



Driven to Predict: **Understanding Why Chicago** **Crashes Happen.**

BY: VICKER IVY

VICTOR ONGAKI

DAISY KERUBO THOMAS

ROSE MATOKE

FELIX MUSAU

OVERVIEW

This presentation focuses on predicting the primary causes of vehicle crashes in Chicago.

The objective is to help the City of Chicago and vehicle safety board understand which factors most contribute to accidents, enabling data-driven prevention strategies.

Goals of the analysis:

- Build a model that can predict the likelihood of accidents based on features.
- Understand key factors behind car accidents.
- Predict the primary contributory cause.
- Provide actionable insights for prevention.





BUSINESS

UNDERSTANDING

- **Challenge:** The rapidly increasing number of reported cases of accidents in the city of Chicago, has raised concerns to the Vehicle Safety Board .
- **Problem Statement:** Chicago is experiencing an increase in the number of crashes over time. Can we predict the main cause to minimize this number.
- **Goal:** The primary interest is reducing traffic accidents in the city of Chicago.

RESEARCH

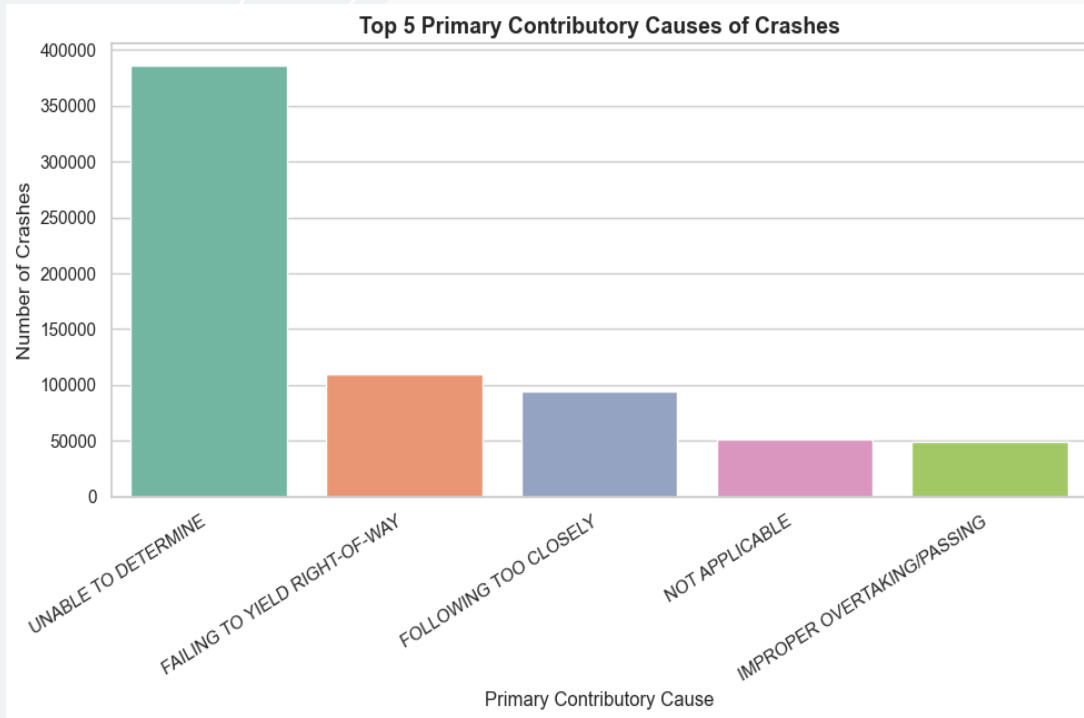
QUESTIONS

- What factors contribute to road accidents?
- How do crash frequencies vary across time (hour of day, day of week, month, or season)?
- What are the most dangerous locations?
- How does speeding correlate with crash severity?
- What are the effects of natural conditions to accidents?



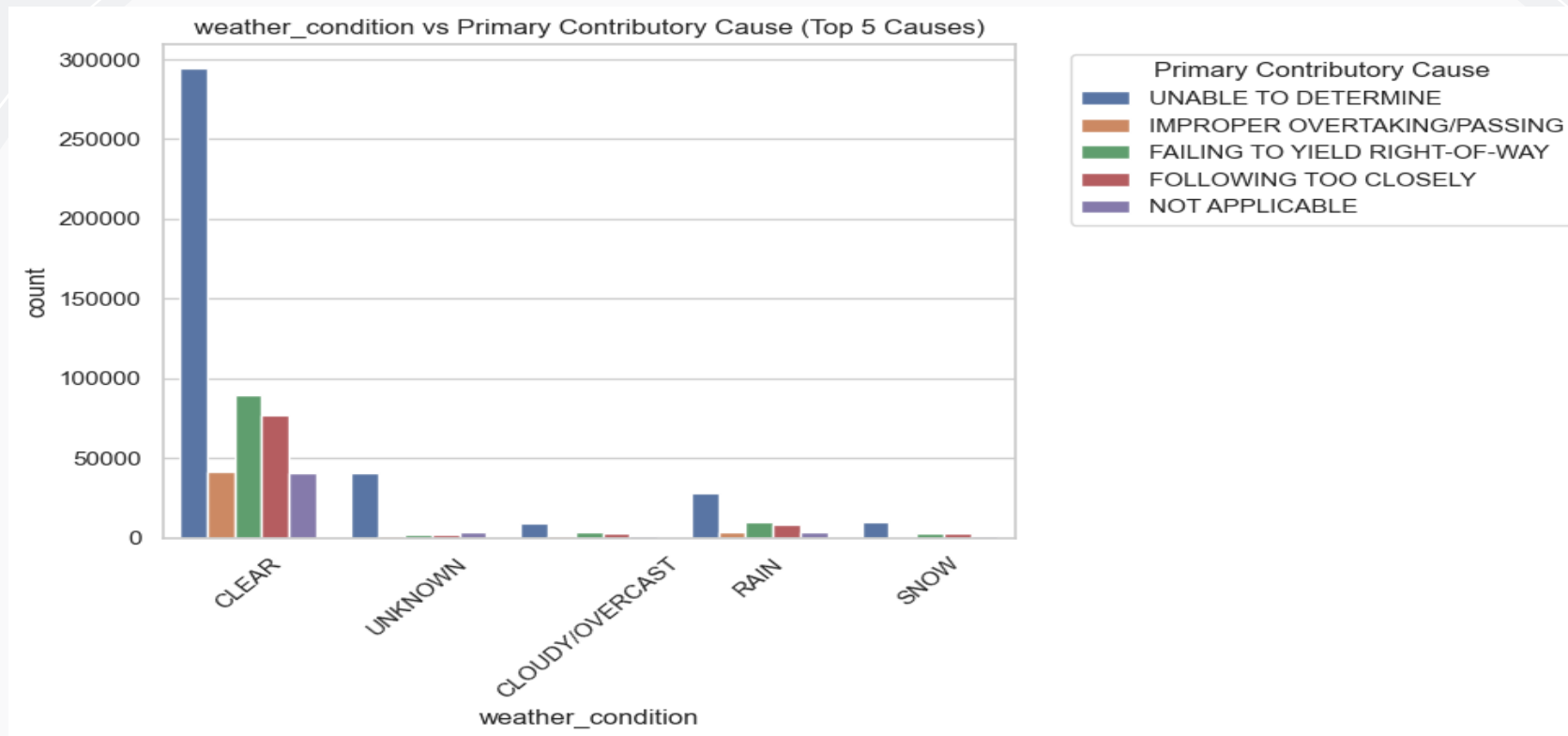
DATA UNDERSTANDING

- **Data source:** City of Chicago Dataset
 - Traffic Crashes Data
 - Vehicle Data
 - Driver/Passenger Data
- **Target Variable:** Cause Groups(Primary Contributory Cause)
- **Key Features:** weather condition, cause groups, lighting condition, location.
- Analysis methods used in Python:
 - Exploratory Data Analysis
 - Modelling
 - Feature Engineering
 - Visualization



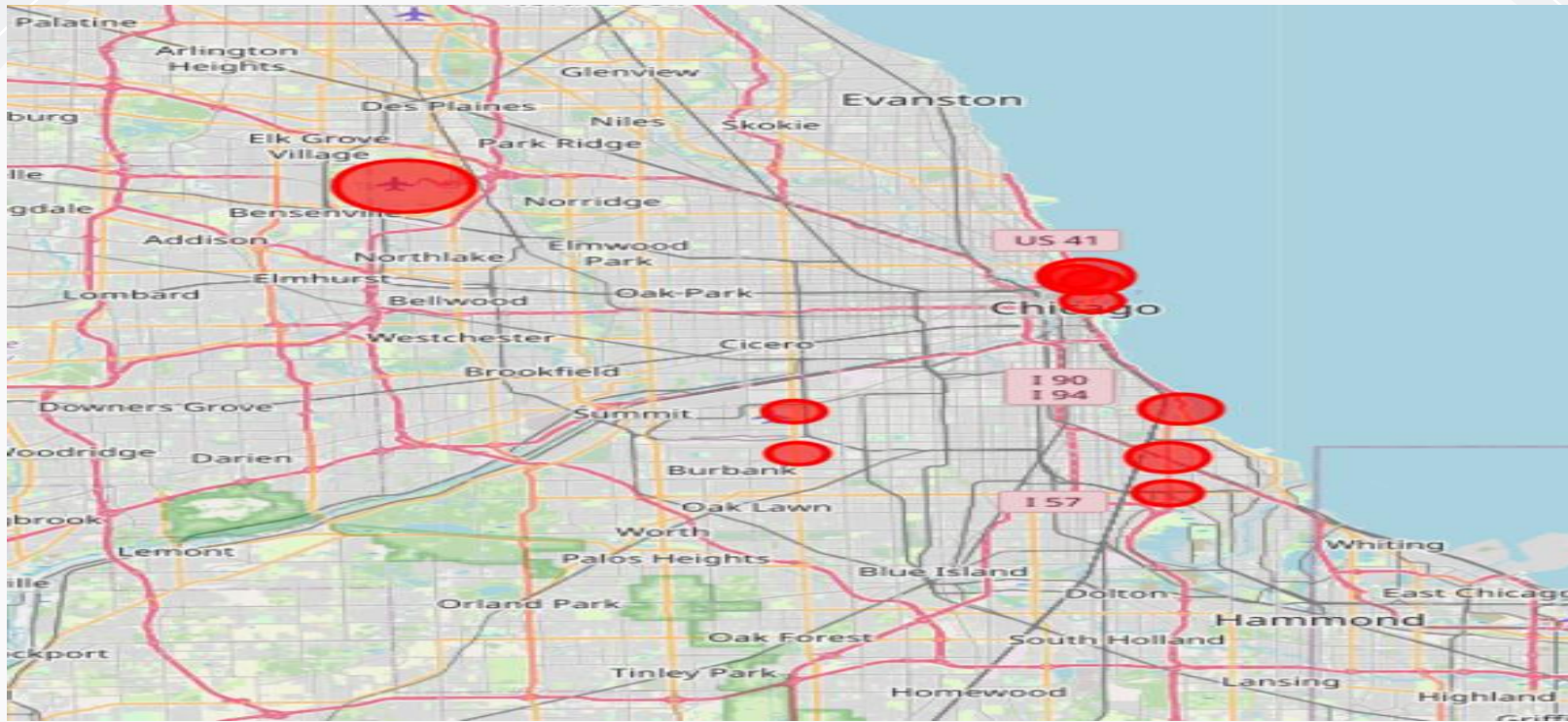
VISUALIZATION 1: Primary Contributory Causes Distribution

- The images above display a **count plot** and a **world cloud** which both highlight the major primary contributory causes of crashes.
- As we can see in the count plot, more than **350,000 crashes were left undetermined** this may suggest data limitations, this is also the largest in word cloud.
- The data is heavily skewed and does not give us clear reasons for the cause of the most common crashes.



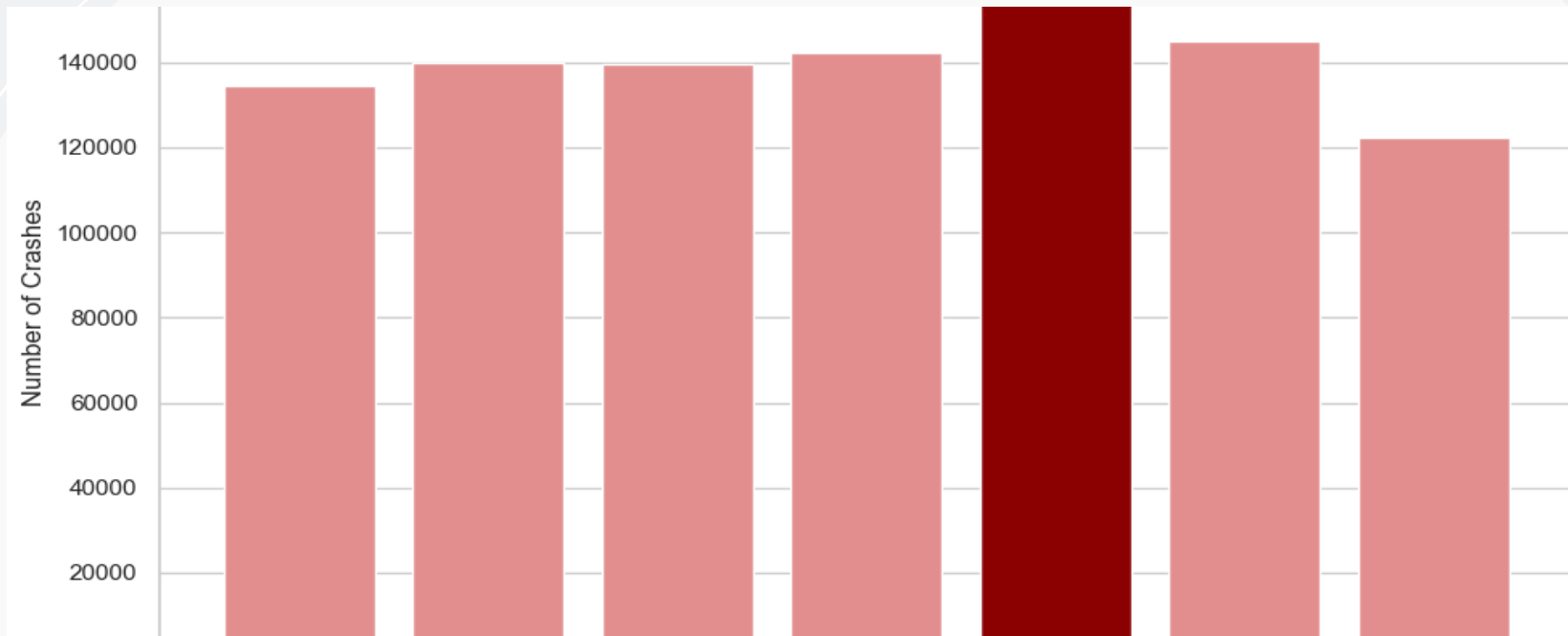
VISUALIZATION 3: Weather Condition vs Top 5 Contributory Causes

- This graph shows that **most crashes happen in good weather**, likely due to **higher traffic volume**.
- It challenges assumptions that **bad weather is the primary risk factor** and highlights the importance of **driver behaviour**.
- **CLEAR** weather has the highest number of incidents, especially for **"UNABLE TO DETERMINE"**.



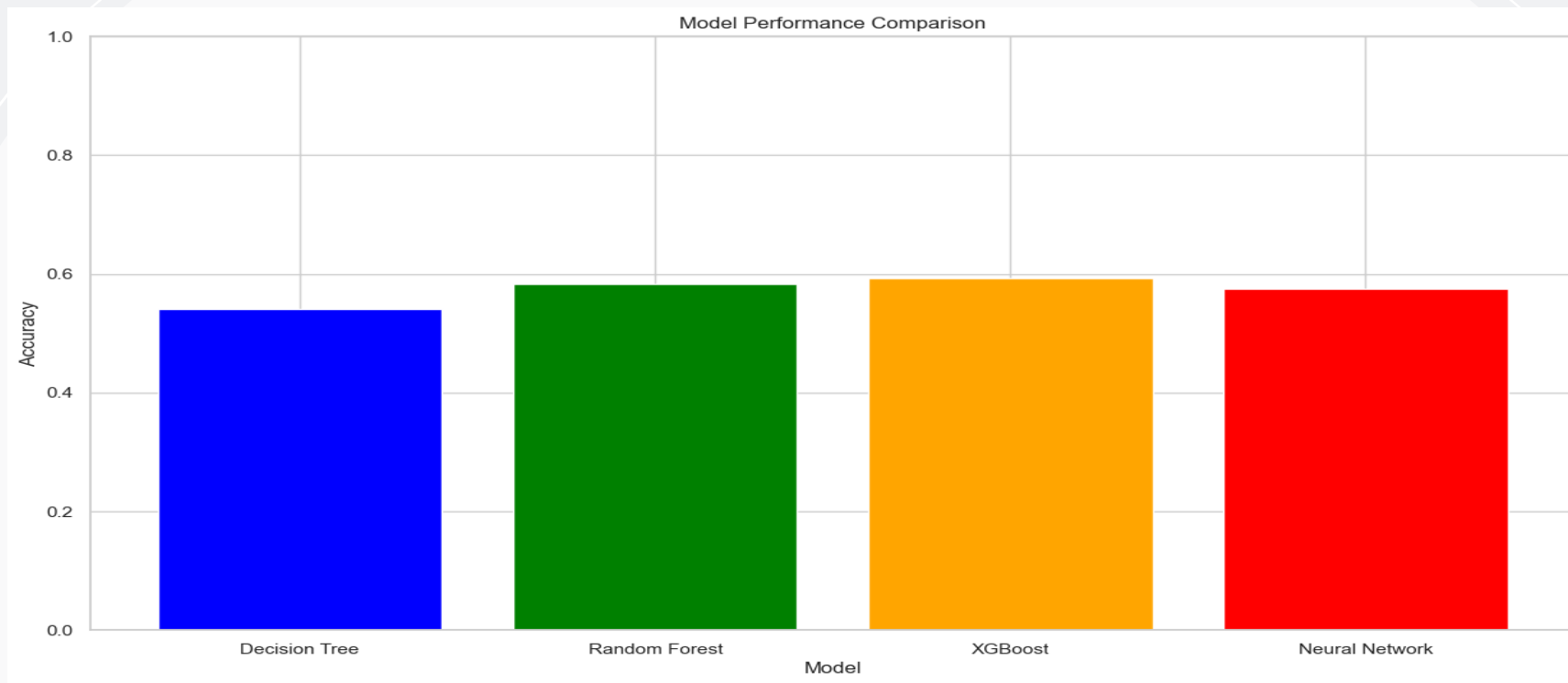
VISUALIZATION 2: Chicago Accidents Hotspots

- This map highlights the **top 10 crash hotspots** in the Chicago metropolitan area based on geographic coordinates and location names:
 1. **Highest crash count:** *O'Hare International Airport* tops the list with **1,610 crashes**, likely due to high traffic volume and complex road layouts.
 2. **Downtown clusters:** Locations like *East Lake Shore Drive* and *North Michigan Avenue* show elevated crash numbers (**884 and 366 crashes**) these are busy commercial and tourist zones.



VISUALIZATION 4: Crash Frequency by Day of the Week

- According to this bar plot the highest number of crashes occurs on **Friday** which are averagely **160,000 crashes**, indicating it as the peak day for traffic accidents during the week.
- This pattern may reflect increased travel, end-of-week fatigue, or social activities leading to higher traffic volumes and risk.



VISUALIZATION 5: Model Performance Comparison

- In this analysis, we evaluated several classification models to determine the best fit for predicting traffic crash contributory causes:
 1. **Top Performer:** XG Boost closely follows with an accuracy of 0.5932, showcasing resilience against overfitting and robust classification capabilities.
 2. **Close Contender:** The Neural Network achieved the highest accuracy(0.5745) among all models and demonstrated strong performance in classifying key categories.

RECOMMENDATIONS

➤ **Hyperparameter Tuning:**

Further refine the Neural Network to address overfitting and unlock additional performance gains.

➤ **Feature Engineering:** Explore new features, to capture more nuanced relationships between accident causes.

➤ **Continuous Learning:** As new data becomes available, retrain models regularly to maintain predictive relevance and adapt to changing traffic patterns





SAFE DRIVE SAVE LIFE

- ## NEXT STEPS
- The stakeholders should put focus more on areas such as traffic control devices, traffic way type.
 - Implement **satellite check-in and decentralized passenger drop-off** with dedicated, enforced lanes for transit and rideshare to reduce private vehicle congestion and queuing within the immediate area.
 - Implement clearer and mandatory protocols for crash reporting officers to minimize "**UNABLE TO DETERMINE**" as a primary cause. This is critical for future data-driven policy decisions.
 - Conduct a full safety engineering review of the **O'Hare International Airport** crash concourse cluster to identify and rectify any underlying road design that contribute to the extremely high crash frequency.

*THANK YOU
FOR YOUR ATTENTION*

