Analysing Vehicle Collision Severity

Elena Gerken
October 2020

Background

- Road vehicle collisions are the leading cause of death for people aged 1-54 in the US.

- We will examine the factors influencing the severity of accidents.

- The goal is to determine which behaviour or conditions regulators and police should focus on.

The Dataset

- The city of Seattle published a dataset of 221,738 records of accidents going back to 2004.

- The target variable 'Severity' has four different outcomes:
  - Property damage only collisions
  - Injury collision
  - Serious injury collision
 - Fatality Collision

Feature selection

Out of 40 possible columns, I chose the ones with the best predictive value:

- Weather condition (clear, raining, overcast, snowing, fog etc.)
- Road condition (dry, wet, standing water, ice etc.)
- Light condition (daylight, dark, streetlights on/off etc.)
- Whether or not the collision was due to inattention. (Y/N)
- Whether or not a driver involved was under the influence of drugs or alcohol.
- Whether of not a driver was speeding.
- The number of pedestrians involved in the collision.
- The number of cyclists involved in the collision.

Balancing the Dataset

- The target variable 'Severity' is heavily balanced towards the outcome 'Property damage only collision'.

- Leaving the dataset unbalanced could result in a model that would never predict a fatal collision.

- I resampled the outcomes of the variables to achieve an even distribution with 3.111 cases per severity category.
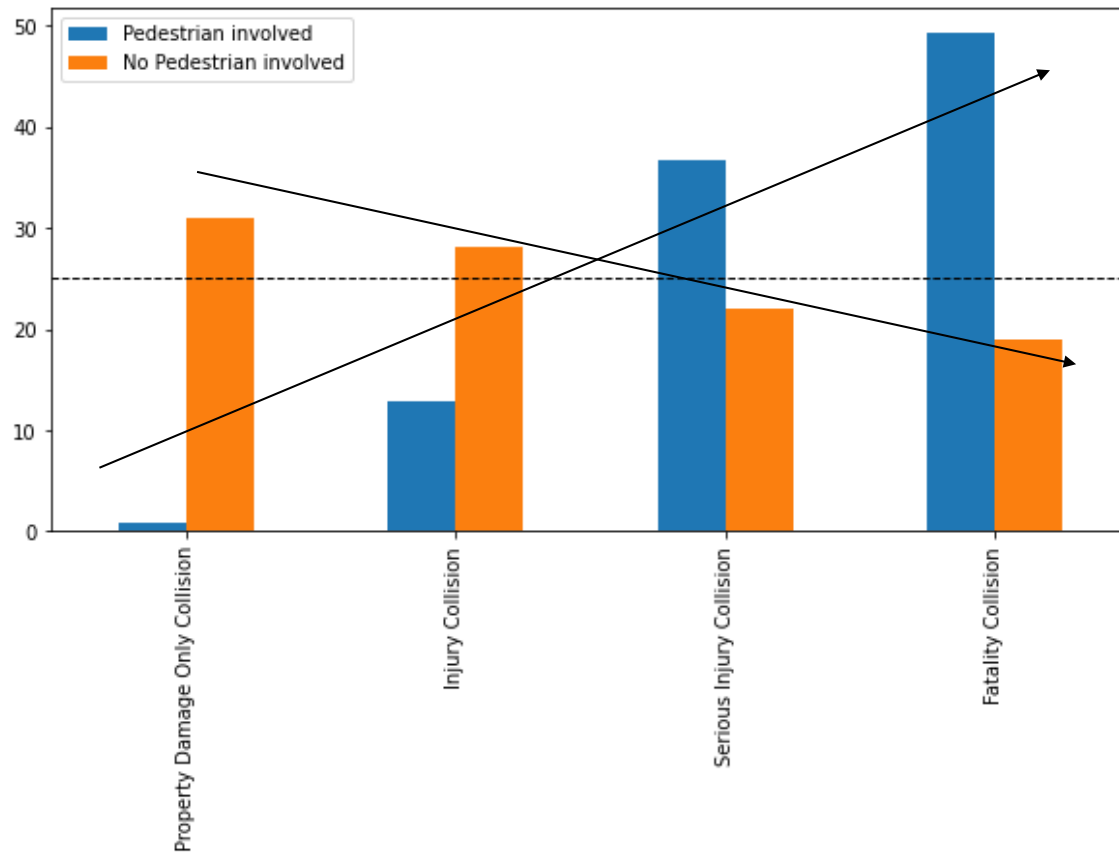
Cyclists
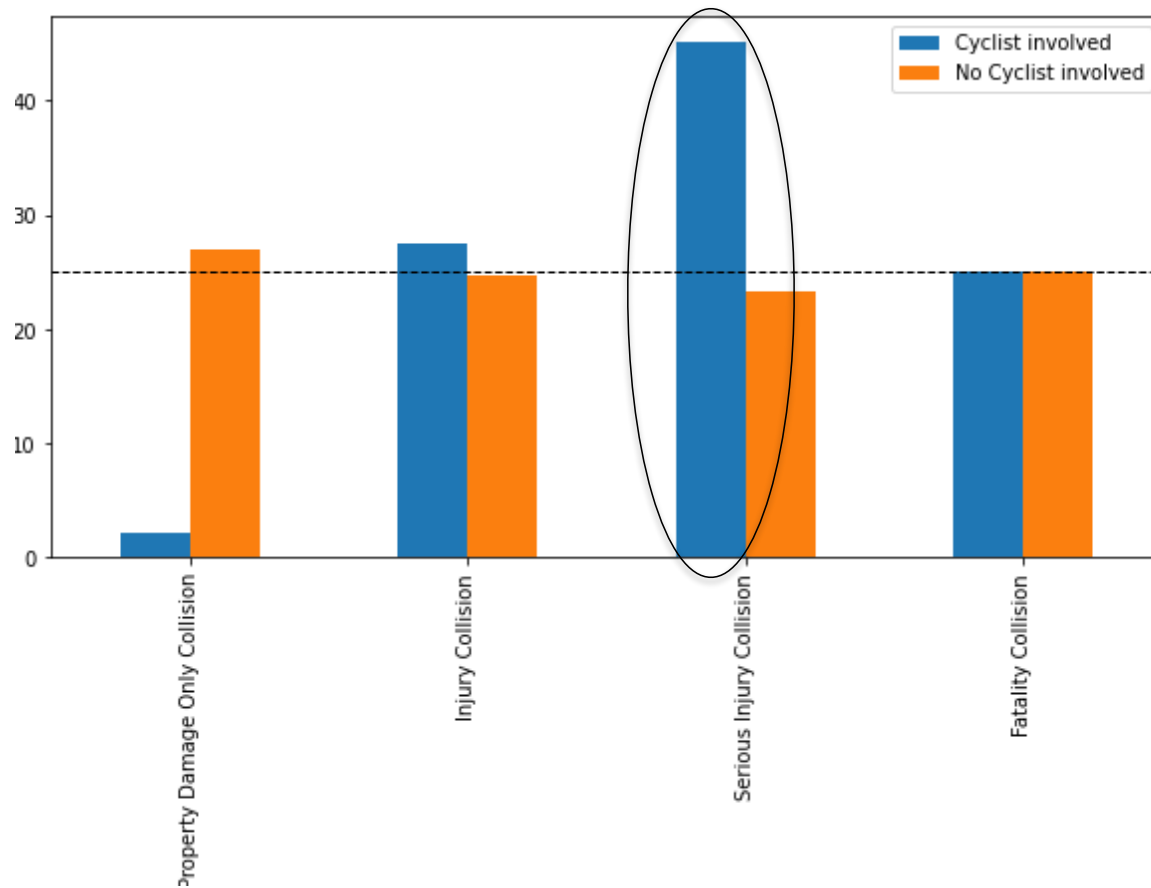
- Cyclist involvement makes an accident prone to end with serious injuries.

- Accidents with cyclists are less likely to produce property damage only.
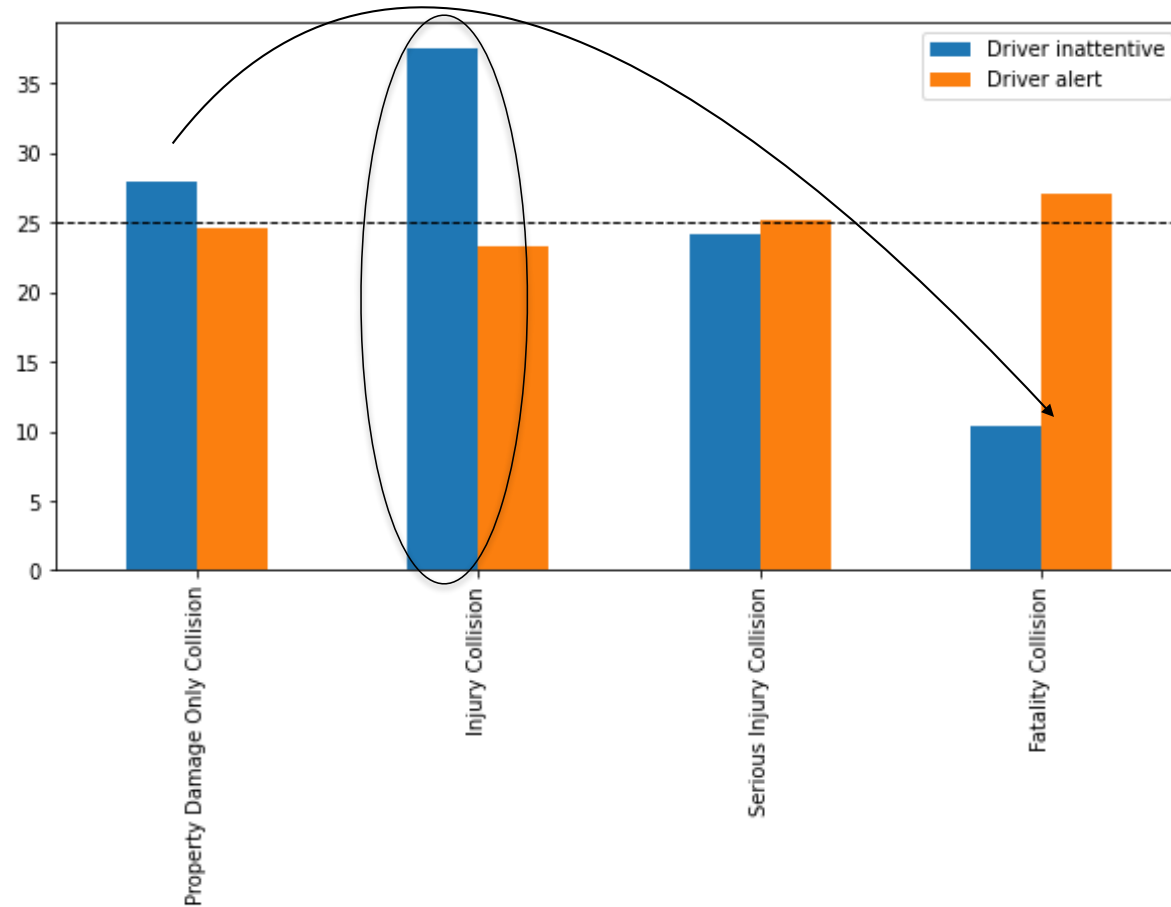
Drugs and Alcohol
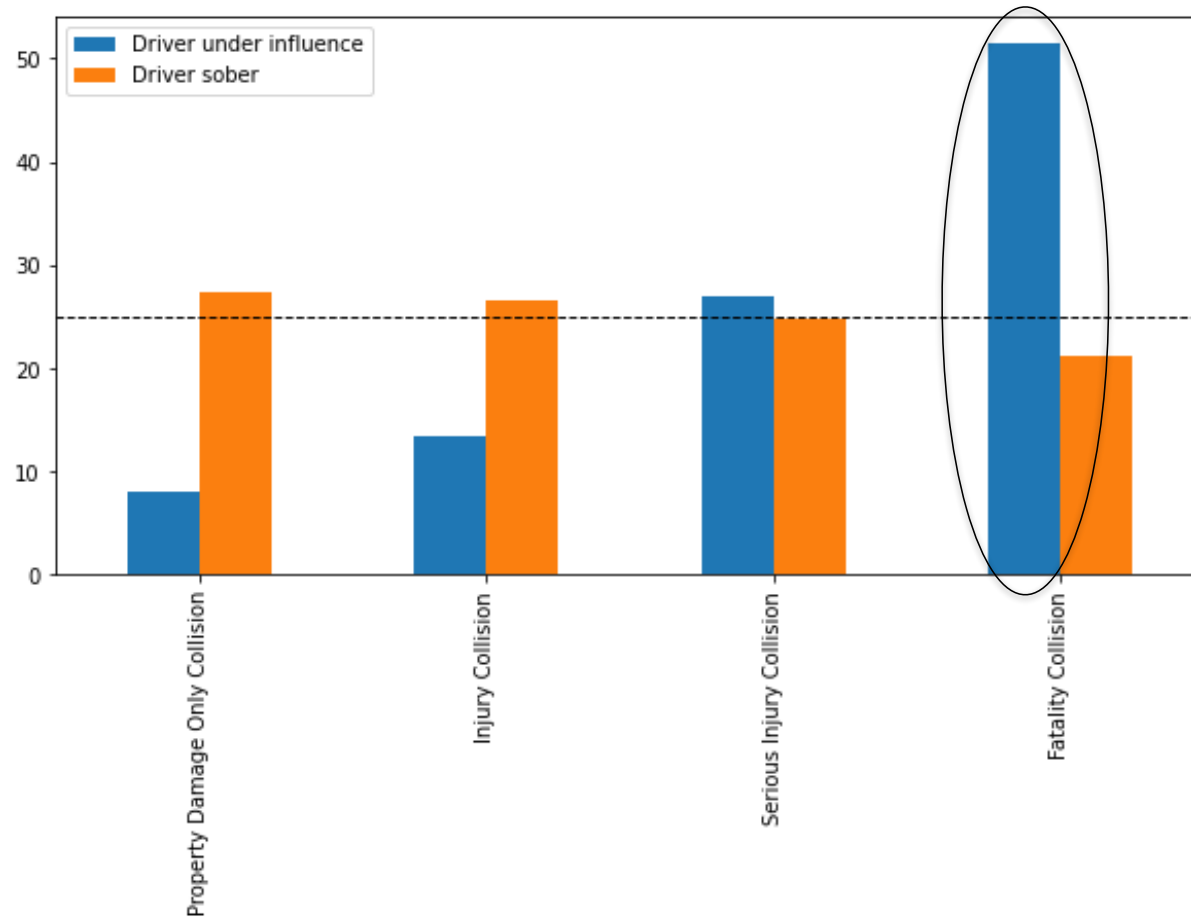
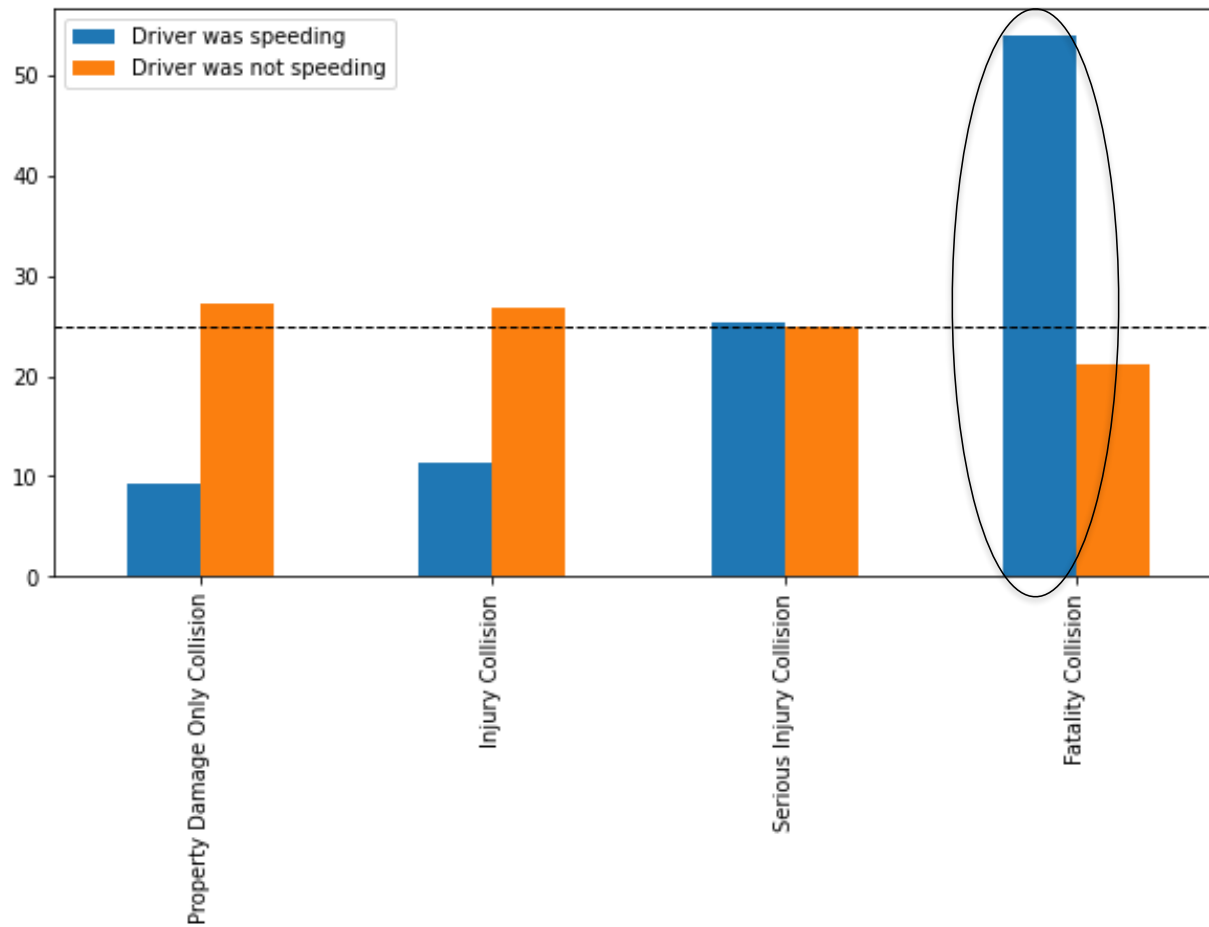- Very strong correlation between intoxicated driving and fatal accidents.

Pedestrians

- Pedestrian involvement makes an accident prone to end with serious injuries or fatalities.

- Accidents without pedestrians are more likely to be less severe.

Driver Inattention

- Driver inattention produces proportionately more injury collisions.

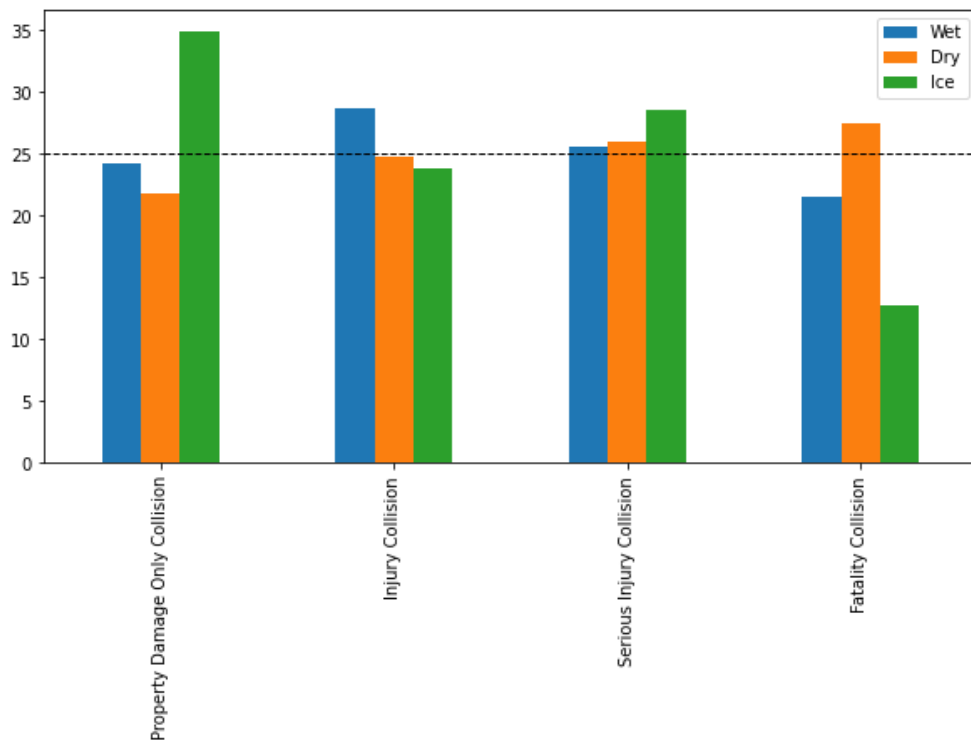- These accidents tend to be less fatal over all.

Speeding

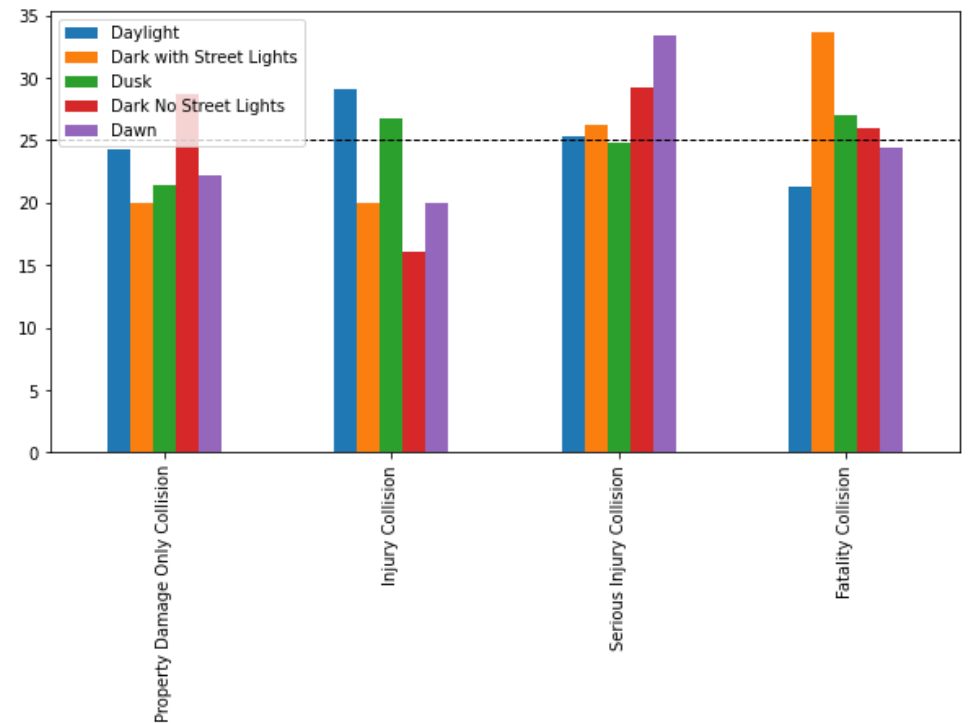- Very strong correlation between speeding and fatal accidents.

External conditions

- Weather, road and light conditions did not have a similarly clear effect as the variables presented above.
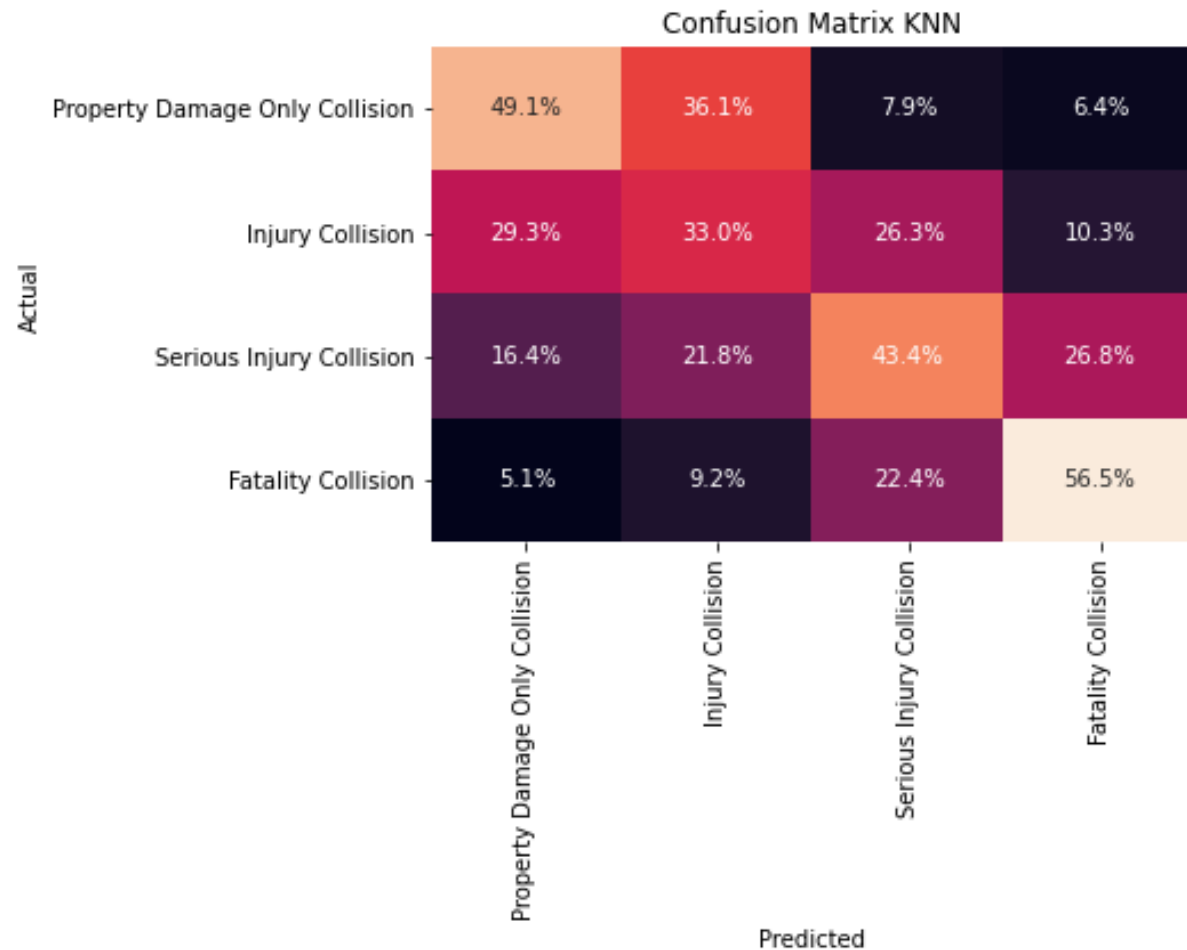

Weather condition

Light condition

Predictive Models

- I developed four classifiers to predict the severity of vehicle collisions:

  - K-nearest-neighbour

  - Decision tree

  - Logistic Regression

  - Support Vector Machine

- All four did not perform great (F1 score $\approx$ 0.4)

Evaluation

Confusion Matrix KNN

	Property Damage Only Collision	Injury Collision	Serious Injury Collision	Fatality Collision
Property Damage Only Collision	49.1%	36.1%	7.9%	6.4%
Injury Collision	29.3%	33.0%	26.3%	10.3%
Serious Injury Collision	16.4%	21.8%	43.4%	26.8%
Fatality Collision	5.1%	9.2%	22.4%	56.5%

Actual (vertical axis) — Predicted (horizontal axis)

- All models performed best at predicting the outcome 'Fatality collision'.

- They struggled with differentiating between accidents that were close in severity, for example:
  - confusing a 'Property Damage Only Collision' with 'Injury Collision' (28-36%)
  - predicting a fatal collision the accident only produced serious injuries (26-30%)

Results

- The three biggest contributors to fatal accidents were pedestrian involvement, speeding and drunk driving.

- Even though the prediction models were not perfect, finding the strong correlation of these three variables can help policy makers and police to direct their efforts in reducing the severity of accidents.