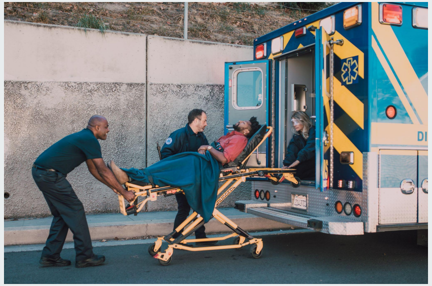
Predicting the Risk of Injury Resulting from Vehicle Collisions in Ottawa, Ontario

Anja Wu,
Ben Nikkel



Introduction

Problem to Solve and Motivation

Problem

- Do collison features contribute to injuries and how?

Motivation

- As a citizen of Ottawa who drives far and often (usually through Hunt Club) curious about
  - how bad intersections are
  - which need to be improved
  - if there are certain features contributing to injuries that can be mitigated by the city

Research Questions

Questions We Asked

1. What were the most dangerous locations?
    a. Which have the most commonly occurring accidents?
    b. Which locations are most dangerous in terms of injury levels?
2. Which accident features are correlated to injuries?
3. Were road/weather conditions correlated to levels of severity of injuries?
4. Did injuries disproportionately affect drivers, pedestrians, bicyclers or motorcycles?
5. Do traffic controls lead to decreases in accidents?
6. Can we predict whether an injury will occur or not, using specific accident features?

Methodology

Data Collection and Cleaning

Data collection:

- https://open.ottawa.ca/ *accessed September, 2023*

Preprocessing:

- Feature distribution analyzed for outliers, etc.
- Missing values for integer columns were changed to zeros
- Latitude and longitude rounded from 5 to 4 decimal points
- Binary injury variable created (0: no injury collision, 1: injury collision)
- Categorical predictor features converted
    - Using scale if ordinal
    - Using one-hot encoding if nominal

Data Analysis

- Collision counts based on grouping truncated lat, long and location
  - Top 10 Locations by Number of Collisions
  - Top 10 Locations by Number of Injuries
  - Top Locations by Number of Major/Fatal Injuries
- Correlation heat map *(Pearson's)*
  - All features with predictor *(removed highly correlated for predictions)*
  - Environmental factors on severity level of injury
- Calculations on proportions:
  - All collisions involving each mode of transportation
  - for each mode of transportation by each level of injury severity
- Collision frequency by type of traffic control

Predictive Model

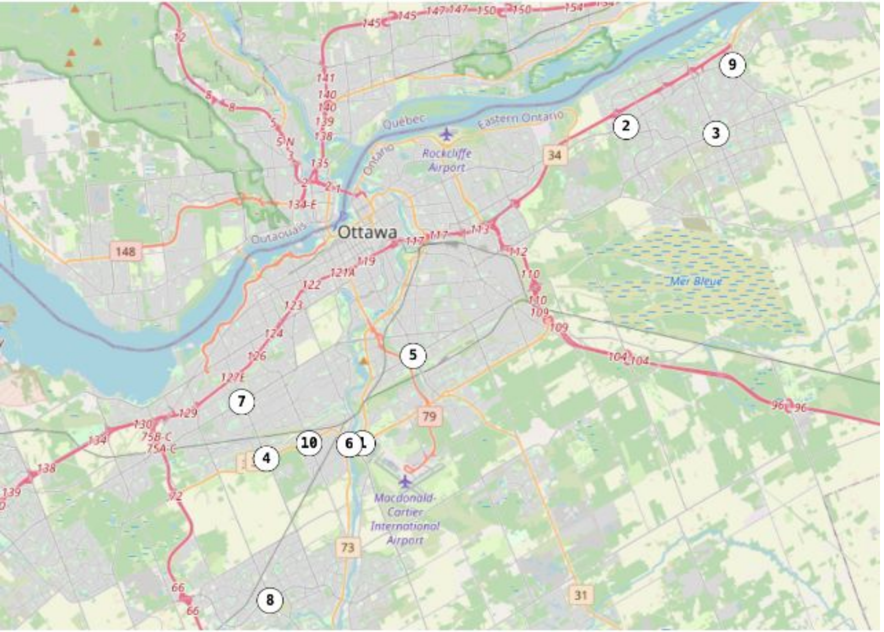
- k-Nearest Neighbours *(scikit-learn)*
  - Random OverSampling (ROS) *(imblearn.over_sampling)*
  - Synthetic Minority Oversampling Technique (SMOTE)
- Decision Tree *(scikit-learn)*
  - Random OverSampling (ROS) *(imblearn.over_sampling)*
  - Random Forest *(scikit-learn)*
- Comparative Methods (*PyCaret*)
  - List on side

Model
Ridge Classifier
Linear Discriminant Analysis
Logistic Regression
Gradient Boosting Classifier
Ada Boost Classifier
Random Forest Classifier
Extra Trees Classifier
K Neighbors Classifier
Dummy Classifier
Decision Tree Classifier
SVM - Linear Kernel
Naive Bayes
Quadratic Discriminant Analysis

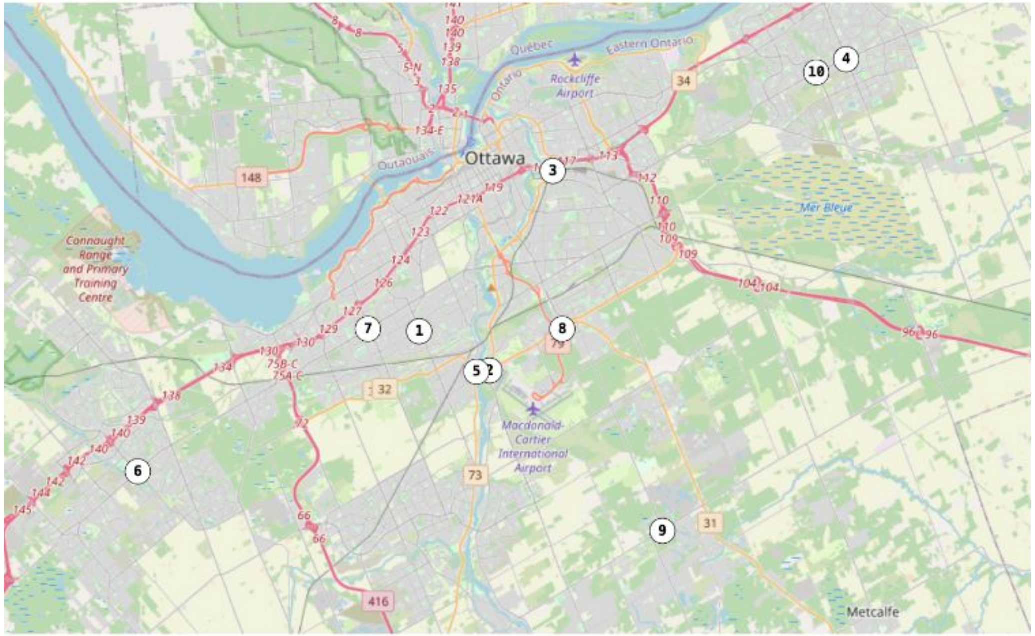
Results

Q1) What were the most dangerous locations?

Top 10 Locations by Count of Accidents



Top 10 Locations by Injury Count



Q1) What were the most dangerous locations?

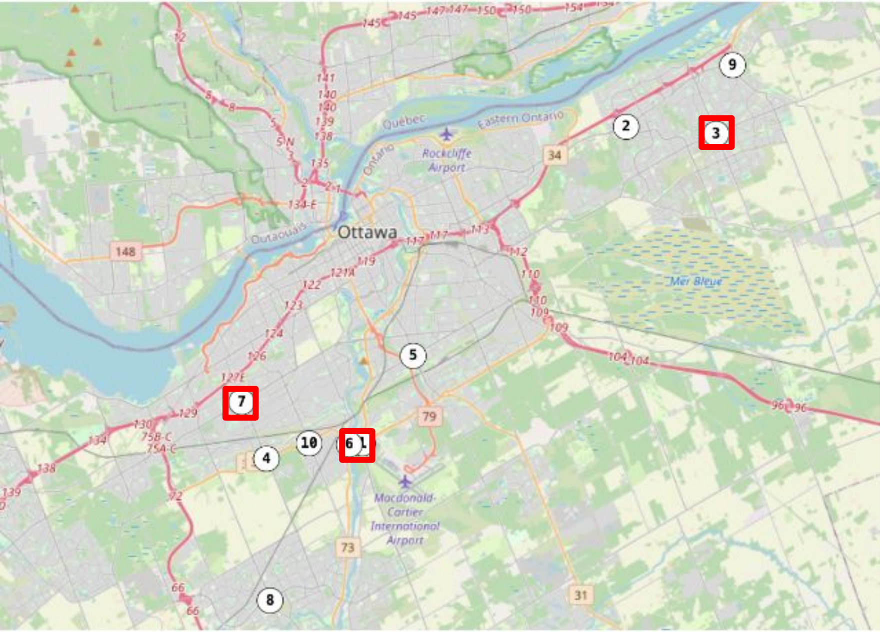
Collisions			Collisions with an Injury		
Rank	**Location**	**n**	**Rank**	**Location**	**n**
1	Hunt Club Rd @ Riverside Dr[a]	232	1	Meadowlands Dr @ Merivale Rd	33
2	St. Joseph Blvd @ Jeanne D'arc Blvd	218	2	Hunt Club Rd @ Riverside Dr[a]	32
3	Innes Rd @ Tenth Line Rd[b]	168	2	Riverside Dr @ Tremblay Rd/Hwy417 Ic117 Ramp52	32
4	West Hunt Club Rd @ Woodroffe Ave	157	2	Innes Rd @ Tenth Line Rd[b]	32
4	Bank St @ Walkley Rd	157	3	Prince Of Wales Dr @ West Hunt Club Rd[c]	29
5	Prince Of Wales Dr @ West Hunt Club Rd[c]	155	4	Hazeldean Rd @ Terry Fox Dr	28
5	Baseline Rd @ Woodroffe Ave[d]	155	4	Baseline Rd @ Woodroffe Ave[d]	28
6	Greenbank Rd @ Strandherd Dr	147	4	Hunt Club Rd @ Bridle Path Dr/Daze St	28
6	St. Joseph Blvd/Old Montreal Rd @ Trim Rd	147	5	Albion Rd @ Mitch Owens Rd	27
7	Merivale Rd @ West Hunt Club Rd	146	5	Innes Rd @ Jeanne D'arc Blvd/Mer Bleue Rd	27

Q1) What were the most dangerous locations?

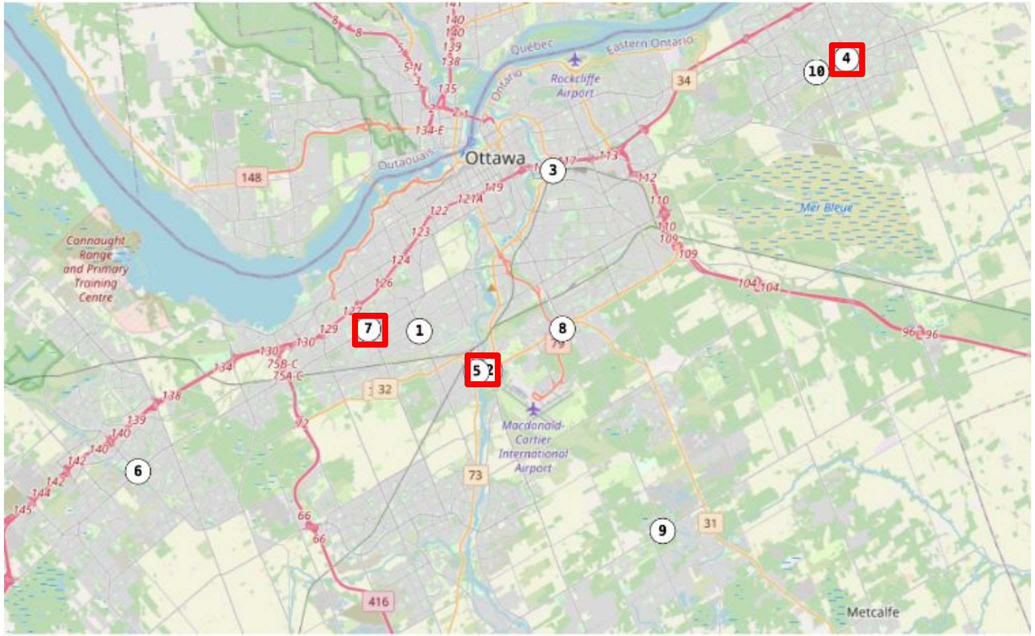
Collisions			Collisions with an Injury		
Rank	**Location**	**n**	**Rank**	**Location**	**n**
1	Hunt Club Rd @ Riverside Dr[a]	232	1	Meadowlands Dr @ Merivale Rd	33
2	St. Joseph Blvd @ Jeanne D'arc Blvd	218	2	Hunt Club Rd @ Riverside Dr[a]	32
3	Innes Rd @ Tenth Line Rd[b]	168	2	Riverside Dr @ Tremblay Rd/Hwy417 Ic117 Ramp52	32
4	West Hunt Club Rd @ Woodroffe Ave	157	2	Innes Rd @ Tenth Line Rd[b]	32
4	Bank St @ Walkley Rd	157	3	Prince Of Wales Dr @ West Hunt Club Rd[c]	29
5	Prince Of Wales Dr @ West Hunt Club Rd[c]	155	4	Hazeldean Rd @ Terry Fox Dr	28
5	Baseline Rd @ Woodroffe Ave[d]	155	4	Baseline Rd @ Woodroffe Ave[d]	28
6	Greenbank Rd @ Strandherd Dr	147	4	Hunt Club Rd @ Bridle Path Dr/Daze St	28
6	St. Joseph Blvd/Old Montreal Rd @ Trim Rd	147	5	Albion Rd @ Mitch Owens Rd	27
7	Merivale Rd @ West Hunt Club Rd	146	5	Innes Rd @ Jeanne D'arc Blvd/Mer Bleue Rd	27

Q1) What were the most dangerous locations?

Top 10 Locations by Count of Accidents



Top 10 Locations by Injury Count



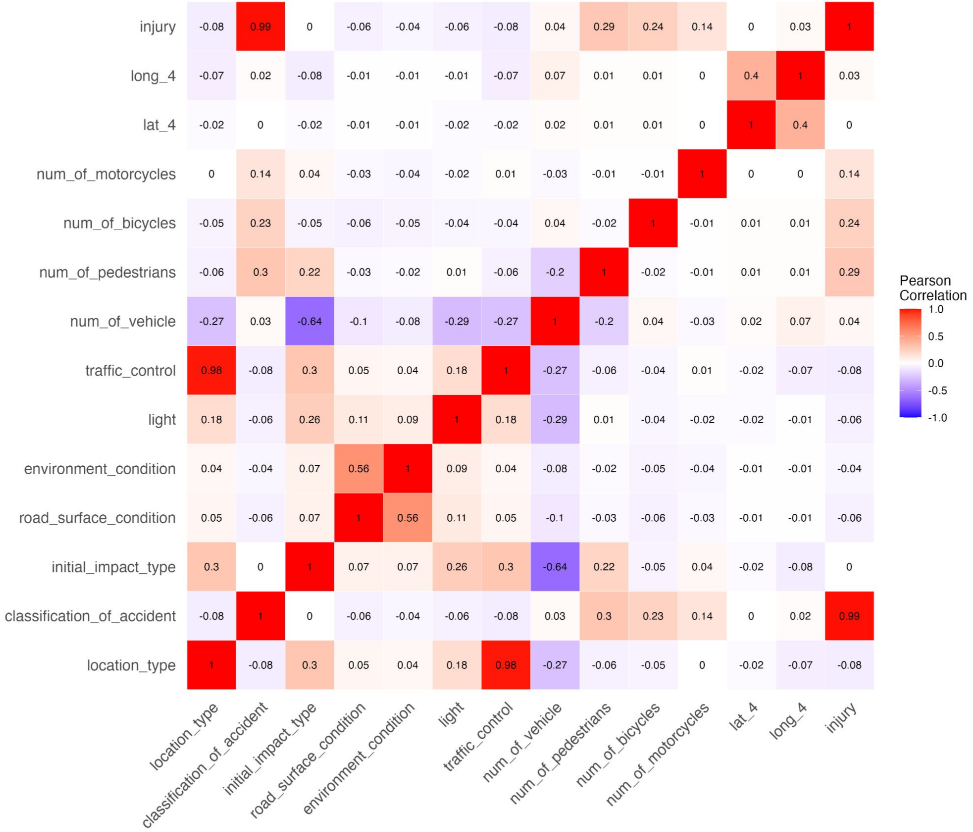
Q1) What were the most dangerous locations?

Major Injuries			Fatal Injuries		
Rank	**Location**	**n**	**Rank**	**Location**	**n**
1	Bank St @ Dalmeny Rd/Marvelville Rd	3	**1**	Mccordick Rd @ Roger Stevens Dr	2
1	8th Line Rd @ Parkway Rd[a]	3	**1**	8th Line Rd @ Parkway Rd[a]	2
1	Hunt Club Rd @ Paul Anka Dr	3	**1**	Russell Rd @ Southvale Cres N	2
1	St. Laurent Blvd @ McArthur Ave	3			
1	King Edward Ave @ Sussex Dr	3			
1	Aviation Pkwy @ Montreal Rd	3			

Q1) What were the most dangerous locations?

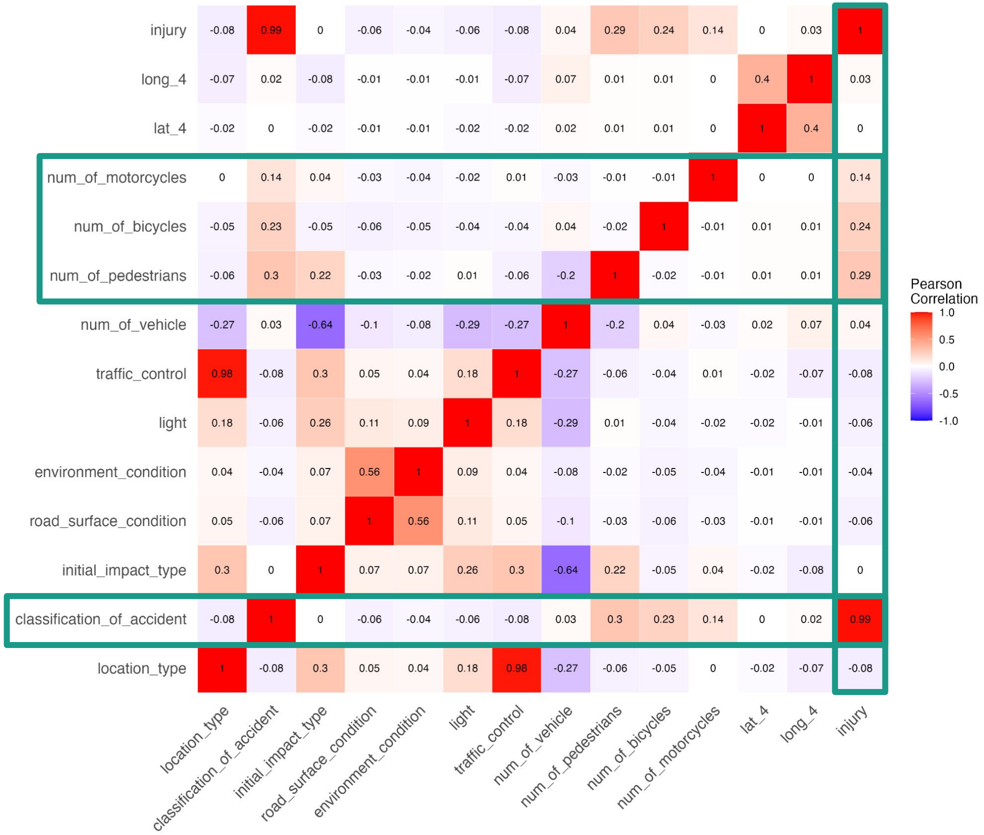
Major Injuries			Fatal Injuries		
Rank	Location	n	Rank	Location	n
1	Bank St @ Dalmeny Rd/Marvelville Rd	3	1	Mccordick Rd @ Roger Stevens Dr	2
1	8th Line Rd @ Parkway Rd[a]	3	1	8th Line Rd @ Parkway Rd[a]	2
1	Hunt Club Rd @ Paul Anka Dr	3	1	Russell Rd @ Southvale Cres N	2
1	St. Laurent Blvd @ McArthur Ave	3			
1	King Edward Ave @ Sussex Dr	3			
1	Aviation Pkwy @ Montreal Rd	3			

Q2) Which accident features are correlated to injuries?

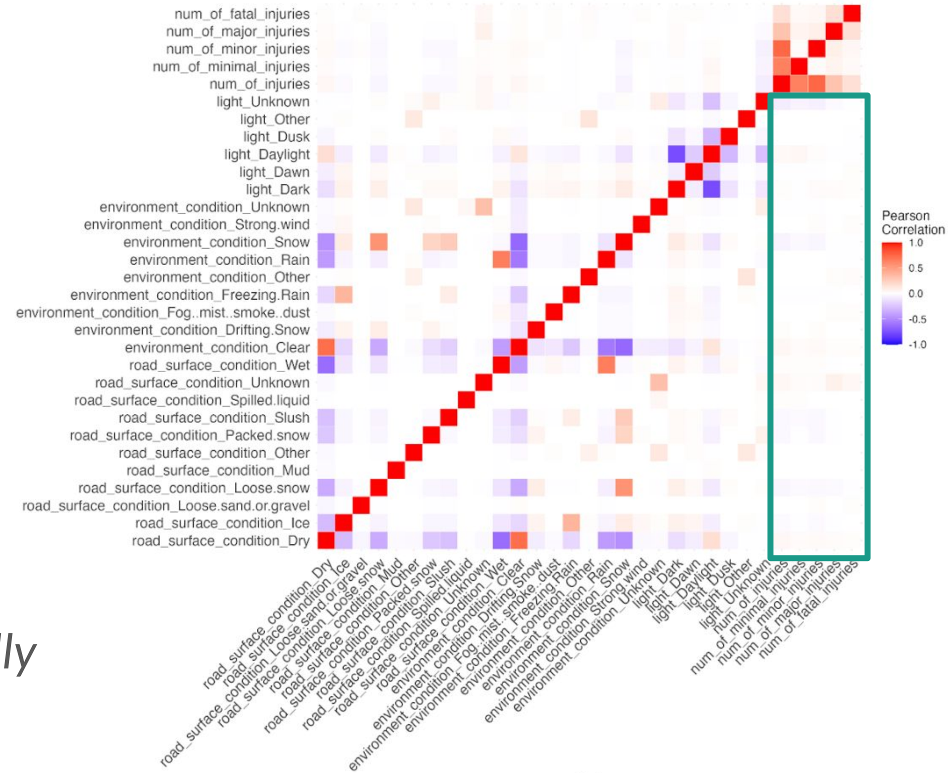


B

Q2) Which accident features are correlated to injuries?



Q3) Were road/weather conditions correlated to levels of severity of injuries?



No correlation

** Weather categorically defined*

Q4) Did injuries disproportionately affect a specific mode of transportation?

		Injury Class				
Mode of Transport	Collision Involvement (%)	None (%)	Minimal (%)	Minor (%)	Major (%)	Fatal (%)
Vehicle	95.28	79.99	8.00	10.87	0.94	0.20
Pedestrian	2.10	0.84	29.41	57.35	9.97	2.43
Bicycle	1.77	12.34	31.95	48.90	6.13	0.68
Motorcycle	0.85	20.13	15.37	46.75	14.42	3.33

Q4) Did injuries disproportionately affect a specific mode of transportation?

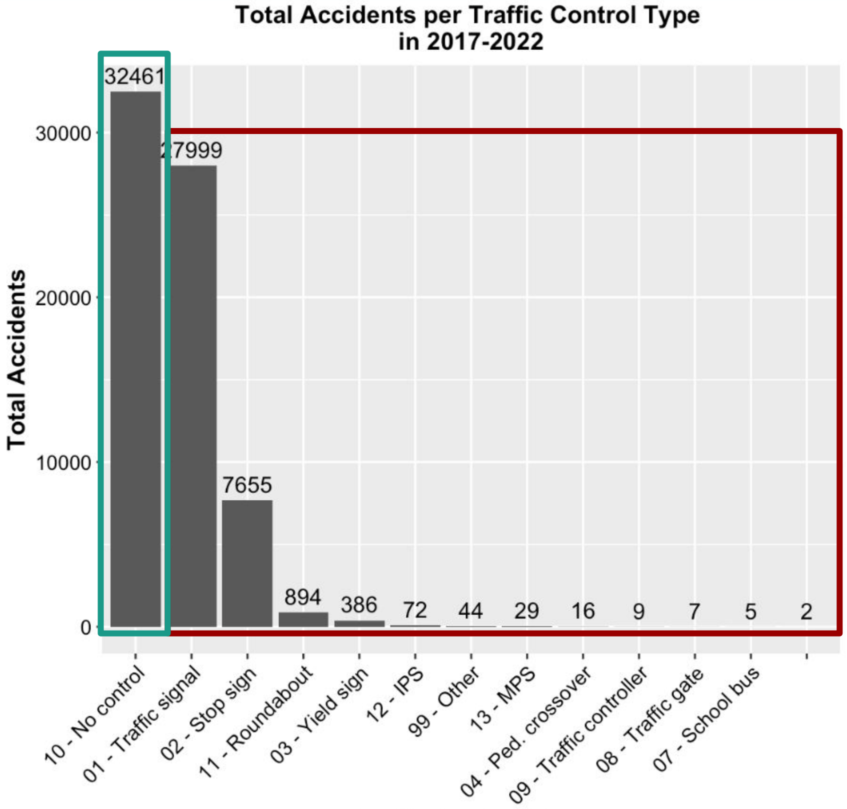
		Injury Class				
Mode of Transport	Collision Involvement (%)	None (%)	Minimal (%)	Minor (%)	Major (%)	Fatal (%)
Vehicle	95.28	79.99	8.00	10.87	0.94	0.20
Pedestrian	2.10	0.84	29.41	57.35	9.97	2.43
Bicycle	1.77	12.34	31.95	48.90	6.13	0.68
Motorcycle	0.85	20.13	15.37	46.75	14.42	3.33

Q4) Did injuries disproportionately affect a specific mode of transportation?

		Injury Class				
Mode of Transport	**Collision Involvement (%)**	**None (%)**	**Minimal (%)**	**Minor (%)**	**Major (%)**	**Fatal (%)**
Vehicle	95.28	79.99	8.00	10.87	0.94	0.20
Pedestrian	2.10	0.84	29.41	$\boxed{57.35}$	9.97	2.43
Bicycle	1.77	12.34	$\boxed{31.95}$	48.90	6.13	0.68
Motorcycle	0.85	20.13	15.37	46.75	$\boxed{14.42}$	3.33

Q5) Do traffic controls lead to decreases in accidents?

All traffic controls together account for a higher number of accidents



**Traffic controls tend to correlate with intersections*

Q6) Can we predict whether an injury will occur or not, using specific accident features?

Decision Tree with maximum depth 6 and criterion entropy

	Precision	Recall	F1-score	Support
0	0.85	0.99	0.92	11552
1	0.88	0.24	0.37	2658
Accuracy	-	-	0.85	14210
Macro Average	0.87	0.61	0.64	14210
Weighted Average	0.86	0.85	0.81	14210

class imbalance ~1:5

Q6) Can we predict whether an injury will occur or not, using specific accident features?

Decision Tree with maximum depth 6 and criterion entropy

	Precision	Recall	F1-score	Support
0	0.85	0.99	0.92	11552
1	0.88	0.24	0.37	2658
Accuracy	-	-	0.85	14210
Macro Average	0.87	0.61	0.64	14210
Weighted Average	0.86	0.85	0.81	14210

class imbalance ~1:5

Q6) Can we predict whether an injury will occur or not, using specific accident features?

Decision Tree with maximum depth 6 and criterion entropy

	Precision	Recall	F1-score	Support
0	0.85	0.99	0.92	11552
1	0.88	0.24	0.37	2658
Accuracy	-	-	0.85	14210
Macro Average	0.87	0.61	0.64	14210
Weighted Average	0.86	0.85	0.81	14210

class imbalance ~1:5

Q6) Can we predict whether an injury will occur or not, using specific accident features?

Decision Tree with maximum depth 6 and criterion entropy

	Precision	Recall	F1-score	Support
0	0.85	0.99	0.92	11552
1	0.88	0.24	$\boxed{0.37}$	2658
Accuracy	-	-	0.85	14210
Macro Average	0.87	0.61	0.64	14210
Weighted Average	0.86	0.85	0.81	14210

class imbalance ~1:5

Q6) Can we predict whether an injury will occur or not, using specific accident features?

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
ridge	Ridge Classifier	0.8412	0.0000	0.2610	0.6931	0.3790	0.3088	0.3576	1.3380
lda	Linear Discriminant Analysis	0.8325	0.7064	0.2913	0.6018	0.3924	0.3086	0.3361	1.0320
lr	Logistic Regression	0.8242	0.7176	0.2872	0.5521	0.3777	0.2871	0.3080	3.9570
gbc	Gradient Boosting Classifier	0.8218	0.6780	0.2551	0.5442	0.3472	0.2593	0.2845	12.9670
ada	Ada Boost Classifier	0.8206	0.6591	0.2558	0.5367	0.3462	0.2571	0.2810	3.6010
rf	Random Forest Classifier	0.8202	0.6976	0.2577	0.5337	0.3473	0.2575	0.2806	4.8450
et	Extra Trees Classifier	0.8191	0.7028	0.2714	0.5256	0.3577	0.2648	0.2845	3.5250
knn	K Neighbors Classifier	0.8153	0.6719	0.2799	0.5058	0.3602	0.2626	0.2784	1.8880
dummy	Dummy Classifier	0.8142	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6000
dt	Decision Tree Classifier	0.7551	0.5914	0.3308	0.3376	0.3340	0.1841	0.1841	0.8070
svm	SVM - Linear Kernel	0.7287	0.0000	0.4309	0.4790	0.3697	0.2346	0.2690	1.0200
nb	Naive Bayes	0.3185	0.7279	0.9568	0.2089	0.3429	0.0545	0.1418	0.6210
qda	Quadratic Discriminant Analysis	0.2301	0.5376	0.9968	0.1940	0.3248	0.0200	0.0968	1.3040
dt_n	**Decision Tree Tuned**	**0.85**	**0.6853**	**0.24**	**0.88**	**0.37**	-	-	-

Q6) Can we predict whether an injury will occur or not, using specific accident features?

	Model	Accuracy	AUC	Recall	Prec.	F1	Kappa	MCC	TT (Sec)
ridge	Ridge Classifier	0.8412	0.0000	0.2610	0.6931	0.3790	0.3088	0.3576	1.3380
lda	Linear Discriminant Analysis	0.8325	0.7064	0.2913	0.6018	0.3924	0.3086	0.3361	1.0320
lr	Logistic Regression	0.8242	0.7176	0.2872	0.5521	0.3777	0.2871	0.3080	3.9570
gbc	Gradient Boosting Classifier	0.8218	0.6780	0.2551	0.5442	0.3472	0.2593	0.2845	12.9670
ada	Ada Boost Classifier	0.8206	0.6591	0.2558	0.5367	0.3462	0.2571	0.2810	3.6010
rf	Random Forest Classifier	0.8202	0.6976	0.2577	0.5337	0.3473	0.2575	0.2806	4.8450
et	Extra Trees Classifier	0.8191	0.7028	0.2714	0.5256	0.3577	0.2648	0.2845	3.5250
knn	K Neighbors Classifier	0.8153	0.6719	0.2799	0.5058	0.3602	0.2626	0.2784	1.8880
dummy	Dummy Classifier	0.8142	0.5000	0.0000	0.0000	0.0000	0.0000	0.0000	0.6000
dt	Decision Tree Classifier	0.7551	0.5914	0.3308	0.3376	0.3340	0.1841	0.1841	0.8070
svm	SVM - Linear Kernel	0.7287	0.0000	0.4309	0.4790	0.3697	0.2346	0.2690	1.0200
nb	Naive Bayes	0.3185	0.7279	0.9568	0.2089	0.3429	0.0545	0.1418	0.6210
qda	Quadratic Discriminant Analysis	0.2301	0.5376	0.9968	0.1940	0.3248	0.0200	0.0968	1.3040
dt_n	**Decision Tree Tuned**	**0.85**	**0.6853**	**0.24**	**0.88**	**0.37**	-	-	-

Comparatively, the fine tuned decision tree did better on most metrics except 2% worse on the recall score

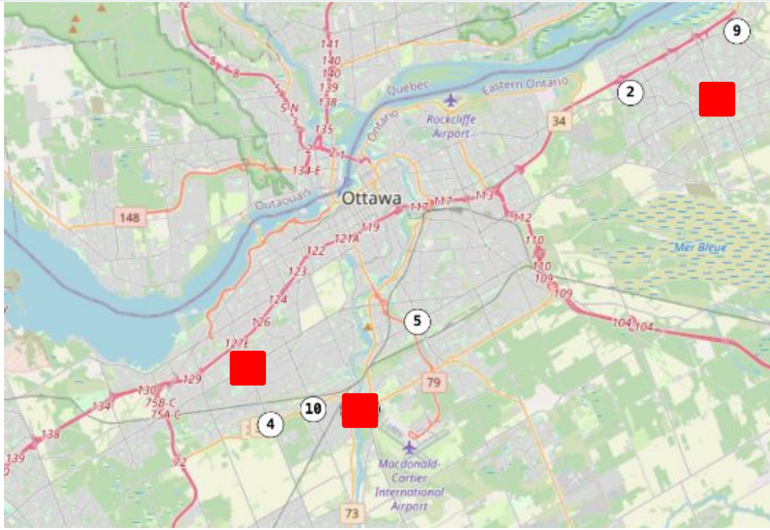
Implications

Why Study This?

- Highlighting dangerous areas
  - For Ottawa residents to be mindful
  - For City of Ottawa officials to look into these
- Uncovering relevant predictors of collision injuries
  - Allow improvements to be made based on features most likely to end in injuries
- With better predictive power → dispatchers could know if there will be an injury or not

Summary

Summary



- Top Accident/Injury Locations
  - Hunt Club and Riverside/Prince of Whales
  - Innes and Tenth Line
  - Baseline and Woodroffe
- Top Major/Fatal injuries Location
  - 8th Line and Parkway
- Correlated Features to Injuries
  - Slight correlation with modes of transportation (pedestrian, bike, motorcycle)
- Modes of Transportation and Injuries
  - Motorcycles < 1% of accidents, but 14% end in major injuries and 3% in fatal
  - Pedestrians account for 2% of accidents, with ~2.5% ending in fatal injuries with 57% in minor injuries
- Predictive Model
  - 85% accuracy (better than chance)

Things the City can do to mitigate injuries

- Unfortunately, there were not highly correlated items from the accident features
- Dataset does not have many features the City can control
  - e.g. does not have speed

BUT...

Main contributions

- Comprehensive list of “dangerous” locations
  - City qualitatively analyze why these are dangerous
- Levels of injury for various modes of transportation
  - Public Service Announcements
- Predictive model (better than dummy)
  - Can further be improve with more data sources (e.g. speed)

Future Work

Future Work

- Incorporating speed to improve predictions
- Looking at volume to improve our analysis to be more comparative between locations
- For the time trends, it seems like there continues to be a decrease on average for the years following COVID. Are the trends proportionally decreasing?
  - e.g. hybrid workers only go to office 2-3 times per week, is there a $\frac{2}{5}$-$\frac{3}{5}$ decrease in accidents compared to other years?
- Look at seasonal trends: are there certain weather/environmental conditions that are correlated to injuries based on season
  - e.g. look at just winter months and see if snow is an influencer of severity of injuries
- Can we predict 2023 injuries?

Thank you!