Bike and Car Accident Prediction

Final Report

Helping Drivers and Bicyclists Share the Road

Many drivers hate sharing the road with bicyclists, but perhaps they could become allies, if road improvements for bikes made the road safer for cars as well and vice versa.

This exploration looks for commonalities between what causes bike and car accidents in the Boston Area

Data Sources

- A database of Boston bike accidents from 2009 to 2012, constructed from police reports during this time
- A database of car accident/crashes for the entire US in 2017
- The Massachusetts Department of Transportation (Mass DOT) road inventory
- The Mass DOT road inventory has a data dictionary

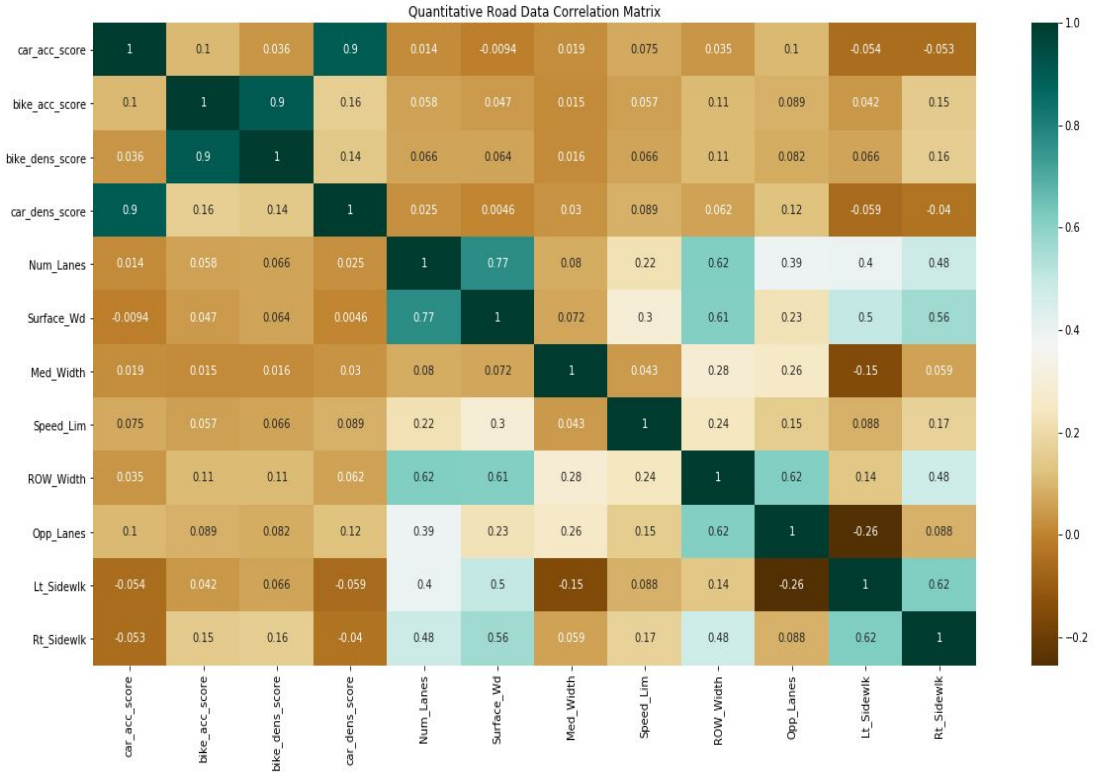
Data Wrangling

- Create clusters of accidents within 50m of one another using shapely and geopandas packages
- Associate road data with those clusters
- Create areas without accidents and associate road data with those

Initial Findings

Many road attributes are not strongly correlated with higher rates of accidents, including:

- Traffic
- Road width
- Number of lanes



Findings With Statistical Significance

- **Facility type:** Facility Types like tunnel, ramp, double deck, rotary and bike lanes all have higher accident density than regular roads
- **Number of Lanes:** roads with 1 and 3 lanes have a higher mean *car* accident density, roads with 4 lanes have a higher mean *bike* accident density
- **Truck Access:** roads that exclude trucks have fewer accidents than roads that allow them
- **Surface Type:** gravel or stone, block road, surface-treated roads and Portland cement roads all have more accidents than regular bituminous road
- **Speed Limit:** raising the speed limit from 20 to 30 increases *bike* accident density

Machine Learning

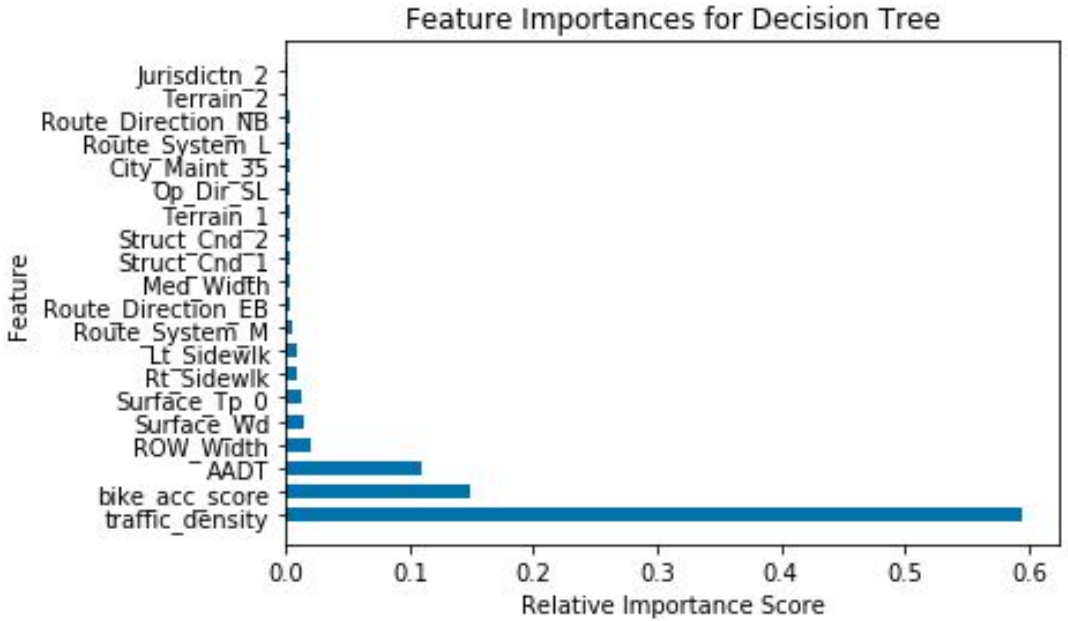
- No regression models effectively predicted the **number** of car and bike accidents
- Classification of areas as being likely to have an accident was much more effective

Model	Vehicle	Accuracy
DecisionTreeClassifier	car	0.886
DecisionTreeClassifier	bike	0.902
RandomForestClassifier	car	0.885
RandomForestClassifier	bike	0.908
BaggingClassifier	car	0.895
BaggingClassifier	bike	0.913
GradientBoostingClassifier	car	0.885
GradientBoostingClassifier	bike	0.895
AdaBoostClassifier	car	0.859
AdaBoostClassifier	bike	0.884

Model	Vehicle	Accuracy
DecisionTreeClassifier	car	0.886
DecisionTreeClassifier	bike	0.902
RandomForestClassifier	car	0.885
RandomForestClassifier	bike	0.908
BaggingClassifier	car	0.895
BaggingClassifier	bike	0.913
GradientBoostingClassifier	car	0.885
GradientBoostingClassifier	bike	0.895
AdaBoostClassifier	car	0.859
AdaBoostClassifier	bike	0.884

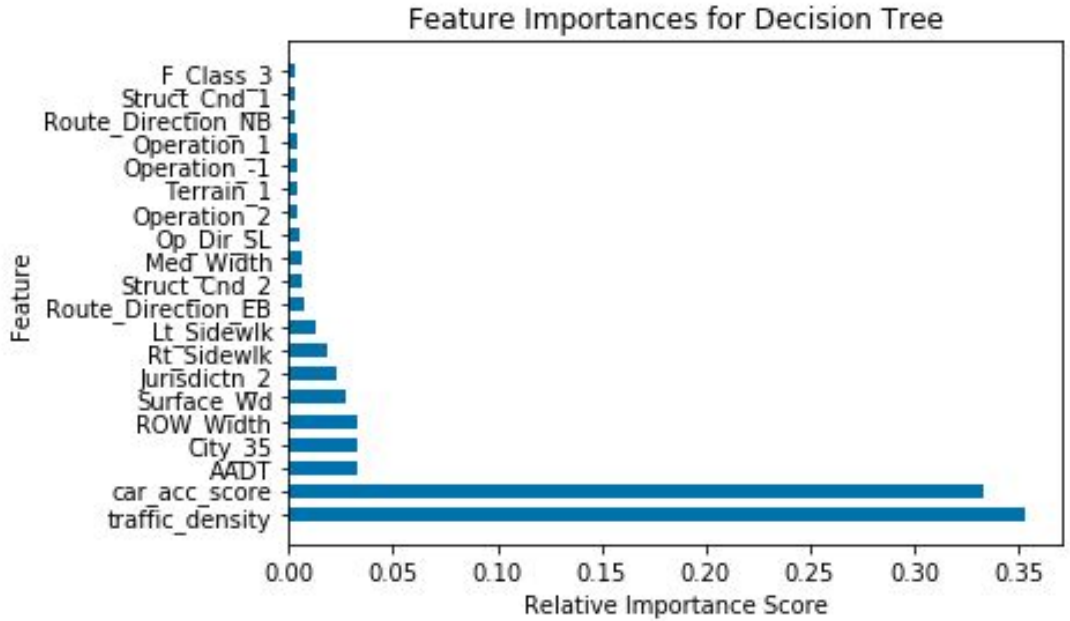
Feature Selection

Most important features for predicting car accidents



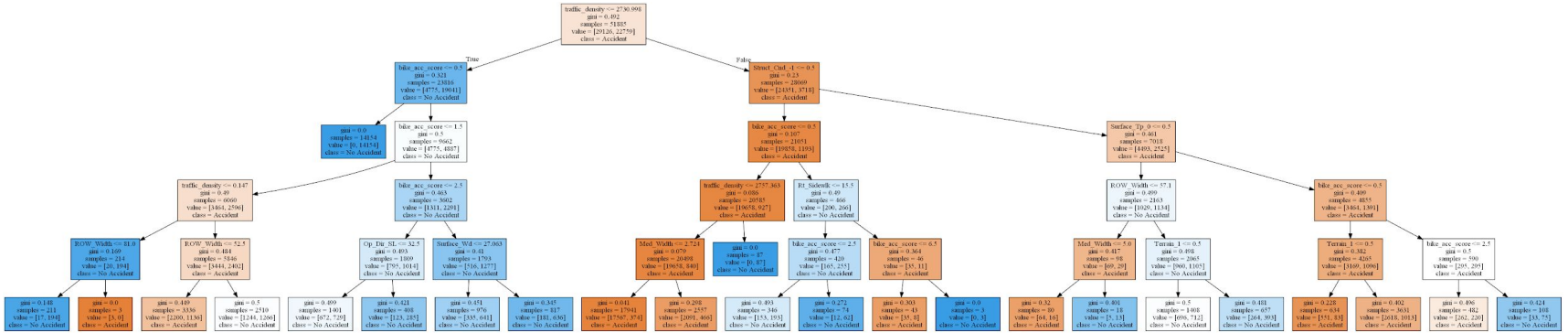
Feature Selection

Most important features for predicting bike accidents



Decision Tree

A simple decision tree predicted car accidents fairly effectively, giving a good guide for road designers



Conclusions and Next Steps

- Bike accidents do help predict car accidents and vice versa, supporting my initial thought that making an area safer for bikes might make it safer for cars
- Traffic volume is the most important feature for predicting both types of accidents. This argues for increasing public transportation to decrease traffic.

Some potential next steps include:

- *Collecting more recent bike accident data, since this data set was collected in 2012*
- *Seeing if the model predicts car accidents for areas outside Boston, or if Boston road conditions are unique*
- *Creating models for other cities and seeing if they support similar conclusions*