CTA Crash Report: Why do drivers crash?

Mitch Allison • FB4DWW • 08.26.2022

Questions to Answer:

- What are the most important variables to determining the type of crash?

- Based on variables, can we predict what type of crash has occurred?
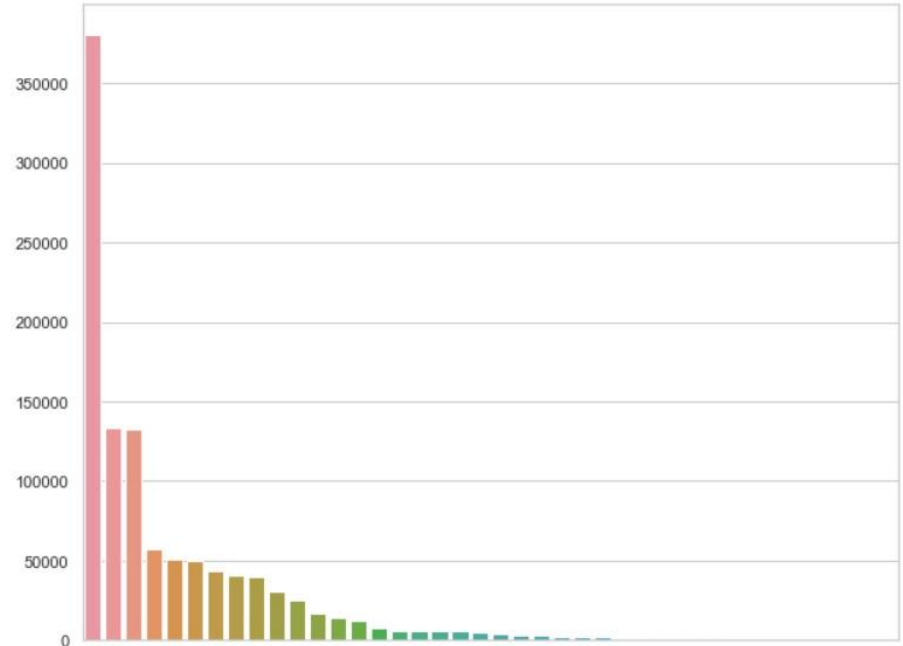
Data

Retrieved from:
[The City of Chicago data portal](#)

Contains:
- Crash condition data
- Car condition data
- Driver condition data

- 1 million+ datapoints

- 2015 to current

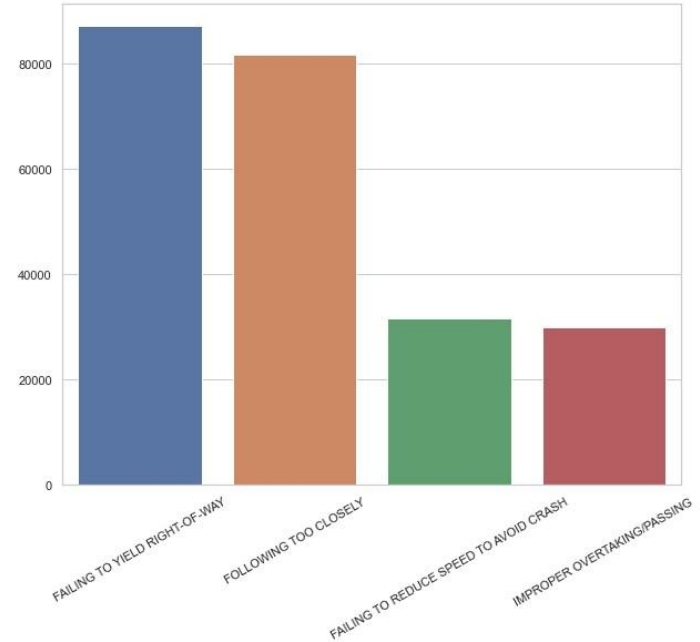- Focused on vehicle/vehicle collisions

Causes of Crashes

- 'Unable to determine' is greatest cause of accidents

- If person is unable to determine cause, model will do the same

Top 4 Causes of Crashes

Focus efforts:

- Limited causes to top 4

- Takes 23 inputs

Most important variables:

- Sex

- Age

Final Model

Performance stats:

- Precision: Predicted crash types match actual types 64%

- Recall: Correctly IDs 64% of crash types

Data Inputs: Simplified

How to make model deployable?

CLEAR
RAIN
UNKNOWN
SNOW
CLOUDY/OVERCAST
OTHER
FREEZING RAIN/DRIZZLE
FOG/SMOKE/HAZE
SLEET/HAIL
BLOWING SNOW
SEVERE CROSS WIND GATE
BLOWING SAND, SOIL, DIRT

VS

clear/unknown
rain
snow
other

Next steps

Continue to iterate upon model

- Model trained on 21% of total crashes.
- Include additional causes: only predicts top 4.
- Improve performance.

Look more into 'Unable to determine' cause

- Can we discover the true cause based on model?

Deploy model

- Answer 23 simple questions to find cause of crash.

MITCH ALLISON

GitHub
Linkedin
Email

If you have any questions/comments/concerns, I can be reached here!