



# Chicago Car Crash Data

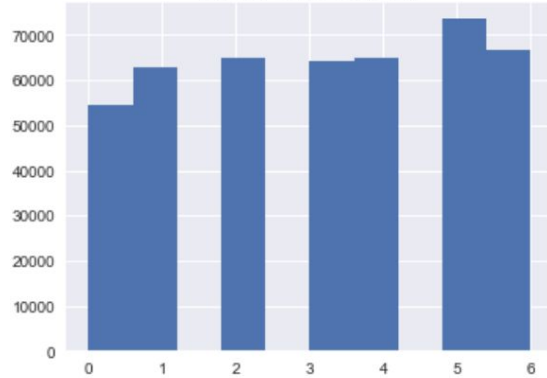
Vehicle Safety Board  
Speaker: Bonny Nichol  
November 17, 2020

## What information is in this dataset?

CRASH\_RECORD\_ID  
RD\_NO  
CRASH\_DATE\_EST\_I  
CRASH\_DATE  
POSTED\_SPEED\_LIMIT  
TRAFFIC\_CONTROL\_DEVICE  
DEVICE\_CONDITION  
WEATHER\_CONDITION  
LIGHTING\_CONDITION  
FIRST\_CRASH\_TYPE  
TRAFFICWAY\_TYPE  
LANE\_CNT  
ALIGNMENT  
ROADWAY\_SURFACE\_COND  
ROAD\_DEFECT  
REPORT\_TYPE  
CRASH\_TYPE  
INTERSECTION\_RELATED\_I  
NOT\_RIGHT\_OF\_WAY\_I  
HIT\_AND\_RUN\_I  
DAMAGE  
DATE\_POLICE\_NOTIFIED  
PRIM\_CONTRIBUTORY\_CAUSE  
SEC\_CONTRIBUTORY\_CAUSE  
STREET\_NO  
STREET\_DIRECTION

STREET\_NAME  
BEAT\_OF\_OCCURRENCE  
PHOTOS\_TAKEN\_I  
STATEMENTS\_TAKEN\_I  
DOORING\_I  
WORK\_ZONE\_I  
WORK\_ZONE\_TYPE  
WORKERS\_PRESENT\_I  
NUM\_UNITS  
MOST\_SEVERE\_INJURY  
INJURIES\_TOTAL  
INJURIES\_FATAL  
INJURIES\_INCAPACITATING  
INJURIES\_NON\_INCAPACITATING  
INJURIES\_REPORTED\_NOT\_EVIDENT  
INJURIES\_NO\_INDICATION  
INJURIES\_UNKNOWN  
CRASH\_HOUR  
CRASH\_DAY\_OF\_WEEK  
CRASH\_MONTH  
LATITUDE  
LONGITUDE  
LOCATION

CRASH\_DAY\_OF\_WEEK



Friday is the day of the week with the tendency to have the most accidents. Many people tend to rush home from work on weekends also increased traffic of people leaving the city for the weekend.

CRASH\_HOUR



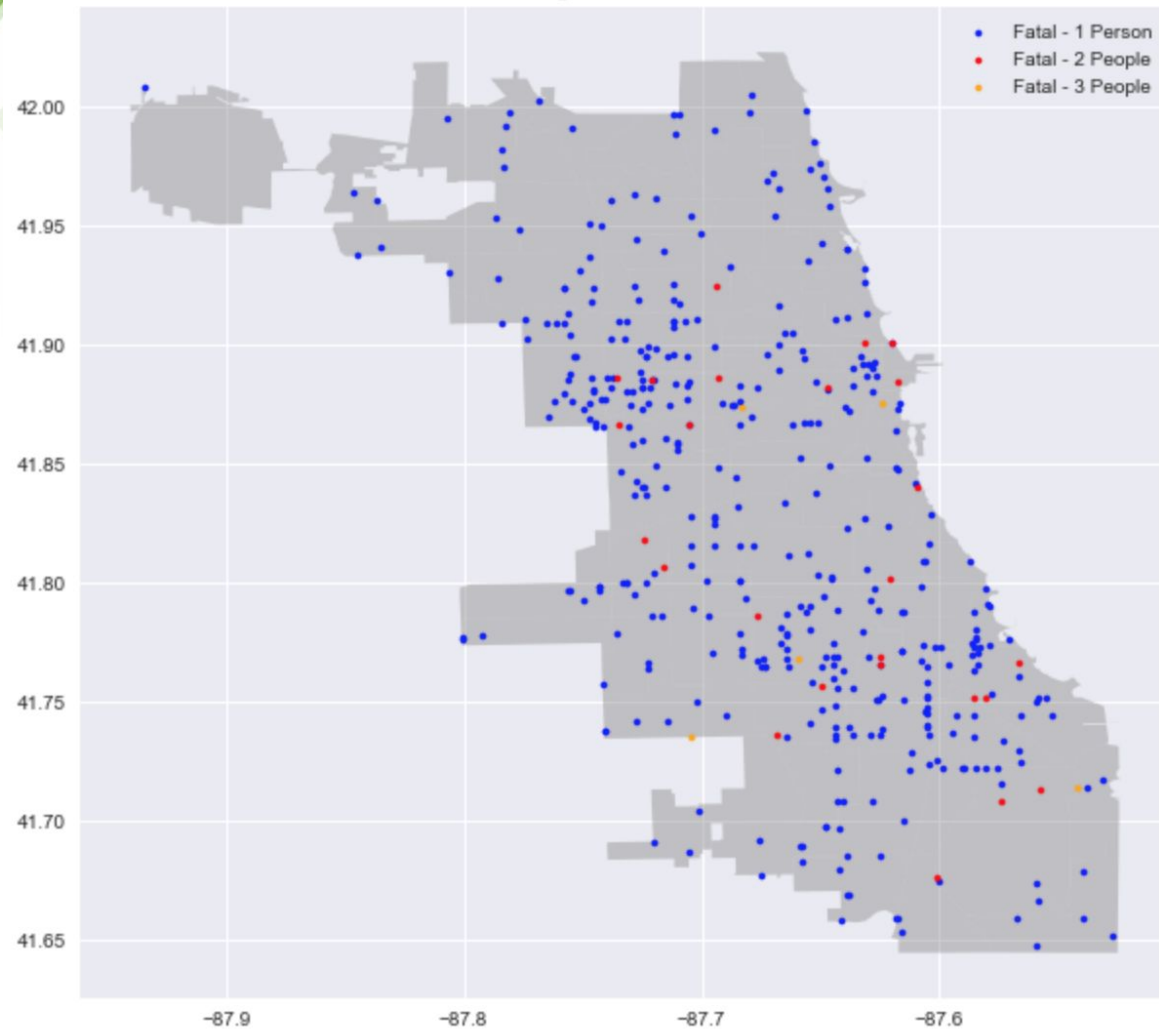
Mid to Late afternoon is the time of day when most crashes occur. This could be explained by rush hour and more traffic on the road at this time.

CRASH\_MONTH

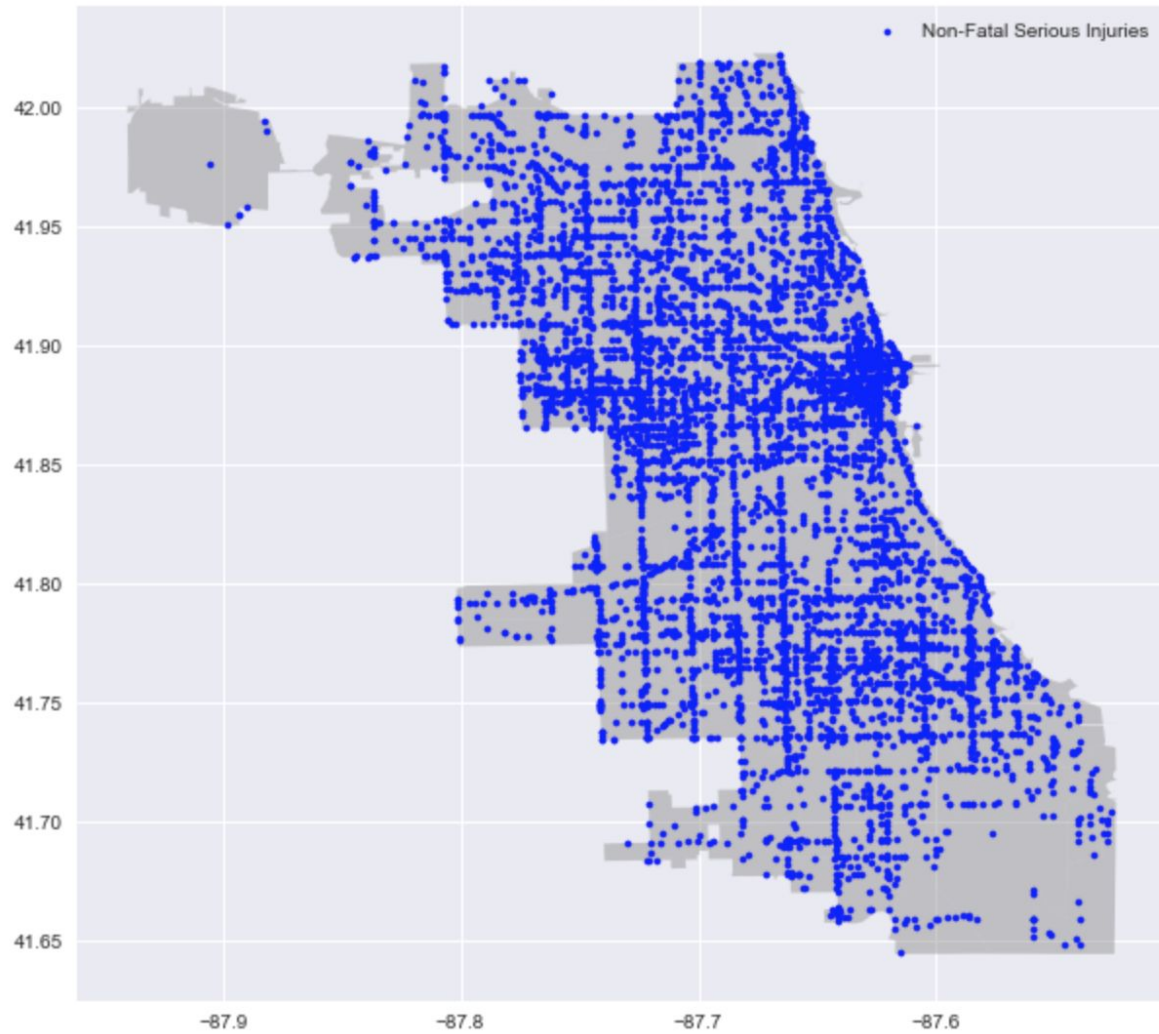


The end and the beginning of the year are the times of year when crashes most often occurred. This could be as a result of winter weather but also we could consider depression and reduced visibility during winter months.

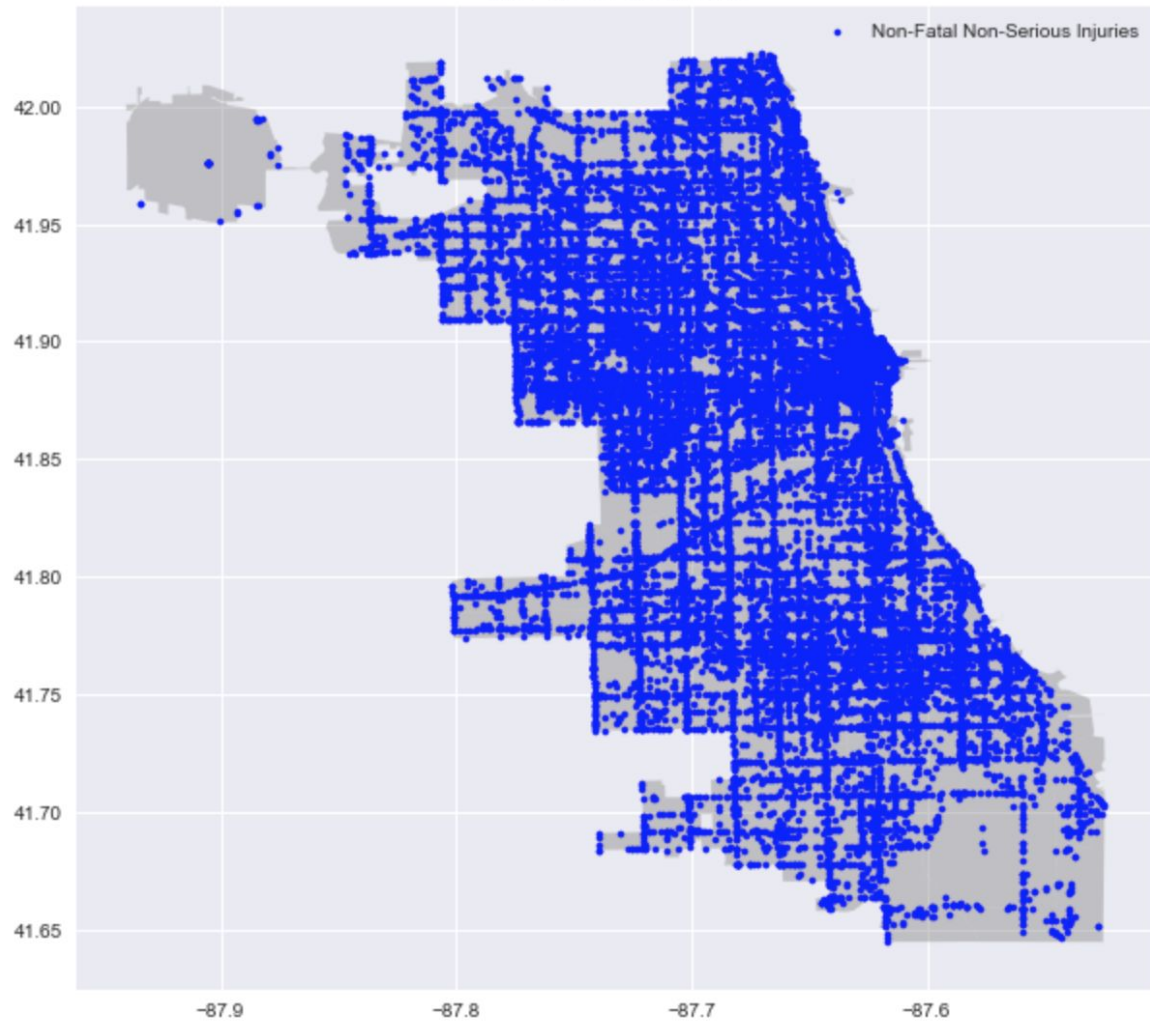
Chicago Car Crash Data



Chicago Car Crash Data



Chicago Car Crash Data





# Methodology

## ANALYSIS

The data science team used several different machine learning models in order to compare the results to find the most accurate model. From this they were able to determine the most important factors of what causes a deadly crash.

### Decision Tree

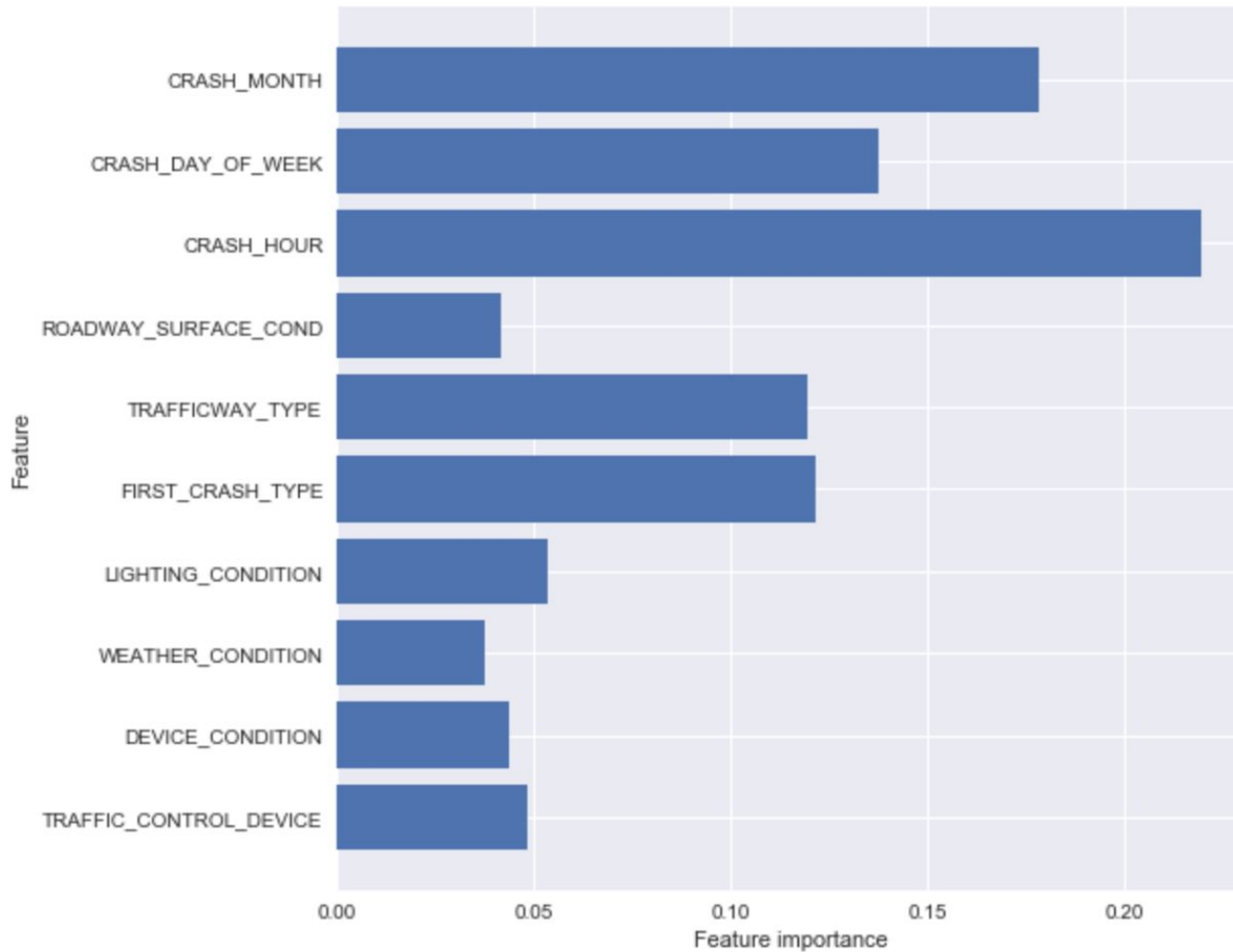
Old but powerful ML model that  
“decides” at each node

### Random Forest

Uses many Decision Trees to  
choose the best solution

### GridsearchCV

Cycles through parameter options  
to find the most optimal model



Results of  
Random Forest  
Classifier with  
Optimal  
Parameters



The slide features a decorative background with a pattern of overlapping squares in various shades of green and purple. The squares are arranged in a way that creates a sense of depth and movement, with some squares appearing more prominent than others. The overall color palette is muted and professional.

# Conclusions and Moving Forward

The Hour of a Crash has the biggest impact on fatal accidents. Potentially, the Vehicle Safety Board could advise for more officers to patrol during the hours or limiting speed limits during these hours to prevent more fatal accidents.

Geographic location indicates that certain areas are more likely to have fatal accidents. Conditions in these areas could be considered to see if there are any specific road problems that can be addressed.

Because this dataset is reliable (created from police reports) and also is constantly changing, I recommend that there be further and continuous review of this dataset to find more patterns that could help explain fatal (and also serious accidents).



**Thank you**