CENTENNIAL COLLEGE

2022 Safe Roads Competition

Alka Bhambhu

Chana Sherrington

Vincent Kwan

Mark Perez

CENTENNIAL COLLEGE

Our Understanding of the Problem

Overview of the Case

"The main objective of this case is to understand the factors that affect traffic safety utilizing the datasets provided by Geotab, the Toronto Police Service, and other open data sources.

The aim of the project is to extract insights from the data and provide specific recommendations to improve traffic safety in your municipality."

- 2022 SAS Safe Roads Case Document

Primary Datasets Applied

TORONTO'S ROAD SAFETY PLAN VISIONZERO

Our Approach

Points of Inquiry

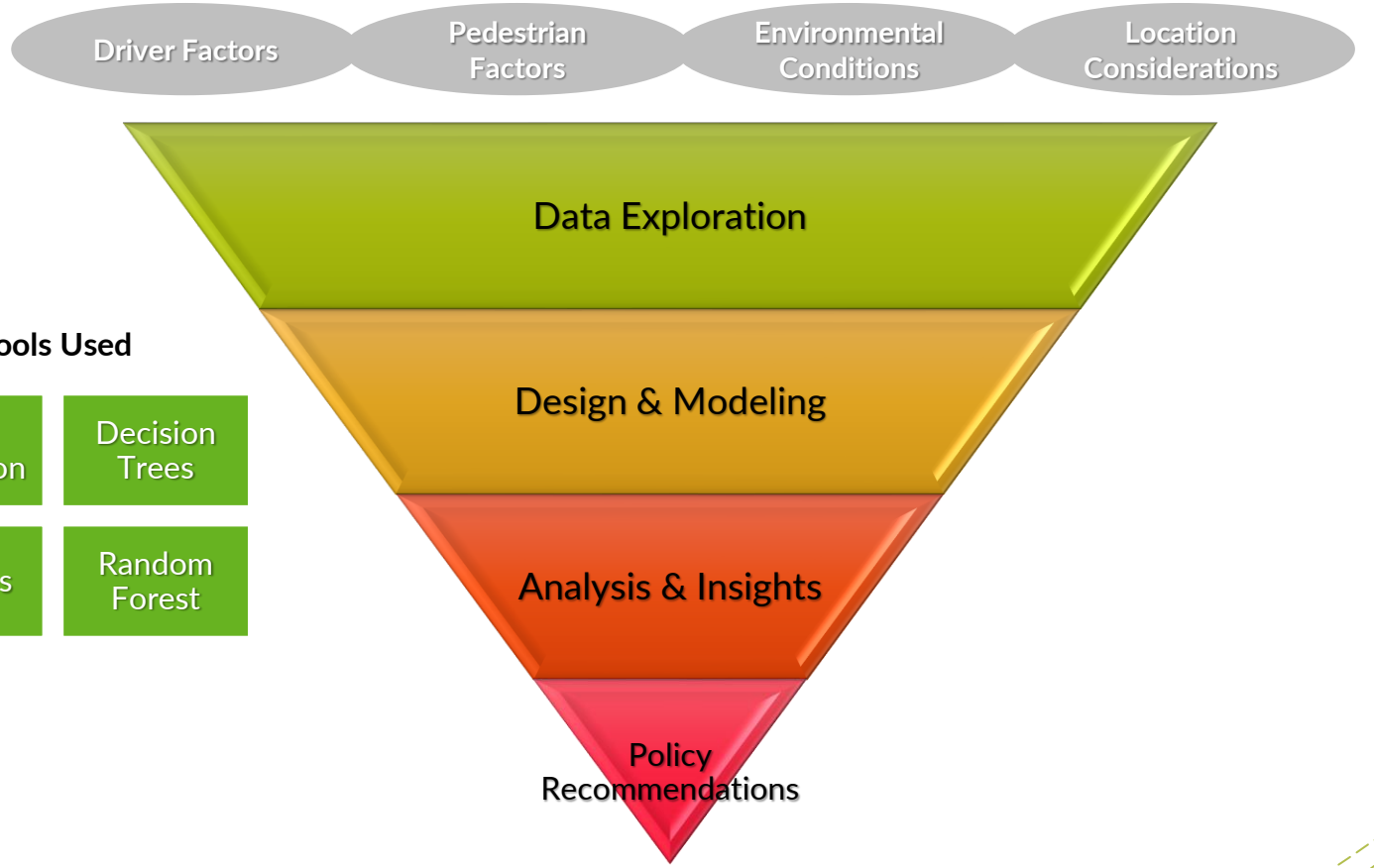
How might we reduce the occurrence of accidents in Toronto?

Which factors drive the occurrence of accidents?

- Which **driver-related** factors contribute the most to the occurrence of accidents?
- Which **pedestrian-related** factors contribute the most to the occurrence of accidents?
- Which **environment-related** factors contribute the most to the occurrence of accidents?
- Are the factors that contribute to the occurrence of accidents the same for **Downtown vs. Non-Downtown** areas?

Our Approach

Solution Operationalization



Tools Used

- Linear Regression
- Decision Trees
- K-modes
- Random Forest

Our Approach

Data Preparation

- Several datasets were created from merging Hazardous Driving Areas, Road Impediments, and KSI data.
- Records were aggregated and merged based on Latitude and Longitude – with the most common factors (pairwise mode of features) being taken.
- Some interval variables were transformed into binary variables prior to creating decision trees.
- Variables were not scaled or normalized to preserve interpretability.
- Since VisionZero datasets were not downloadable, data from the site was used as an overlay to inform observations.

Our Approach

Issues & Limitations (and how they were addressed)

- Records were aggregated on the basis of Latitude and Longitude truncated to 3 decimal places since not truncating would have led to a sparse dataset. This gave us an accuracy of about 100 meters.
- Curse of Dimensionality: Few records in comparison to the many variables in the dataset. Feature engineering was very important. But in the process, we lost certain aspects of the dataset.
- Blind spots in the data: There are known unknowns and unknown unknowns. Finding the elusive unknown unknowns is a matter of great skill and luck!
- Lack of complete causal factors: Our dataset did not contain the complete set of causal factors that lead to KSI events.
- Multicollinearity was a major issue that had to be dealt with in the dataset.

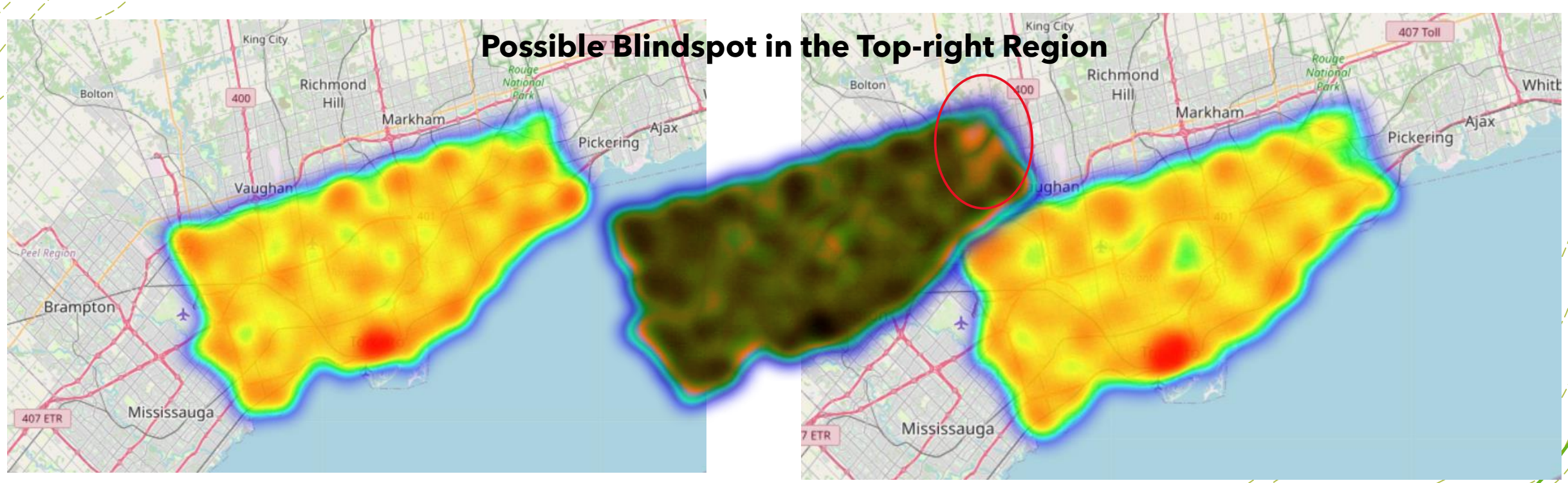
What We Found

Visualizing Incident Occurrence

Non-fatal Accidents

KSI events

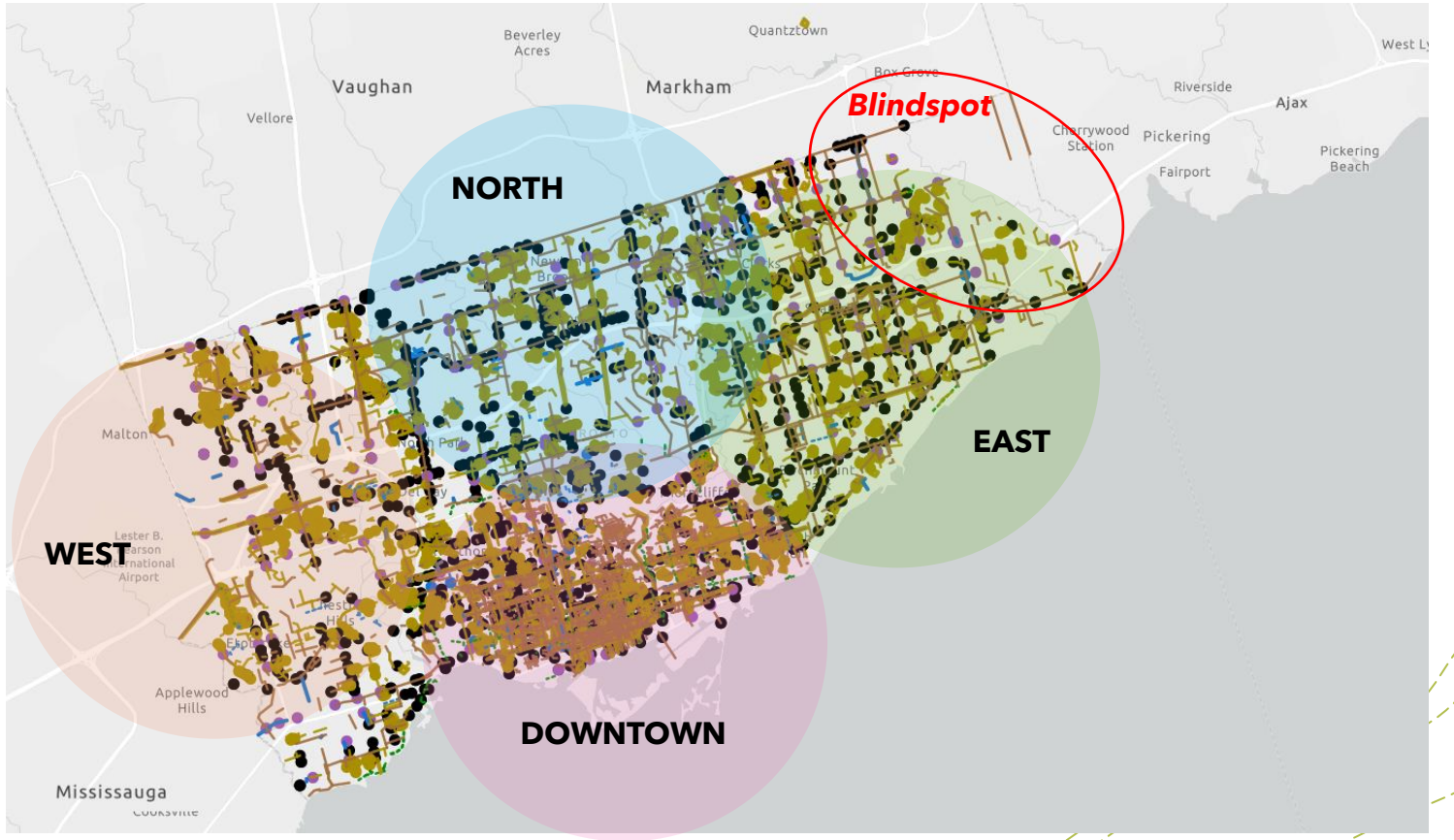
Possible Blindspot in the Top-right Region



What We Found

Visualizing Traffic Improvement Measures

- Traffic improvement measures appear to have been highly concentrated in the DOWNTOWN and EAST areas which generally matches the high accident incidence areas.
- There appears, however, some sparsity of interventions in the NORTH and WEST portions of the map except for some key areas.
- There also appears to be a blindspot in the Northeast corner perhaps due to the relatively fewer number of KSI incidents found in that area.



What We Found

Behavioural & Environmental Factors

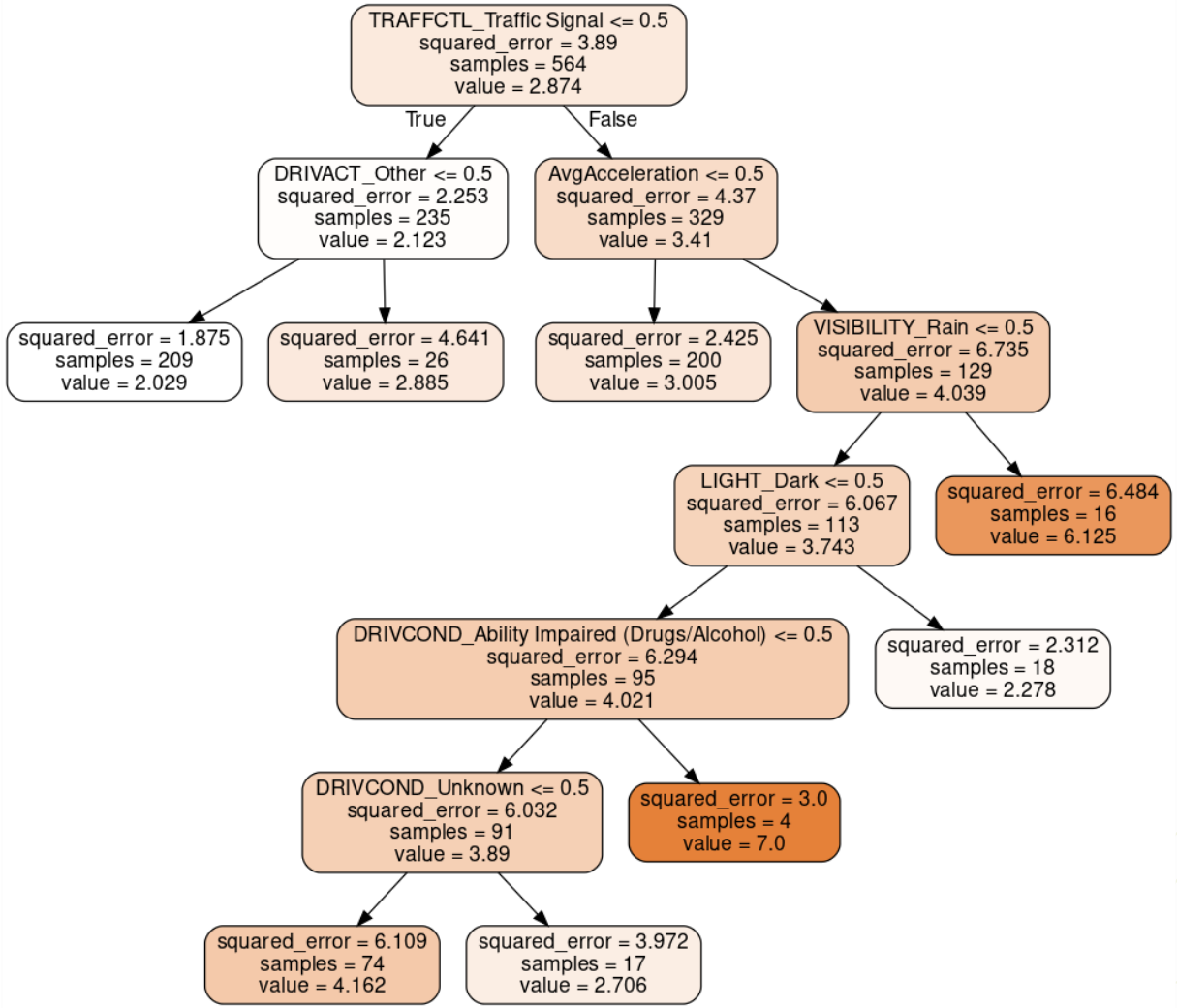
- Dataset Used = Hazardous Driving Areas (Geotab) + Road Impediments (Geotab) + KSI
- Predicting Count of KSI Events
- Using Tree-based methods and Linear Regression

Variable Name	Description
SeverityScore	Score given based on frequency and severity of Harsh Braking Incidents Transformed to a binary variable for Trees
DRIVACT	Driver's action before accident
DRIVCOND	Driver's condition before accident
TRAFFCTL	Traffic Control measures at place of accident
INVAGE	Age of Involved Party (Driver)
AvgAcceleration	Transformed to a binary variable for Trees
VEHTYPE	Categorized based on size
VISIBILITY	Environment Condition
PercentOfVehicles	Percentage of Vehicles affected by the Impediment Transformed to a binary variable for Trees

What We Found

All Behavior & Environment Factors

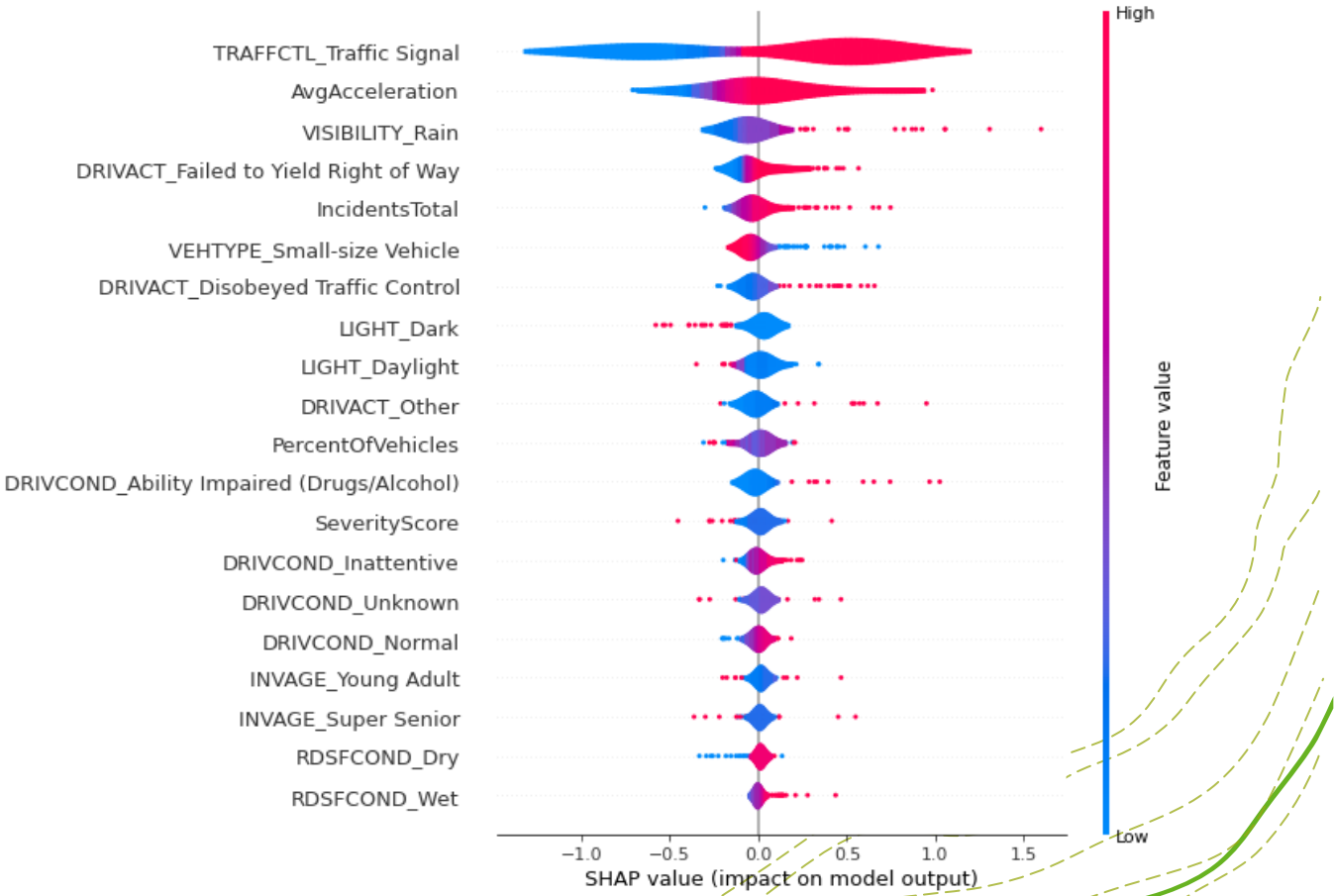
This Decision Tree displays the major driver behavioural and environmental factors in accidents. Most accidents occur at locations with traffic signals installed, and during the day. Intoxicated and ability-impaired driving leads to greater KSI events.



What We Found

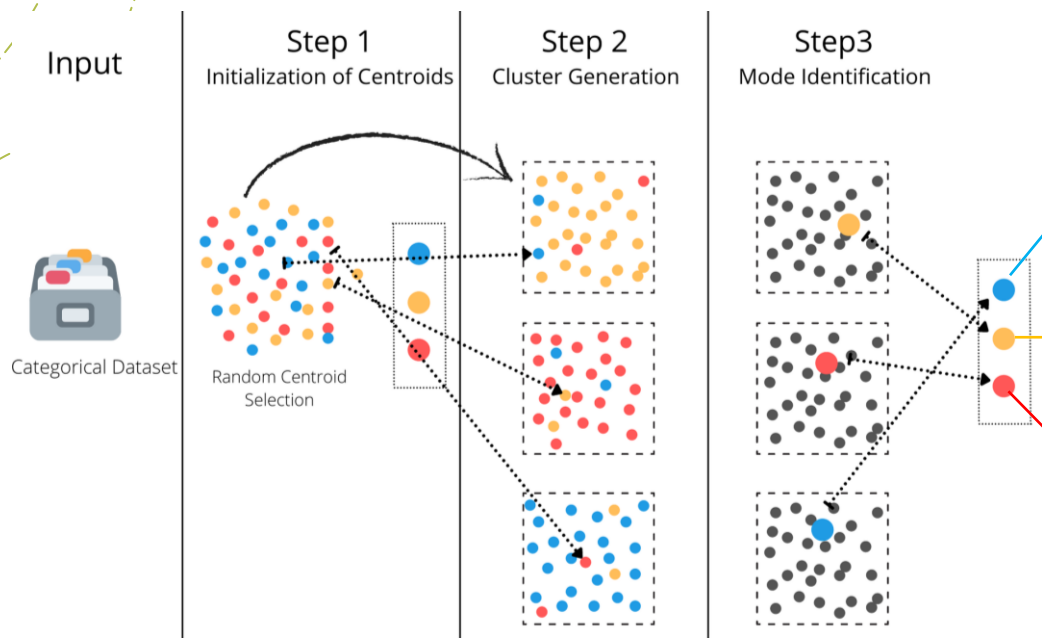
All Behavior & Environment Factors

- Most accidents occur at intersections - thus, the model predicts greater counts of KSI when traffic signals are present.
- AvgAcceleration is a major factor leading to greater KSI events.
- Adverse road conditions like Rain and wet roads lead to more accidents.
- Inadequate lighting conditions lead to greater KSI events.
- Impaired Ability of Drivers leads to greater KSI events.



What We Found

K-modes Clustering to Group Driver's behaviour



Sloppy Drivers
- Inattentive
- Improper Turns
- **Senior & Young drivers**

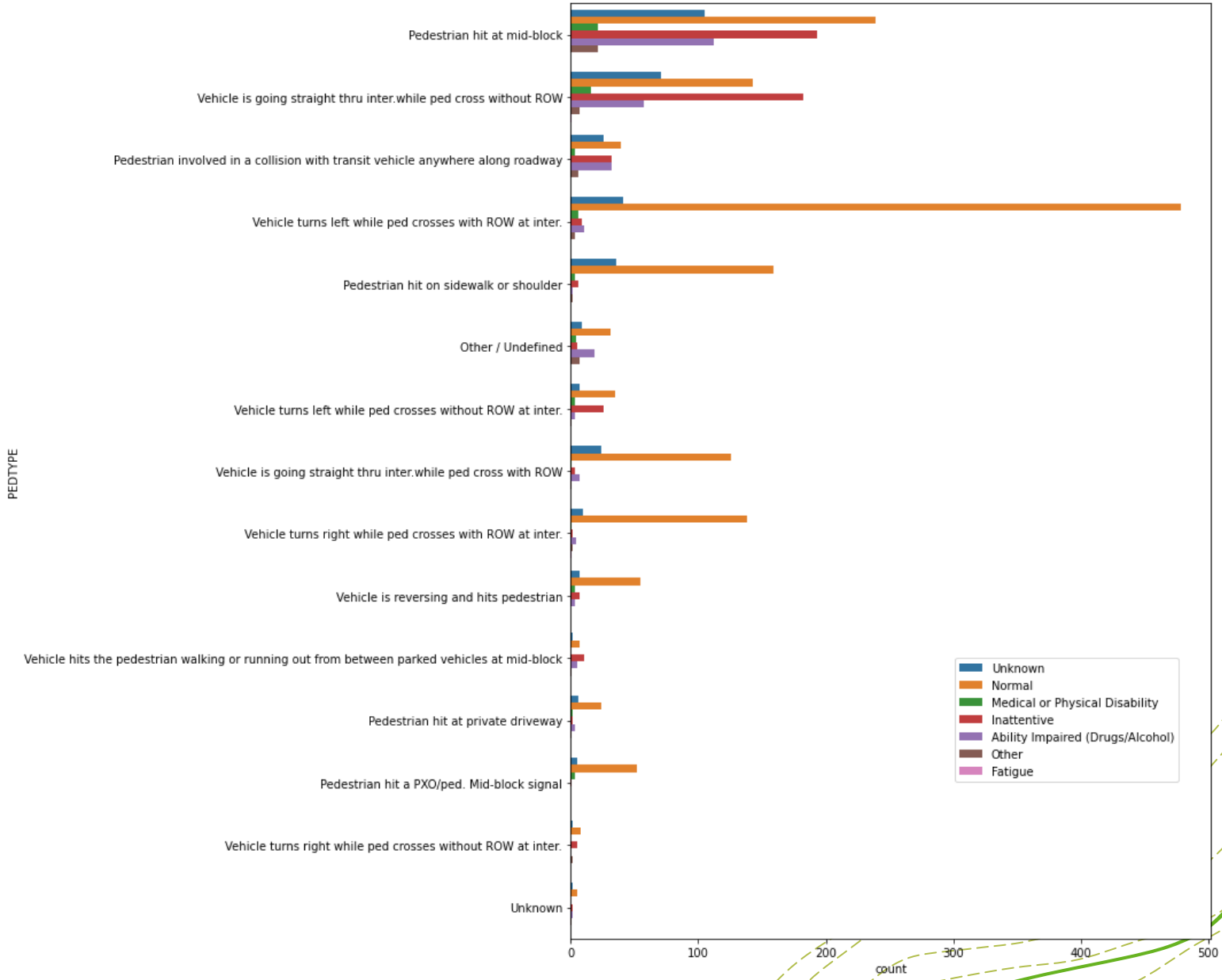
Hot Heads
- Aggressive Driving
- Lost Control
- **Young drivers**

Ability Impaired
- Failed to Yield Right of Way
- Speeding
- **Adult drivers**

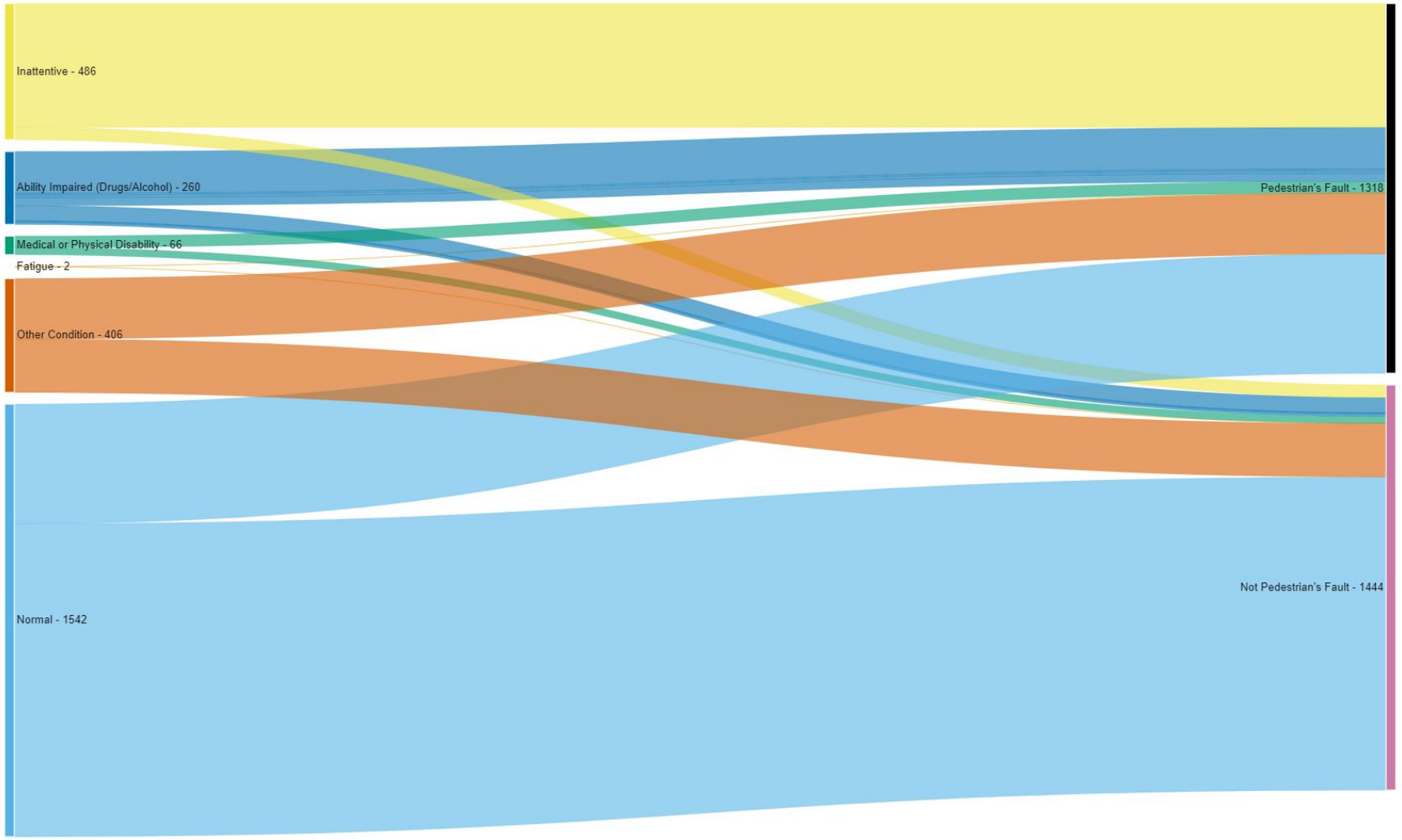
What We Found

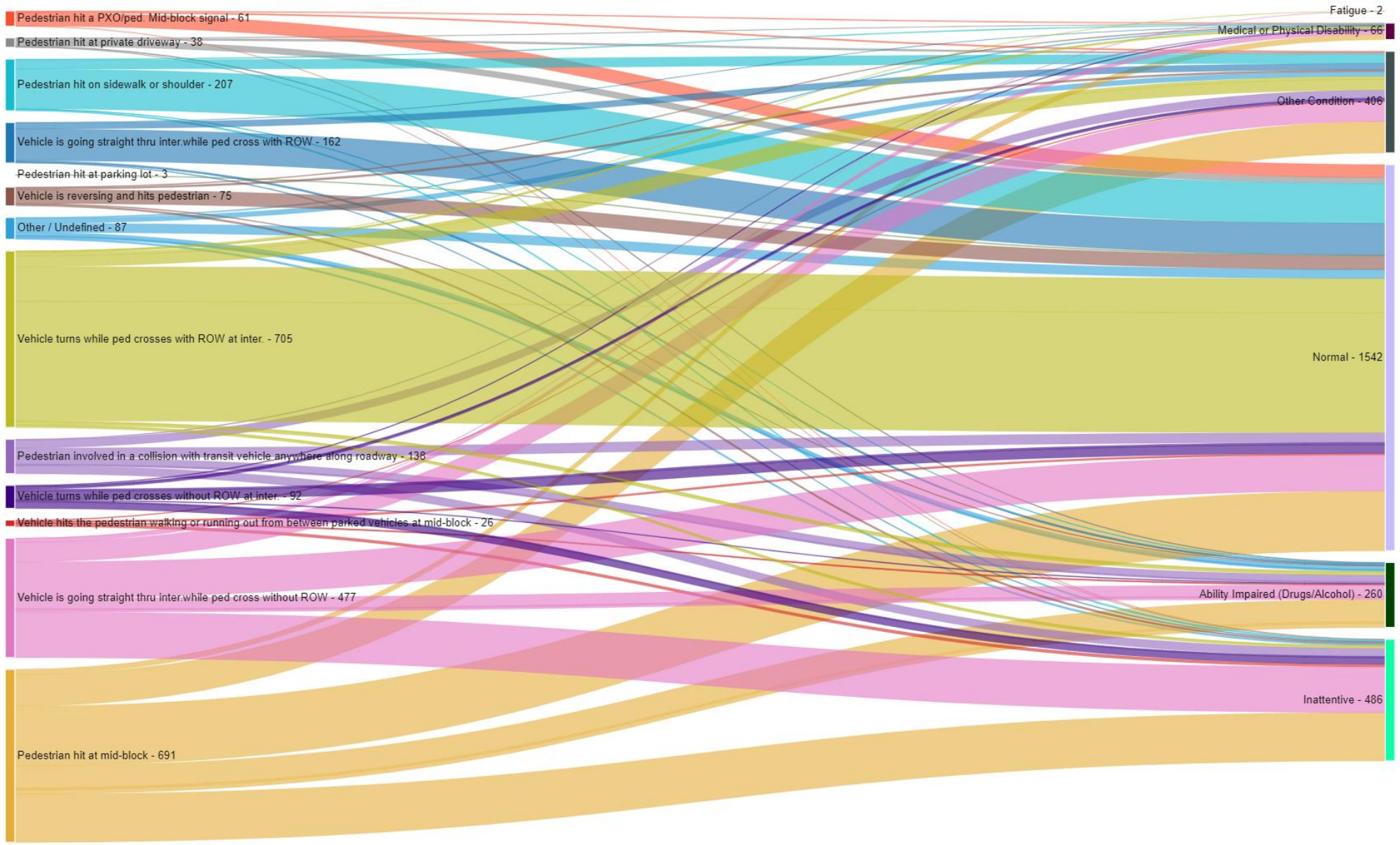
Pedestrian Factors

- Pedestrian inattentiveness is the top cause of accidents where a vehicle is going straight through an intersection and the pedestrian has no ROW. It is the second top cause in accidents where pedestrians are hit at midblock.
- In all other scenarios, pedestrians were tagged as ‘normal’, indicating that the pedestrian was not the cause of the accident.



Pedestrian Condition vs Fault




Pedestrian hit a PXO/ped. Mid-block signal - 61
Pedestrian hit at private driveway - 38
Pedestrian hit on sidewalk or shoulder - 207
Vehicle is going straight thru inter.while ped cross with ROW - 162
Pedestrian hit at parking lot - 3
Vehicle is reversing and hits pedestrian - 75
Other / Undefined - 87
Vehicle turns while ped crosses with ROW at inter. - 705
Pedestrian involved in a collision with transit vehicle anywhere along roadway - 138
Vehicle turns while ped crosses without ROW at inter. - 92
Vehicle hits the pedestrian walking or running out from between parked vehicles at mid-block - 26
Vehicle is going straight thru inter.while ped cross without ROW - 477
Pedestrian hit at mid-block - 691
Fatigue - 2
Medical or Physical Disability - 66
Other Condition - 406
Normal - 1542
Ability Impaired (Drugs/Alcohol) - 260
Inattentive - 486

CENTENNIAL COLLEGE

What We Found

Pedestrian Factors

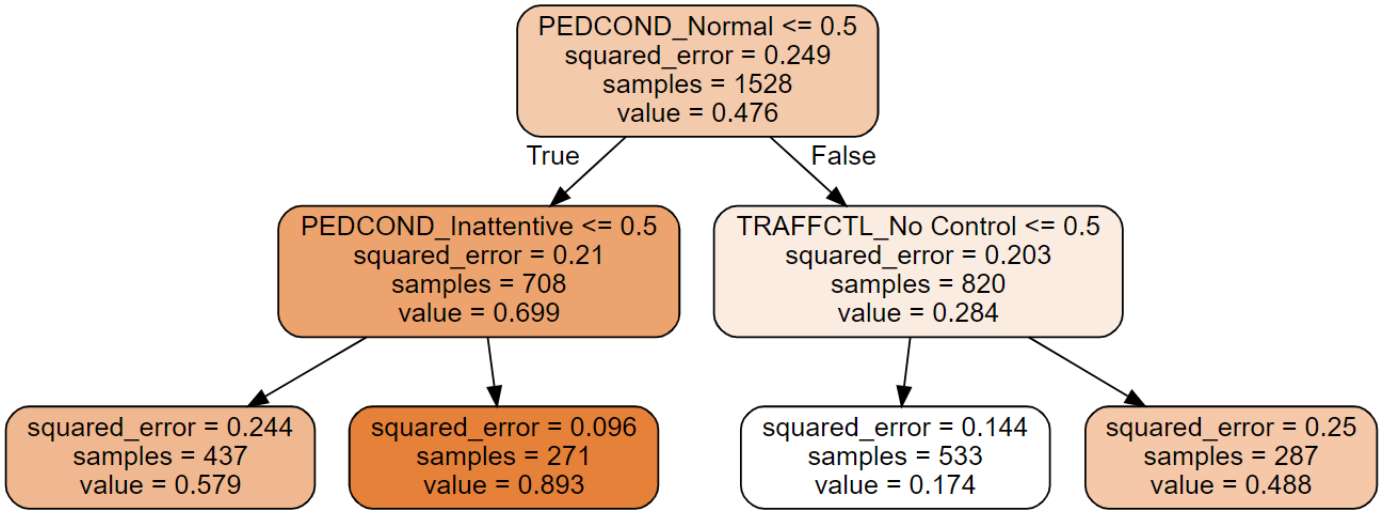
- A mix of pedestrian, and environmental related variables were used in modeling for accidents with pedestrian factors in mind.

Variable Name	Description
COUNT	Number of accidents
PEDTYPE	Driver's action that led to pedestrian collision
PEDACT	Pedestrian's action before accident
PEDCOND	Pedestrian's condition before accident
TRAFFCTL	Traffic Control measures at place of accident
LIGHT	Environment condition
VISIBILITY	Environment Condition

What We Found

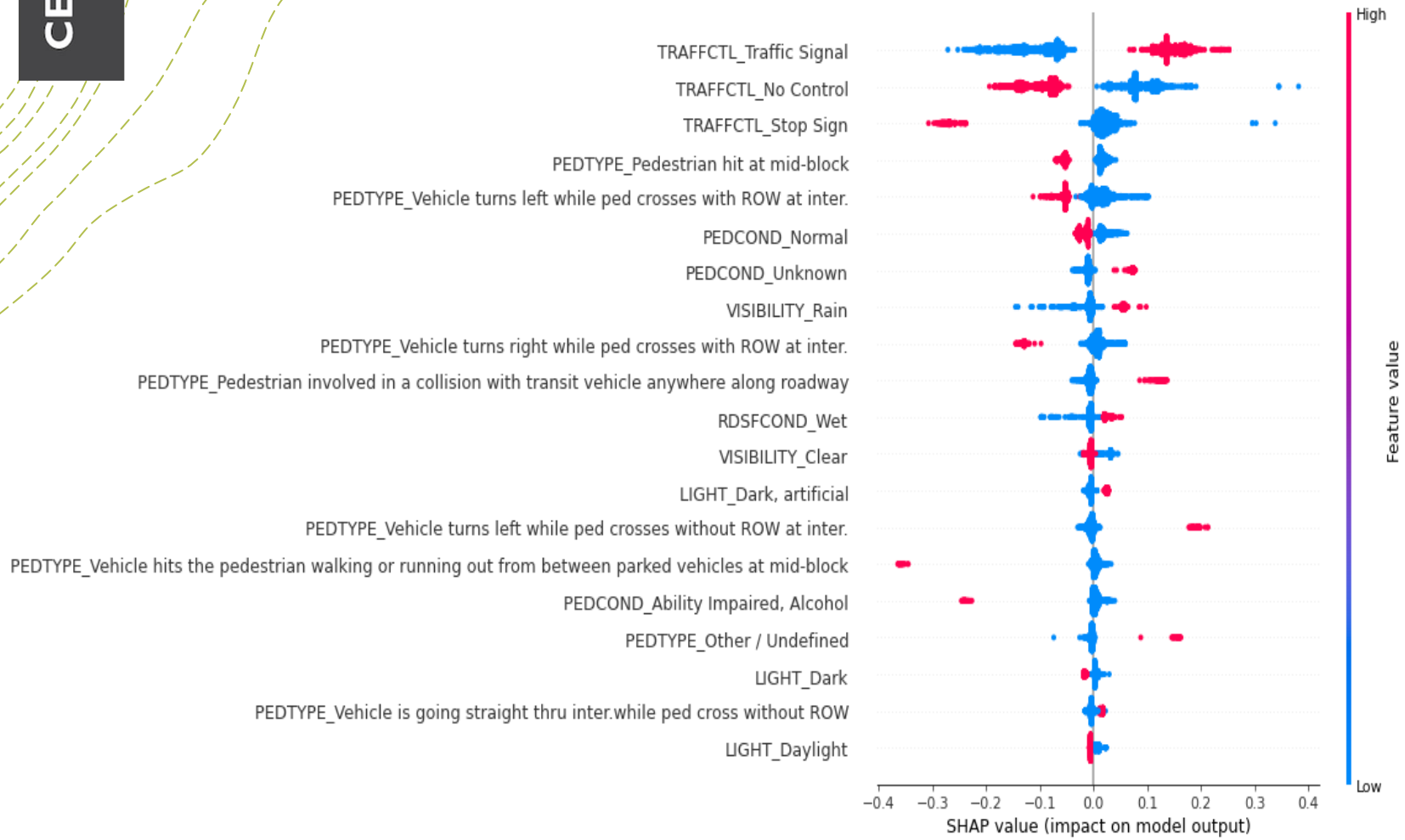
Pedestrian Factors

This decision tree depicts the major behavioral and environmental factors in which pedestrians cause road accidents. Main causes are inattentiveness and crossing streets where there is no traffic control.



What We Found

Pedestrian Factors



+ Traffic Control signals have the highest impact on pedestrian safety indicating the primary dependence/trust of pedestrians on these traffic controls.

Analysis of Accidents in Downtown vs Non-downtown

- Dataset Used = Road Impediments (Geotab) + KSI
- Why not use Hazardous Impediments? Hazardous Impediments dataset is heavily clustered along the 401 Highway. As a result, during our merge procedure – a lot of KSI events across the city are dropped. This leads to a very sparse dataset, with not enough results for prediction/inference to take place.
- Using Tree-based methods and Linear Regression
- Interpreting outputs using Shapley values
- Comparing results to our Null Model – a naïve guess based on the mean of COUNT

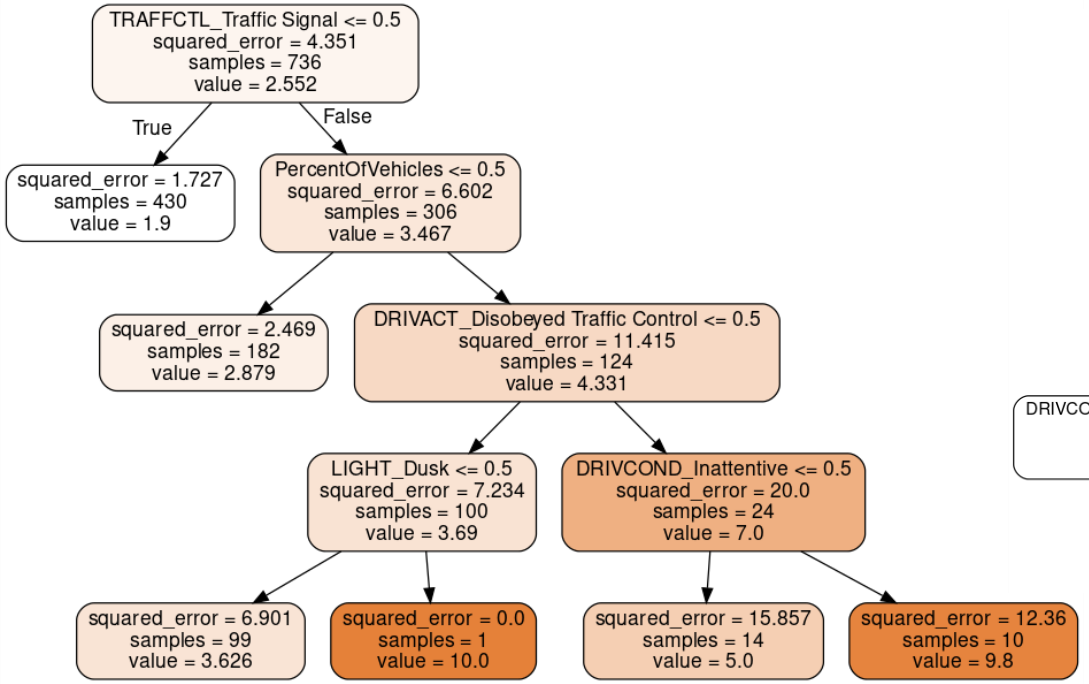
Variable Name	Description
DRIVACT	Driver's action before accident
DRIVCOND	Driver's condition before accident
TRAFFCTL	Traffic Control measures at place of accident
INVAGE	Age of Involved Party (Driver)
AvgAcceleration	Transformed to a Binary variable for Tree models
VEHTYPE	Categorized based on size
VISIBILITY	Environment Condition
PercentOfVehicles	Percentage of Vehicles affected by the Impediment Transformed to a Binary variable for Tree models

CENTENNIAL COLLEGE

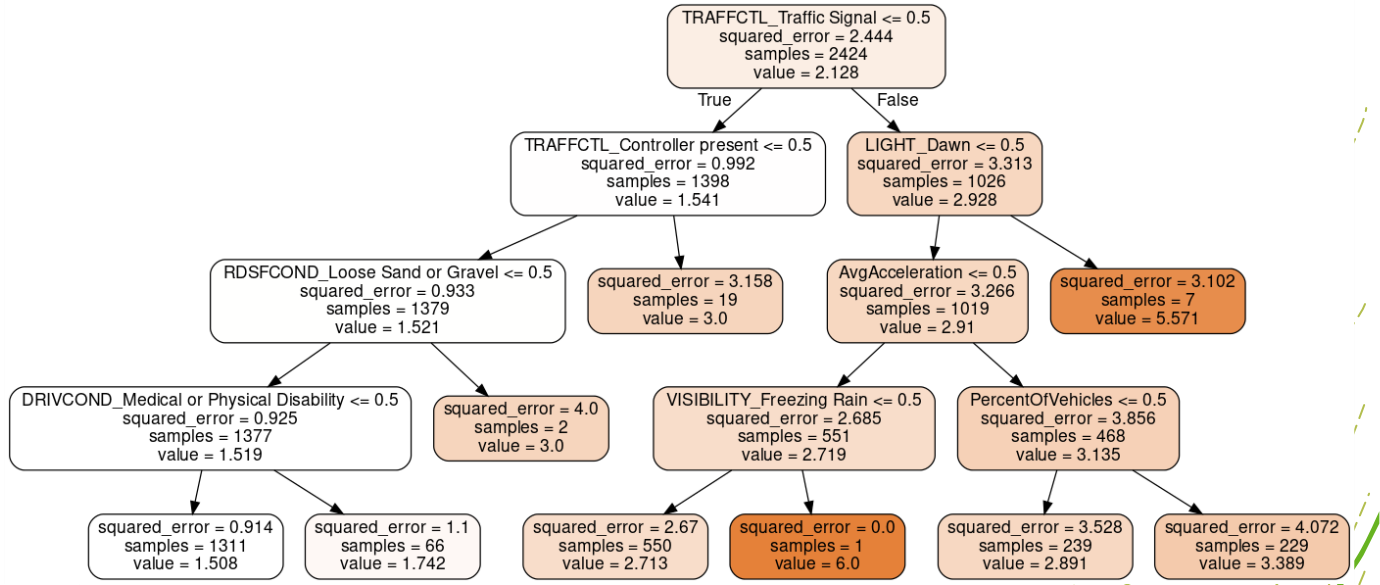
What We Found

Downtown vs. Non-downtown

Heavy Traffic Flow combined with Inattentiveness leads to greater accidents. Lastly, more accidents occur during the night.



Weather and Road Surface Conditions are much more important. Speeding is also an important factor. Lastly, more accidents occur during the day.

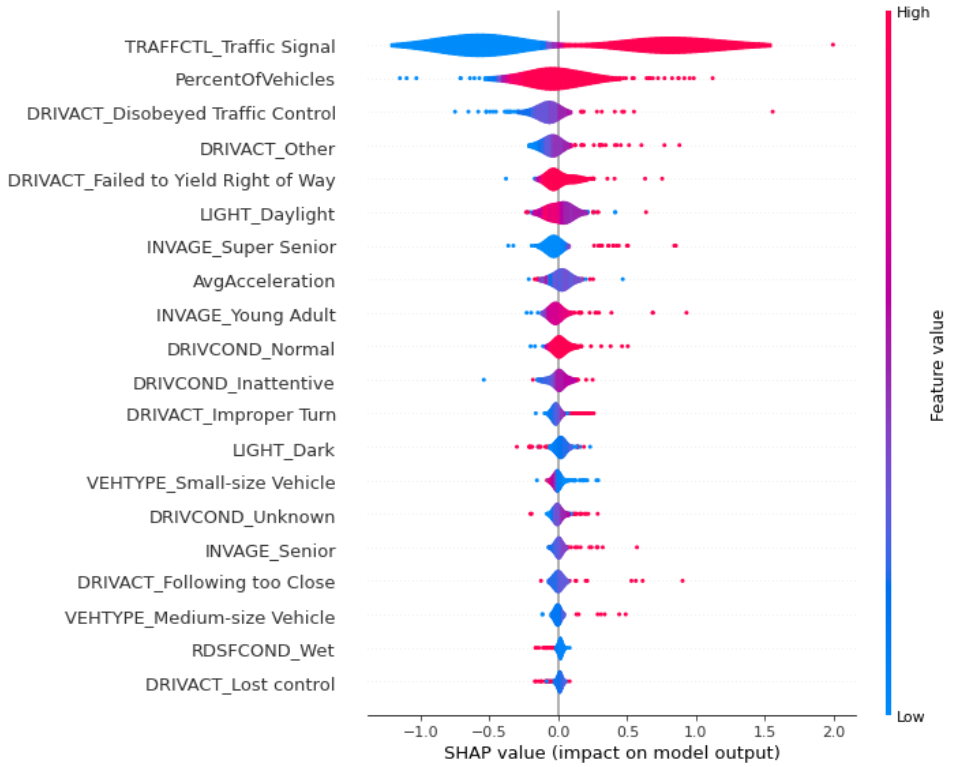


CENTENNIAL COLLEGE

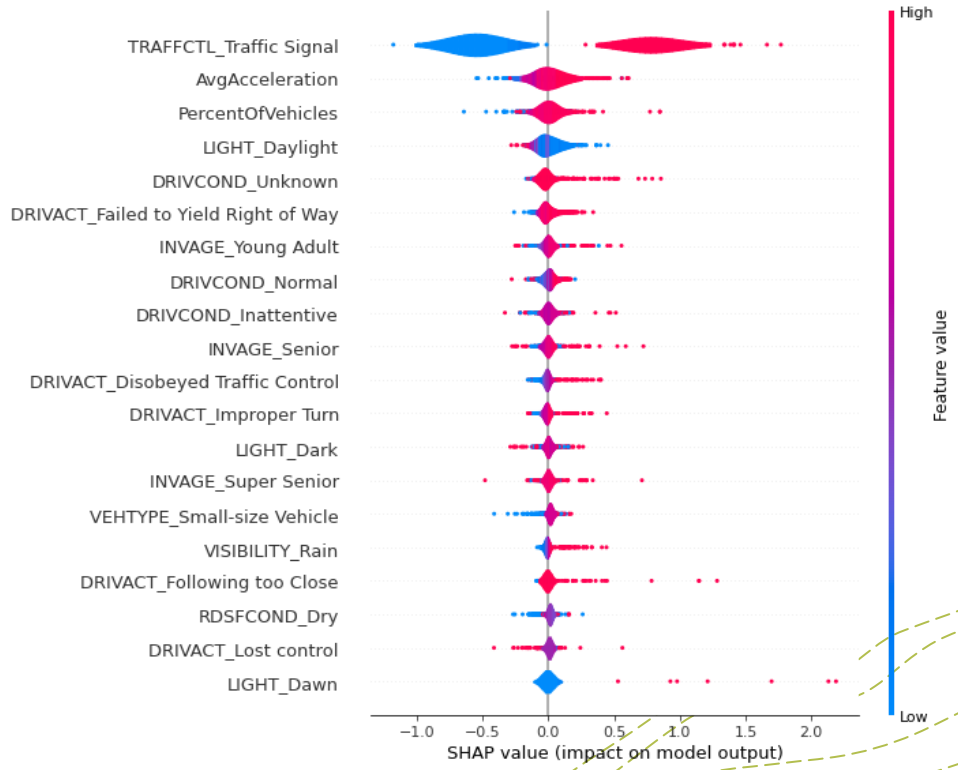
What We Found

Downtown vs. Non-downtown

Drivers disobeying traffic controls and heavier traffic flow lead to more accidents downtown.



Speed (AvgAcceleration) is an important factor for Non-downtown accidents. Road Surface Condition is a bigger factor for non-downtown areas.



CENTENNIAL COLLEGE

What We Found

Emerging Themes and Key Insights

- Inattentiveness of drivers and pedestrians is a major issue. This is a key factor leading to more accidents.
- Heavy traffic flow combined with rough driving conditions lead to greater KSI events.
- Speeding (AvgAcceleration) is not as important for Downtown areas compared to Non-Downtown areas.
- Weather and Road Surface Conditions are much more important factors for Non-Downtown areas.
- Our findings indicate a high degree of dependence on traffic controls on the part of pedestrians which may be overriding the prudence of practicing standard situational observation skills

Recommendations for Policy Consideration

- Environmental and Behavioural Factors
  - Our analysis confirms that current interventions taken have been appropriate insofar as remediation for general environmental and behavioral factors are concerned.
- Pedestrian Factors
  - Special consideration may need to be made in related to interventions that improve pedestrian compliance to avoid mishaps. Some actions might include:
    - Explore creating more crosswalks - In many parts of Toronto, crosswalks are 1 km or more away from each other.
    - Explore the use of underground walkways in Downtown Toronto – High density cities like Hong Kong, Singapore, and Seoul extensively use underground walkways to reduce pedestrian-related incidents. In Toronto, it would serve a double purpose of providing pedestrians with the bonus of a climate-controlled environment.
    - Explore stiffer penalties for pedestrians – While penalties for misbehaving drivers are regularly applied, penalties for erring pedestrians are inconsistently applied if at all.
    - Speed bumpers near intersections.
- Downtown vs. Non-downtown
  - Our analysis has found that there are differences in downtown and non-downtown causes of accidents leading us to believe that interventions ought to account for these differences.
  - Given the huge influence that downtown data has over conclusions, we observe that many interventions applied tend to favor factors that are actually more relevant to downtown than to other areas where the reasons for accidents are more heterogenous.

Areas for Further Inquiry

- Further analysis using data from Uber, Google Maps and Waze.
- We identified the lack of a full set of causal factors in the dataset. A more comprehensive analysis combining more datasets would lead to a more complete analysis.
- The dataset doesn't include the effect of construction, and other road congestion issues on traffic incidents.

Related Research

- Fridman, L., Ling, R., Rothman, L. et al. Effect of reducing the posted speed limit to 30 km per hour on pedestrian motor vehicle collisions in Toronto, Canada - a quasi experimental, pre-post study. BMC Public Health 20, 56 (2020). https://doi.org/10.1186/s12889-019-8139-5
- Leung Kevin, Iu Jerry, Gelgor Gabriel, and Halili Arbri. Collision Statistics: A Study in Toronto Road Safety. STEM Fellowship Journal. 2(1): 49-54. https://doi.org/10.17975/sfj-2016-009
- Mao, Yaoquan & Zhang, J & Robbins, Glenn & Clarke, K & Lam, Miu & Pickett, William. (1997). Factors affecting the severity of motor vehicle traffic crashes involving young drivers in Ontario. Injury prevention : journal of the International Society for Child and Adolescent Injury Prevention. 3. 183-9. 10.1136/ip.3.3.183.
- Bavcevic, Zeljko & Harinam, Vincent. (2020). Targeting Fatal Traffic Collision Risk from Prior Non-Fatal Collisions in Toronto. Cambridge Journal of Evidence-Based Policing. 4. 1-15. 10.1007/s41887-020-00054-z.

You may find our code, datasets, and analysis in the link below:

https://github.com/Vincent-Kwan/Safe-Roads-Competition

End of Presentation. Nothing Follows.