



# In-Depth Analysis of Chicago Traffic Data

Providing insights to  
improve road safety



Name: Jackson Weisner  
Email: [jweis4@uic.edu](mailto:jweis4@uic.edu)  
Git: [jweis4](https://github.com/jweis4)

Erica Hammond  
[ehammo2@uic.edu](mailto:ehammo2@uic.edu)  
Hammond-EC

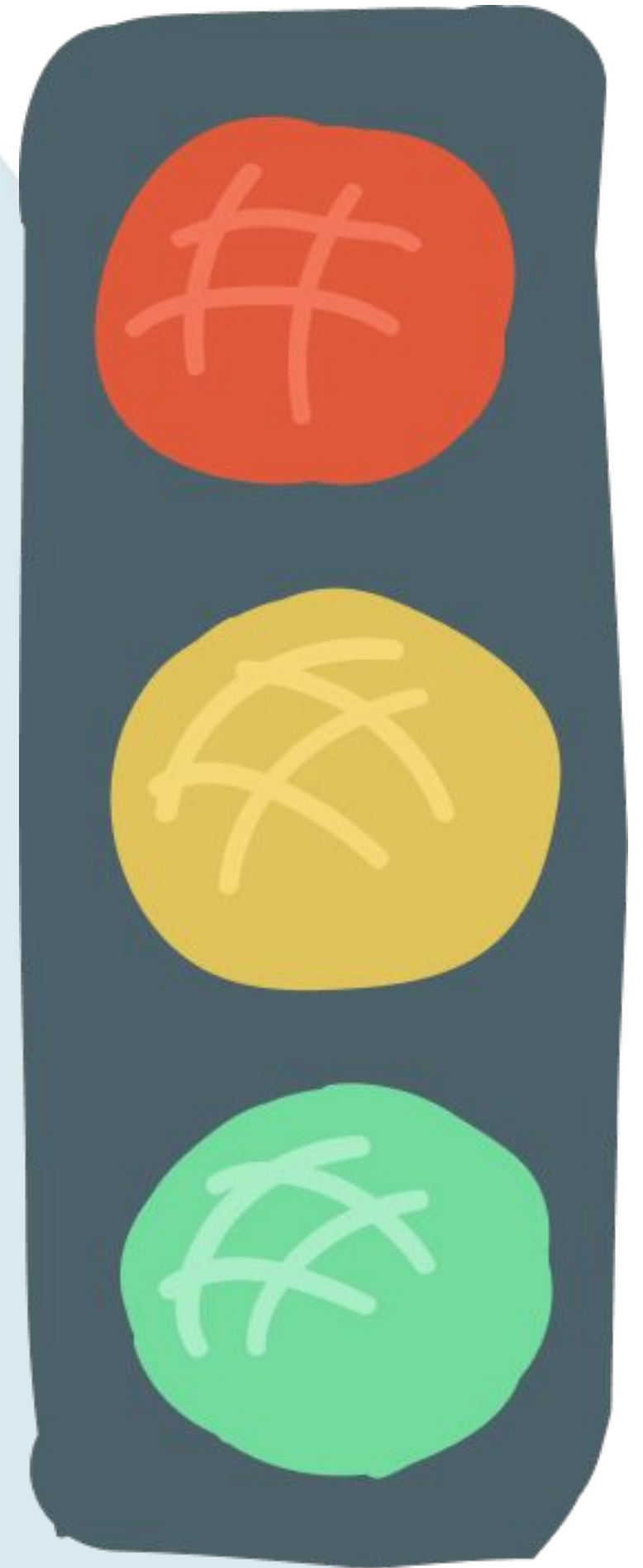
Junha Liu  
[jliu238@uic.edu](mailto:jliu238@uic.edu)  
[junhaliu](https://github.com/junhaliu)

Mohammed Riaz  
[mraiz5@uic.edu](mailto:mraiz5@uic.edu)  
[riazmb01](https://github.com/riazmb01)

Prit Patel  
[ppate425@uic.edu](mailto:ppate425@uic.edu)  
[pppate425](https://github.com/pppate425)

# Introduction

- **Our project is a comprehensive analysis of chicago traffic collisions**
- **We changed the scope to just Chicago, instead of three cities**
- **Our findings will help drivers know the risks of being on the road, and will help keep the community safe**



# Data

A blue traffic light with three lenses: red (top), yellow (middle), and green (bottom).

01

## **Traffic Crashes – Crashes**

General information about the crash

02

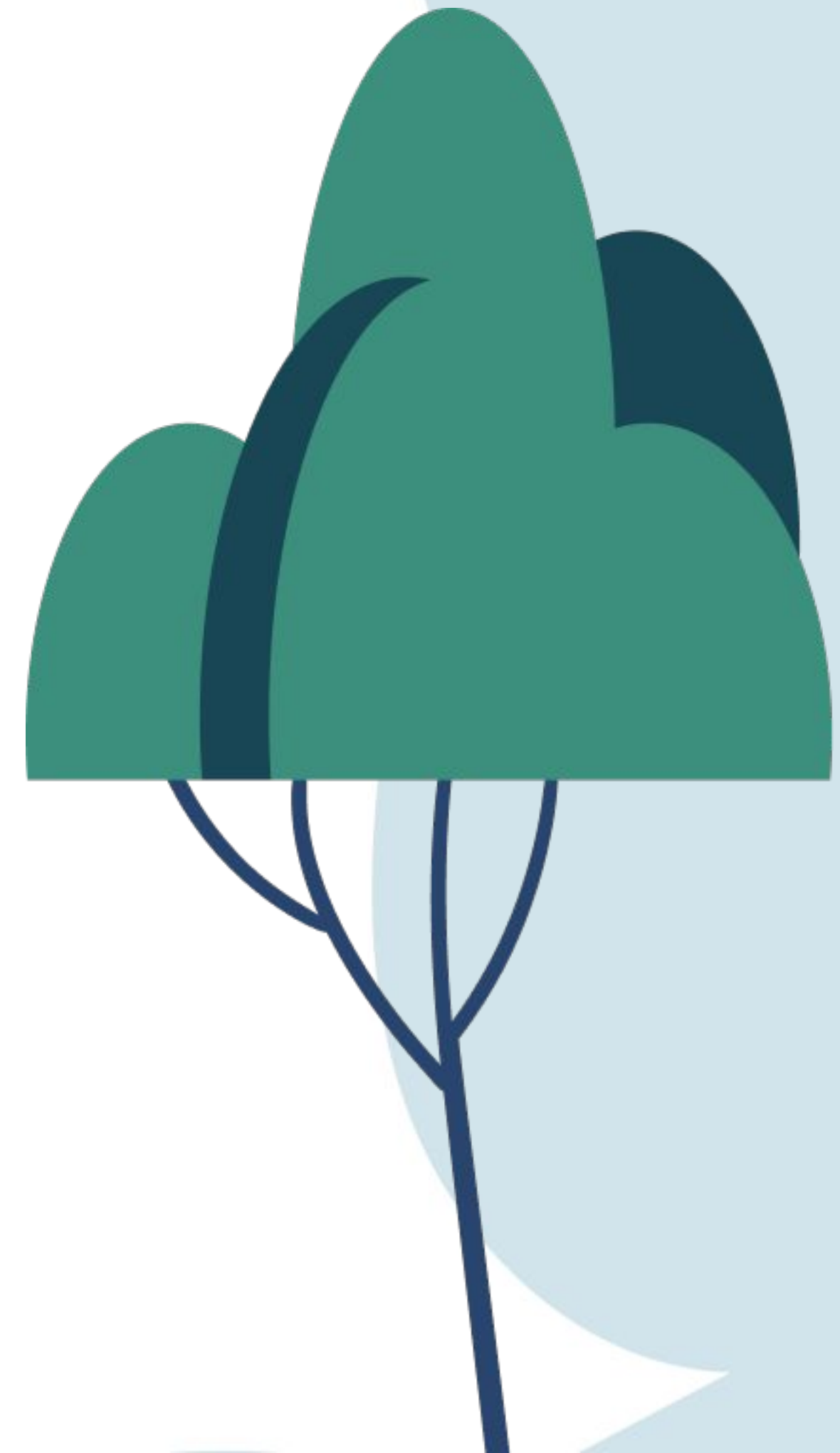
## **Traffic Crashes – People**

Details about the people involved in a crash

03

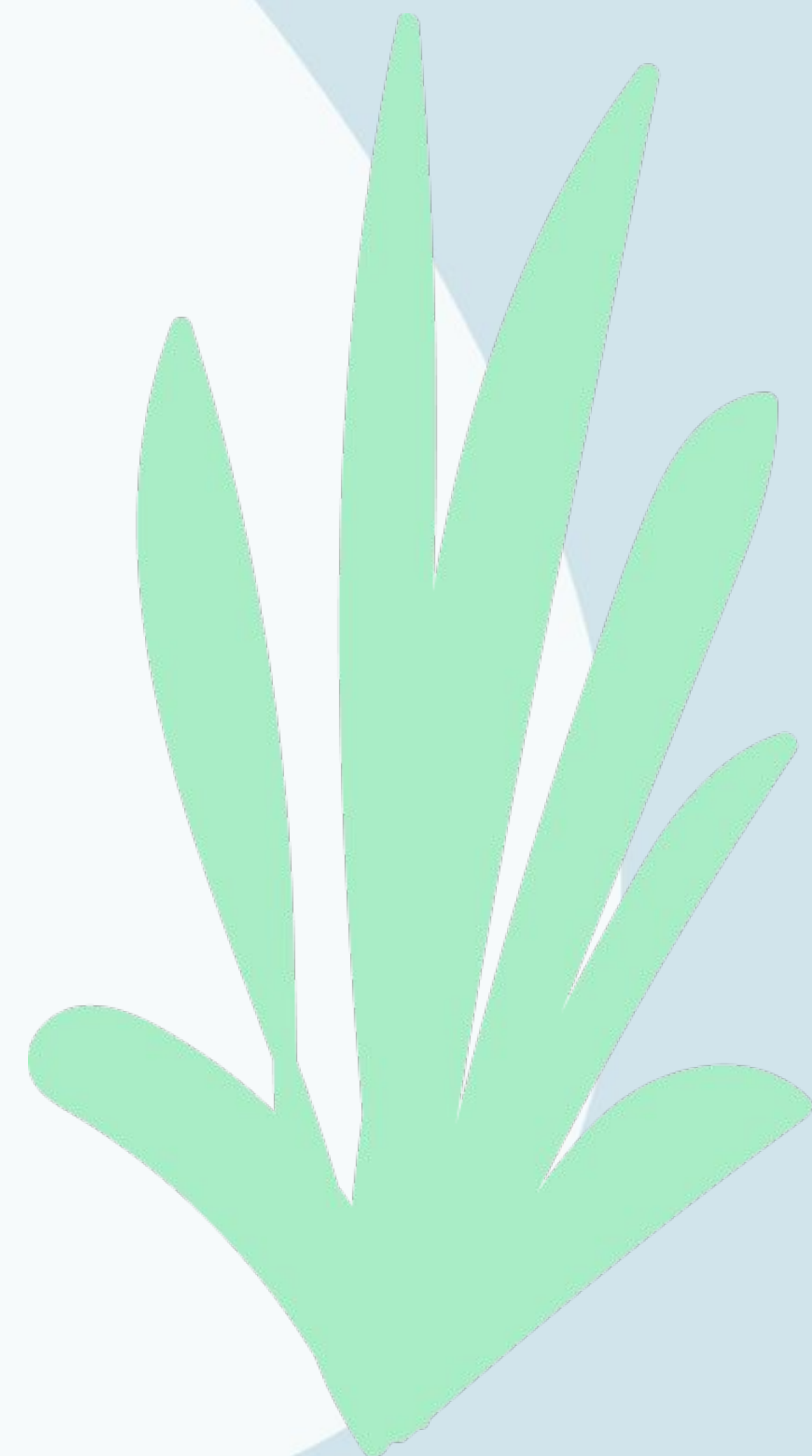
## **Traffic Crashes – Vehicles**

Information about vehicles and damage



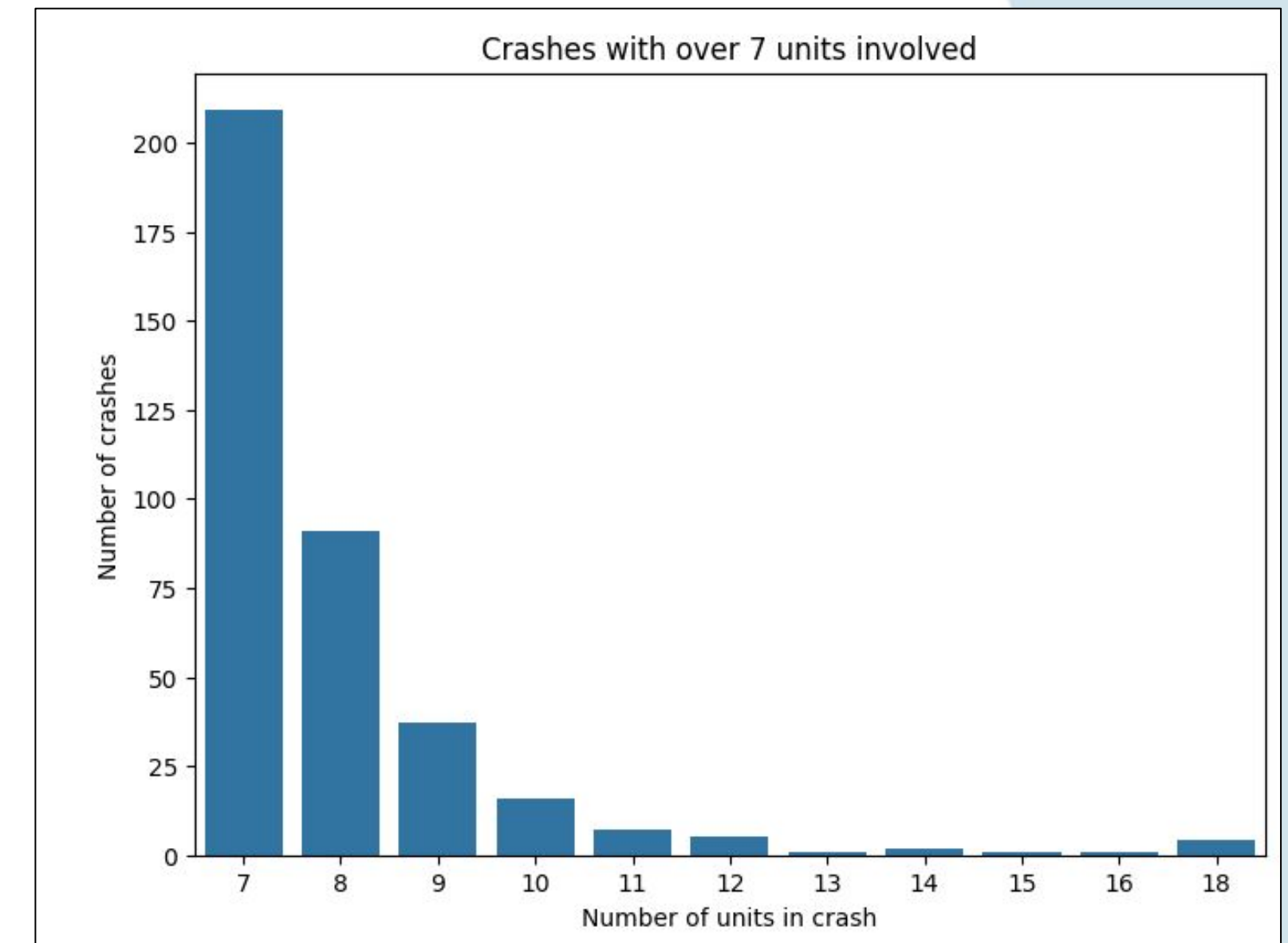
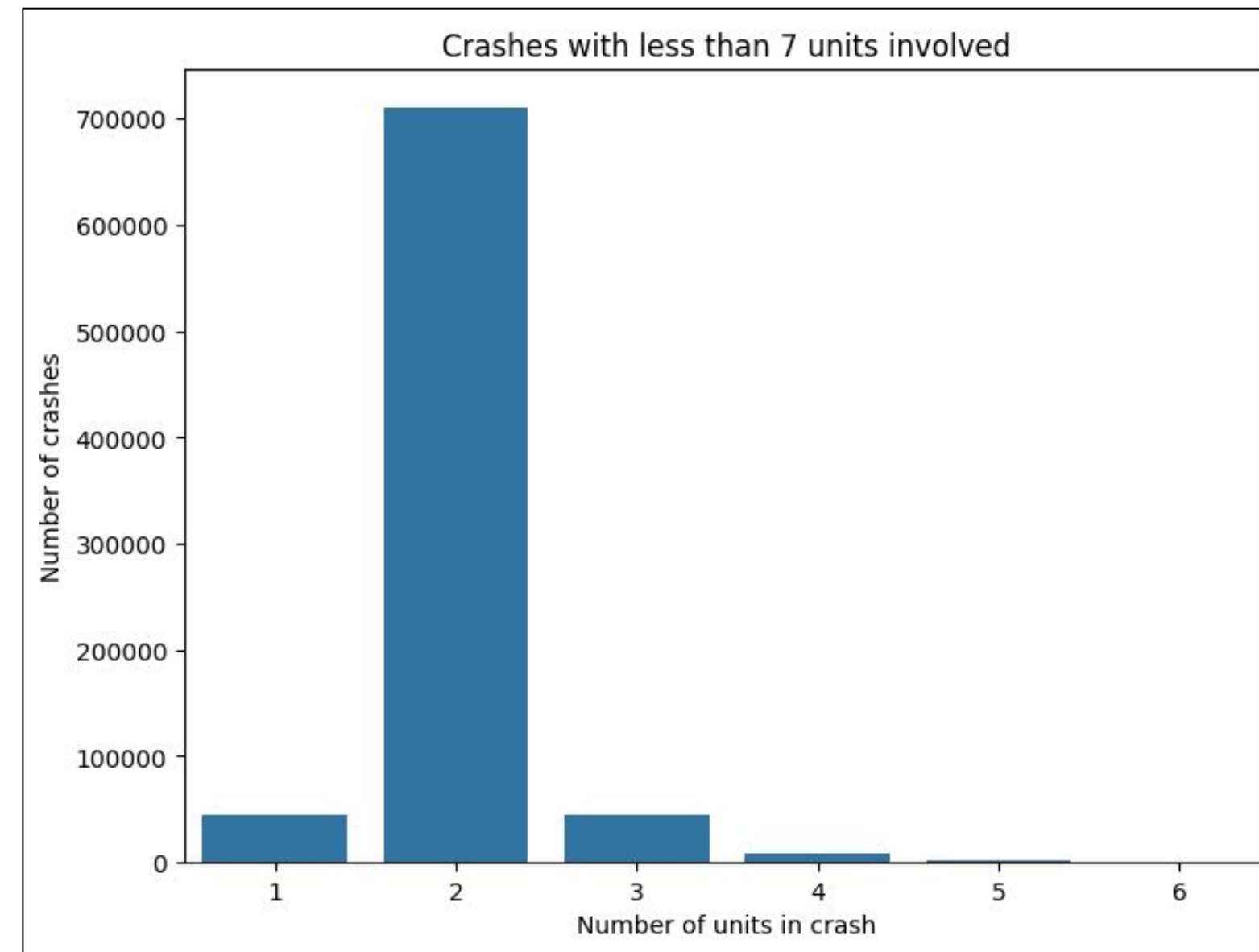
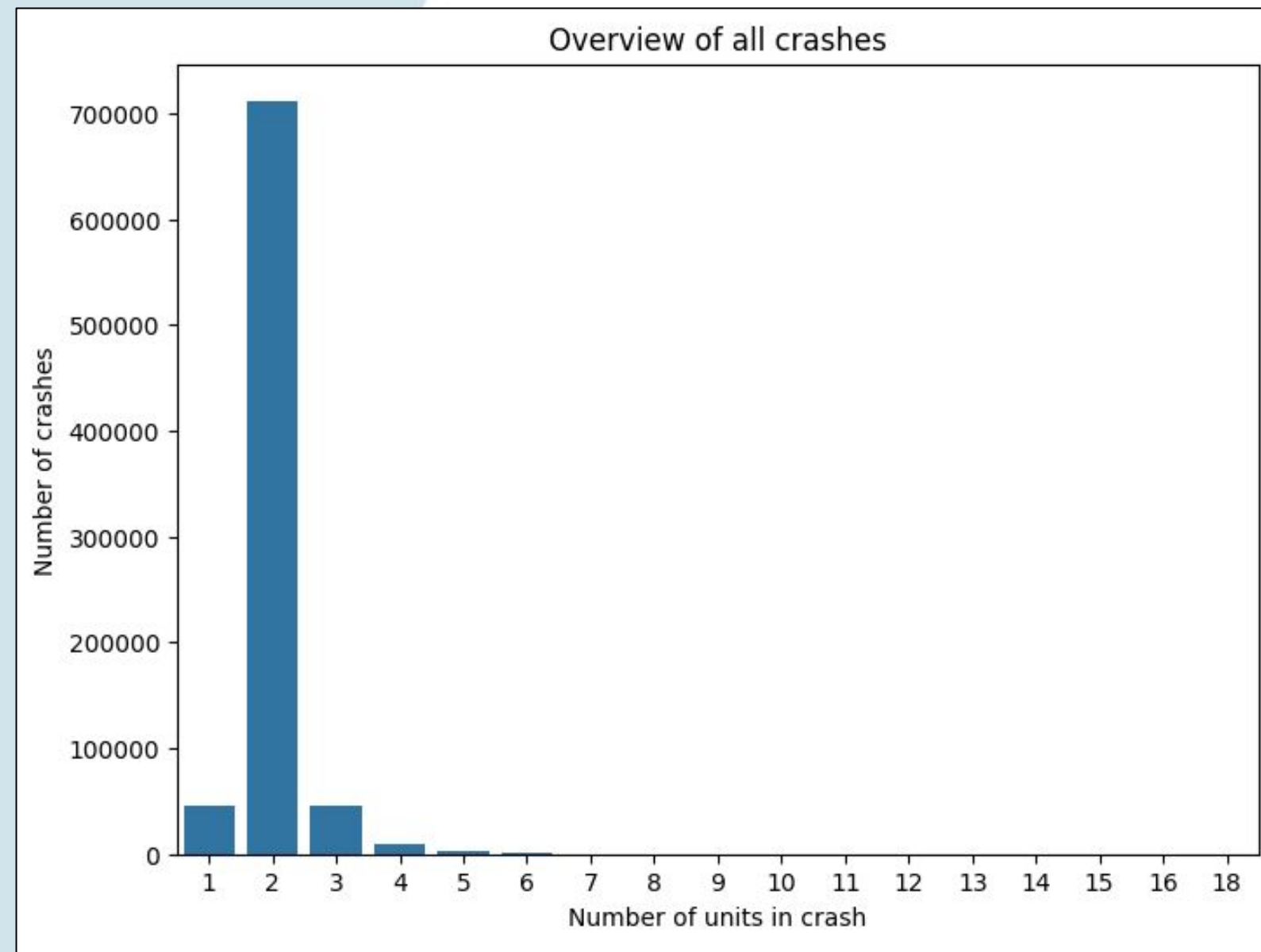


**EDA**





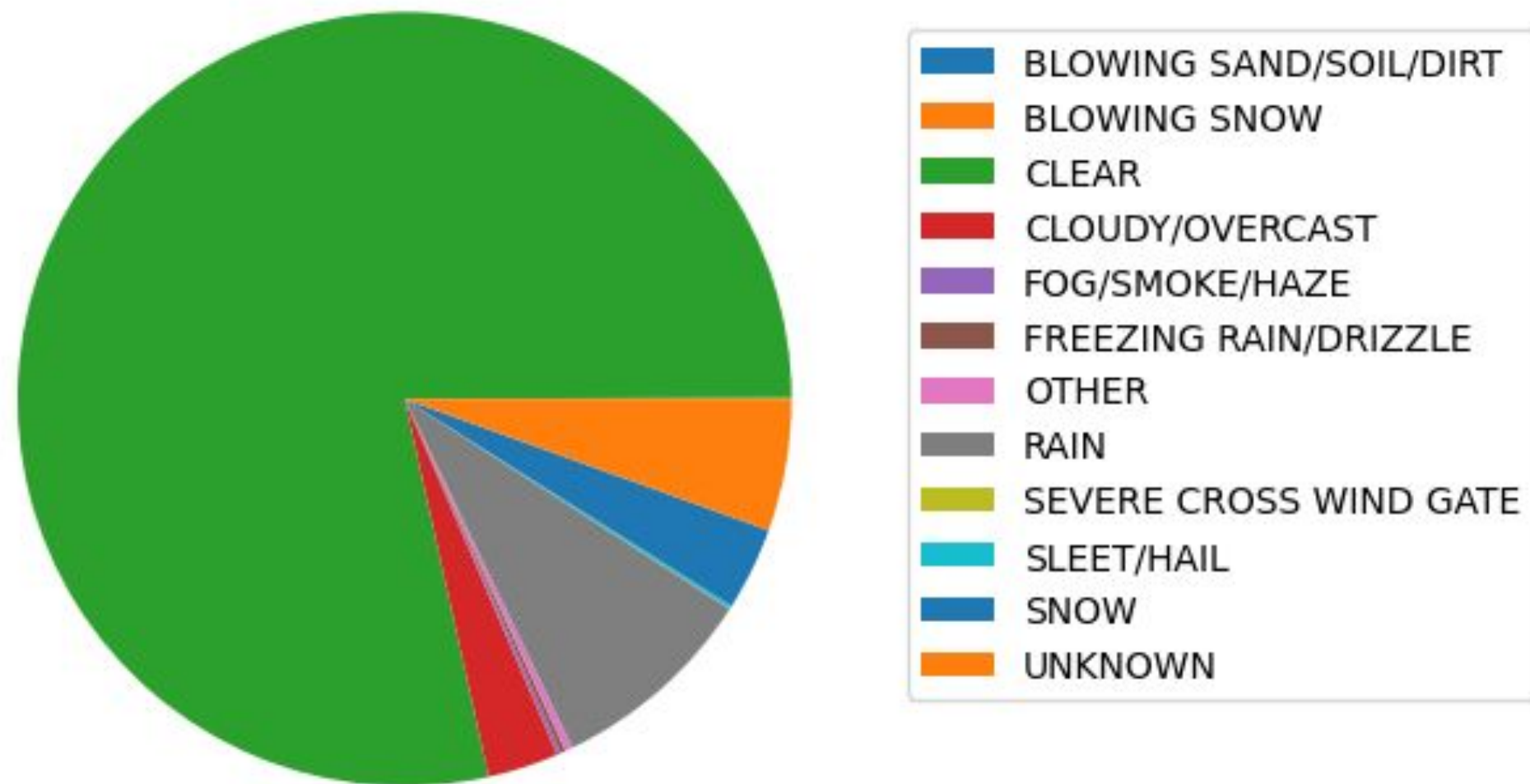
# EDA – Jackson



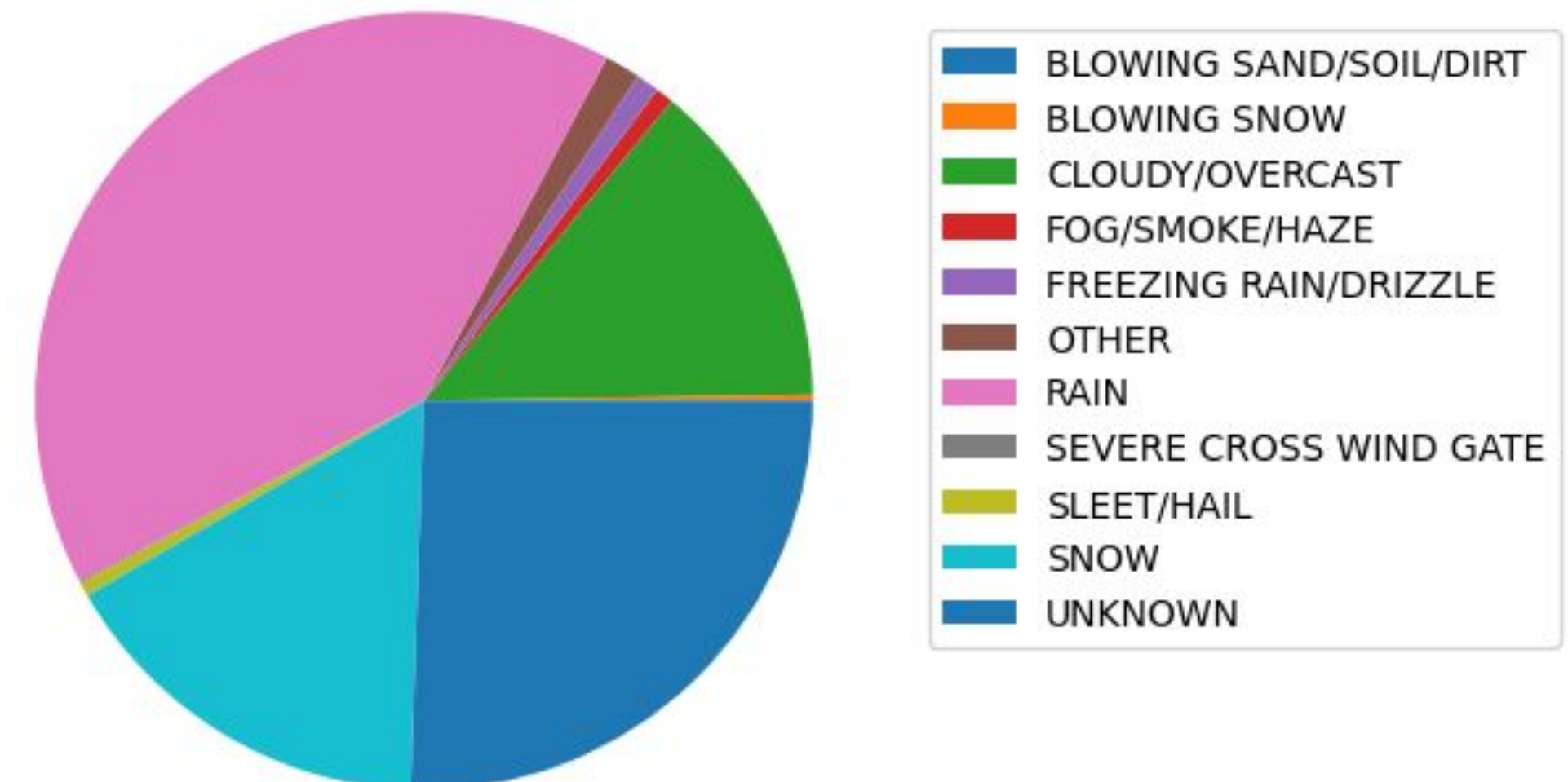
- **Majority of crashes have 2 units involved**
- **1 and 3 units are around the same number**
- **The number of crashes with over 3 units gets lower with each number**

# EDA – Jackson

Weather Conditions For Crashes



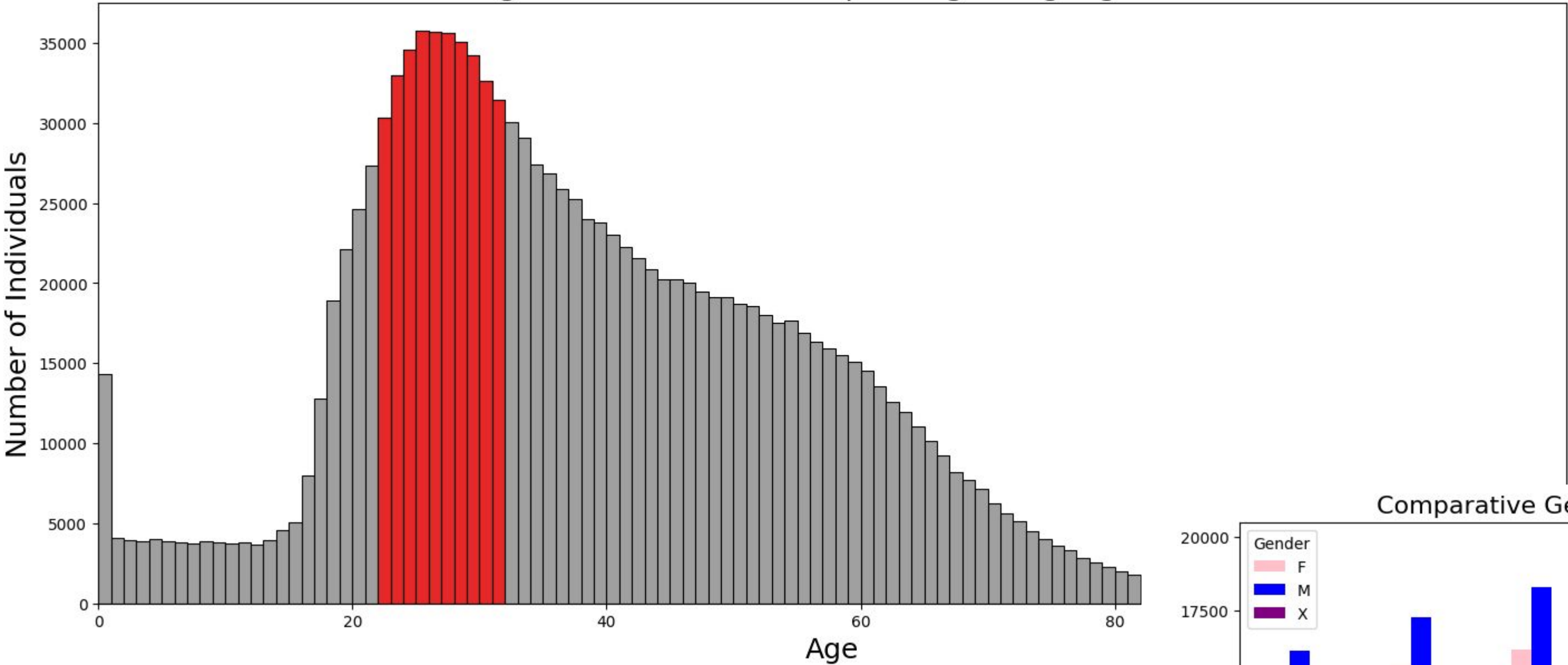
Weather Conditions For Crashes, Excluding 'CLEAR'



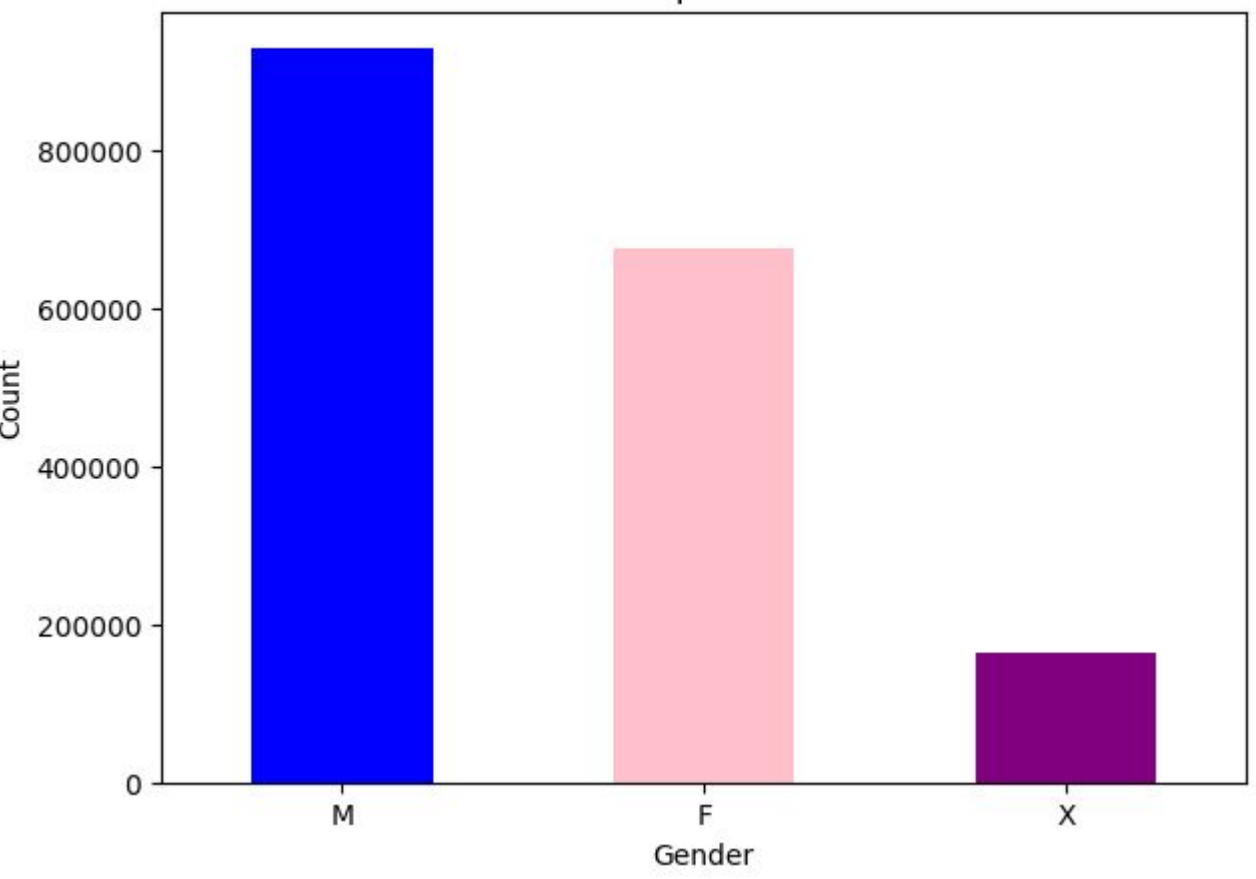
- Majority of crashes were in clear weather
- The second most common weather condition was rain

# EDA – Erica

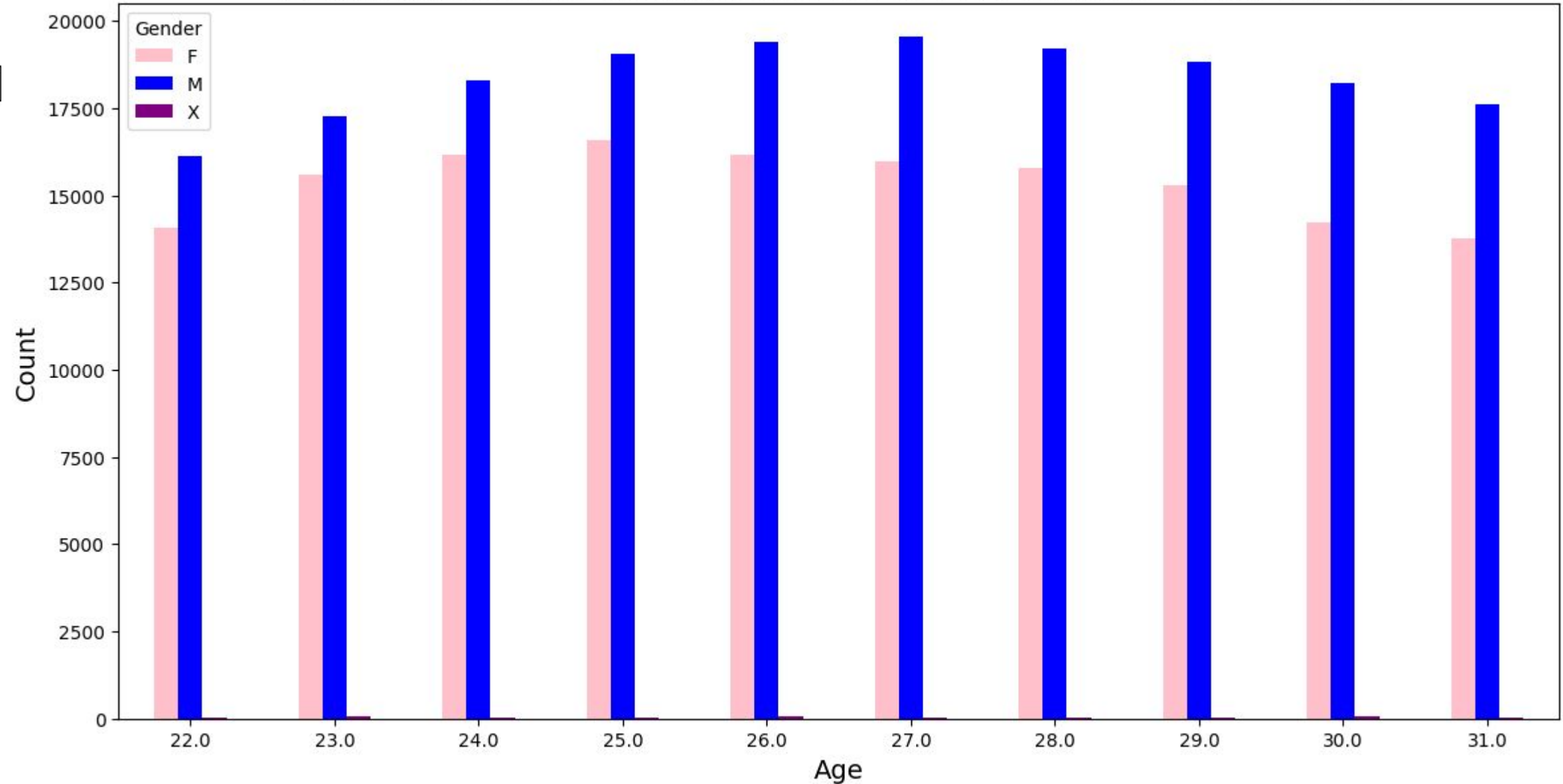
Age Distribution with Top 10 Ages Highlighted



Gender Distribution of People Involved in Traffic Crashes

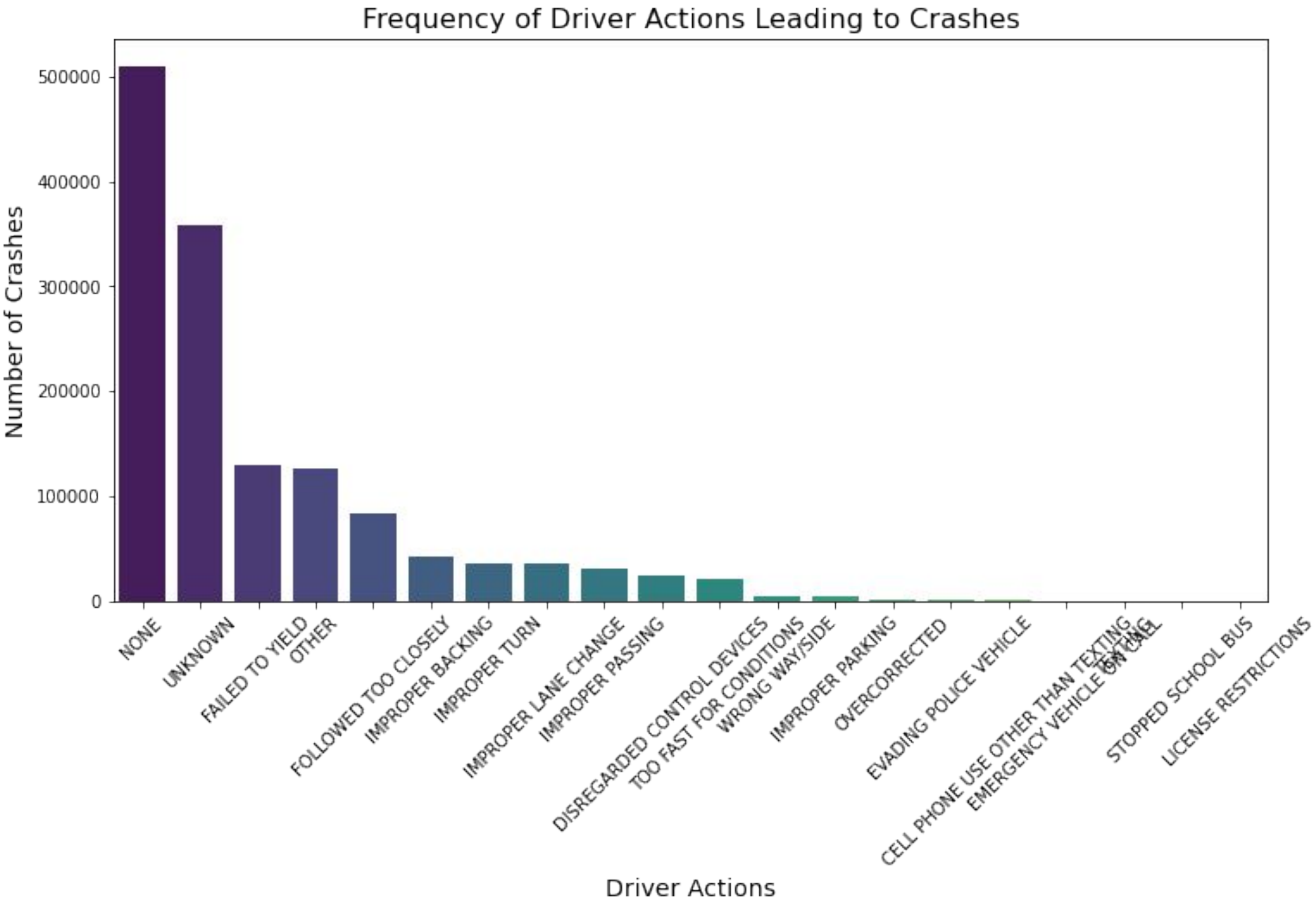


Comparative Gender Distribution for Top 10 Most Frequent Ages in Crashes



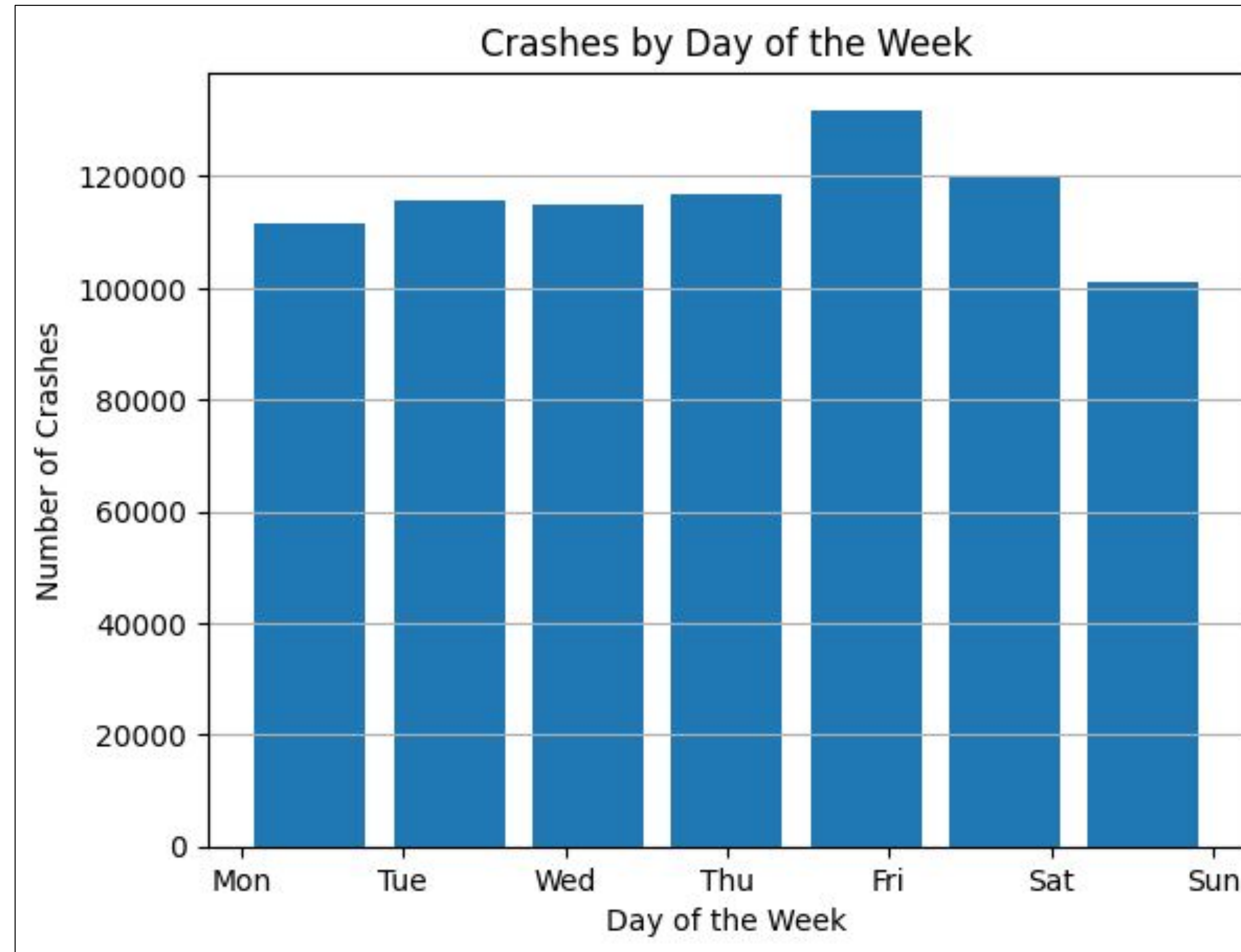


# EDA – Erica



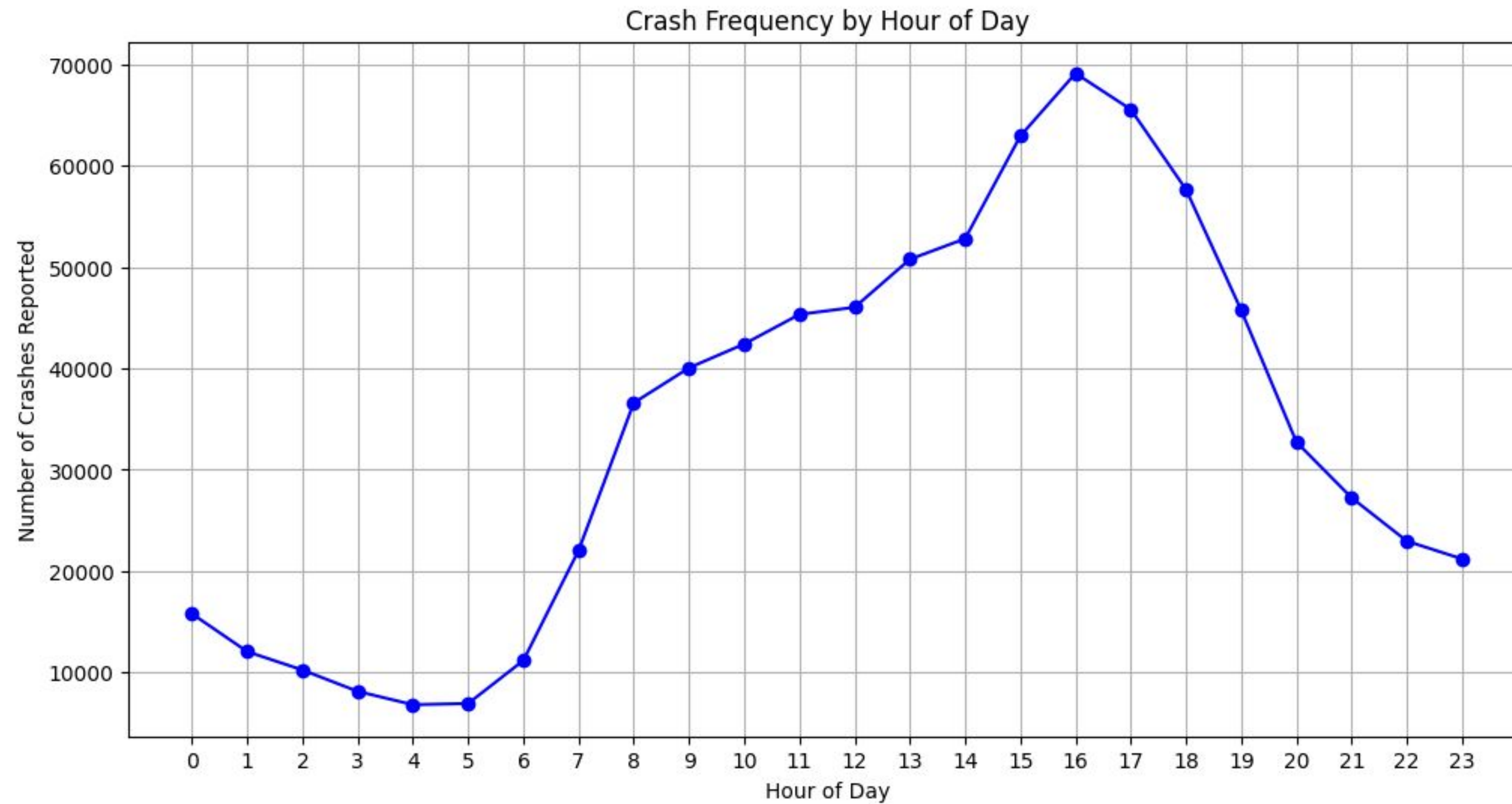


# EDA – Mohammed



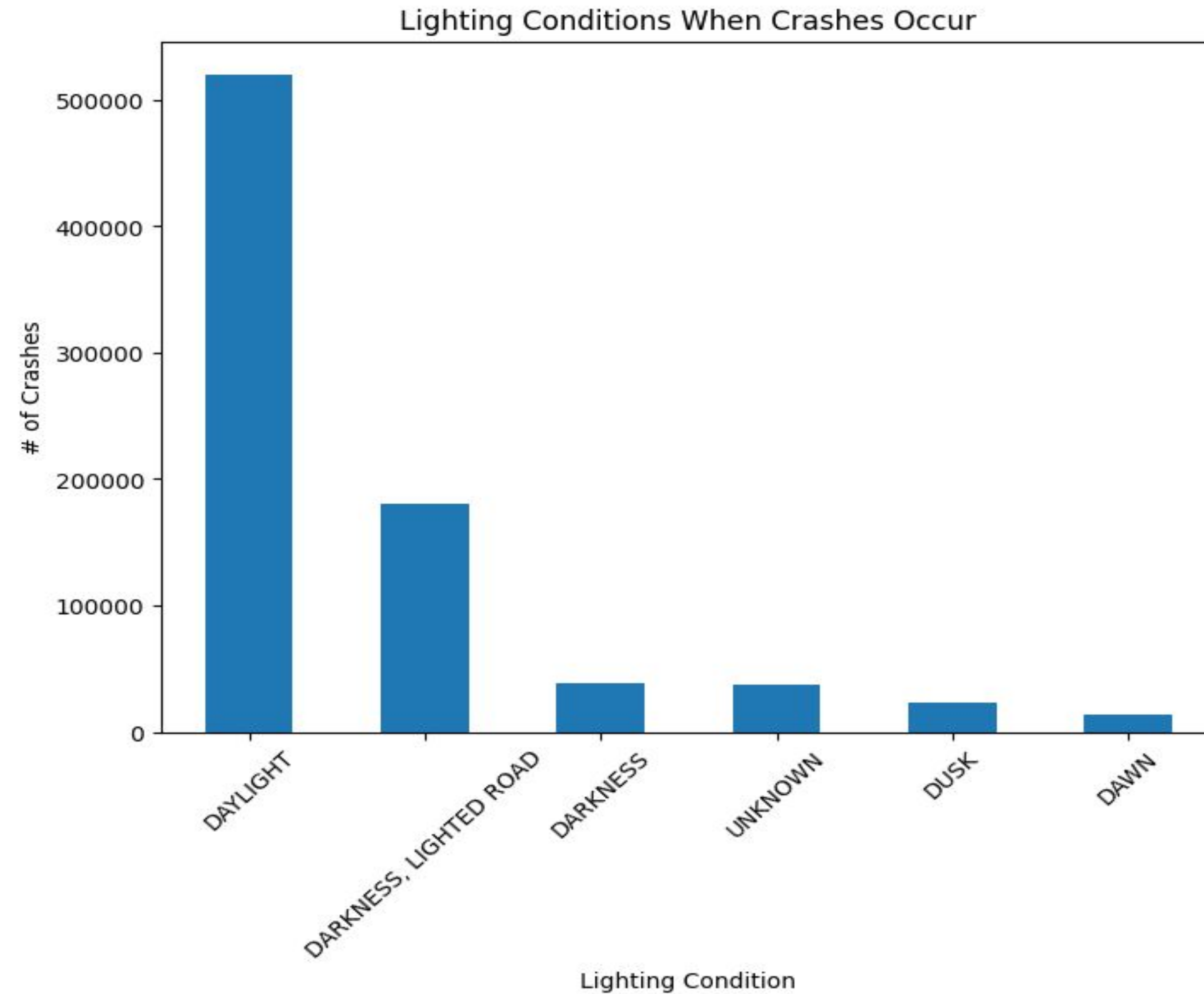
- Increases steadily, peaking on Friday, with the exception of a slight dip on Wednesday
- Sunday has the lowest rate of accidents, with a strong dip

# EDA – Junha



- Rising sharply from early morning
- Peaking between 3 pm – 5 pm
- Declining during the night

# EDA – Prit



- Majority of all crashes happen in the daylight which is expected as a large percentage of people drive in the daylight
- Darkness accounts for almost the rest of the crashes
- Dusk and dawn are in last with small percentages of crashes

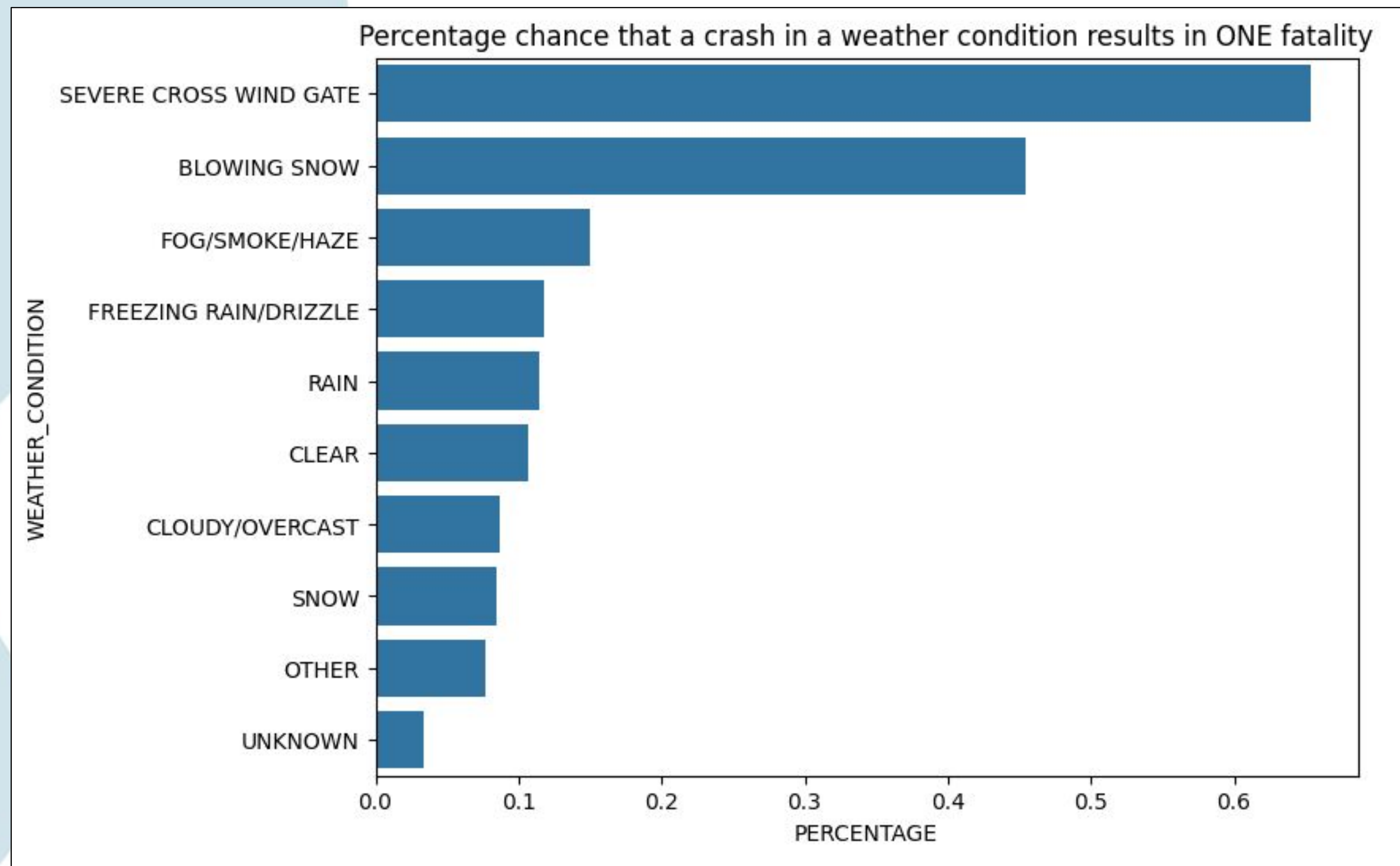


# Hypotheses

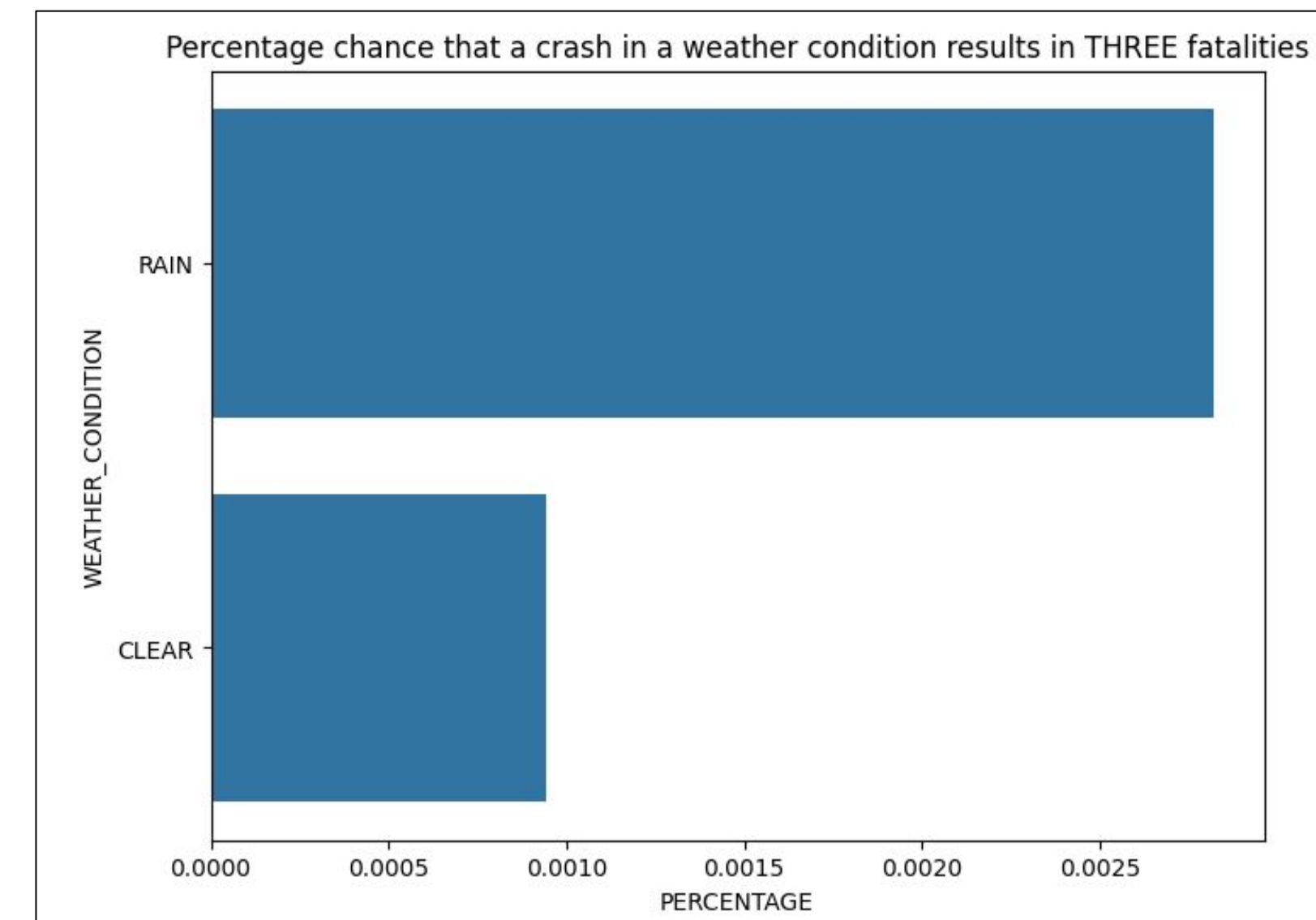
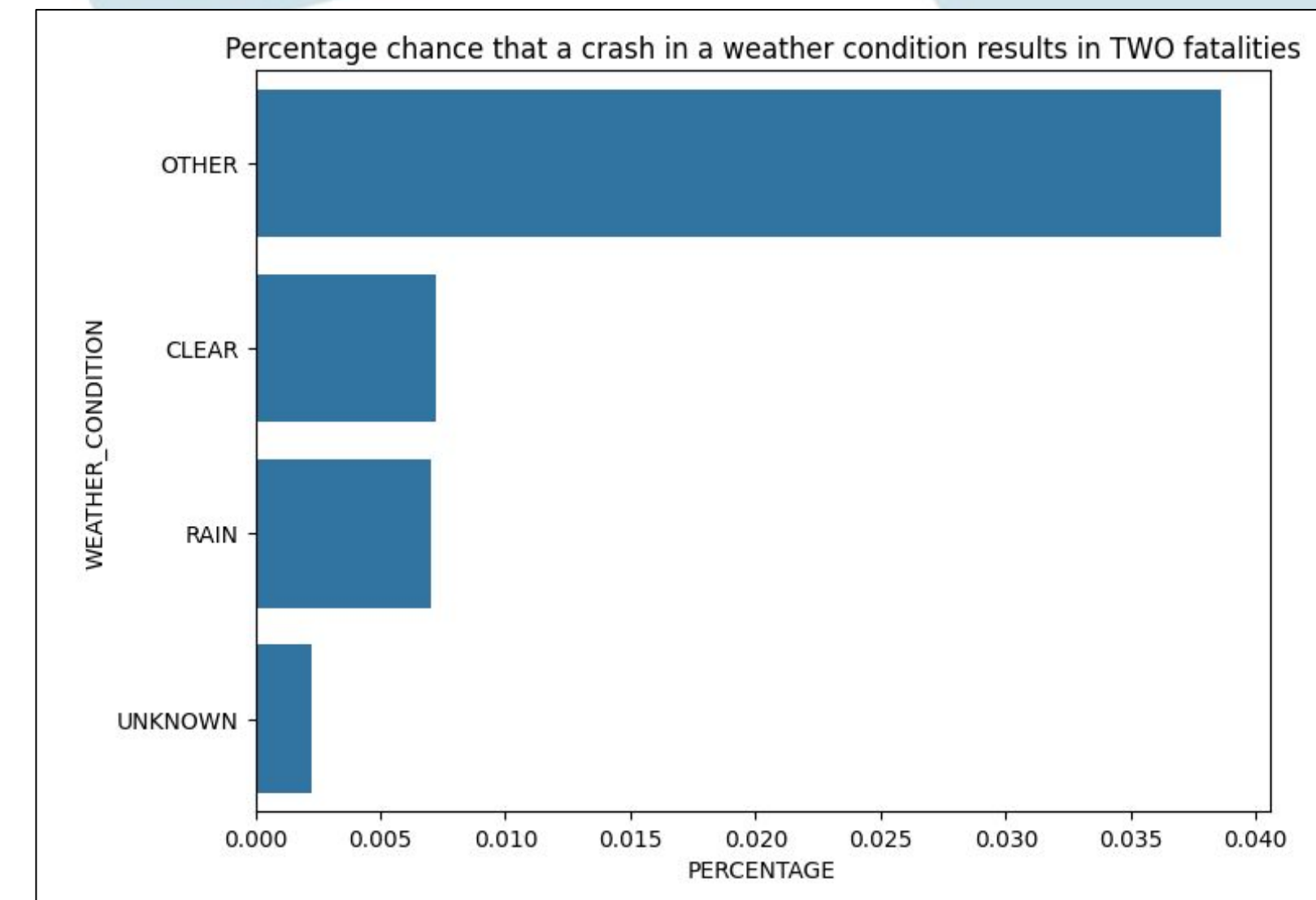




# Hypothesis - Jackson

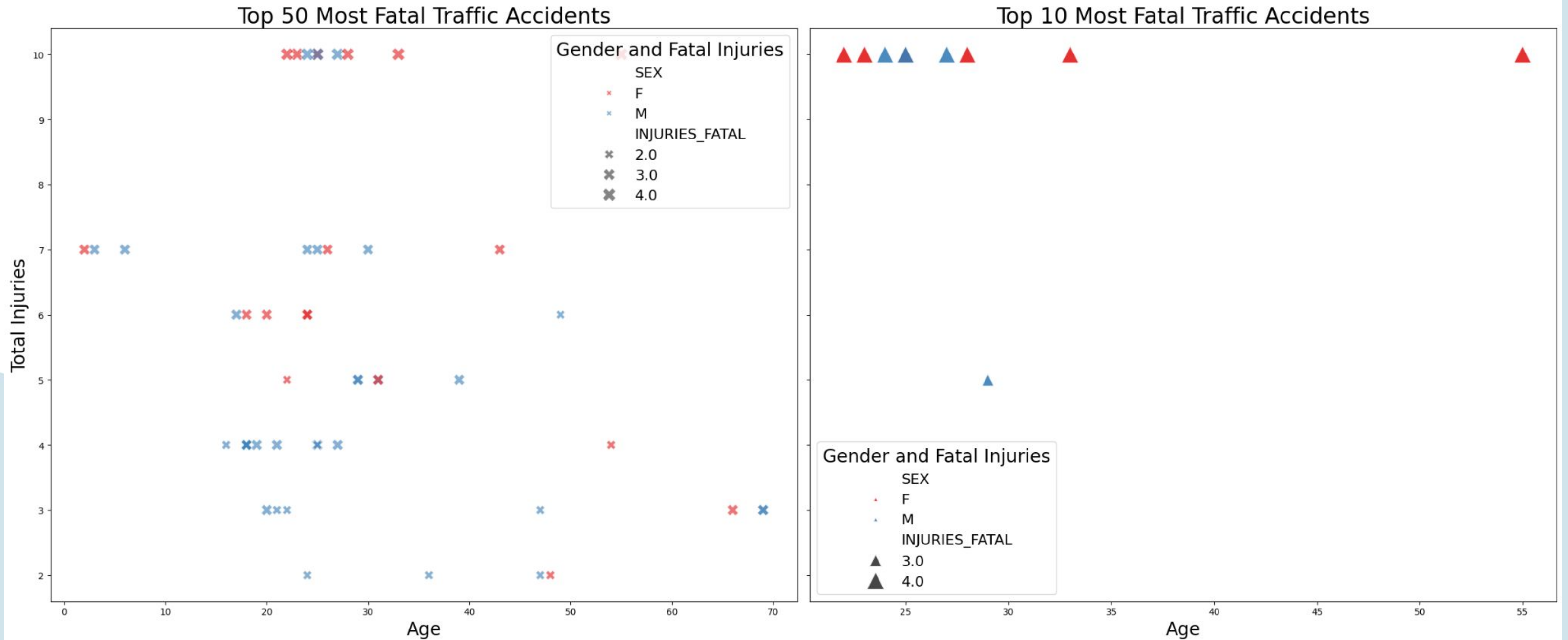


- Percent chance that a weather condition will result in a certain fatality number
- Overall low chance of dying
- Highest was "Severe Cross Wind Gate" and "Blowing Snow" for one fatal injury



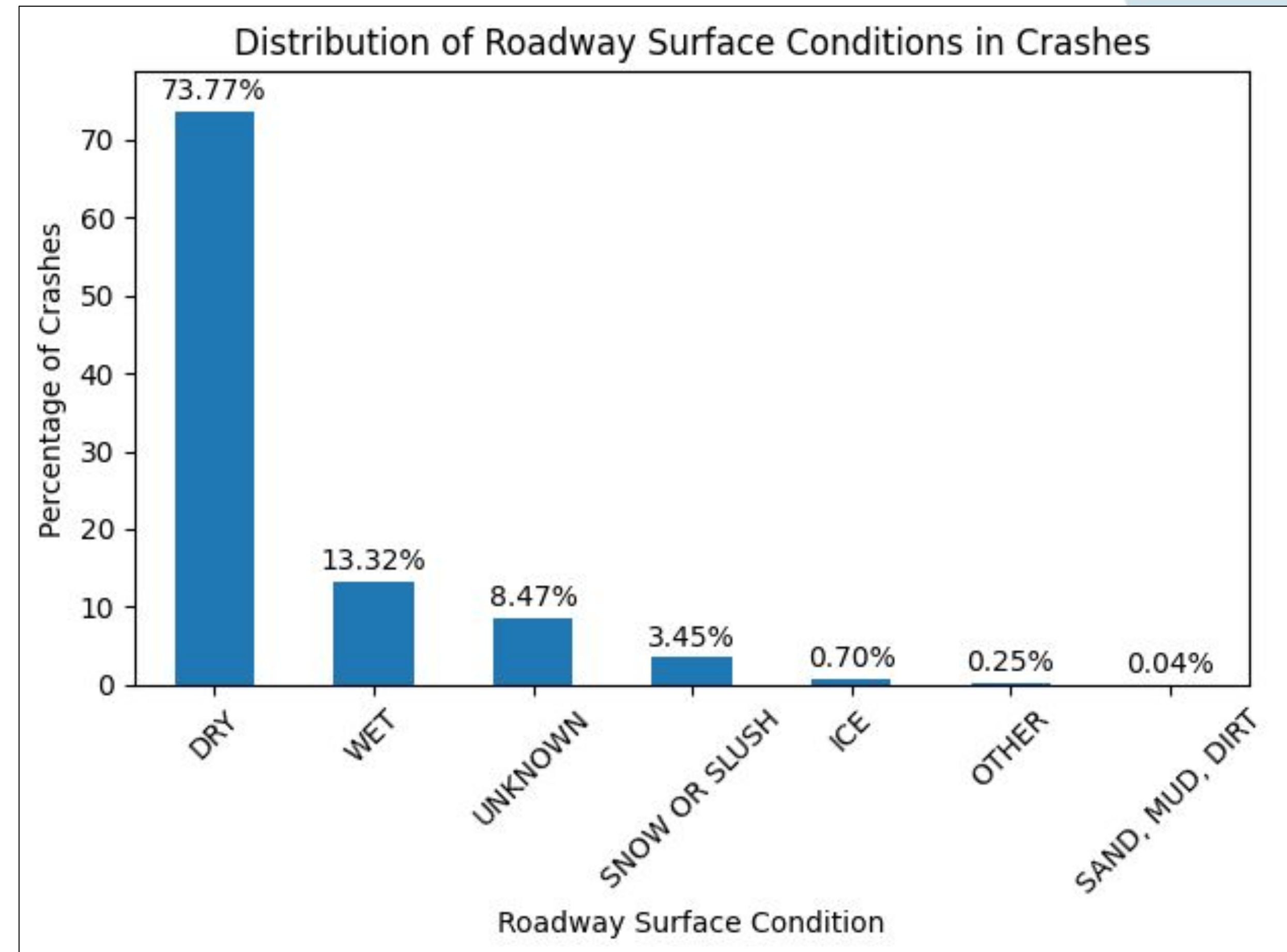
# Hypothesis – Erica

**Men between the ages of 22 and 31 are the most likely to be involved in the greatest number of fatal accidents.**



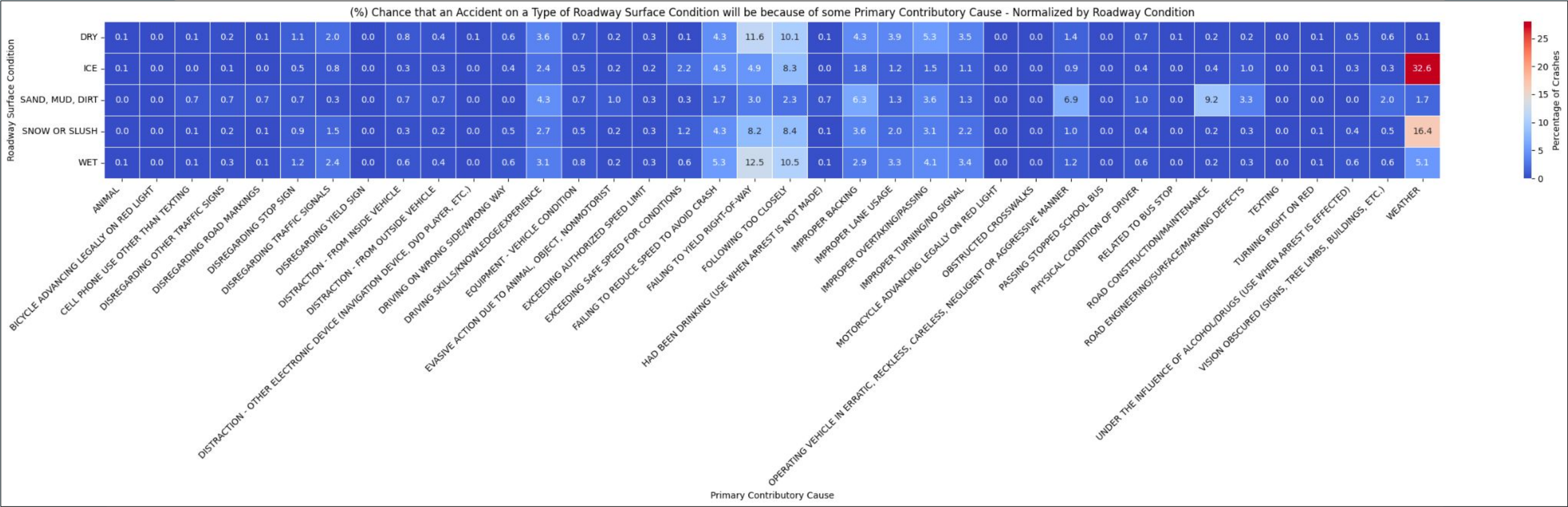
# Hypothesis – Mohammed

- **Hypothesis: Adverse roadway conditions will increase the likeliness that an accident takes place due to some primary contributing cause**
- **Majority of incidents take place in dry conditions**
  - **Normalized by road condition**
- **Conclusion: The data supports that some primary contributing causes of traffic accidents are in fact more frequent in certain roadway surface conditions**





# Hypothesis – Mohammed

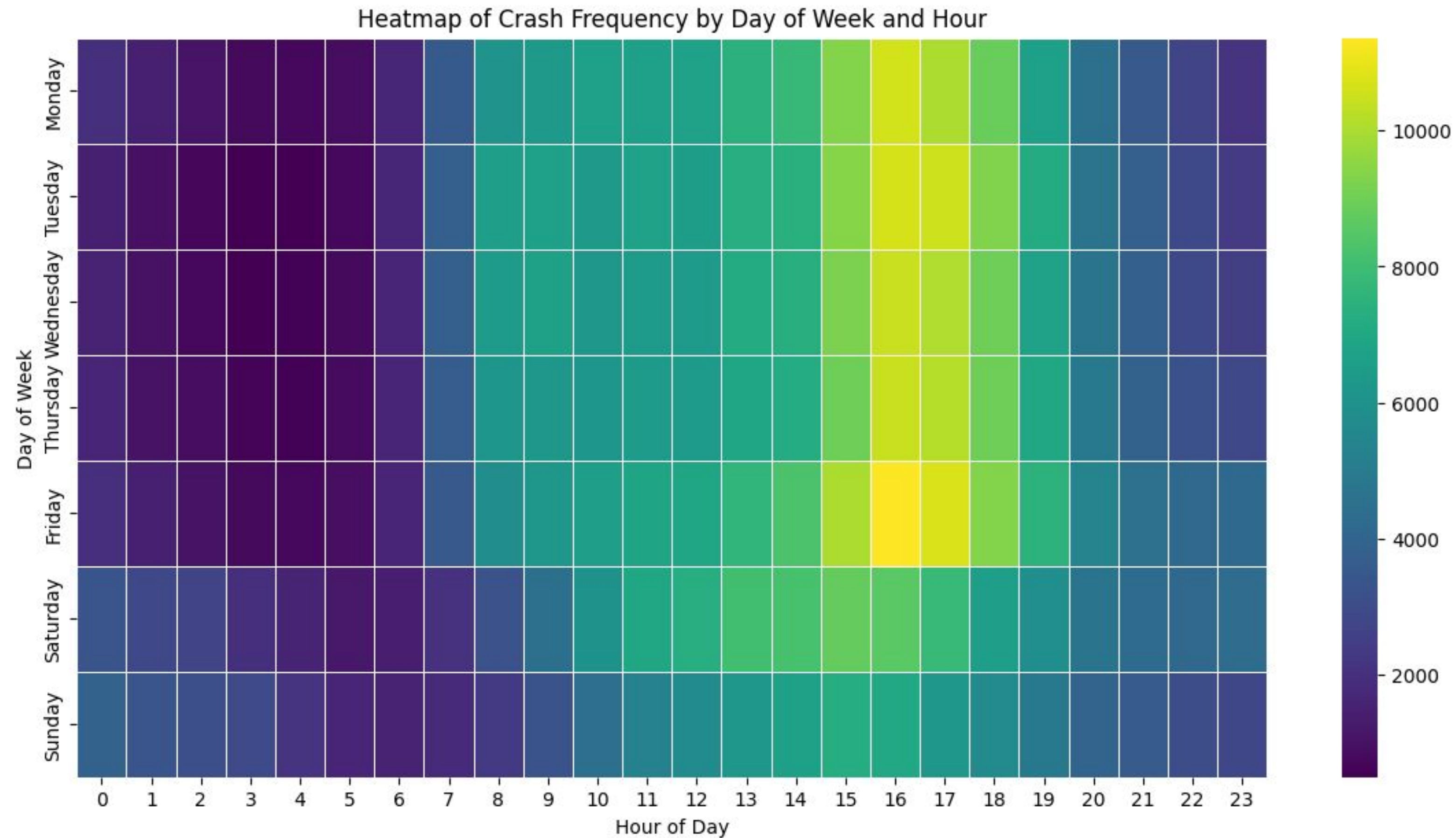


**The analysis confirms my hypothesis, where some primary contributing causes to accidents are increased in adverse conditions.**

**For example, failure to reduce speed, following too closely, and failing to yield right of way are all more likely to cause an accident in wet conditions, rather than dry.**

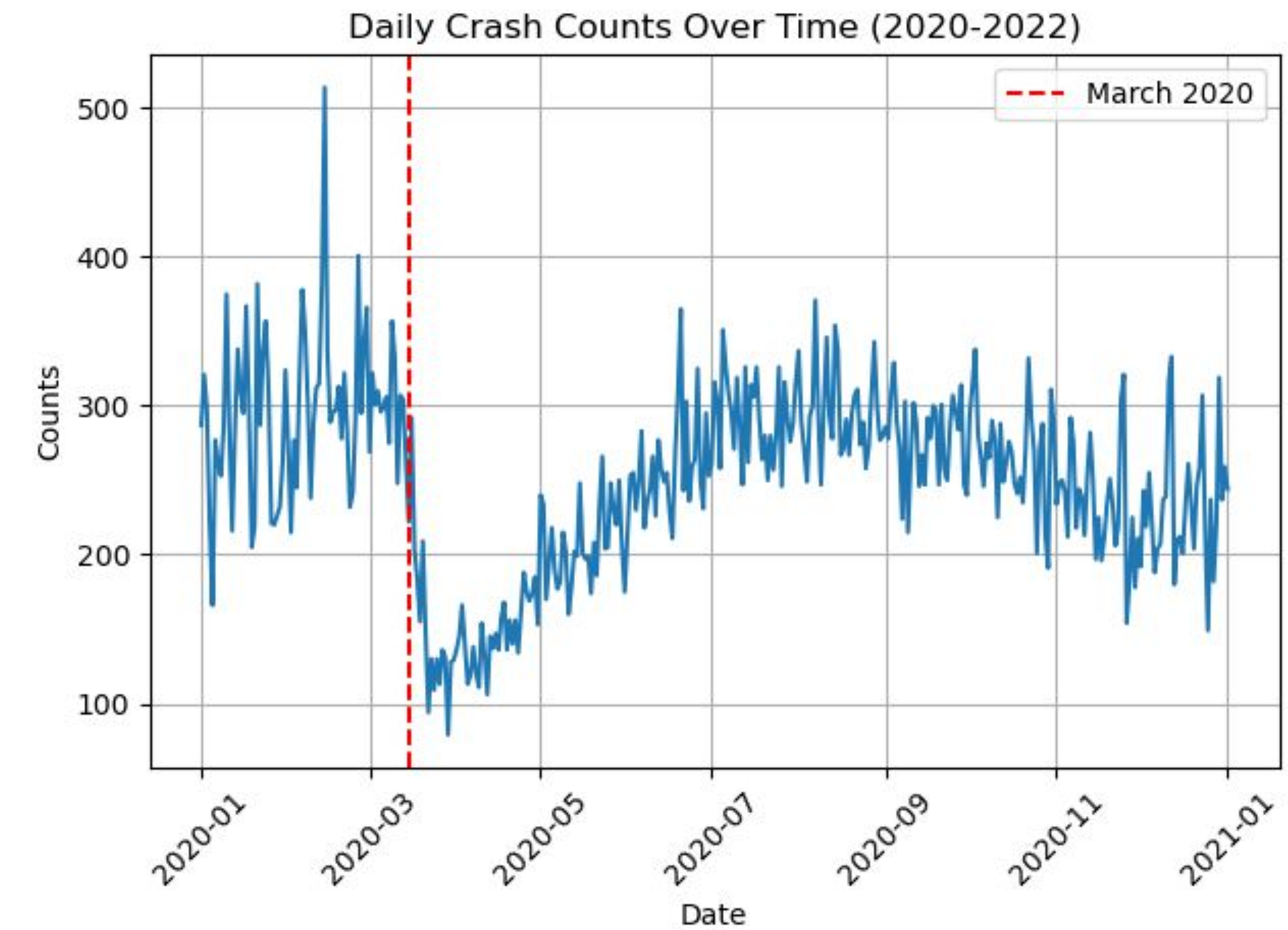
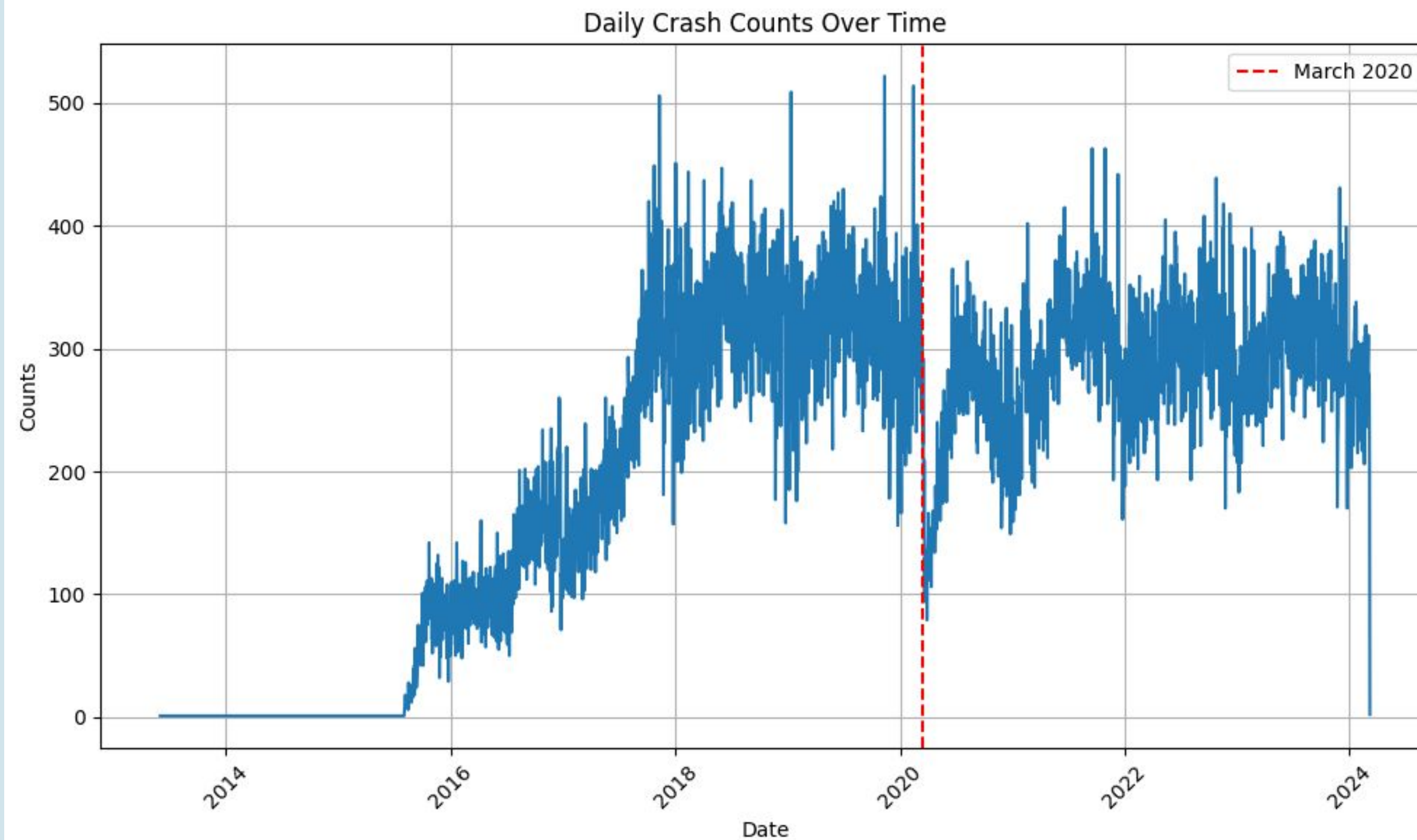


# Hypothesis – Junha



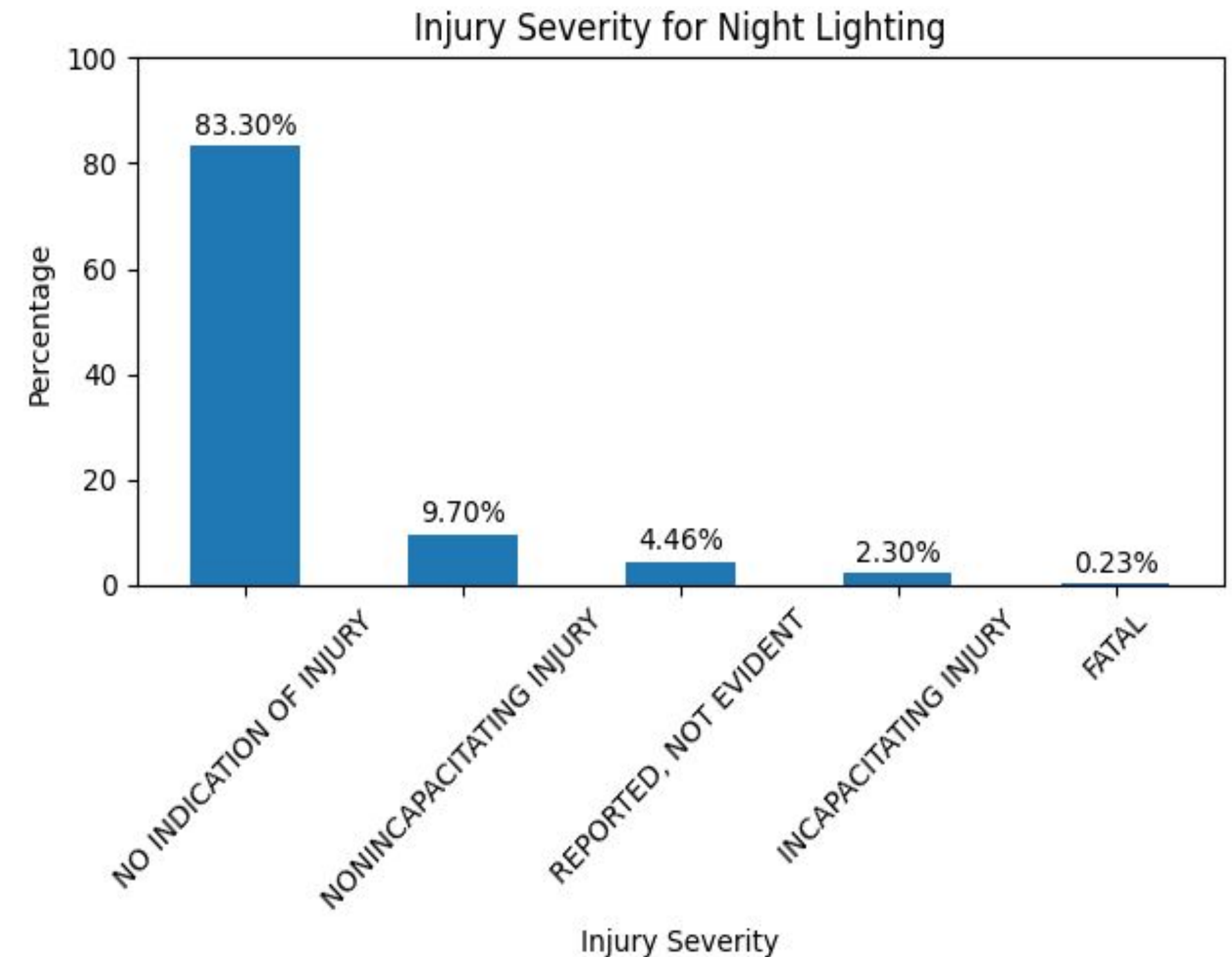
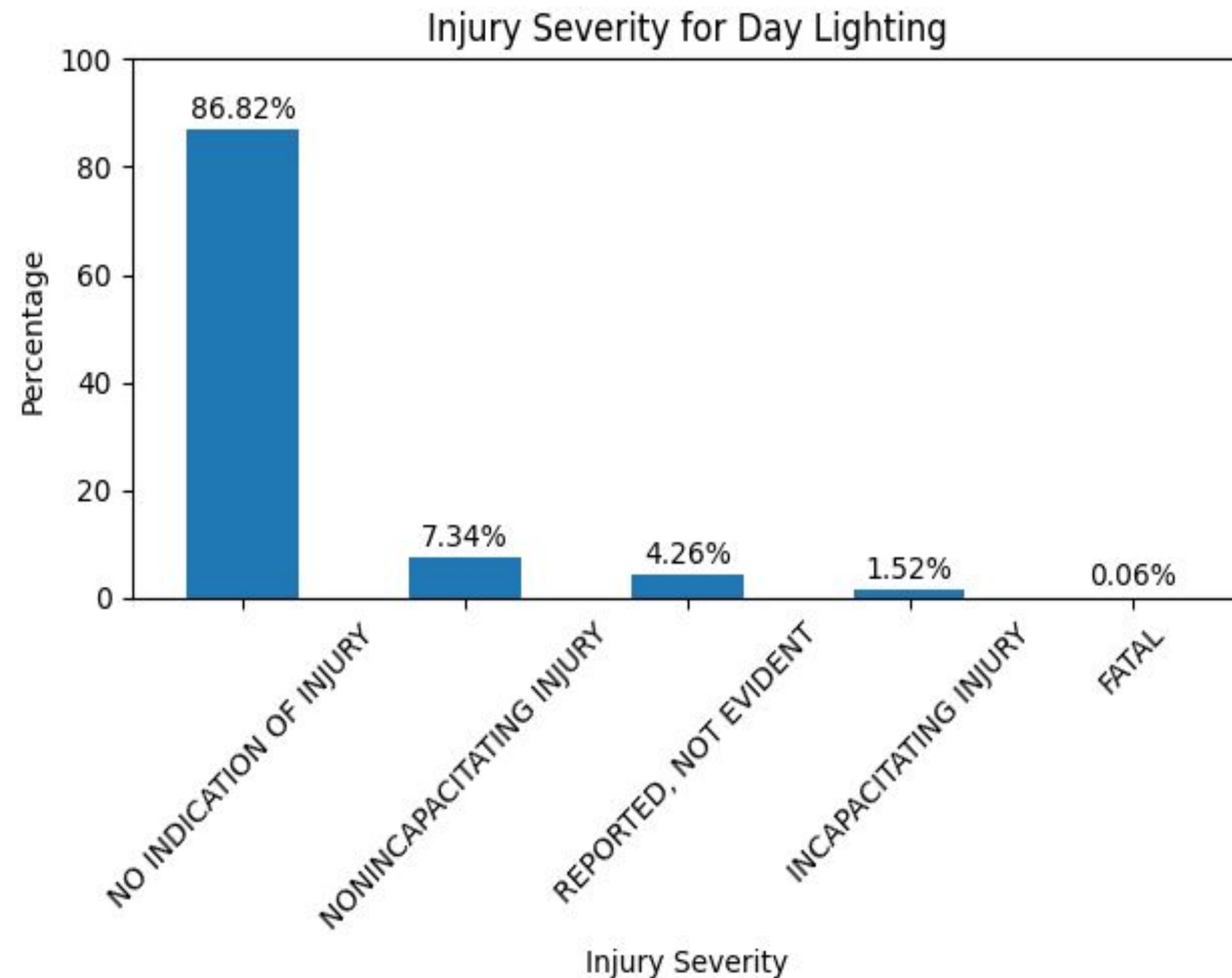
- Peaking on Weekday near 4 pm
- Early mornings across all days show the lowest frequencies

# Hypothesis – Junha



- **Significant decrease in traffic crashes around March 2020 (COVID-19 shutdown)**
- **The COVID-19 pandemic led to a global shift towards remote work and reduced travel, resulting in fewer traffic collisions**

# Hypothesis – Prit



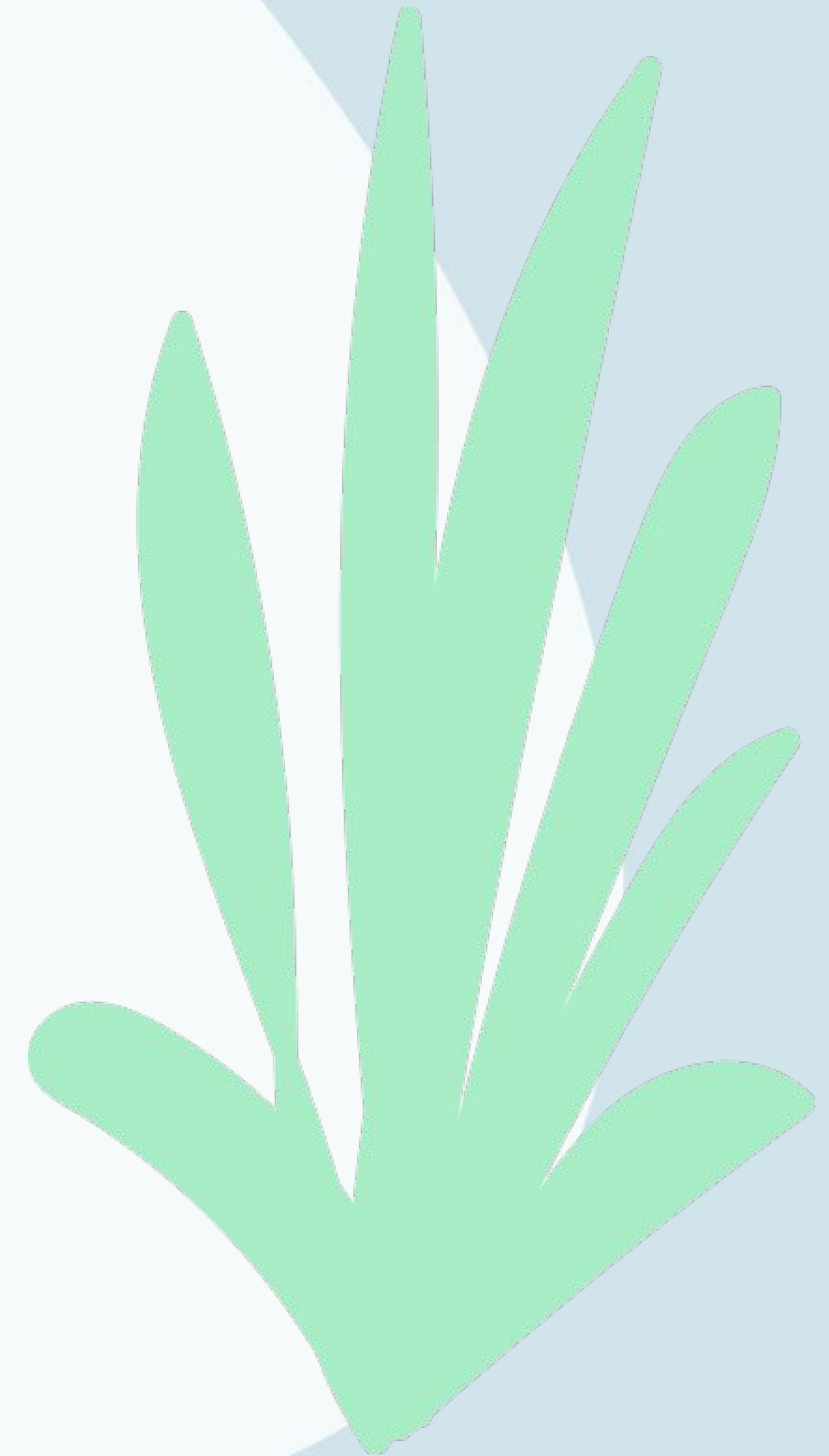
Hypothesis: Collisions occurring in **night light** conditions result in **more severe injuries** compared to collisions during daylight hours.

- Injury severity **rises** in the night lighting.
- Fatal crashes are almost **4 times more likely** in the dark compared to day lighting.
- Incapacitating injury is **2 times more likely**.





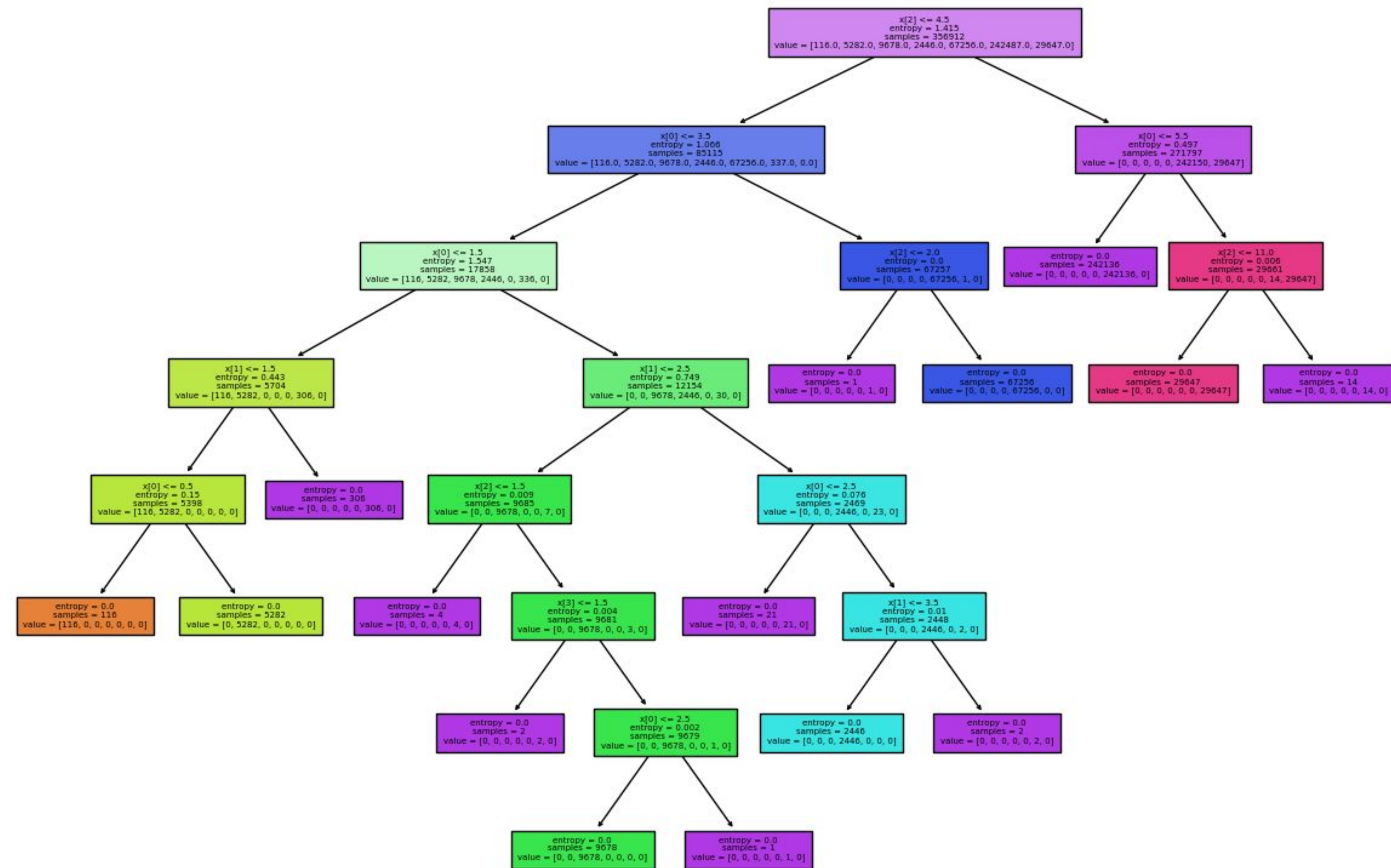
# Machine Learning Analyses





# Machine Learning Analysis – Jackson

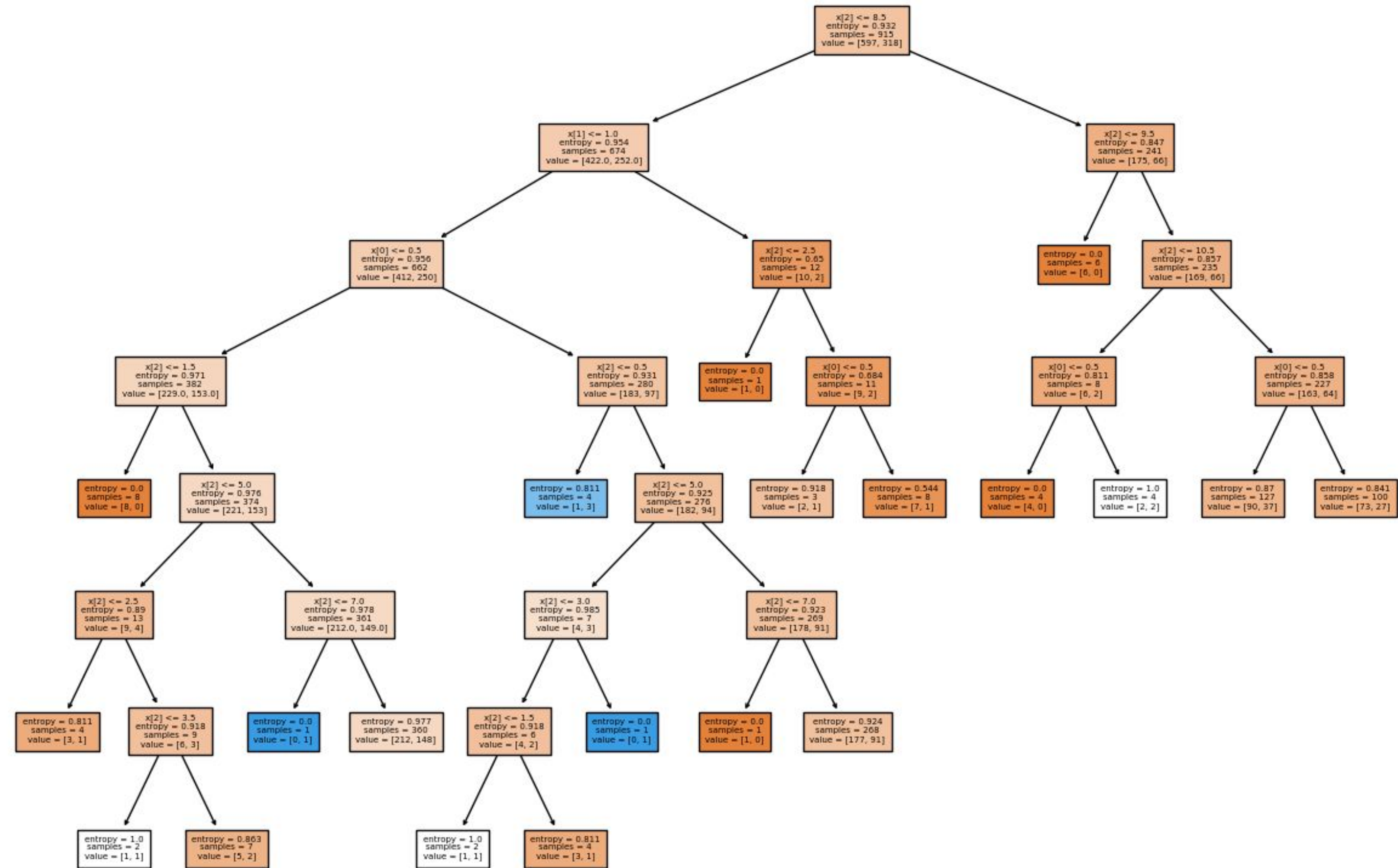
- Decision tree classifier
- Features
  - Total injuries
  - Number of units
  - Crash hour
  - Lane count
- Class (7 classes)
  - Air bag status
    - **'DEPLOYED OTHER (KNEE, AIR, BELT, ETC.)'**, **'DEPLOYED, COMBINATION'**, **'DEPLOYED, FRONT'**, **'DEPLOYED, SIDE'**, **'DEPLOYMENT UNKNOWN'**, **'DID NOT DEPLOY'**, **'NOT APPLICABLE'**



99% Accuracy

# Machine Learning Analysis – Jackson

- Decision tree classifier
- Features
  - Blood alcohol result value
  - Physical condition
  - Sex
- Class
  - Cell phone use
- ~62% accuracy score
- Majority of leaves were classified as “Yes”

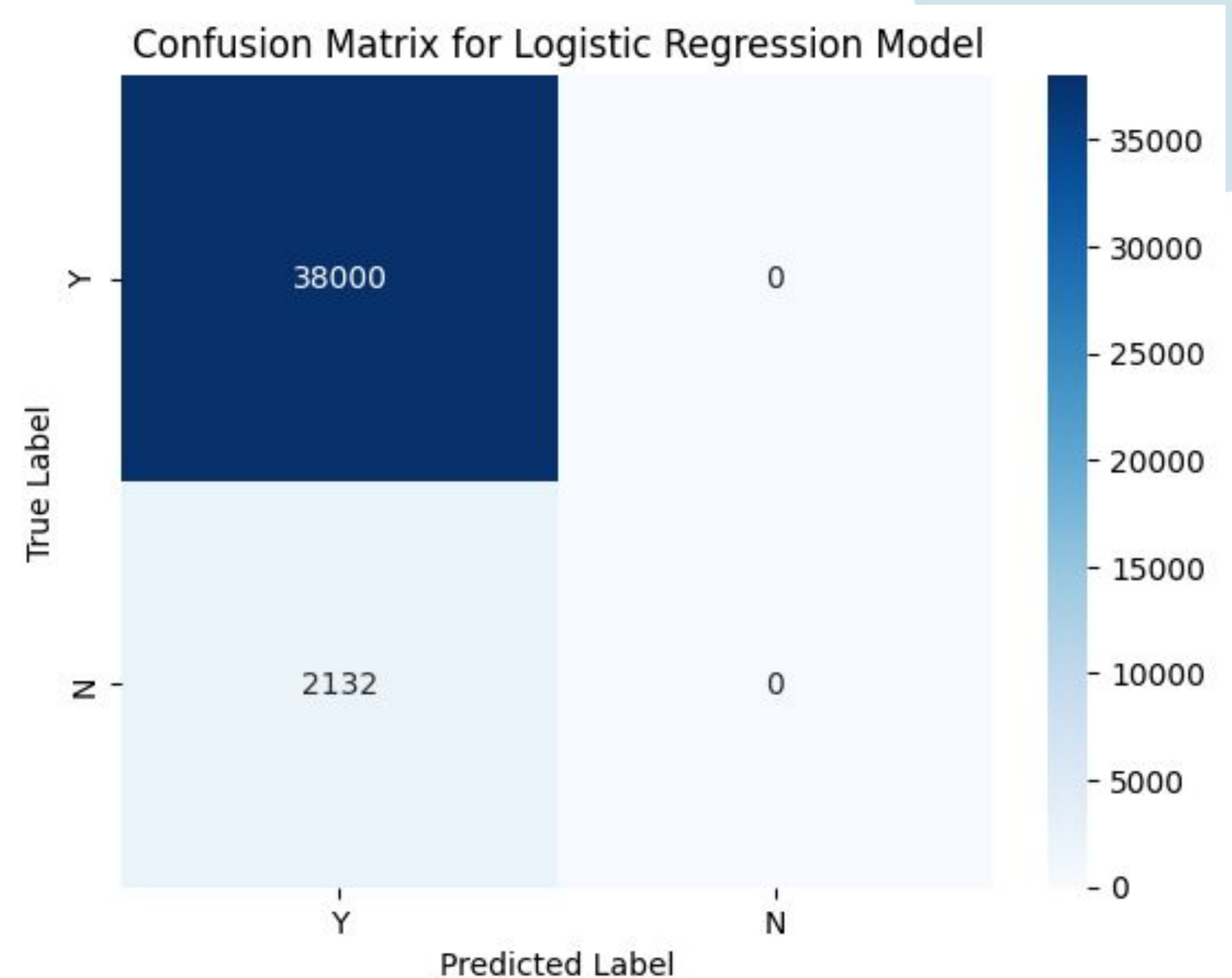


62% Accuracy



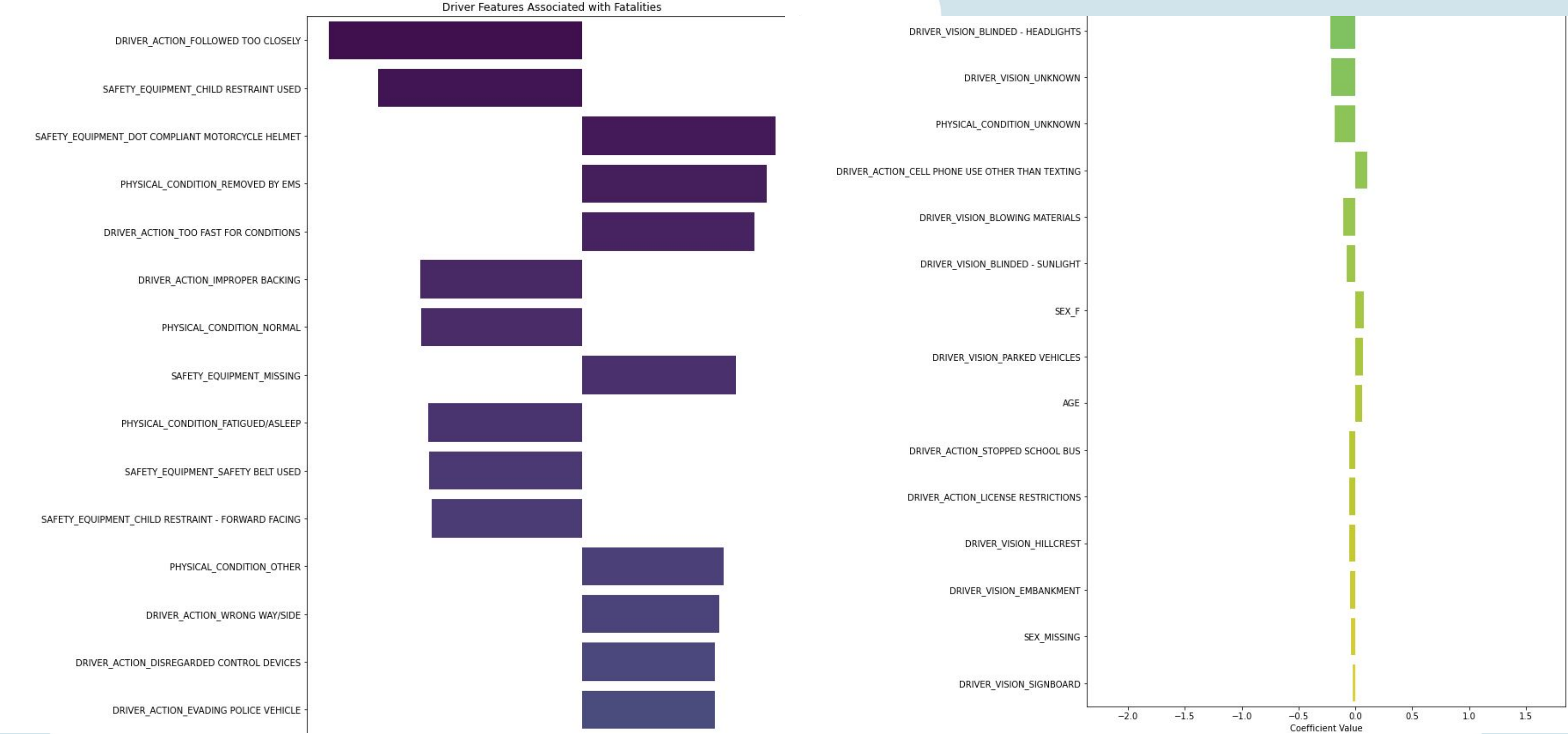
# Machine Learning Analysis – Jackson

- **Logistic regression**
- **Feature: Vehicle year**
- **Class: Was towed?**
- **~94% accuracy score**
- **Only 5% are false positives**
- **Majority of vehicles were towed**



**94% Accuracy**

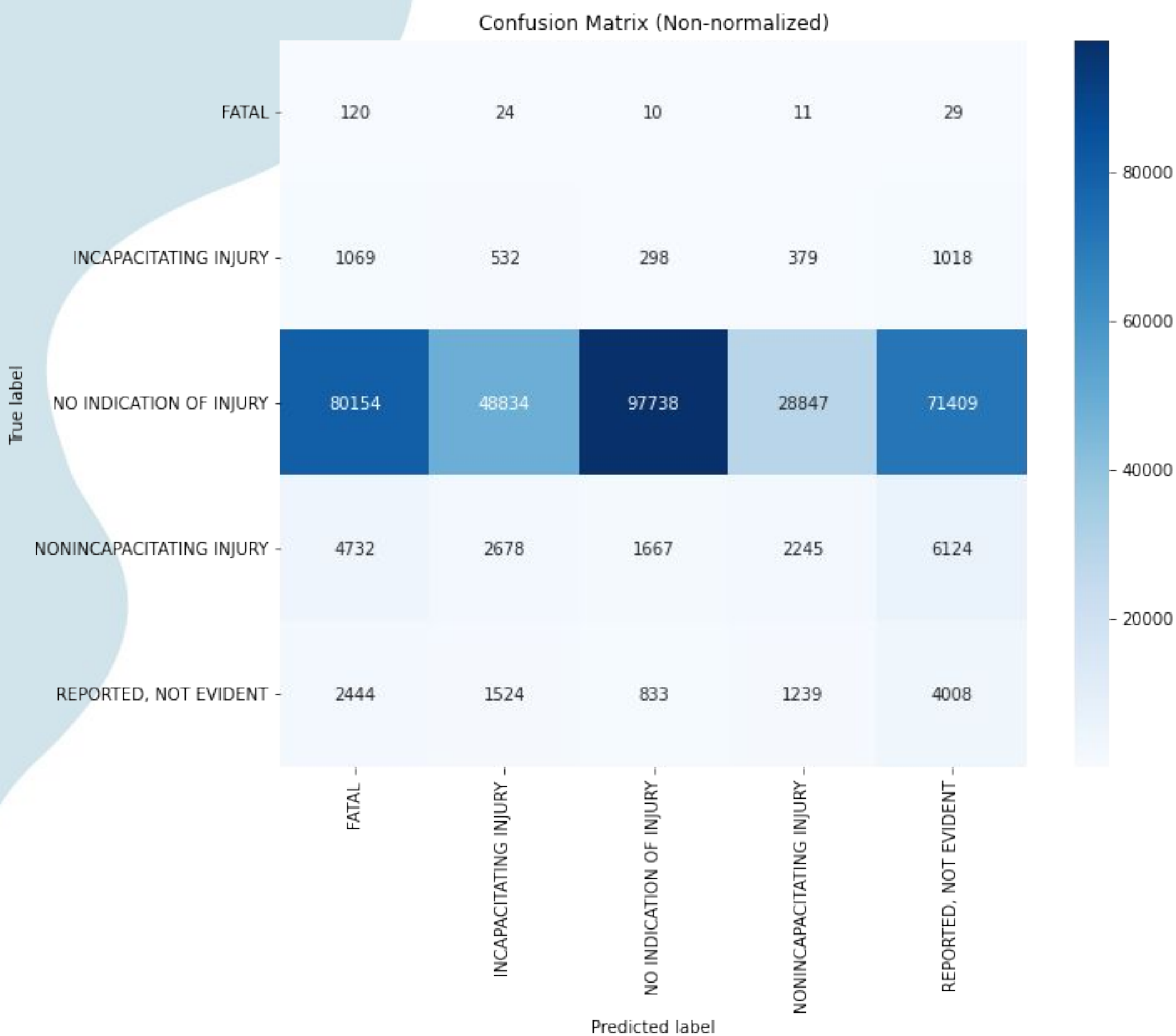
# Machine Learning Analysis - Erica



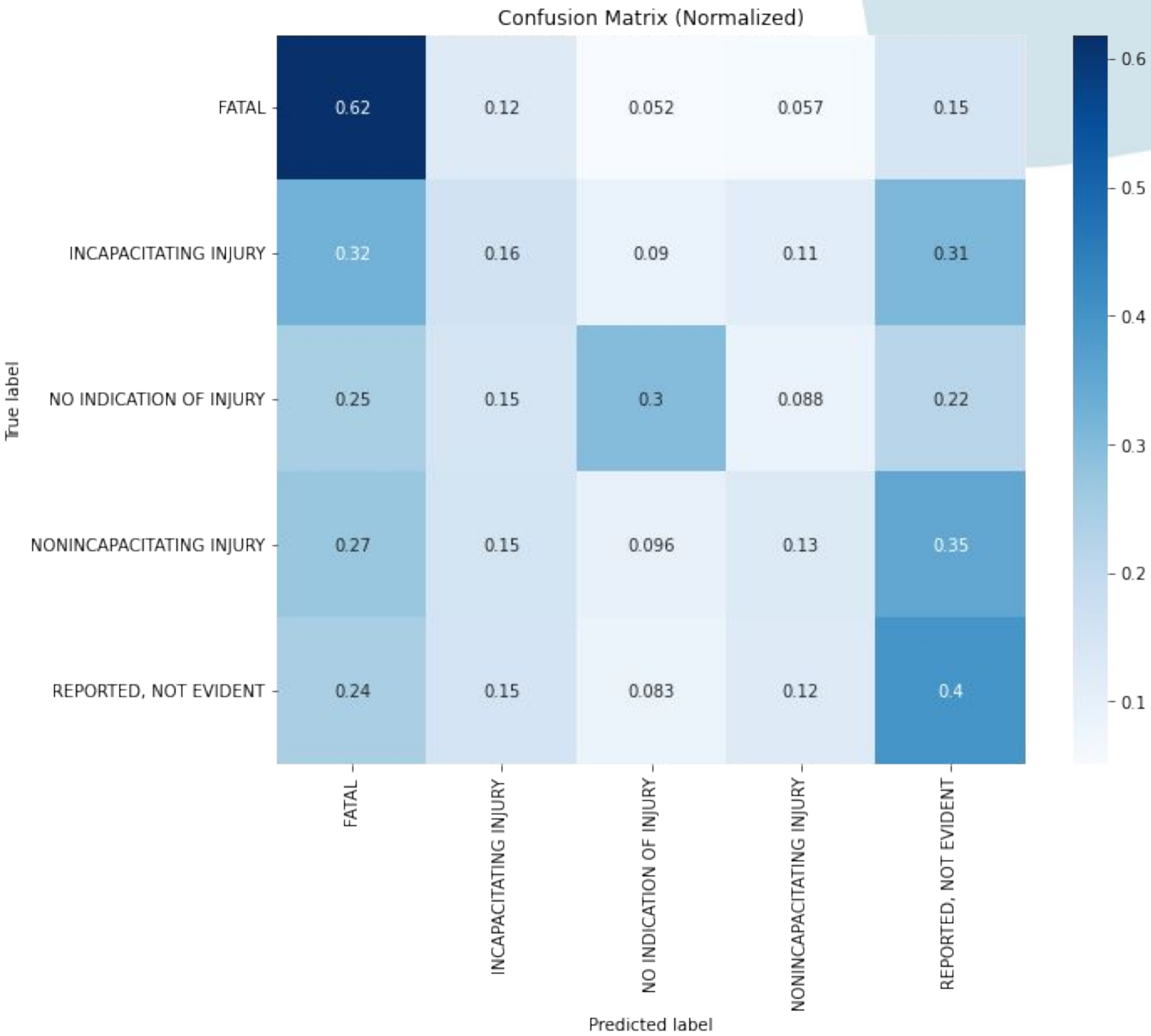


# Confusion Matrices

## Non-normalized

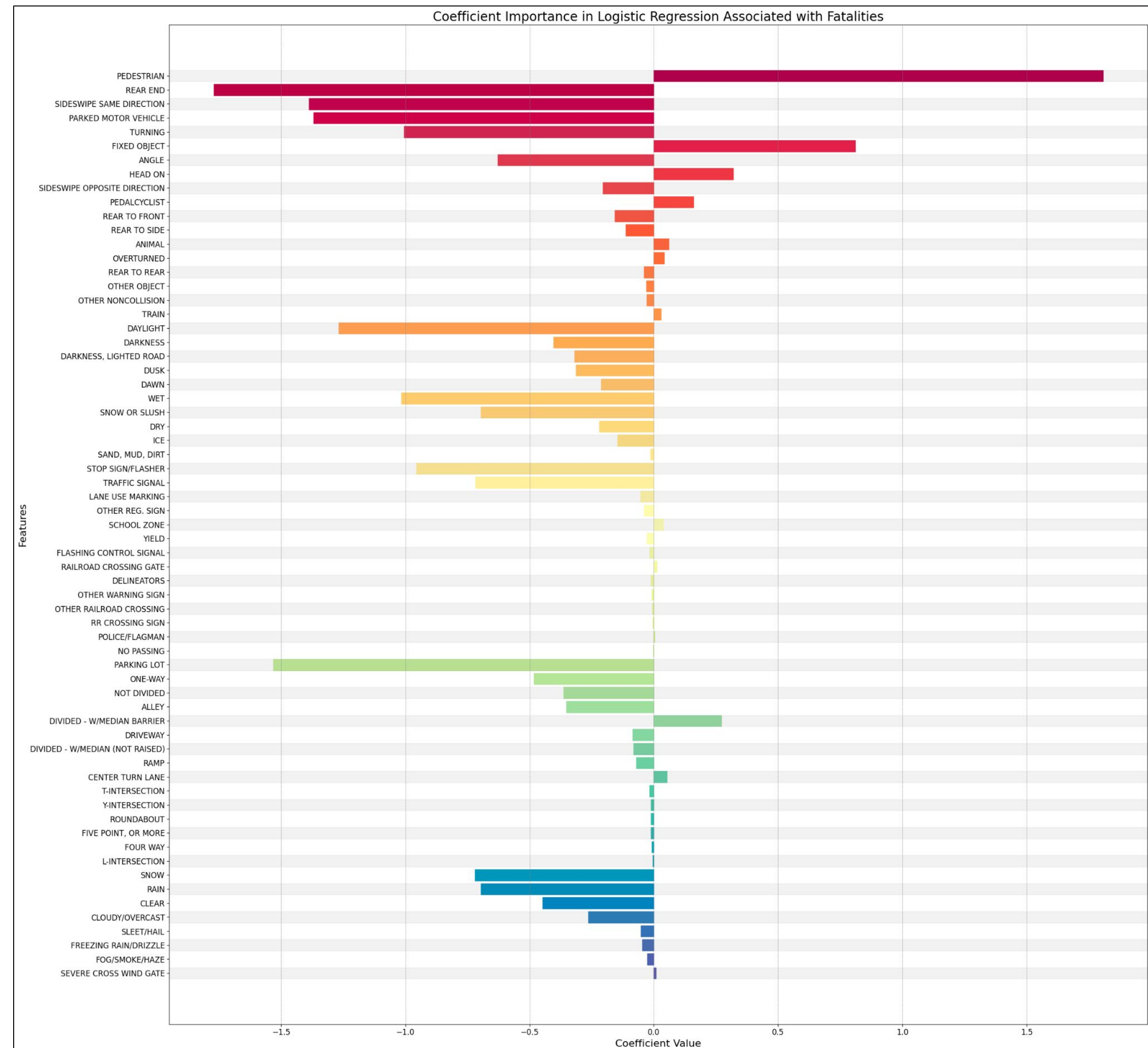


## Normalized



# Machine Learning Analysis - Mohammed

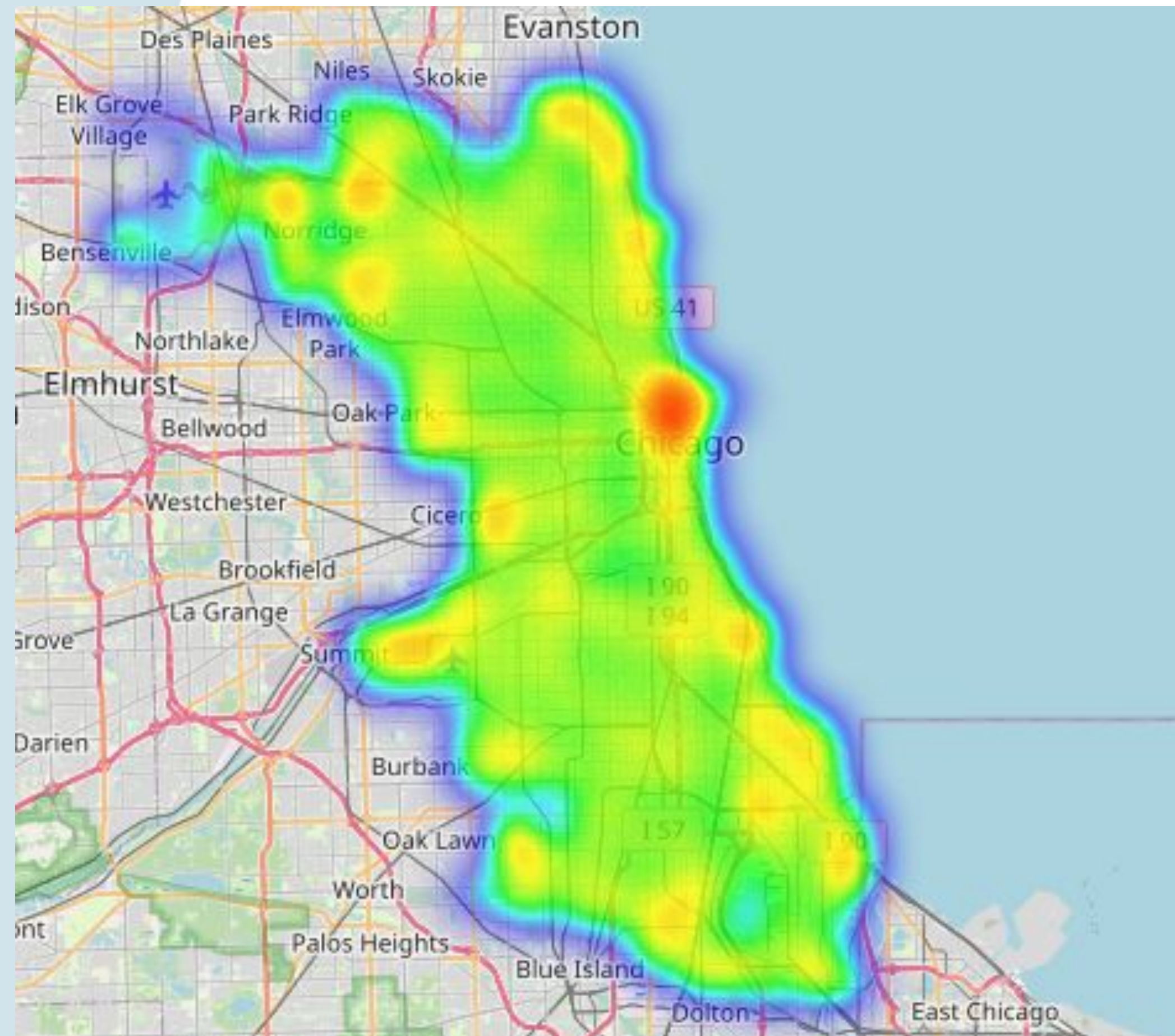
- **Hypothesis: Certain features are strongly correlated with fatal outcomes than others**
- **Logistic regression on what features are associated with fatal outcomes**
- **Conclusions:**
  - **Any accident involving a pedestrian is highly likely to produce a fatal outcome**
    - **Lower risk for pedalcyclist**
  - **Hitting a fixed object is second most likely to lead to a fatality**
  - **A head-on collision is positively correlated with a fatality, where rear-end accidents are significantly negatively correlated with a fatality**
  - **Parking lots are statistically safe**
  - **Severe cross wind gates are the only weather-related condition to have a positive correlation with fatalities**





# Machine Learning Analysis – Junha

## Traffic Crash Hotspots in Chicago

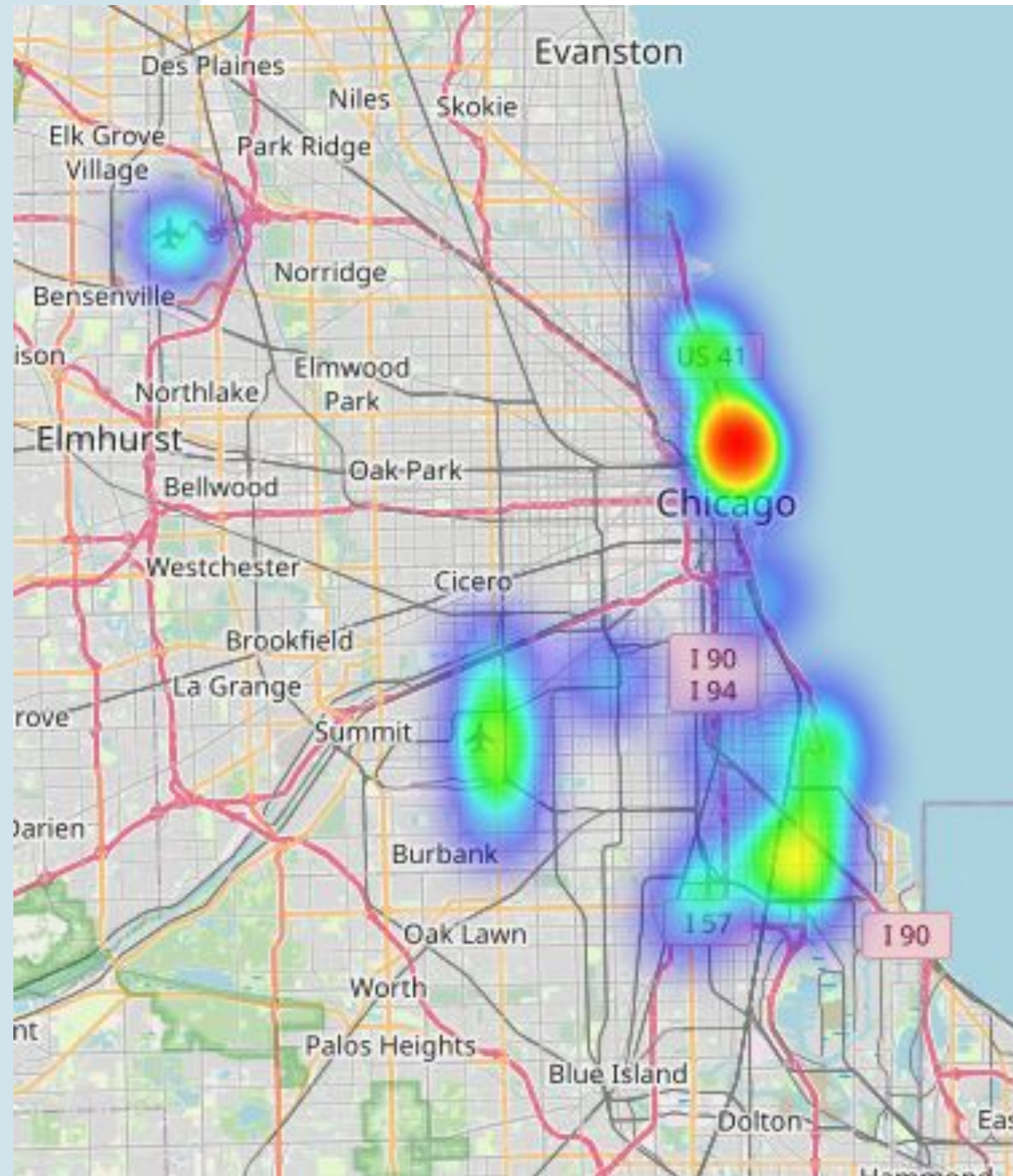


- **Used Geographical data (latitude and longitude)**
- **Our goal is to find traffic crash areas with high density**
- **DBSCAN**
  - **Unsupervised Learning Algorithm**
  - **Density-based algorithm**
  - **It groups together points that are close to each other based on density**
  - **Clustering algorithm**
  - **Used it to find areas with high concentration by removing as many noise points as possible**



# Machine Learning Analysis – Junha

## Traffic Crash Hotspots in Chicago

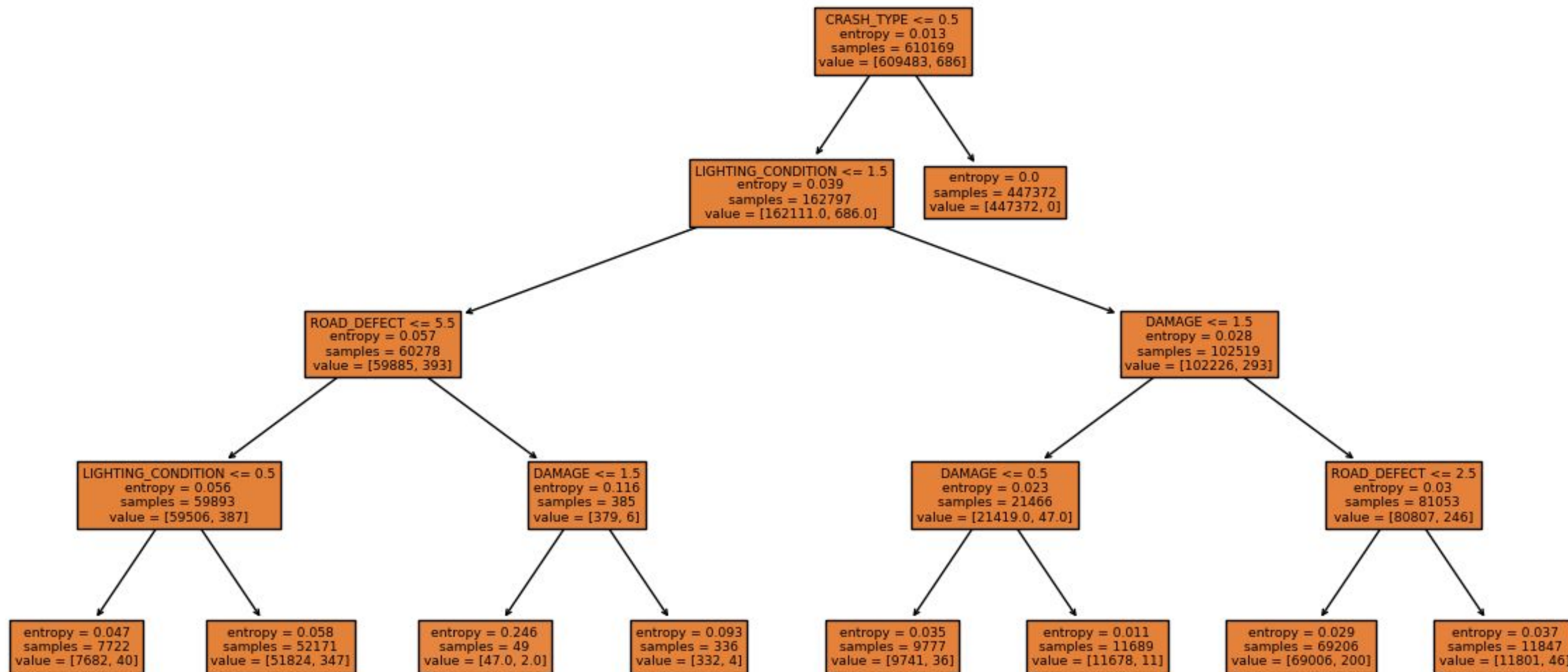


- **Significant traffic incidents around major landmarks and areas**
  - **O'Hare Airport**
  - **Midway Airport**
  - **South Side of Chicago**
  - **Downtown**
  - **North Side Area (Magnificent Mile and Navy Pier)**
- **High traffic volumes and urban density contribute to the increased crash occurrences**
- **It highlights urban planning interventions are needed to mitigate crash risks**



# Machine Learning Analysis – Prit

Predicting Fatal Crashes



Will the collision result in a fatal or non-fatal injury?

Deciding Factors:

- 'LIGHTING\_CONDITION'
  - Daylight
  - Darkness
- 'WEATHER\_CONDITION'
  - Snow
  - Rain
- 'CRASH\_TYPE'
  - Tow Required
- 'ROAD\_DEFECT'
  - Debris
  - Potholes
  - Shoulder

Accuracy Score: 99%+

# Takeaways

- **We uncovered some interesting insights from our project**
- **Learned how to use ML and other techniques to analyse large datasets**

## What went well

- **Team communication**
- **Amount of work completed**

## Challenges

- **Challenging to work with large data**
- **Challenging to decide what to do our analyses on**



