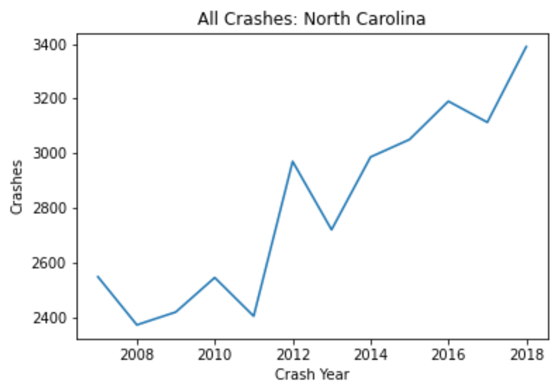
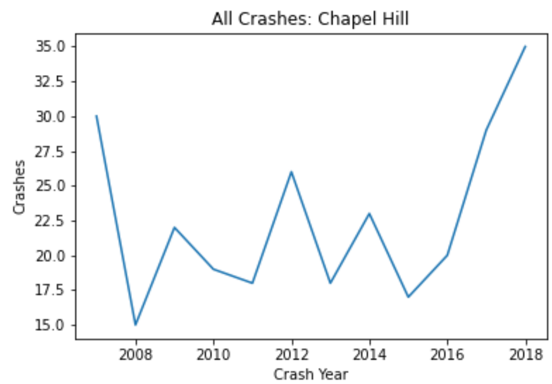
Examining Proposed Public Transportation Changes in Chapel Hill, NC in the Interest of Public Safety

A Fictional North Carolina Pedestrian Crash Case Study

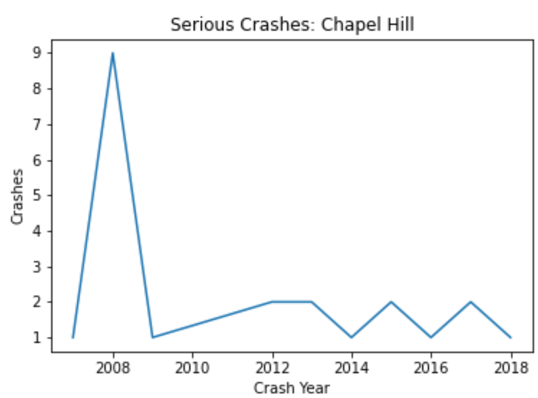
Introduction

The number of pedestrian crashes has been steadily rising y/y in the state of North Carolina. The city of Chapel Hill, NC has seen a sharp interest in overall numbers of pedestrian involved traffic accidents. Although the number of serious accidents in Chapel Hill has remained low, the growing number of overall accidents has the city concerned about potential increases in more serious accidents.





Chapel Hill, NC is home to the campus of the University of North Carolina (UNC) and is concerned with keeping its citizens and students safe. In a pre-emptive measure, the city is investigating a number of possible changes to decrease the number of serious accidents including: decreasing speed limits by 10 mph, placing traffic controls where none currently exist, stricter impaired driving measures, and discouraging young drivership (through improved public transportation/infrastructure).



Methods

Data Sources. Data for this analysis was taken from the North Carolina Department of Transportation's data set through the Town of Chapel Hill's GIS/Analytics portal which consists of 33707 records with 52 features. City population data was obtained from a publicly available source.

Data Treatment – Selected Features. The following features were used for modeling: city, whether or not an ambulance was required, driver intoxication, crash time/date, pedestrian age, driver age, lighting conditions, hit-and-run status, pedestrian intoxication, type of road, speed limits, presence of traffic control, weather, vehicle size, and city size. Other features were excluded either due to their intermittent or low prevalence, such as work zone designations, or redundant features, such as membership to other categories. Some complex features, such as road configuration, were excluded. Crash Severity was selected as the target label, which indicates the extent of injuries resulting from the crash.

Data Treatment – Encoding/Scaling. With the exception of city population, all selected features were categorical values. The categories for each feature were manually examined. The treatment of each feature is described below.

Crash Month	These were encoded to a numerical value (i.e. Jan = 1)
Crash Day	These were encoded to a numerical value (i.e. Sunday =1)
Vehicle Type	There were over 20 unique entries. These were encoded to either 'small' or 'large' vehicle types. See documentation for specific encodings. One-hot encoded.
Daylight	Used existing categories. One-hot encoded.
Age Group	Age groups were not consistent and had groupings such as 21-25 *and* 20-24. These groupings were combined. See documentation for specific groups that were combined. One-hot encoded.
Road Class	Used existing categories. One-hot encoded.
Speed Limit	Used existing categories. One-hot encoded.
Traffic Control	Categories were collapsed to a binary value of whether or not traffic control was present.
Weather	Used existing categories.
Alc/Drug Status	Categories were collapsed to a binary value of whether or not drugs/alcohol were suspected/present or not.
Hit and Run	Binary encoding.
Crash Severity	Categories were collapsed to either a serious injury or not. Serious injuries were "suspected serious injury" or "killed".
City population	City populations were scaled using a standard scaler on the training split.

Data Treatment – Missing Values. Not all categories had missing values. Categories with missing values were treated as described below.

Vehicle Type	Imputed with most common vehicle
Crash Day/Time	Only 1 value was a time and was dropped.
Lighting Condition	Lighting conditions were imputed based on the hour of the incident. Between 8AM and 8PM was designated as 'Daylight' and between 8PM and 8AM was designated as 'Dark- Unknown Lighting'.
Age Group	Only 258 values were missing and values were imputed using the most common value (30-39).
Speed Limit	There were 3656 missing speed limits. All missing speed limits had a road class value. Missing speed limit values were imputed using the most common

	speed limit for that road class present in the data set (i.e. the record is designated as a local street with a missing speed limit – the speed limit is imputed to be the most common speed limit for local streets).
Crash Severity	This was the label for modeling. 336 Missing values were dropped. The missing labels were examined for any sort of pattern. Though values with hit and runs were more present than in the labeled set, they did not constitute the majority of missing data.

Modeling. Approaches based on supervised learning were explored, using the crash severity as the data labels. Severe injuries or death were labeled as the 'positive class' and non-serious or no injuries were labeled as the 'negative class'. Classifiers explored were: KMeans, SVM, Randomized Forest, and Gradient Boosting. Data reduction through principal component analysis were explored, but PCA implementations gave inferior results compared to using the features as prepared. After initial evaluation of the performance of the various models, a Gradient Boosting classifier was chosen as the final model based on a balance of accuracy, recall and prediction time. Hyper-parameters were chosen using a randomized search method, with 'roc_auc_ovo_weighted' as scorer. A metric that focuses more on precision/recall was chosen due to the weight imbalance between classes and the utility of properly identifying the positive class. See documentation for the various performance of the that were explored or not selected.

Selected Model. The Gradient Boosting Classifier hyperparameters were selected using a semi-manual approach. The selected model and parameters had acceptable levels of accuracy, f-score, and minimal overfitting.

Table 1. Model Training and Testing Performance

Train	Precision	Recall	F1-Score	Test	Precision	Recall	F1-Score
Non-Serious	0.95	0.77	0.85		0.93	0.76	0.84
Serious	0.34	0.74	0.47		0.33	0.69	0.44
Accuracy			**0.77**				**0.75**

Results/Discussion

Prediction Method and Predictions. In order to get insight into the effects of policy changes specifically within Chapel Hill's demographic, application of the trained model was focused specifically on Chapel Hill. The model's performance on the Chapel Hill section of the data set are given in Table 2.

Table 2. Model Performance on Past Chapel Hill Incidents

	Precision	Recall	F1-Score
Non-Serious	0.95	0.92	0.94
Serious	0.35	0.50	0.42
Accuracy			**0.89**

A selection of 10,000 incidents were bootstrapped sampled from the Chapel Hill for the purposes of policy effect prediction. The baseline prediction with no policy changes predicts that approximately 12% of all incidents are serious. Interestingly, the model predicts a slight *increase* in serious incidents with the installation of traffic controls; however counterintuitive, this phenomenon does have precedence with intersections that had traffic controls being removed to created so-called shared spaces. Reducing speed limits in on roads that range from 30-45 mph by 10 mph reduced the occurrence of serious incidents, with a greater effect observed for lowering 30/35 mph zones to 20/25 mph. Reducing young drivership or drug/alcohol related policies had little predicted effect. Implementing all of the proposed policy changes resulted in a reduction of serious incidents down to 4.54% of all incidents.

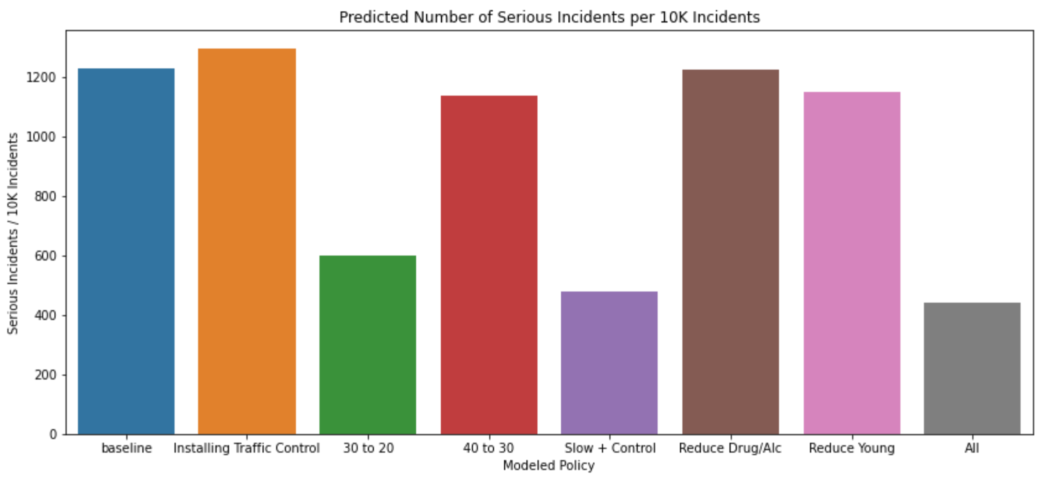


Figure 1. Predicted number of serious traffic incidents given implementation of new local policies. **'Baseline'** indicates the number of serious accidents predicted per 10,000 incidents in Chapel Hill. **'Installing Traffic Control'** indicates placing traffic control at intersections/locations where none was previously present. **'30 to 20'** and **'40 to 30'** indicate lowering speed limits on roads from 35/30 to 25/20 and 45/40 to 35/30 respectively. **'Slow + Control'** indicates implementing the proposed speed limit reductions *and* installing traffic controls. **'Reduce Drug/Alc'** indicates implementing policies to *eliminate* impaired drivers. **'Reduce Young'** indicates reducing young drivership (24 yrs and younger) by 20%. 'All' indicates implementing all of the policies proposed.

Combined Effects. Though installing traffic controls was predicted to slightly increase the rate of serious accidents, when combined with other measures, traffic controls had a synergistic effect. Including traffic controls in combination with lower speed limits had a greater benefit than lowering speed limits alone. However, the differences observed when including traffic control measures is relatively minor compared to the larger effect of speed reduction.

Policy	Serious Rate
Baseline	**12.3%**
Traffic Controls Only	13.0%
Reducing 40's to 30's Only	11.4%
Reducing 40's to 30's + Traffic Controls	11.8%
Reducing 30's to 20's Only	6.0%
Reducing 30's to 20's Only + Traffic Controls	6.0%
Reducing both 40's and 30's to 30's/20's	5.1%
Both Reductions + Traffic Controls	4.81%

Conclusion

General. The case study addresses the fundamental questions regarding the expected effectiveness of various transportation policy changes at curbing the rate of serious traffic accidents with pedestrians in the locality of Chapel Hill. Of the measures proposed: increasing traffic controls, reducing speed limits, reducing impaired driving, reducing young drivership, speed limit reductions had the greatest predicted effect on the rate of serious pedestrian accidents involving motor vehicles. The most effective predicted policy would be to reduce the speed limits of roads with 35 and 30 mph limits to 25 and 20 mph respectively. The proposed change would reduce the rate of serious incidence by 50%. If multiple policies are considered, there appears to be a synergistic effect, through with diminishing returns.

Limitations. The model implemented utilized traffic data for the entire state of North Carolina and assumes that the driving habits and accident characteristics are generalizable for the entire region. The city of Chapel Hill is home to a flagship University, which makes it unusual compared to other townships of similar population sizes.