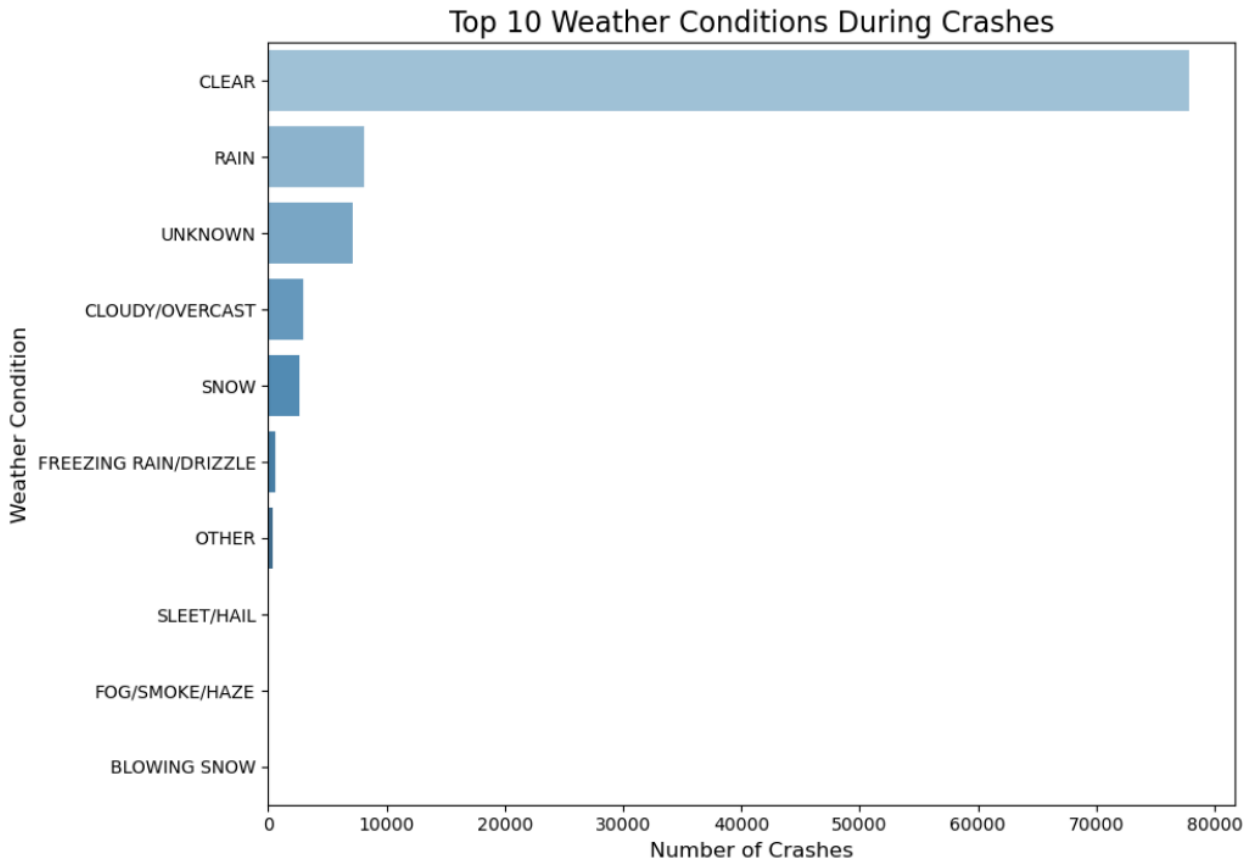
- *Explanation:* This chart shows when crashes happen most frequently throughout the day. Our predictive tool uses this insight to identify high-risk hours, enabling the Traffic Management Department to deploy resources more effectively during peak times.

- **Visual 3: Number of Crashes by Day of Week**



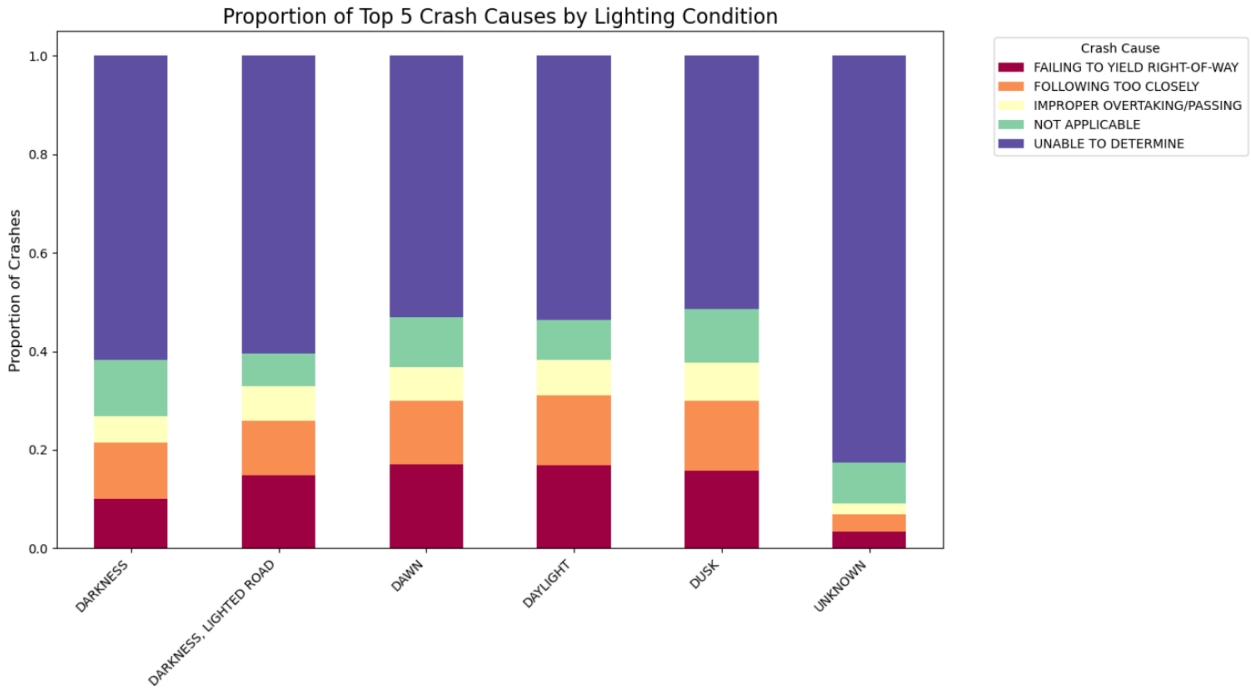
- *Explanation:* This chart breaks down crashes by the day of the week. This information is crucial for our predictive tool to identify specific days that are typically safer or have more incidents, guiding weekly traffic management strategies.

- **Visual 4: Top 10 Weather Conditions During Crashes**



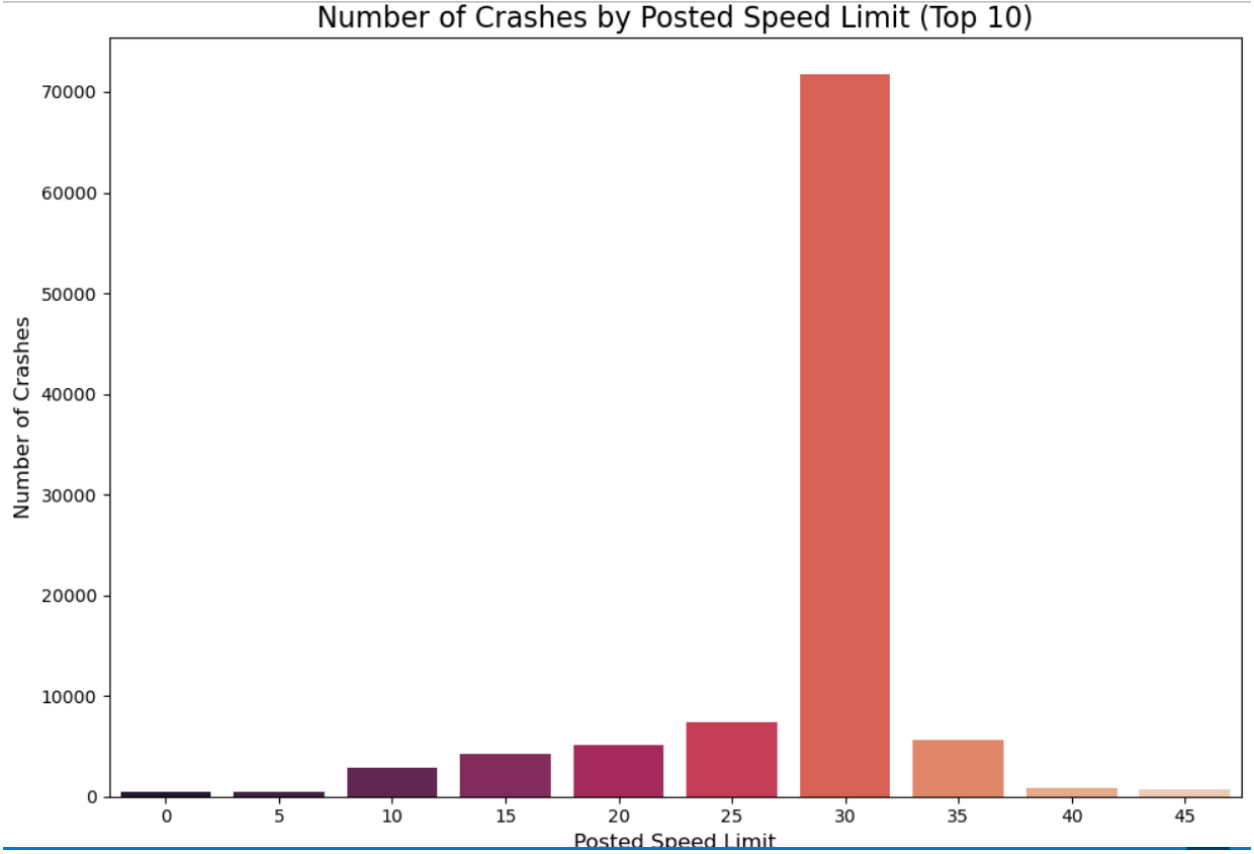
- *Explanation:* This chart shows the most common weather conditions present during crashes. Our predictive tool considers these environmental factors to anticipate crash causes, allowing for weather-aware safety advisories.

- **Visual 5: Proportion of Top 5 Crash Causes by Lighting Condition**



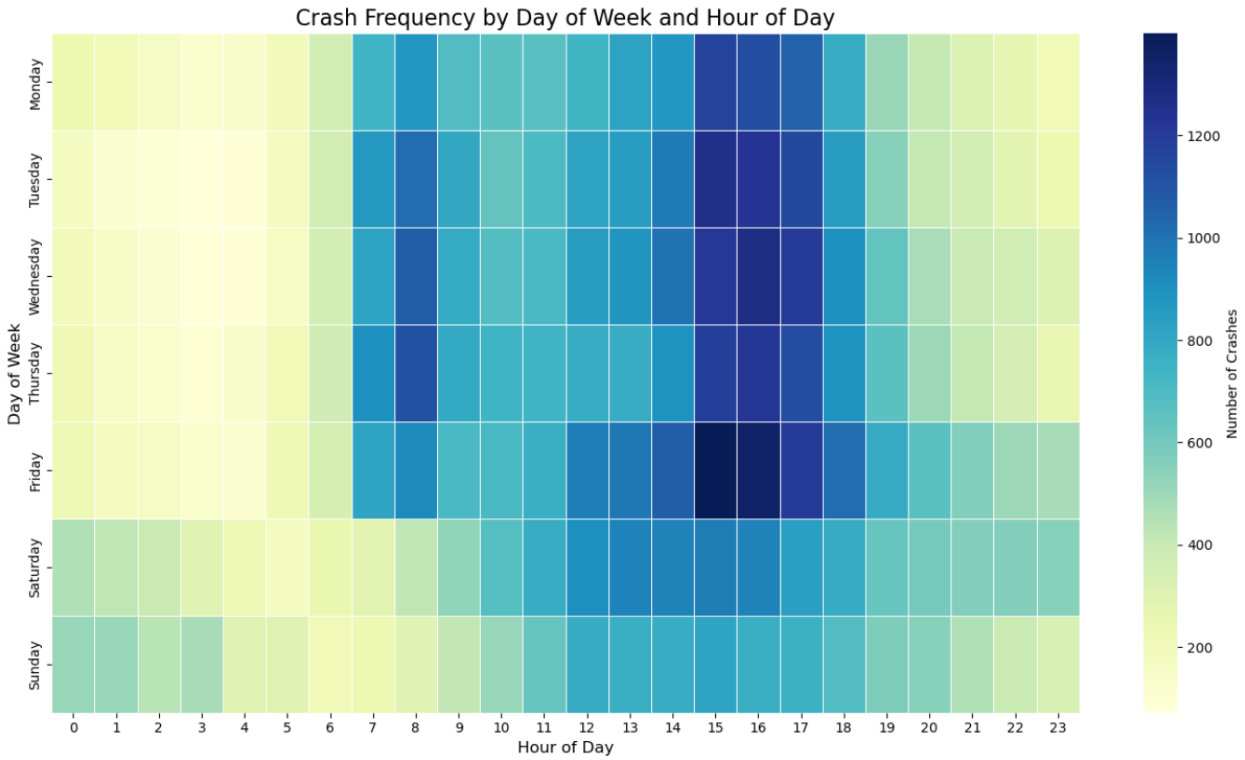
- *Explanation:* This chart shows how different lighting conditions influence the types of crash causes. Our predictive tool leverages this to understand specific risks, for instance, if certain causes are more prevalent at night or during the day.

- **Visual 6: Number of Crashes by Posted Speed Limit (Top 10)**



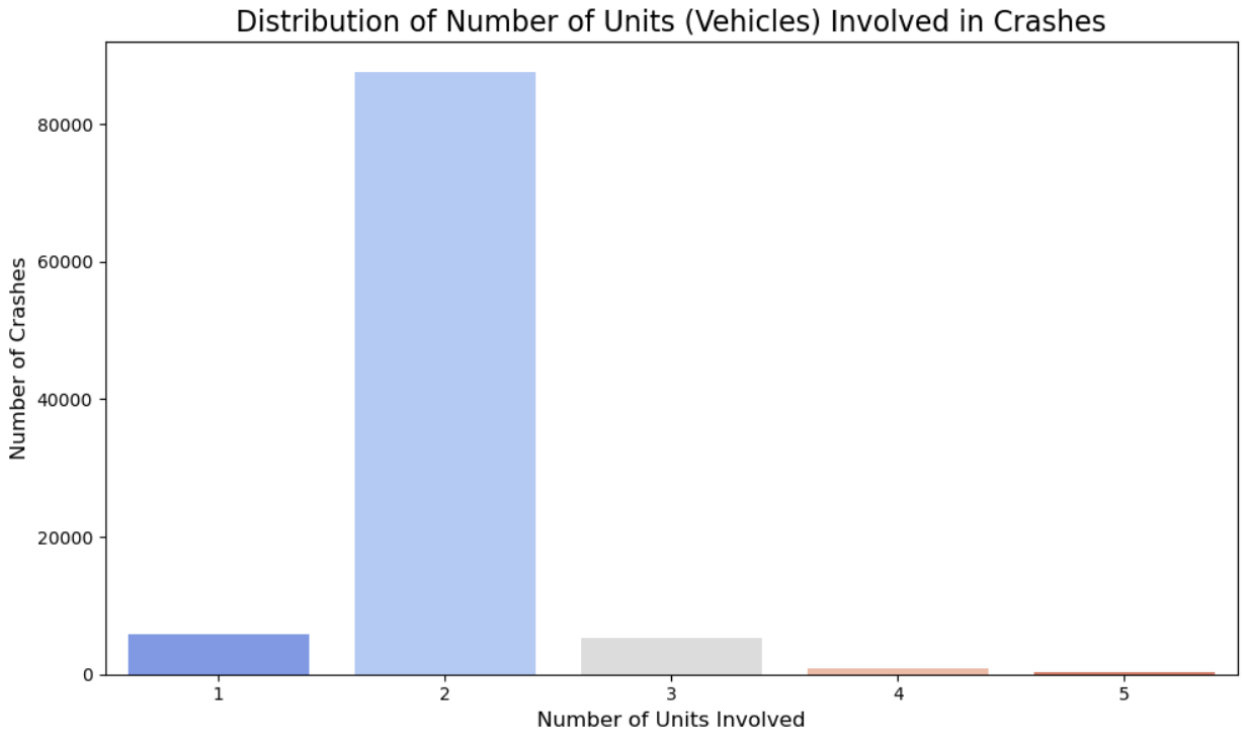
- *Explanation:* This chart shows which posted speed limits are most frequently associated with crashes. The predictive tool incorporates this to identify if certain speed zones are more problematic, informing targeted enforcement or engineering efforts.

- **Visual 7: Crash Frequency by Day of Week and Hour of Day (Heatmap)**



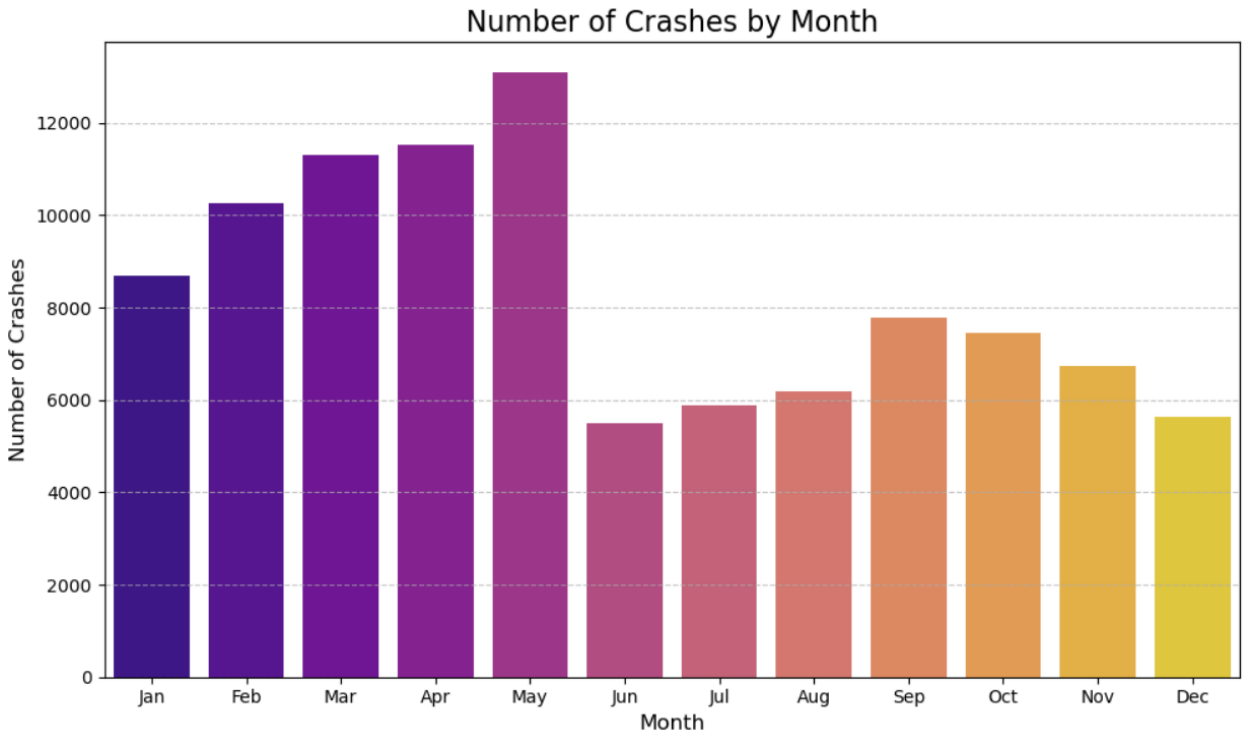
- *Explanation:* This heatmap illustrates when crashes are most frequent, showing patterns across different hours and days of the week. Darker colors indicate more crashes. This visual directly supports the importance of 'CRASH_HOUR' and 'CRASH_DAY_OF_WEEK' as top features for our predictive tool, guiding precise intervention timing.

- **Visual 8: Distribution of Number of Units (Vehicles) Involved in Crashes**



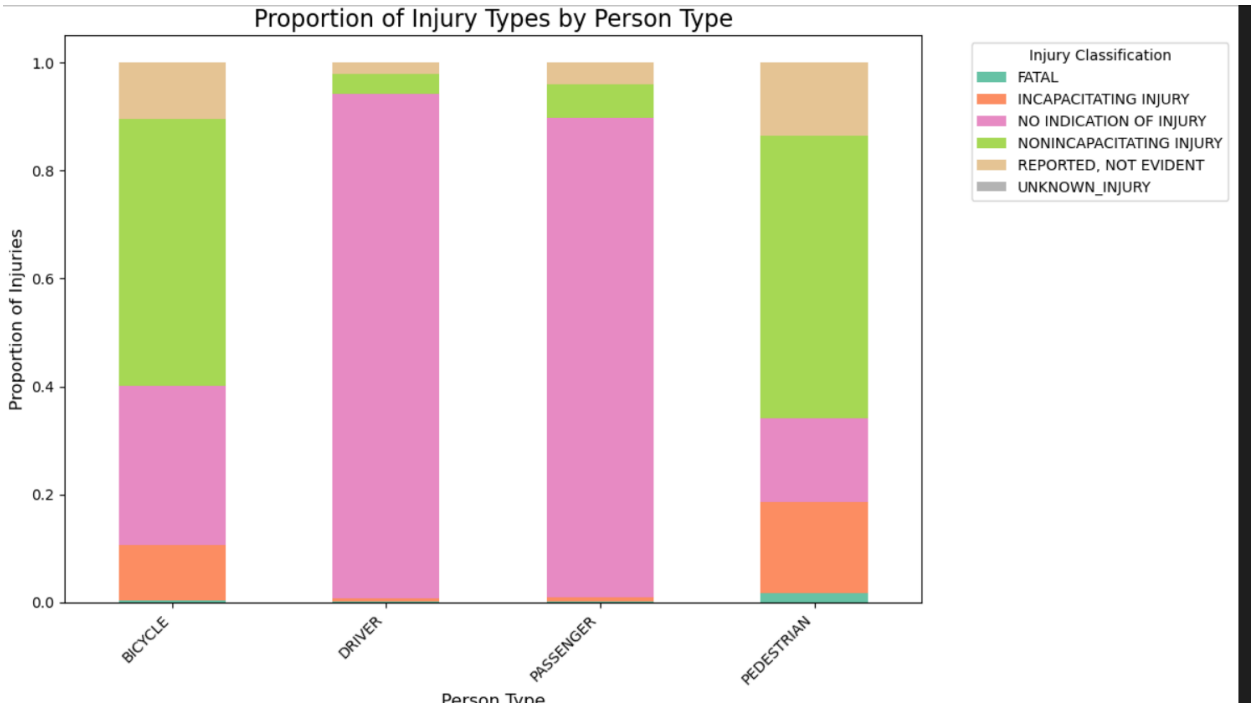
- *Explanation:* This chart shows how many vehicles are typically involved in crashes. Our predictive tool uses this information to understand the complexity of incidents, whether they are single-vehicle, two-vehicle, or more complex collisions.

- **Visual 9: Number of Crashes by Month**



- *Explanation:* This chart illustrates the monthly distribution of crashes, helping to identify any seasonal patterns or months with higher crash frequencies. This insight is integrated into our predictive tool to account for seasonal variations in crash causes.

- **Visual 10: Proportion of Injury Types by Person Type**



- *Explanation:* This chart shows the proportion of different injury types for various individuals involved in crashes (drivers, pedestrians, cyclists). While not a direct input for predicting *causes*, it highlights the human impact of crashes and underscores the importance of our predictive tool in preventing these outcomes.

Slide 6: Recommendations - Actionable Insights from Our Predictive Tool

Our predictive tool enables the Traffic Management Department to implement **proactive, targeted interventions**:

- **Focused Resource Allocation:** Use the tool's predictions and feature importances to direct resources to specific high-risk hours, days, and geographical areas (police beats).
- **Data-Driven Campaigns:** Design public safety campaigns that directly address the most common and predictable crash causes, leveraging insights from the tool's key drivers.
- **Strategic Infrastructure Planning:** Identify road segments or intersections where specific predictable causes are prevalent, informing targeted infrastructure improvements.

Slide 7: Next Steps - Evolving Our Predictive Capability

To further enhance this predictive tool:

- **Deeper Optimization:** Fine-tune the model's internal settings to improve prediction accuracy and generalization.

- **Smarter Data Handling:** Explore advanced techniques to better predict rare crash causes, making the tool more comprehensive.
- **Feature Expansion:** Look for new, creative ways to combine existing data or integrate external information (like real-time traffic data) to give the tool even richer information for prediction.
- **Prediction Validation:** Conduct detailed analysis of the tool's mispredictions to uncover new patterns and refine its learning.

Slide 8: Thank You

- **Questions & Discussion**