



# PREDICTING PRIMARY CAUSES OF CHICAGO AUTO ACCIDENTS

Dominic Garcia



# Agenda

- Business Problem
- Data
- Recommendations
- Future Work

# Business Problem

- USAA wants to better understand the liability associated with their customers' accidents in order to determine the premiums they should be charging.
- As one of the country's biggest metropolitan areas with a variety of weather conditions, Chicago is a prime candidate for studying auto accidents.
- I am tasked with building a classifier that will let the company know, given the facts about a customer's accident, whether that customer is due for an increased premium.

# Data

- Data for this project is provided by the City of Chicago and is sourced from police reports.
- Over 70,000 auto accidents were analyzed for model-building.
- More on the data at <https://data.cityofchicago.org/Transportation/Traffic-Crashes-Crashes/85ca-t3if>

# Data

- Target variable of dataset & study is 'Primary Cause,' which has four categories:
  - *Outside Hazard*
  - *Ignoring Traffic Signs & Warnings*
  - *Reckless Driving*
  - *Impairment/Distraction*

# Results & Recommendations

- The model's prediction values are, in order from most to least dependable (based on performance metrics): 'Ignoring Traffic Signs & Warnings', 'Outside Hazard', 'Reckless Driving', & 'Impairment/Distraction'
- If a client's accident cause is determined to be 'Ignoring Traffic Signs & Warnings': Consider increasing their premium.
- If a client's accident cause is determined to be 'Outside Hazard': Consider minimal or no increase to the premium, as the accident has been caused by something out of the client's control.
- If a client's accident cause is determined to be 'Reckless Driving': Consider increasing their premium. Additionally, flag the client as a candidate to drop if the reckless driving continues. It is harder to write off such causes as accidental or momentary, like one could with 'Ignoring Traffic Signs & Warnings.'
- If a client's accident cause is determined to be 'Impairment/Distraction': Do not make any conclusions about the case and use the prediction as a prompt to investigate more closely. This is due to both the class' poor performance metrics & its relative ambiguity ('Impairment/Distraction' covers everything from drunk driving to medical emergencies while operating a vehicle).

# Results & Recommendations (cont.)

- These are the factors that consistently were most relevant in determining the primary cause of a given accident:
  - First crash type: at an angle, during a turn, collision with a parked vehicle, rear ending
  - Outdoor conditions: clear, snowy
  - Whether there was a traffic control device at the scene
  - Whether the accident involved a personal injury or a towed vehicle
  - Whether the accident was a hit-and-run
  - Road surface conditions: ice, snow/slush, wet

# Future Work

- Look more closely at coefficients of my model and determine which factors are most impactful for each specific outcome.
- Inspect the city's dataset of people involved in these same accidents. Using common indices across the two datasets, I could create a column containing the number of passengers in each accident and explore the impact that this variable has on the model's predictions.
- Look into whether I could break down the 'Impairment/Distraction' category any further. Perhaps, as this class undergoes specification, the performance for both the new class(es) and overall model could improve significantly.





# QUESTIONS?

Github: <https://github.com/dl-gd>

Email: [dlgarcia.017@gmail.com](mailto:dlgarcia.017@gmail.com)

