

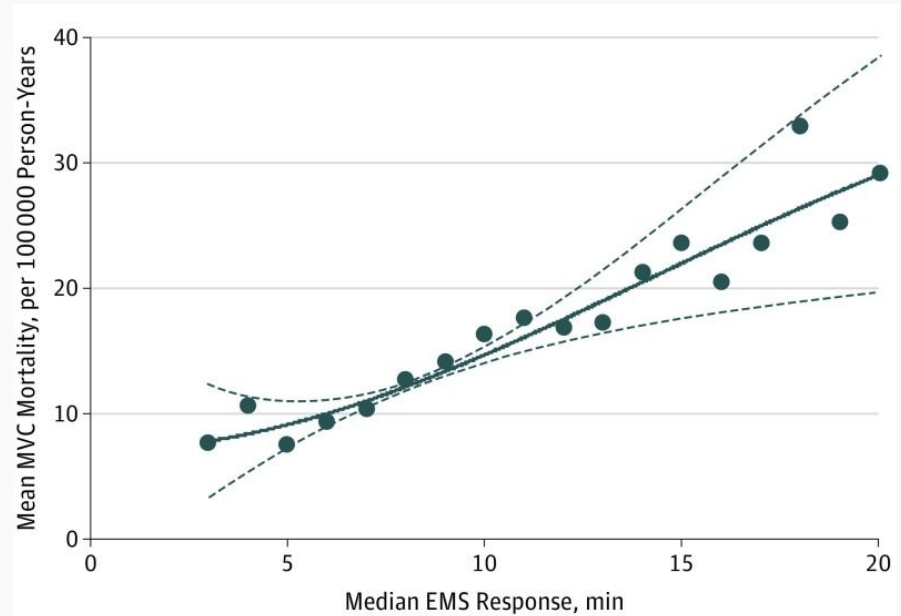
Vehicle Collision Classification

Chris Gardner



Vehicle Collisions: Leading cause of death

- Vehicle collisions accounted for ~35,000 deaths in 2015 (US)
- Slow response to incidents is associated with higher mortality rates



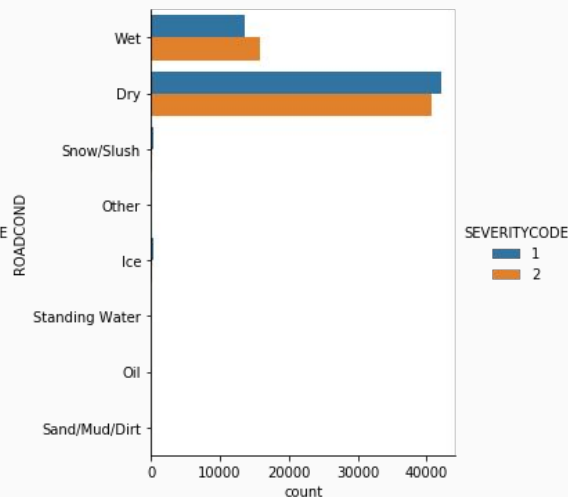
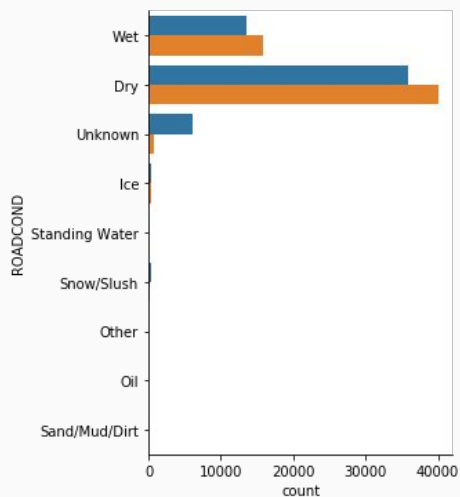
(Byrne JP, Mann NC, Dai M, Mason SA, Karanickolas P, Rizoli S, Nathens AB. Association Between Emergency Medical Service Response Time and Motor Vehicle Crash Mortality in the United States.)

Data Acquisition

- GISWEB: Collision - All Years Dataset
 - NOAA: NY Weather Station Data (by date)
 - Data.gov: Ny Collisions data
-
1. Clean and merge NOAA and Data.gov -> Collisions + Weather Conditions
 - a. All features have low correlations. Need more/better data.
 2. GISWEB
 - a. Only need to drop unnecessary columns.

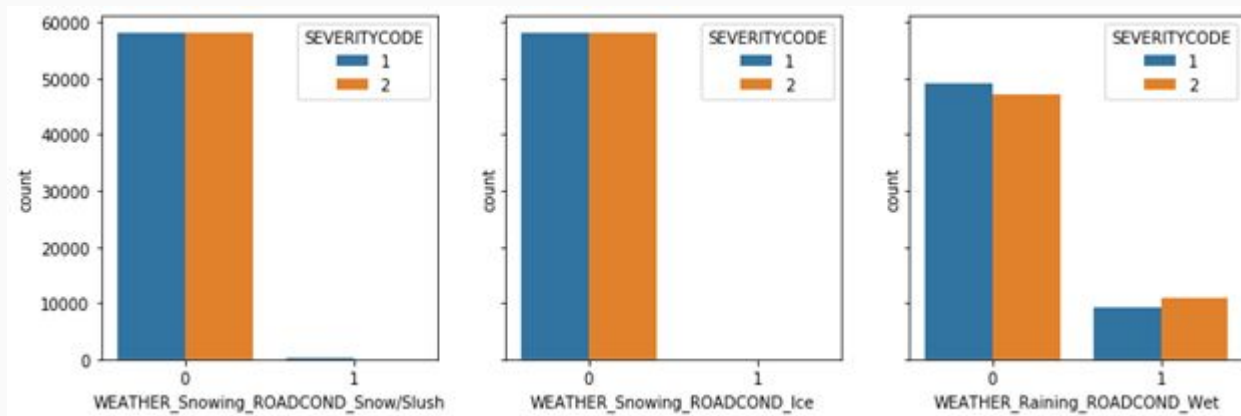
Data Cleaning: Unknowns

Unknowns were very prevalent in the data and provided a significant impact on correlation/predictions. Replaced unknown data values with the mode of the data.



Data Features

Using feature interactions to assess combinations of weather and road conditions



Model Evaluation

After comparing logistic regression, decision tree and k-nn. We found logistic regression gives us the most useful model.

Class	Precision	Recall	f1-score
Property Damage	0.58	0.73	0.65
Injury	0.64	0.48	0.55

We would like to have higher recall, the model is not capturing most of the incidents with injuries.

Findings

1. Incident location/type has the largest impact on crash severity. With high correlation between collisions occurring at intersection and injuries.
2. Weather and road conditions are also good predictors of crash severity. Especially with the interaction of rainy days with wet roads correlating to injury while clear day + dry roads do not.

Future Work

1. Need to gather more data and more specific data. Since we had a large issue with 'Unknowns' in the data set we need to collect more data without this issue. This is vital for building a good model
2. Want to investigate other data features such as location and time of day
3. Combine severity classification with collision frequency regression.