REDUCING TRAFFIC ACCIDENTS

DERIC WILLIAMSON

INTRODUCTION

For the Chicago Department of Transportation, as a data scientist, my job is to go over the city's traffic accident reports and create a model that will predict the worst incidences. In doing so, I will find some underlying causes and possible advice for the city.

For the Chicago Department of Transportation, this data will help reduce the number of traffic accidents that result in injury.

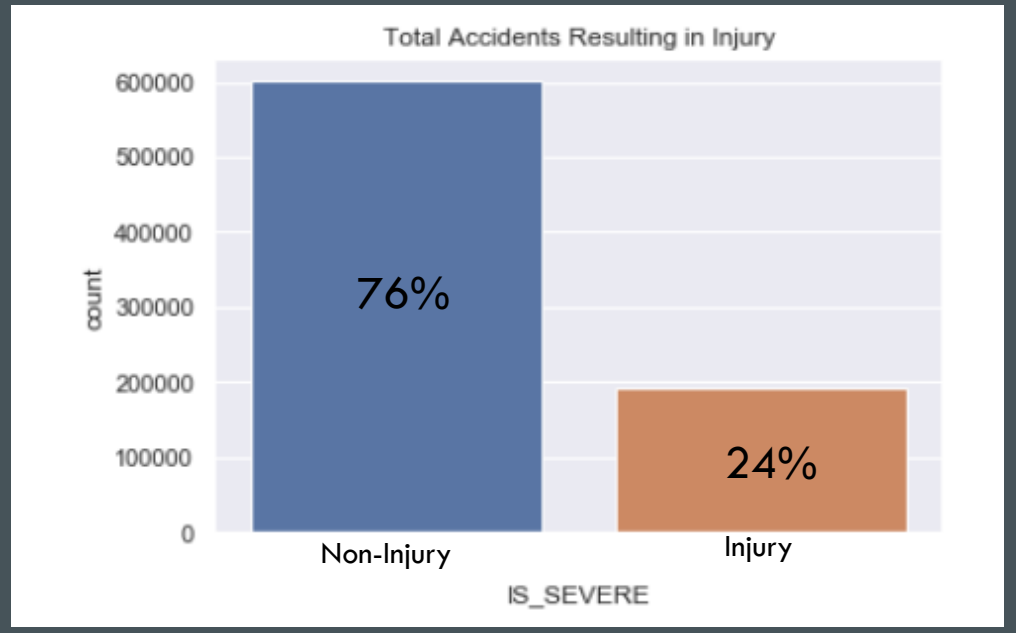
Drivers will always be distracted while driving, so what can we do to help reduce injury in accidents?

THE DATA

The data was taken from cityofchicago.org/transportation and was created from the city of Chicago's electronic crash reporting system from 2015 to present.

There were three different datasets that I derived information from. The datasets were about the people, the vehicles, and the specific crashes involved which I combined into one dataset for this project.

The data was used to predict whether an accident resulted in an injury or not, as shown by the graph. This shows that the data was **imbalanced** at 24 percent resulting in some type of injury.



THE MODEL

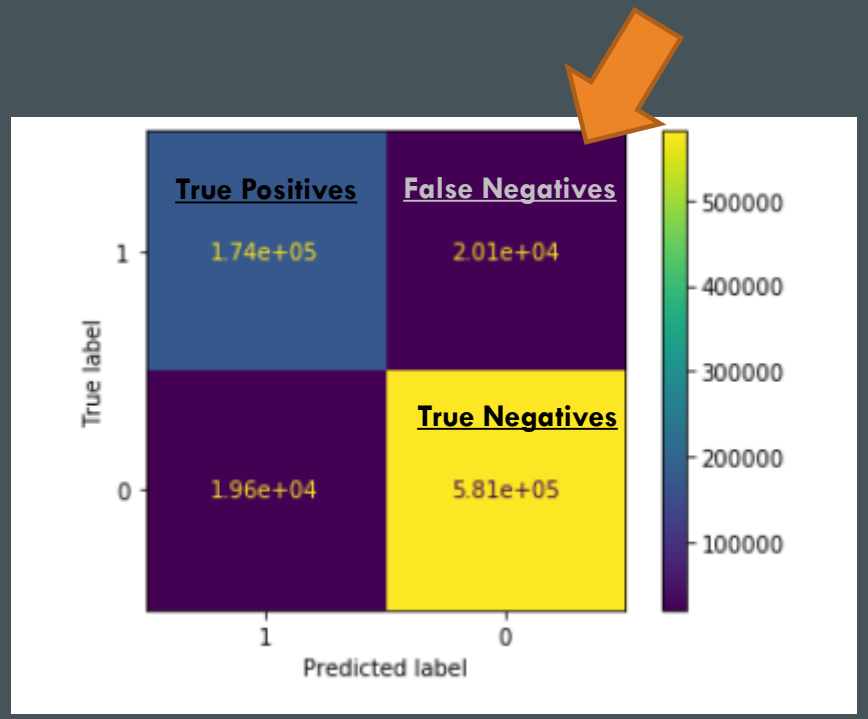
After fitting the data into several different models, I ended up using the Decision Tree model.

I chose the Decision Tree because it scored the highest on Accuracy, Recall, and AUC (Area Under Curve)

Recall is important because if a crash results in an injury, we need to know about it. We need to lower false negatives thus raising recall scores.

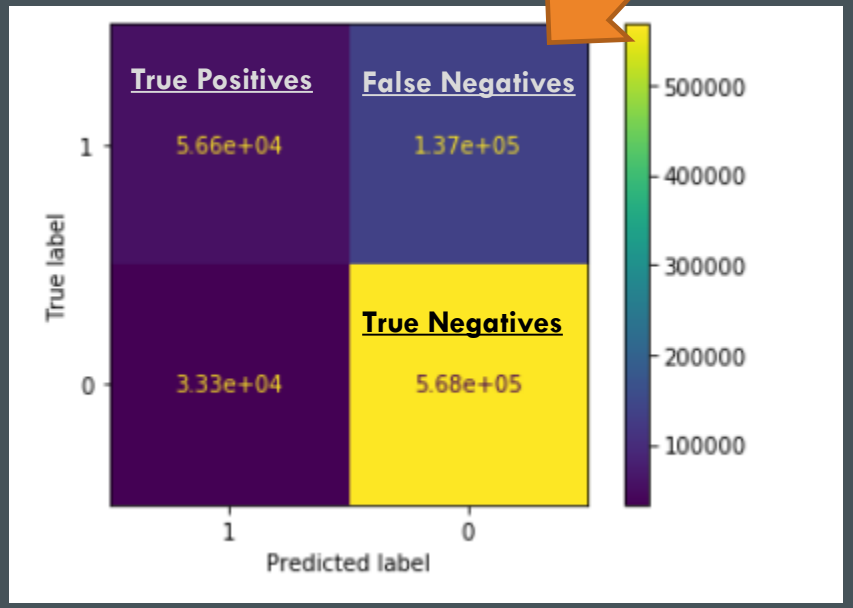
…the next step in the model process is to fine tune the model to fix over fitting…

	precision	recall	f1-score	support
0	0.84	0.84	0.84	120197
1	0.51	0.52	0.51	38722
accuracy			0.76	158919
macro avg	0.68	0.68	0.68	158919
weighted avg	0.76	0.76	0.76	158919



THE FINAL MODEL

	precision	recall	f1-score	support
0	0.81	0.94	0.87	120197
1	0.63	0.29	0.40	38722
accuracy			0.79	158919
macro avg	0.72	0.62	0.63	158919
weighted avg	0.76	0.79	0.75	158919



…After fine tuning, the model will be able to handle unknown data more efficiently by eliminating over fitting.

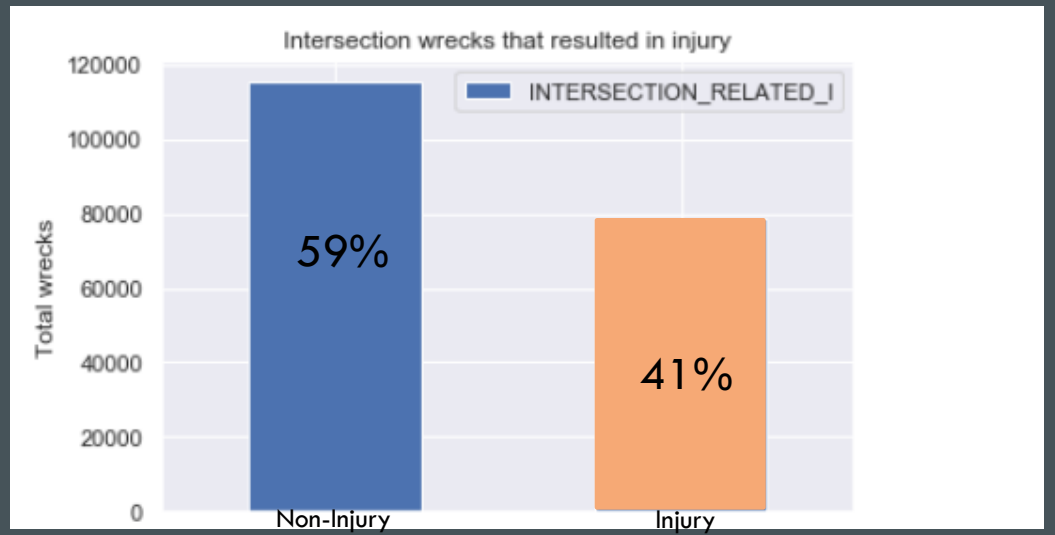
For my final model, my recall score went down while the number of false negatives increased.

For future work on this model, I would work on raising the ROC (receiver operating characteristic) threshold to aide in lowering false negatives, thus raising the recall score of the model.

IMPORTANT ACCIDENT FEATURES

	features	importance
0	INTERSECTION_RELATED_I_Y	0.304175
1	FIRST_CRASH_TYPE_PEDESTRIAN	0.150219
2	FIRST_CRASH_TYPE_REAR END	0.094665
3	FIRST_CRASH_TYPE_SIDESWIPE SAME DIRECTION	0.087648
4	FIRST_CRASH_TYPE_FIXED OBJECT	0.075503
5	FIRST_CRASH_TYPE_TURNING	0.059234
6	NUM_PASSENGERS	0.052805
7	PRIM_CONTRIBUTORY_CAUSE_FAILING TO REDUCE SPEE...	0.036606
8	FIRST_CRASH_TYPE_PEDALCYCLIST	0.036019
9	FIRST_CONTACT_POINT_OTHER	0.019419

The model ranked the most important features contributing to injuries, with the top 3 being related to **intersections, pedestrians, and accidents involving rear-ending.** This table shows us that the highest importance was intersection related.



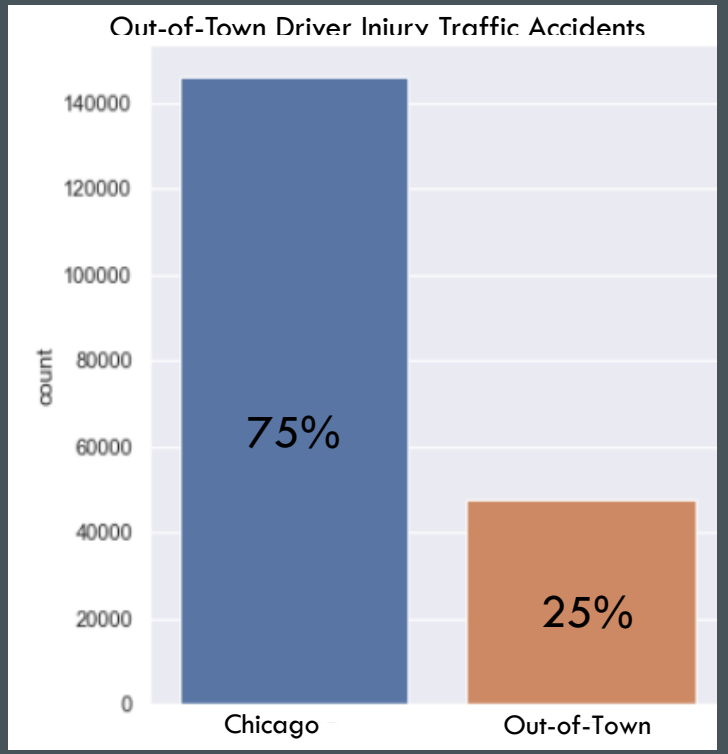
The graph shows that 41% of intersection related accidents resulted in some type of injury.

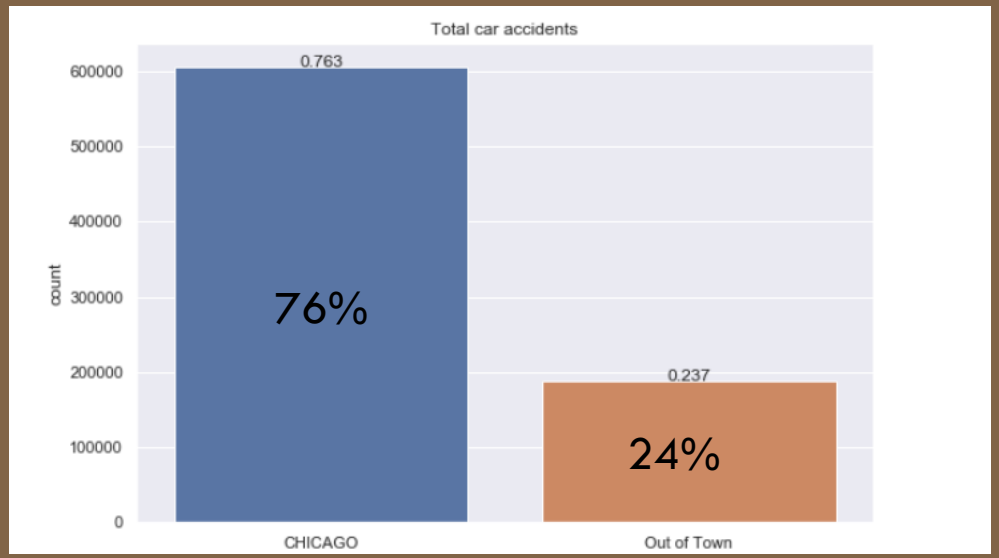
My recommendations to reduce intersection related accidents such as rear-ending, and failing to reduce speed include:

- Widening traffic lanes to allow for a clear line of sight for drivers turning right on red
- Placing more roundabouts in high traffic intersections- this makes you pay more attention to what you're doing and less dependent on traffic lights

CITY VERSES OUT OF TOWN DRIVERS

Driving in Chicago's traffic can be difficult for anyone, especially for drivers that are unfamiliar with the city's complicated freeway systems and busy rush hour traffic. The dataset, as reflected in the *Total Car Accidents* graph below, showed that 24% of total accidents were involving out-of-town drivers. The graph on the right, titled *Out-of-Town Driver Injury Traffic Accidents*, shows that 25% of injury car accidents were also caused by out-of-town drivers.





The city will need to investigate ways to make Chicago's roadways easier for out-of-town drivers to navigate, whether it be to make the roadways less complicated or to be more diligent on good road sign placement.

RECOMMENDATIONS

Intersections and Rear-ending

- Decrease speed limits
- Placing more roundabouts in high traffic intersections

Pedestrians

- Widen lanes and add more right sided shoulders
- Make roadways more pedestrian friendly by adding additional sidewalks and using more signs on the sidewalks
- Increase the distance between the road and the sidewalk
- Increase the usage of raised pedestrian crosswalks

FUTURE WORK

Most of my future work will be focused on improving my model by:

- Raising ROC curve thresholds as previously mentioned to raise recall score
- Running a scatter plot on misclassified data to find a common theme as to why it was misclassified
- Looking at feature importance on the chosen model and drop the lowest to speed up future works

THANK YOU!

Thank you to my audience for giving me your time and attention during my presentation today.