Predicting the Likelihood of a Severe Accident

Dona Ray

The Problem

Auto Road Accidents

Consequences:

Emotional loss

Medical bills

Loss of Productivity

Property Damage

Increase in Insurance Costs

Taxpayer costs

Time spent by Police Authorities

Traffic delay

Question

What are the factors that cause a fatal or severe accident?

Weather

Road Surface Conditions

Road Type

Junction Type

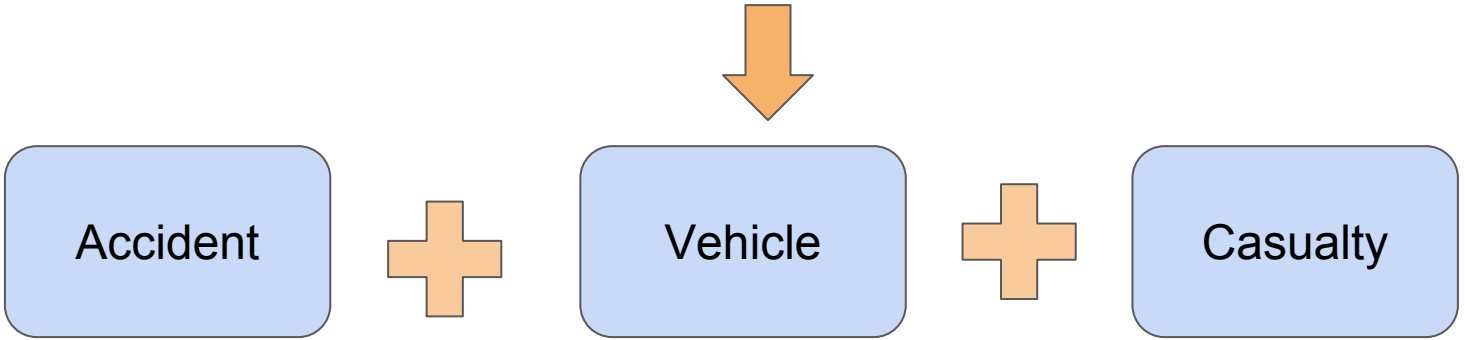
Day and Time

Sex of Driver

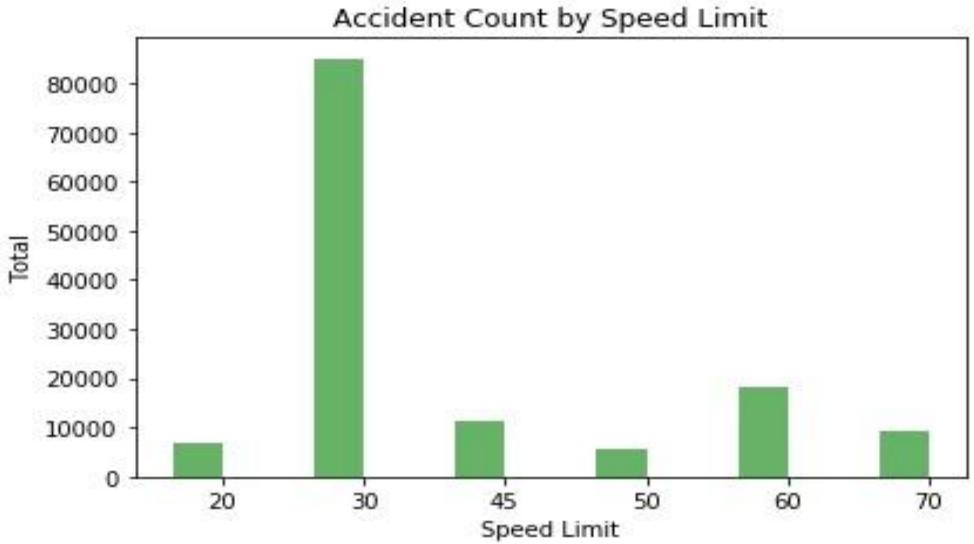
Age of the Driver

Speed Limit

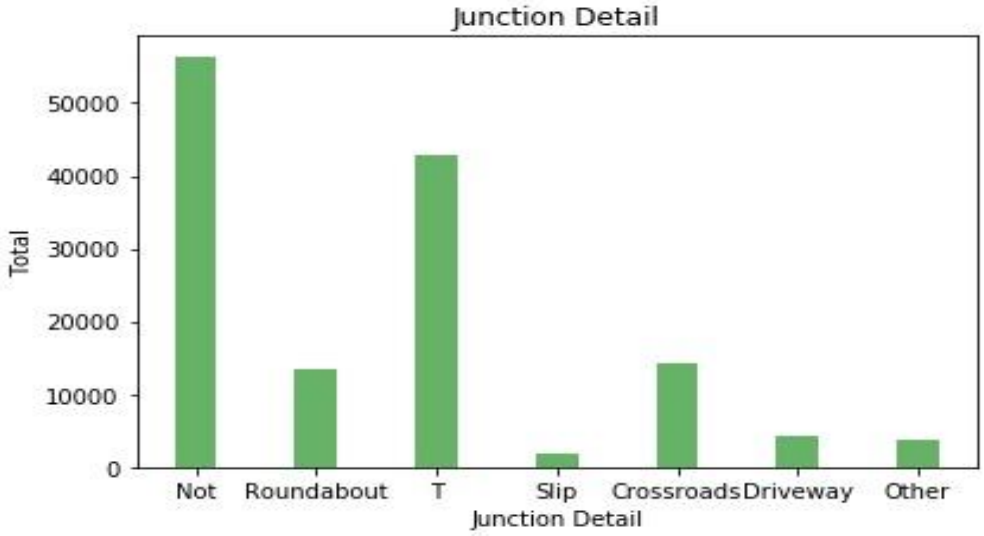
Data: Government of UK, 2016



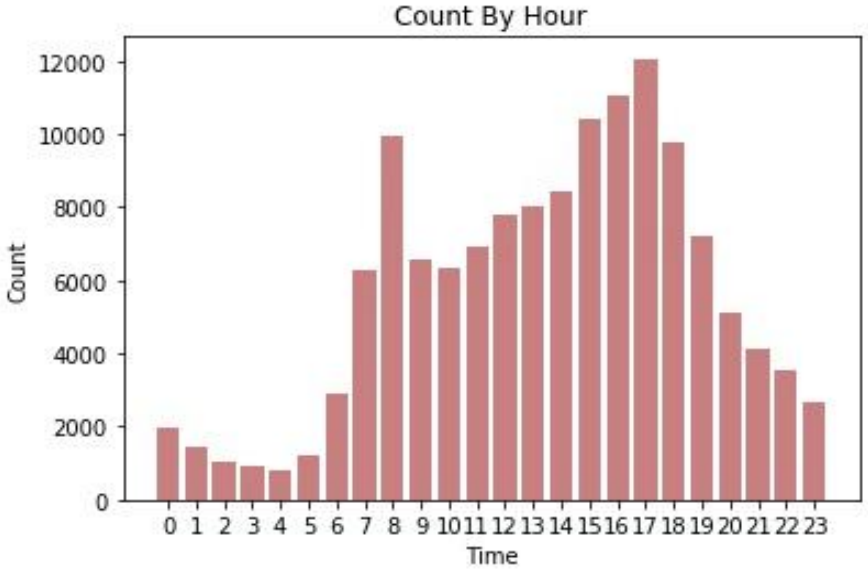
CSV Format
Merge by Accident ID, Vehicle ID
252, 481 observations

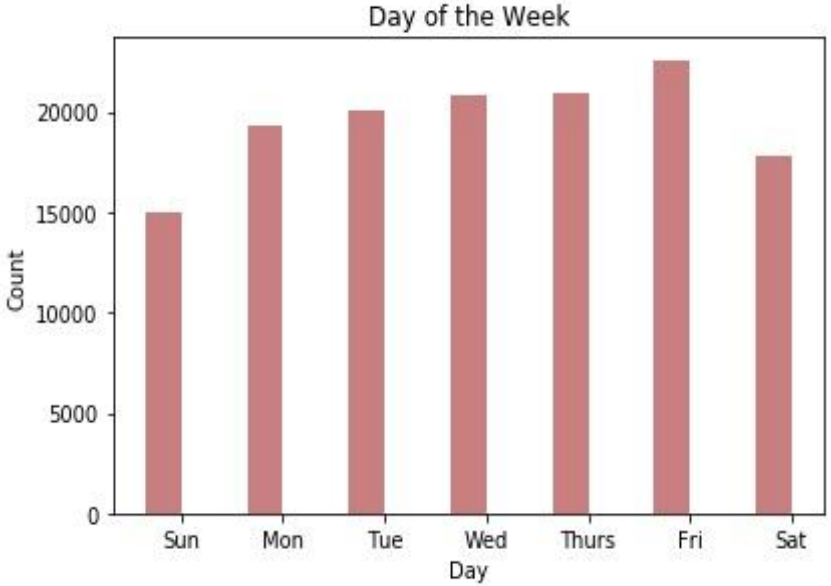


Mode: Speed Limit 30

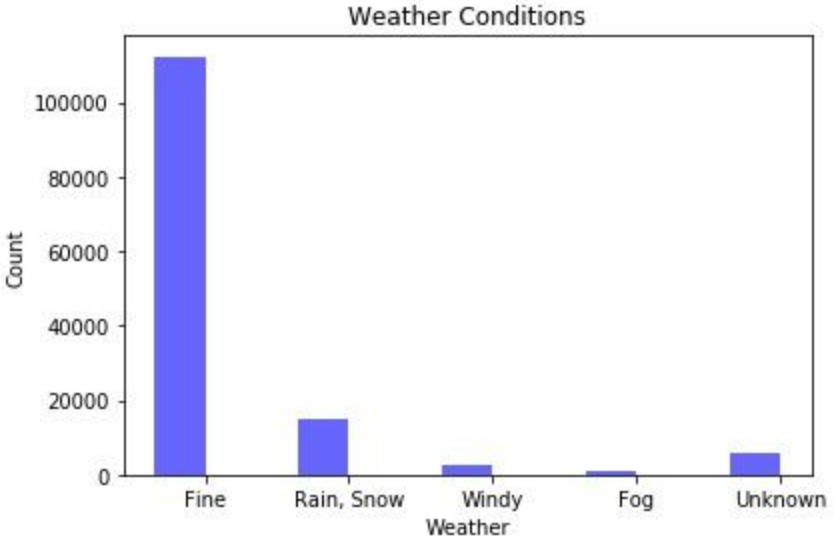


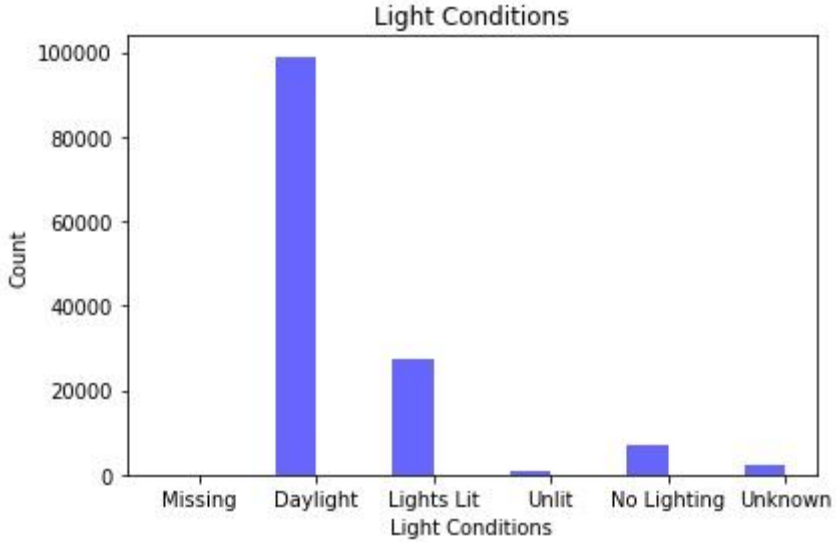
T-sections
Roundabouts
Crossroads



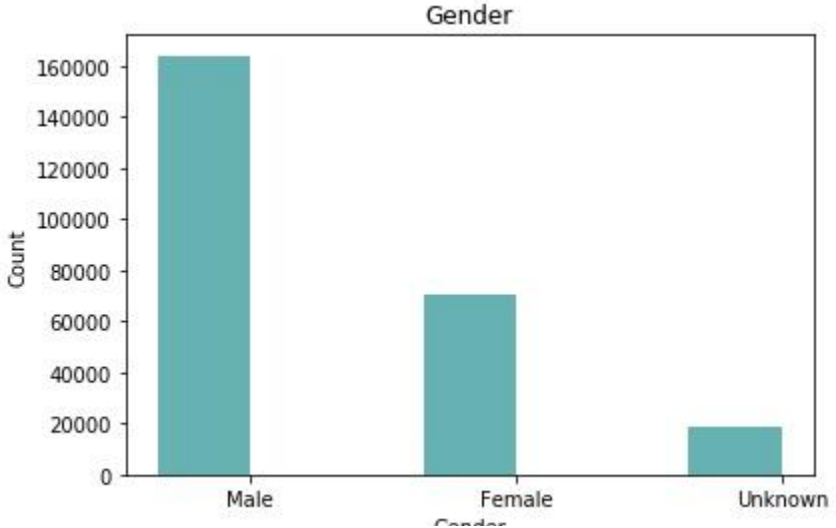


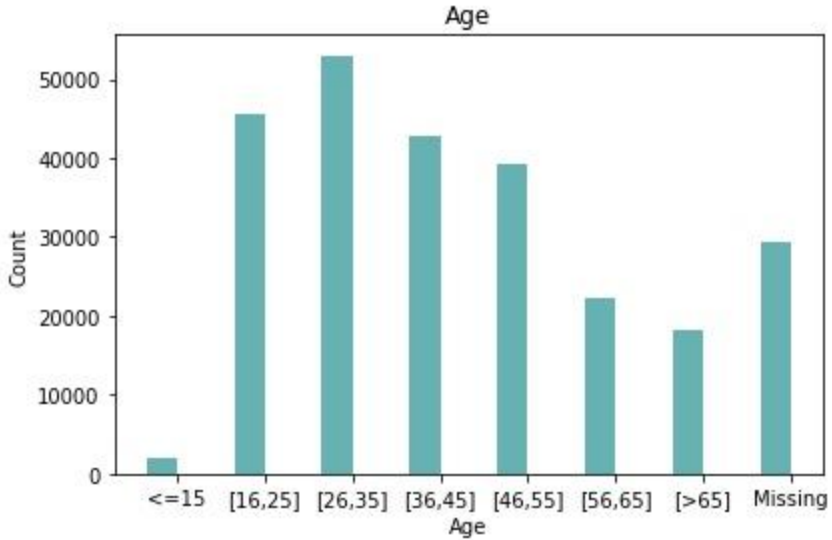
Rush Hour Traffic leads to more accidents !!





Bad Weather and Poor lighting have no effect!!





Males in age band [26, 35] have higher incidences!!

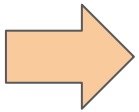
- Data Preprocessing
  - Balance the data
- Cross-Validation
  - 80% Training and 20% Test data
  - K-Fold Cross-Validation
  - GridSearch for Hyperparameter Tuning
  - Scoring Metric
    - F-score
- Modeling Steps
  - Supervised Learning
    - Naive Bayes
    - K Nearest Neighbors (KNN)
    - Logistic Regression
    - Decision Tree
    - Random Forest

Model Comparisons

Classifier	Unoptimized			Optimized		
	Accuracy	F-score	ROC score	Accuracy	F-score	ROC score
Naive Bayes	-	-	-	0.66		0.58
k-NN	0.59	0.21	0.52	0.64	0.22	0.54
Logistic Regression	0.61	0.25	0.58	0.61	0.25	0.58
Decision Tree	0.64	0.22	0.56	0.70	0.25	0.57
Random Forest	0.70	0.21	0.54	0.68	0.26	0.59

Feature Importance

Feature	Gini Importance
Age	0.4225
Hour	0.2335
Speed Limit 60	0.0656
Dry Road	0.0503
Good Weather	0.0361
Speed Limit 30	0.0268
Single Road	0.0258



Age and Time of day are the key predictors of a severe accident.

Policy Recommendation

Staggered work timings to help reduce traffic on the roads

Encourage work from home

Additional signs at T-sections and Crossroads

Consider possibility of reducing 30 and 60 speed limits

Billboard signs cautioning drivers to be alert

Insurance premium discrimination by age so costs are fair