Predicting Severe Traffic Outcomes - City of Seattle

September 2020

Opportunity exists for city governments to save lives and benefit society by focusing effort on the key drivers of severe accidents

- Reducing severe traffic accidents can provide tangible benefits to individuals, families, and society
- Negative impact of severe accidents is profound. In the US, over the course of 2019 alone:
  - ~39K lost their lives to car crashes
  - 4.4M *seriously* injured in traffic accidents
- City governments should endeavor to keep their streets safe. They benefit from reducing traffic accidents in their jurisdictions.
- However, government resources are often limited. Understanding where to focus time and money to drive the most impact can help!

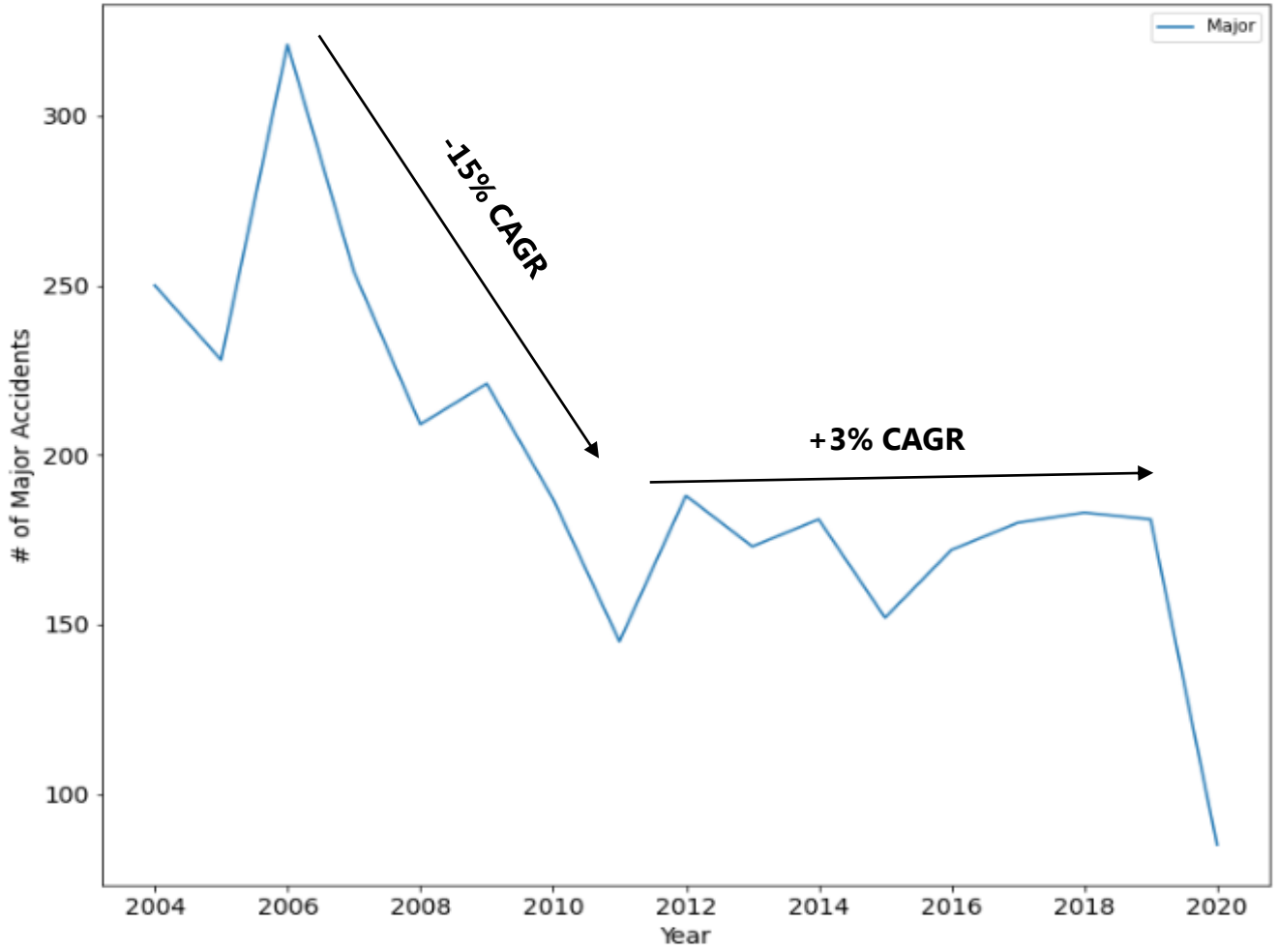
A data set provided by the City of Seattle was cleaned and used to understand key factors that contribute to severe accidents

- The City of Seattle and it's Department of Transportation (SDOT) maintain a database of historical accidents and associated severity
- The data set contains rich detail on ~220,000 accidents that occurred between 2004 and September 2020
- Data was appropriately cleaned and incomplete records were purged before sampling data for analysis
- Target variable defined as Major / Severe Accidents - defined as accidents resulting in either death or serious injury
- Given the desire to reliably predict relatively less frequent, severe accidents, data leveraged for analysis was balanced

While major / severe accidents decreased in Seattle between 2006 and 2011, the city has not materially reduced accident severity in recent years

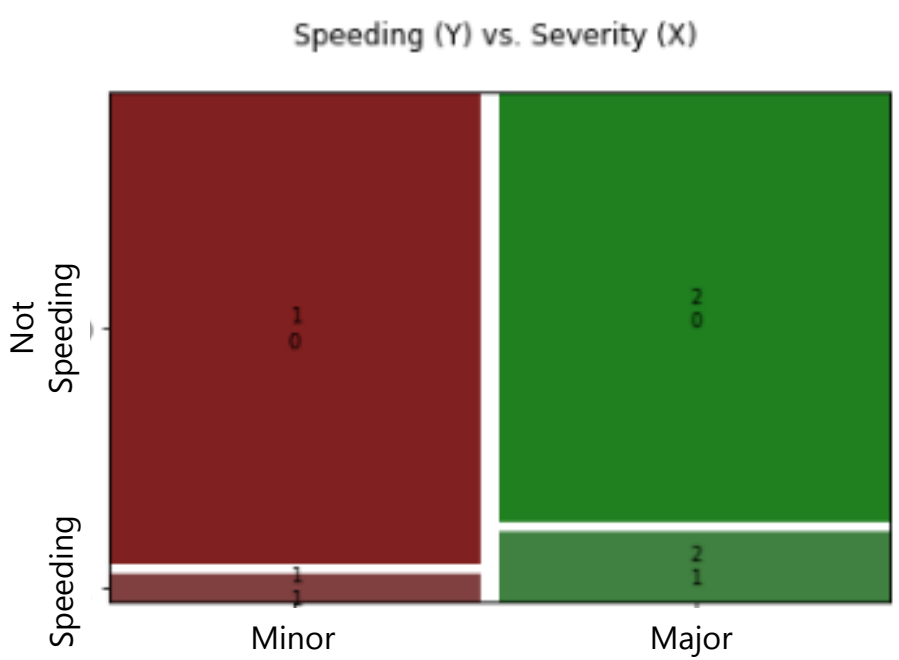
- The number of severe accidents in Seattle peaked in 2006, when 321 fatal and/or severe injury accidents were reported
- Between 2006 and 2011, the city saw an average annual reduction of severe accidents of 15%
- However, between 2011 and 2019, the city has not been able to significantly drive further reductions in severe accidents (severe accidents increased ~3% per year)
- Additional, new strategies may be needed to drive a decrease in severe accident incidence in Seattle

Number of Major / Severe Accidents per Year
2004 – YTD 2020



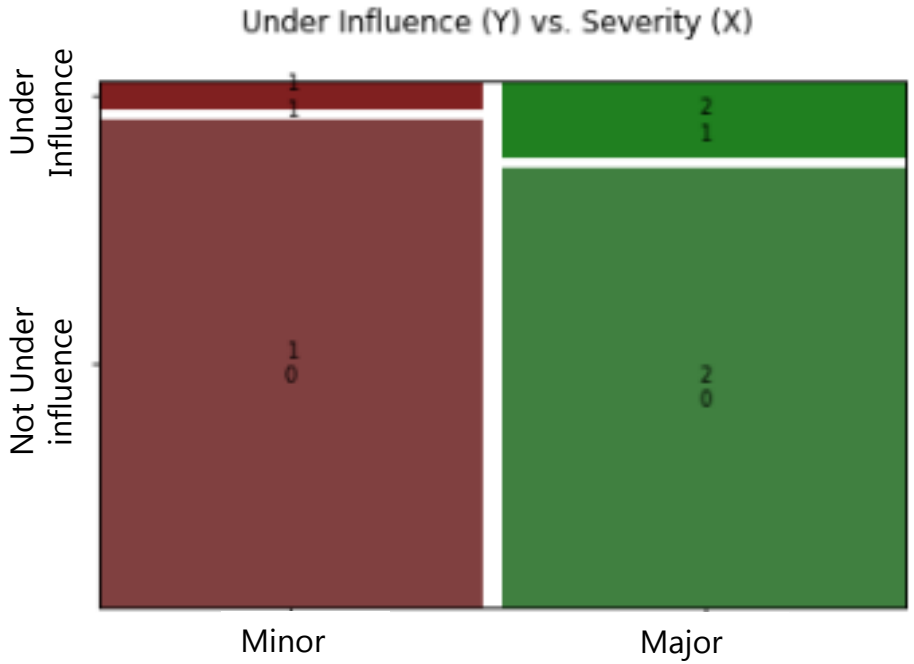
Initial data exploration suggested that variables like *speeding* and *under the influence* might be useful in predicting accident severity

Relationship between Speeding and Severity
Chi-Square Score = 426; p-value <.00001



% of accidents that involve speeding:	6%	14%

Relationship between Under the Influence and Severity
Chi-Square Score = 550; p-value <.00001

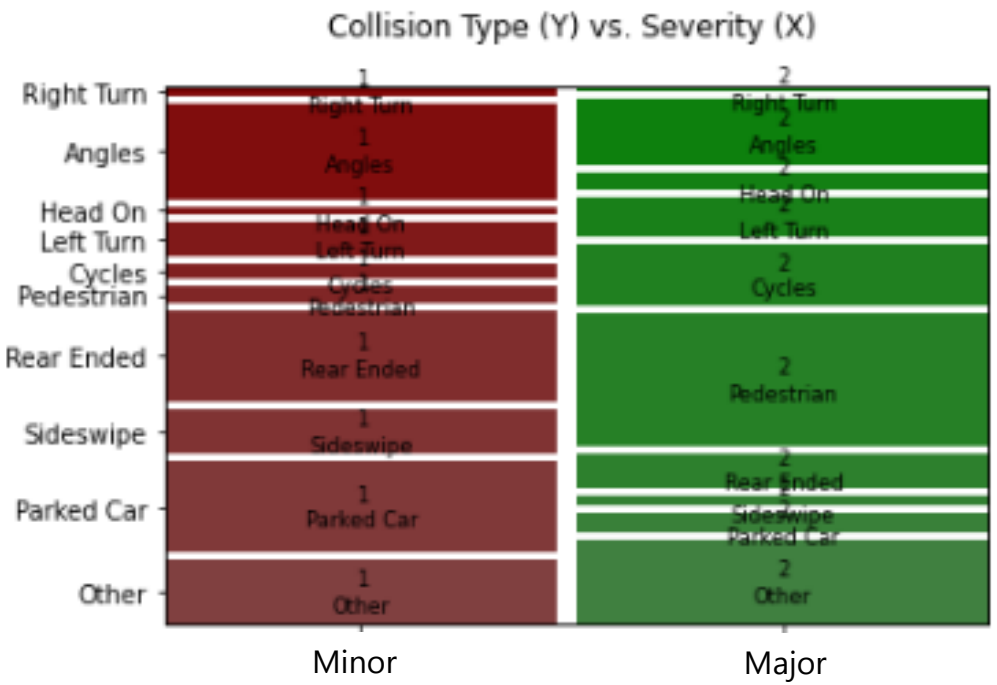


% of accidents that involve drinking/drugs:	5%	15%

*Note: Similar analysis was conducted for many other categorical variables in the SDOT data; however, only a few representative examples are provided

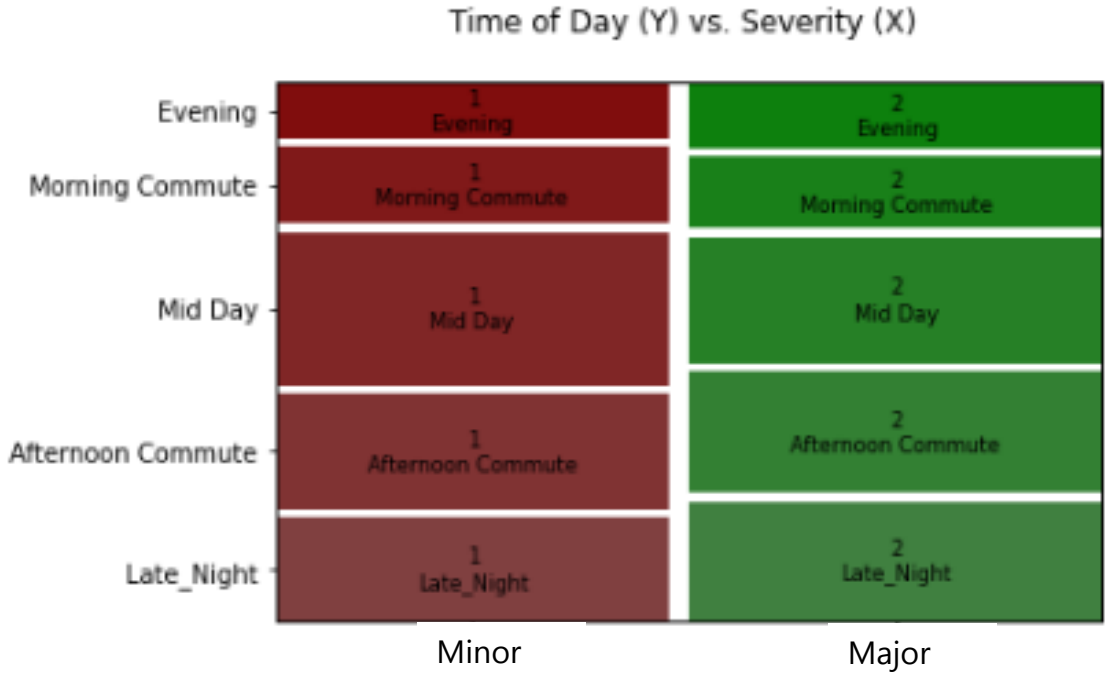
Additionally, *collision type* and *time of day* might be helpful in predicting accident severity in the Seattle area

Relationship between Collision Type and Severity



% of accidents that involve:	Minor	Major
Pedestrians -	3%	29%
Cycles -	3%	13%

Relationship between Time of Day and Severity



% of accidents that involve:	Minor	Major
Late night –	20%	24%
Mid day –	30%	25%

*Note: Similar analysis was conducted for many other categorical variables in the SDOT data; however, only a few representative examples are provided

Four machine learning models were created and evaluated to predict accident severity for the Seattle data

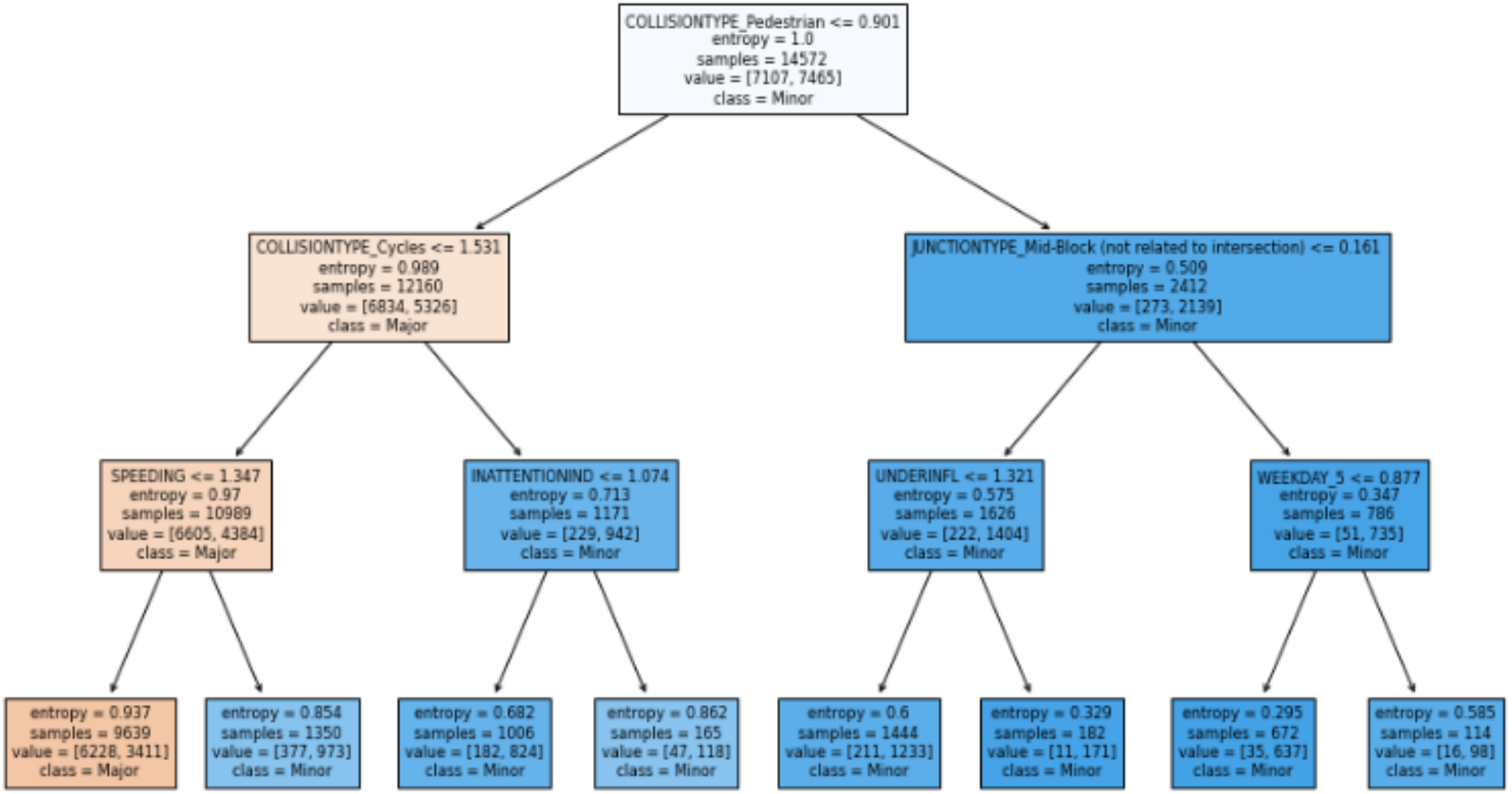
- Models all generated roughly similar levels of accuracy on the balanced dataset (i.e., accuracy ranged from 0.70-0.75 depending on the model)
- The decision tree and logistic regression models provide the most interpretable results
    - Both models provide increased understanding of the specific variables that are most associated with severe accidents

Accident Severity Model Results

Model	Jaccard Similarity Score	F1-Score
K-Nearest Neighbors (n=6)	0.74	0.74
Support Vector Machines (kernel = 'rbf')	0.75	0.75
Decision Tree (max depth = 3)	0.70	0.70
Logistic Regression	0.74	0.74

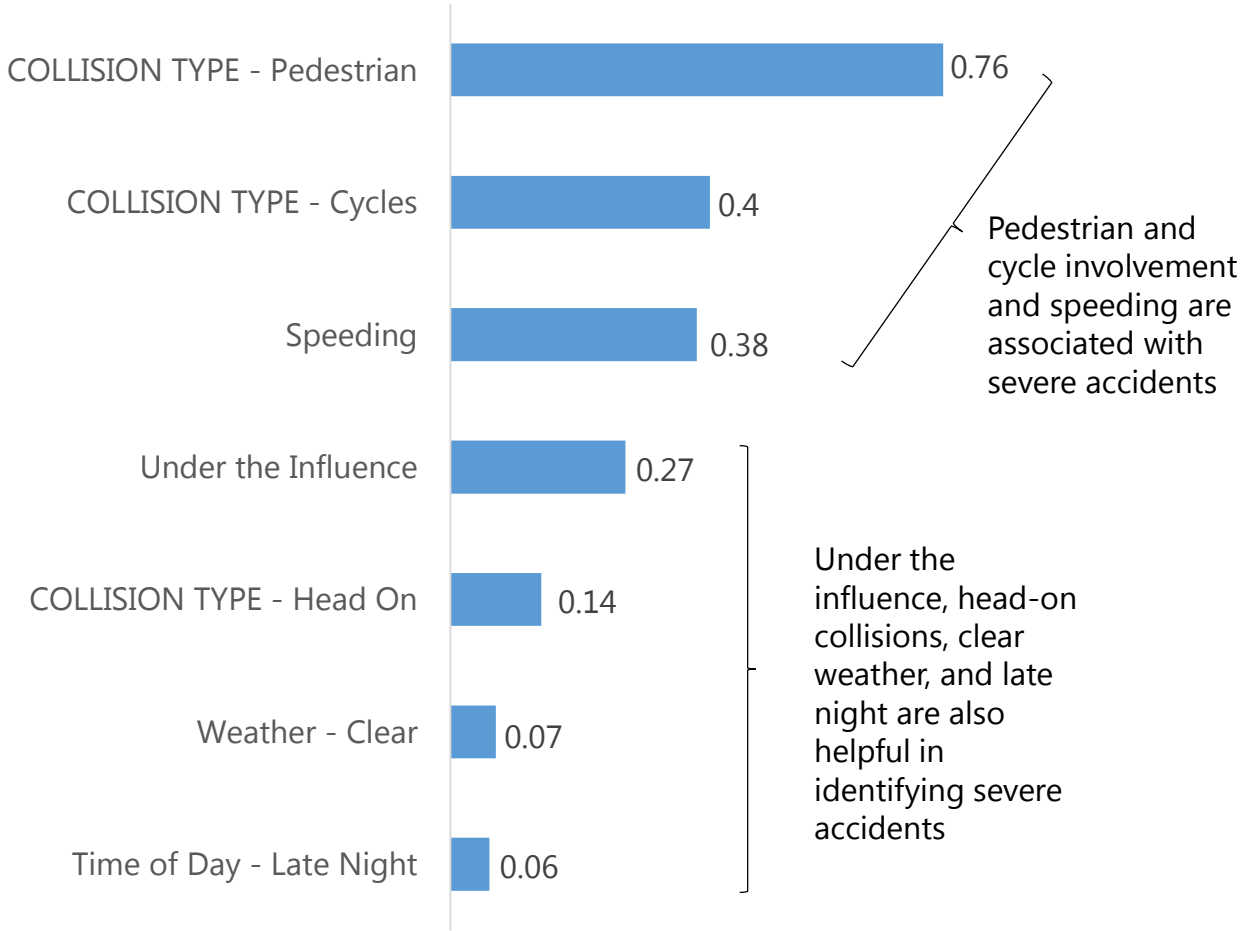
The decision-tree model suggests that pedestrian involvement, cycle involvement, and speeding are highly associated with severe accidents

Accident Severity Decision Tree Model (max-depth = 3)

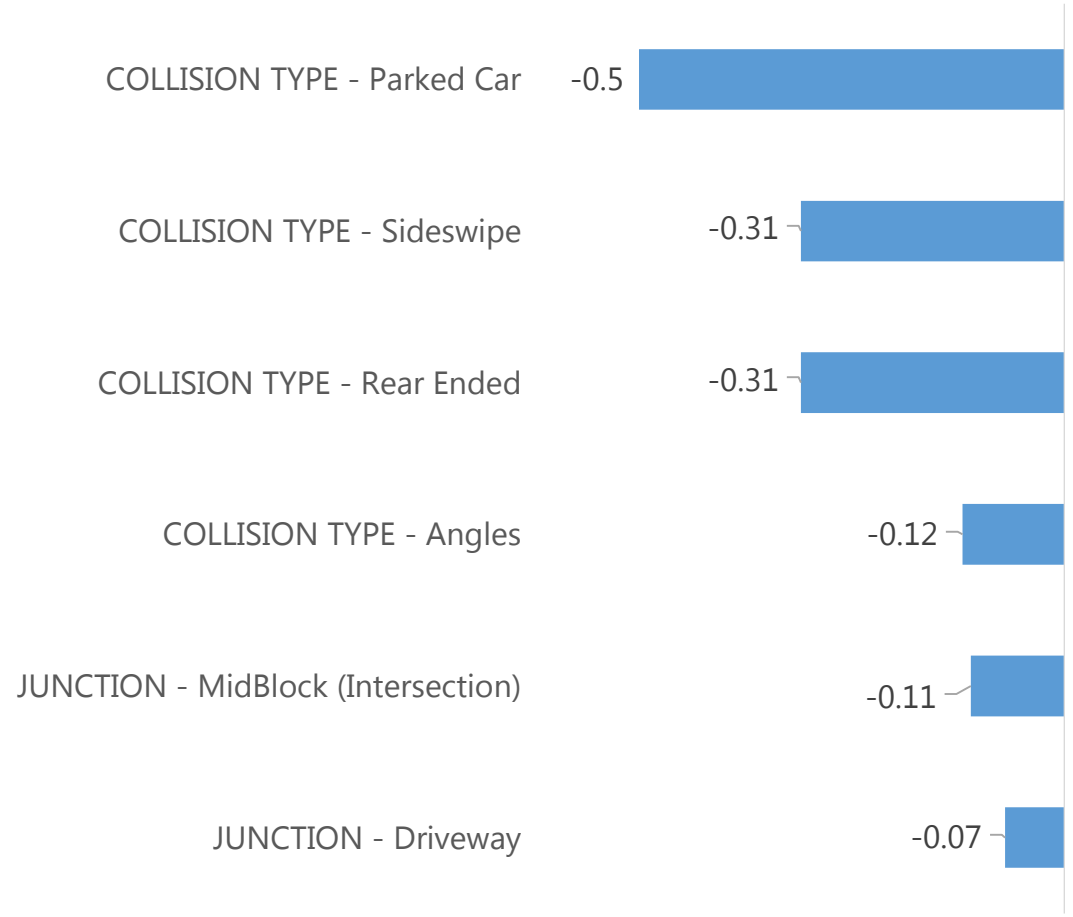


Similarly, the logistic regression model coefficients provide context on variables that are heavily associated with severe vs. minor accidents

Logistic Regression - Top 7 Largest Positive Coefficients



Logistic Regression - Top 6 Most Negative Coefficients



The logistic and decision tree machine learning models provide insight on potential areas of focus to reduce severe accidents going forward

	Potential Focus Area	Description	Sample Ideas*
1	**Pedestrian / Cycle Involvement.**	Given that accidents involving pedestrians and cycles are much more likely to be severe (all else equal), the city may want to consider prioritizing efforts aimed at protecting pedestrians / cyclists	• Limiting traffic in select high risk areas • Building out barrier protected bike lanes in high risk areas • Funding enforcement of bicycle and pedestrian safety rules (e.g., helmet laws / jaywalking fines)
2	**Speeding.**	Speeding is highly associated with severe accident outcomes. Consequently, the city may want to consider prioritizing efforts aimed at reducing excess speeding	• Reduce speed limits in high risk areas • Better communicate existing speed limits in high risk areas
3	**Under the Influence Driving.**	Programs aimed at reducing the likelihood that drivers will get behind the wheel when under the influence of alcohol or drugs may be helpful	• Support and/or fund programs to provide free or at cost rides to under the influence individuals at key times of the day (e.g., 11pm - 2am)
4	**Head on Collisions.**	Head-on-collisions are associated with more severe accident outcomes. Lowering risk of head-on collisions may be beneficial in mitigating accident severity	• Double down on efforts to install median barriers, reflectors, etc. in high risk areas

*Not exhaustive and for discussion only

Thanks!