Predicting the Severity of an Accident

SATISH KUMAR BIDWAIK

Problem Statement

In traffic situations, passengers at Seattle are prone to accidents on the roads. This can be due to different factors such as the weather conditions, the road conditions, the light conditions amongst other factors. It is highly recommended to be able to predict the severity of an accident based on the factors available to prepare for the casualty before the accident occurs.

Data Set

The dataset provided for the Seattle city contains a total of 194673 observations and 37 attributes (relating to the accidents that occur on the road) with the labelled data (SEVERITYCODE) which describes the fatality of an incident. Given this dataset, the aim of this project is to select the necessary attributes that will be used to build a model that will help to predict the severity of an accident.

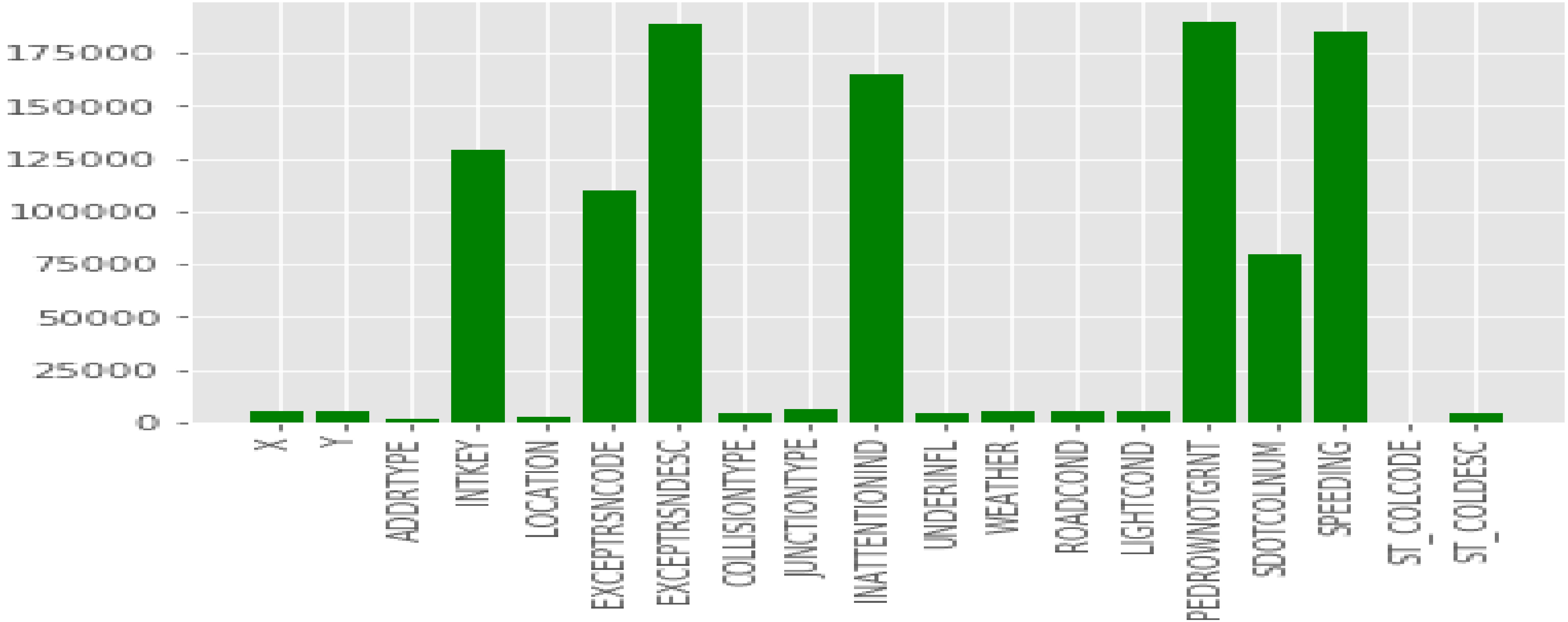
Data Source:

https://s3.us.cloud-object-storage.appdomain.cloud/cf-courses-data/CognitiveClass/DP0701EN/version-2/Data-Collisions.csv

Data Set understanding

- SEVERITYCODE present of all records
- There are columns entries with NaN (Null) values, need conditioning
- Identification of correct features to prepared effective model for accident severity prediction

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 194673 entries, 0 to 194672
Data columns (total 38 columns):
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 0   SEVERITYCODE    194673 non-null  int64
 1   X               189339 non-null  float64
 2   Y               189339 non-null  float64
 3   OBJECTID        194673 non-null  int64
 4   INCKEY          194673 non-null  int64
 5   COLDETKEY       194673 non-null  int64
 6   REPORTNO        194673 non-null  object
 7   STATUS          194673 non-null  object
 8   ADDRTYPE        192747 non-null  object
 9   INTKEY          65070 non-null   float64
 10  LOCATION        191996 non-null  object
 11  EXCEPTRSNCODE   84811 non-null   object
 12  EXCEPTRSNDESC   5638 non-null    object
 13  SEVERITYCODE.1  194673 non-null  int64
 14  SEVERITYDESC    194673 non-null  object
 15  COLLISIONTYPE   189769 non-null  object
 16  PERSONCOUNT     194673 non-null  int64
 17  PEDCOUNT        194673 non-null  int64
 18  PEDCYLCOUNT     194673 non-null  int64
 #   Column          Non-Null Count   Dtype
---  ------          --------------   -----
 19  VEHCOUNT        194673 non-null  int64
 20  INCDATE         194673 non-null  object
 21  INCDTTM         194673 non-null  object
 22  JUNCTIONTYPE    188344 non-null  object
 23  SDOT_COLCODE    194673 non-null  int64
 24  SDOT_COLDESC    194673 non-null  object
 25  INATTENTIONIND  29805 non-null   object
 26  UNDERINFL       189789 non-null  object
 27  WEATHER         189592 non-null  object
 28  ROADCOND        189661 non-null  object
 29  LIGHTCOND       189503 non-null  object
 30  PEDROWNOTGRNT   4667 non-null    object
 31  SDOTCOLNUM      114936 non-null  float64
 32  SPEEDING        9333 non-null    object
 33  ST_COLCODE      194655 non-null  object
 34  ST_COLDESC      189769 non-null  object
 35  SEGLANEKEY      194673 non-null  int64
 36  CROSSWALKKEY    194673 non-null  int64
 37  HITPARKEDCAR    194673 non-null  object
dtypes: float64(4), int64(12), object(22)
memory usage: 56.4+ MB
```



Data Set – Ratio of NaN values

Data Processing and cleaning

- Selected features
  - ADDRTYPE,
  - COLLISIONTYPE
  - PERSONCOUNT
  - VEHCOUNT
  - JUNCTIONTYPE
  - UNDERINFL
  - WEATHER
  - ROADCOND
  - LIGHTCOND
  - SPEEDING
  - HITPARKEDCAR

- The other attributes were dropped either because they do not relate to the target variable or because they have a lot of missing values.

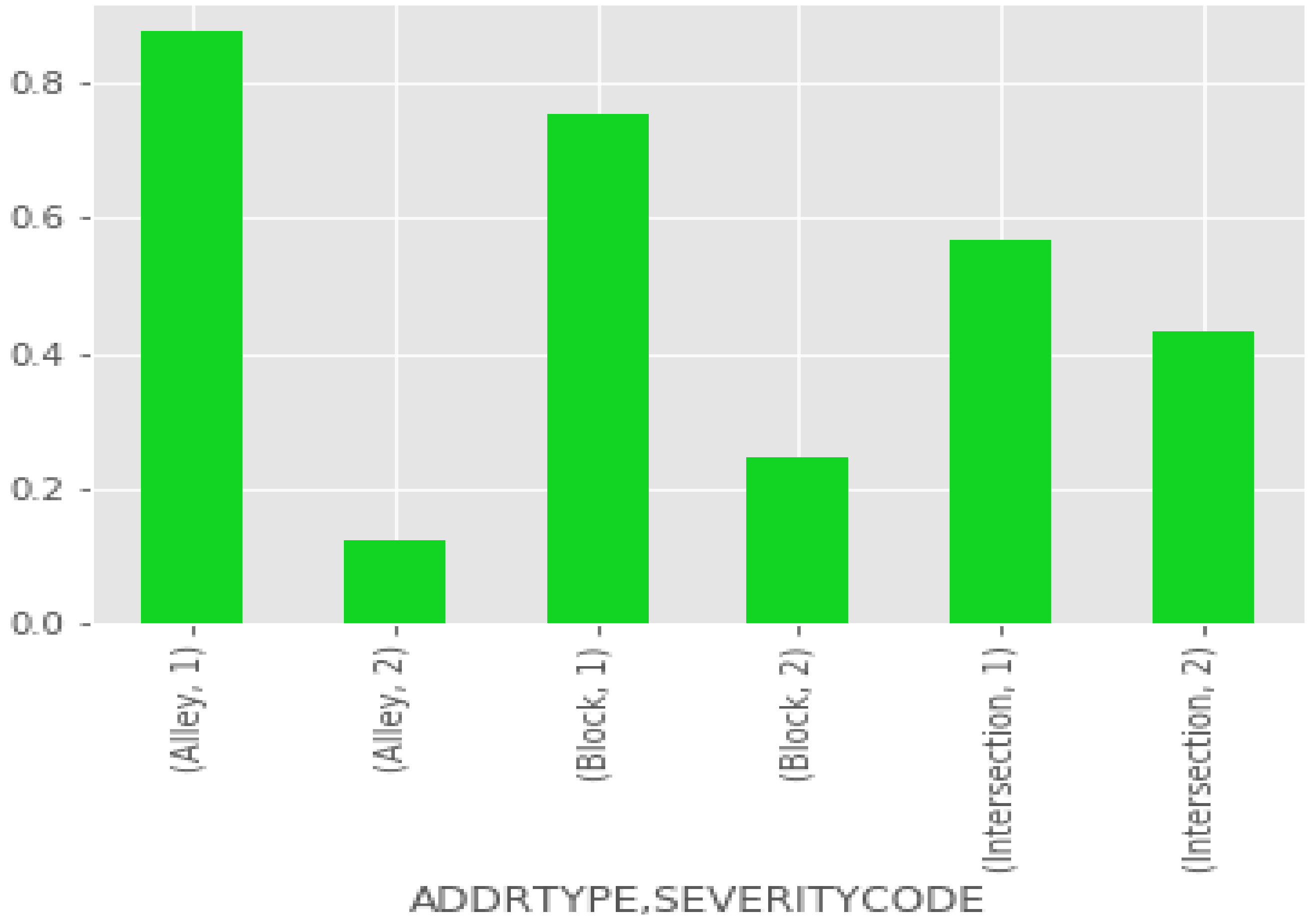
- The following attributes are categorical values and needed to be changed to numerical values
  - ADDRTYPE
  - COLLISIONTYPE
  - JUNCTIONTYPE
  - UNDERINFL
  - WEATHER
  - ROADCOND
  - LIGHTCOND
  - SPEEDING
  - HITPARKEDCAR

Exploratory Analysis

A VIEW ON SELECTED ATTRIBUTE’S RELATIONSHIP WITH SEVERITY CODE

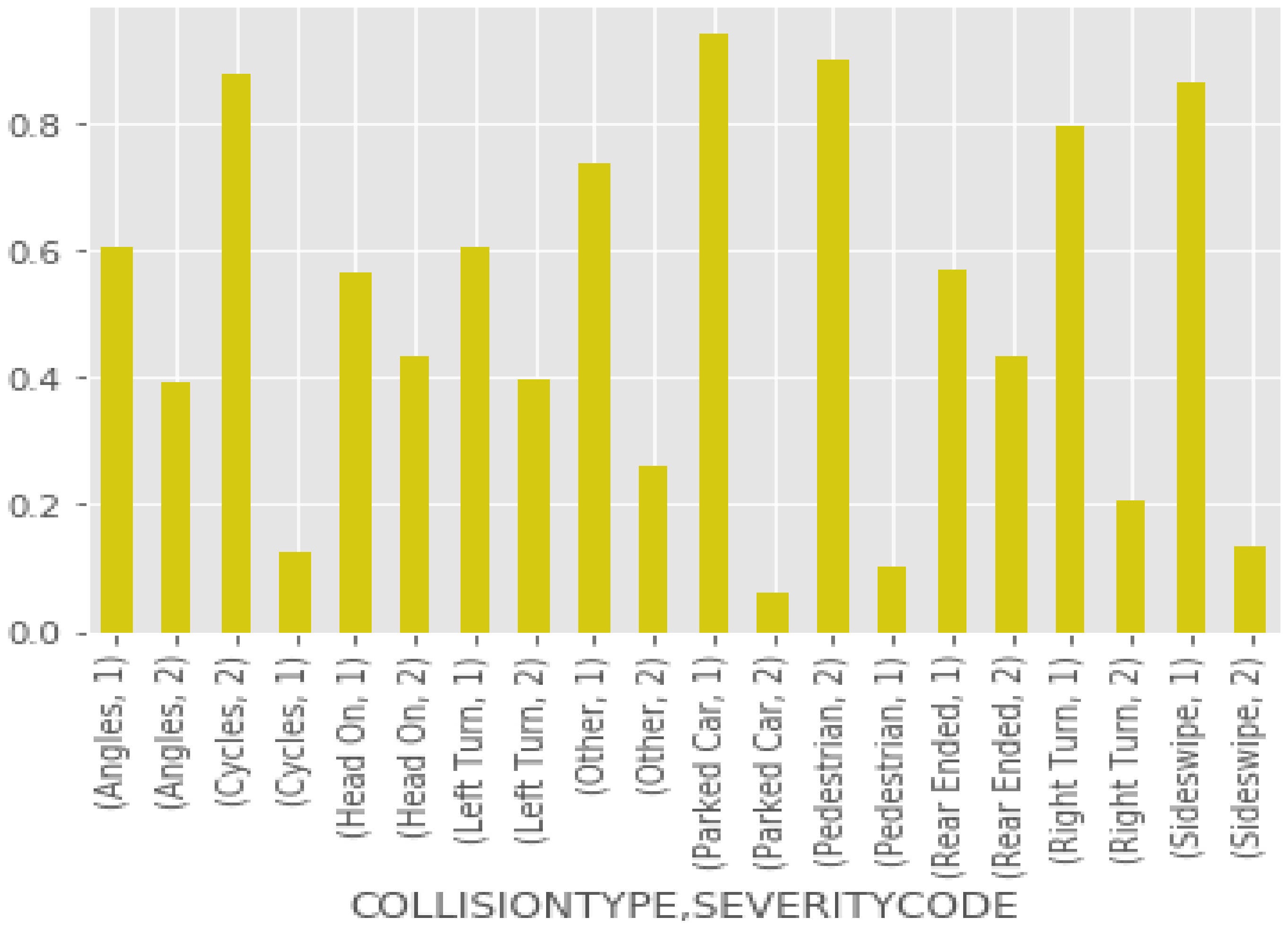
ADDRTYPE - SEVERITYCODE

ADDRTYPE	SEVERITYCODE	
Alley	1	0.876596
	2	0.123404
Block	1	0.754930
	2	0.245070
Intersection	1	0.568012
	2	0.431988



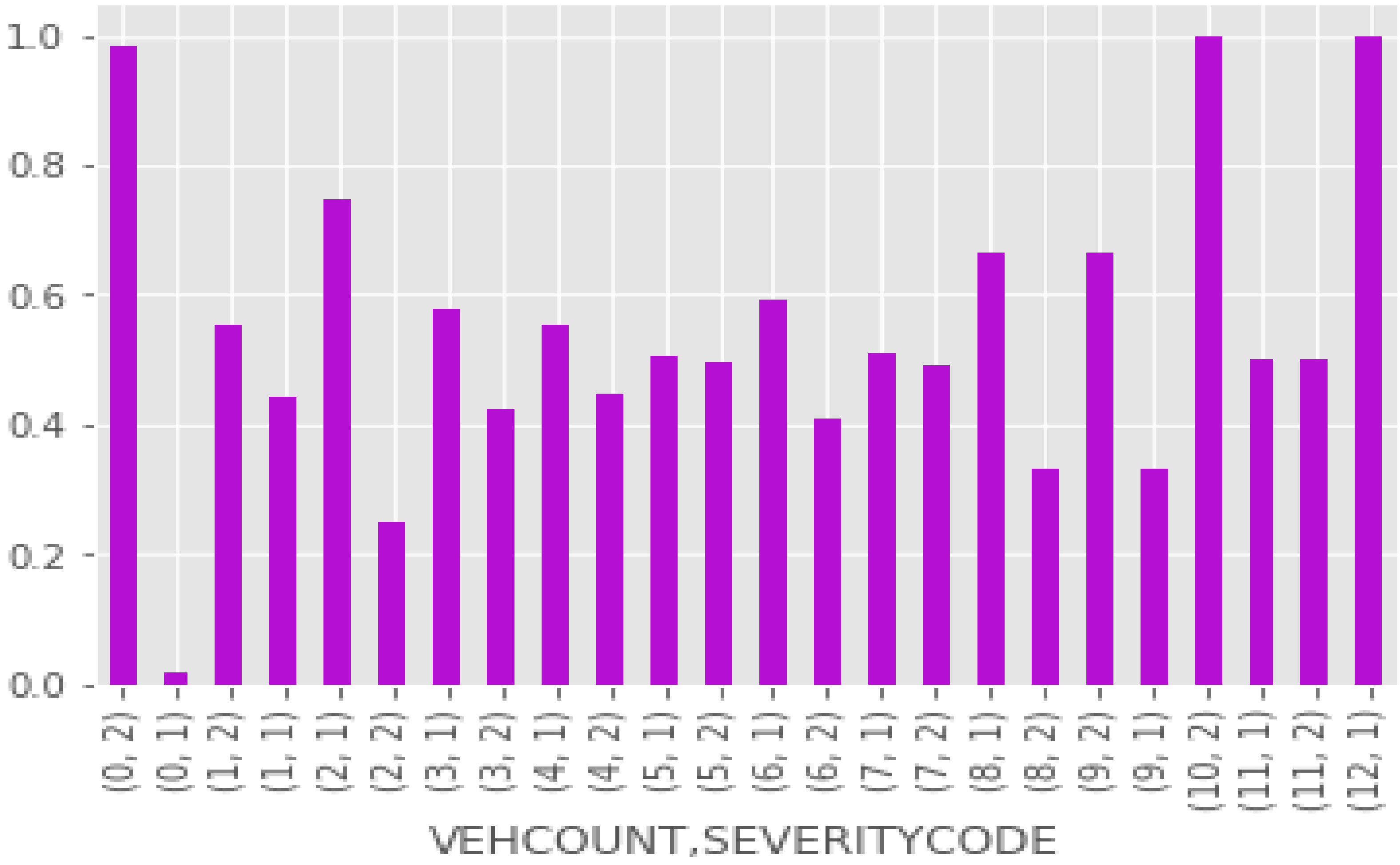
COLLISIONTYPE - SEVERITYCODE

COLLISIONTYPE	SEVERITYCODE	
Angles	1	0.606101
	2	0.393899
Cycles	2	0.877098
	1	0.122902
Head On	1	0.566132
	2	0.433868
Left Turn	1	0.604312
	2	0.395688
Other	1	0.738371
	2	0.261629
Parked Car	1	0.938960
	2	0.061040
Pedestrian	2	0.898511
	1	0.101489
Rear Ended	1	0.568205
	2	0.431795
Right Turn	1	0.793786
	2	0.206214
Sideswipe	1	0.865026
	2	0.134974



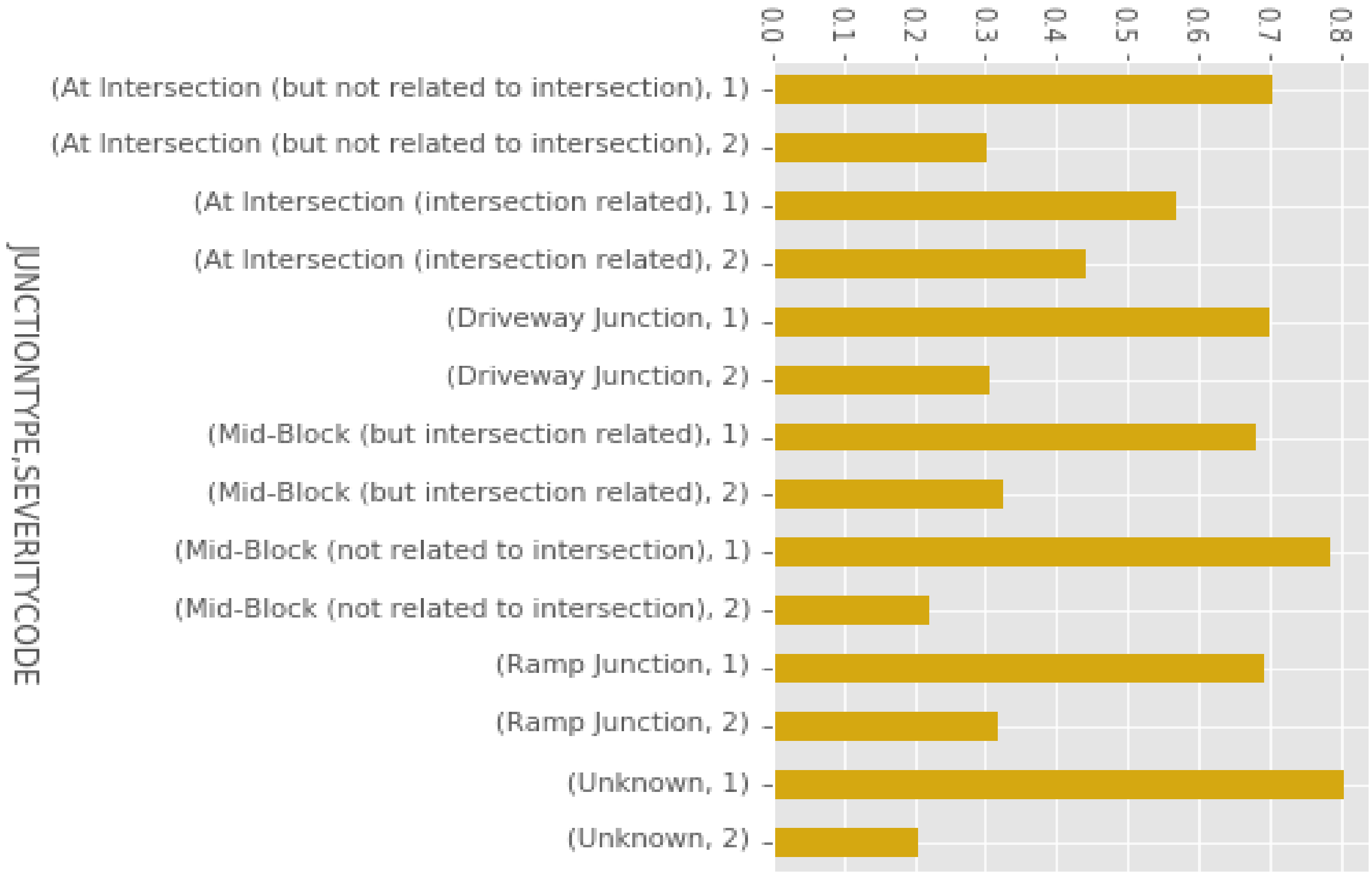
VEHCOUNT - SEVERITYCODE

VEHCOUNT	SEVERITYCODE	
0	2	0.984615
	1	0.015385
1	2	0.555076
	1	0.444924
2	1	0.748156
	2	0.251844
3	1	0.577512
	2	0.422488
4	1	0.554632
	2	0.445368
5	1	0.503802
	2	0.496198
6	1	0.590278
	2	0.409722
7	1	0.511111
	2	0.488889
8	1	0.666667
	2	0.333333
9	2	0.666667
	1	0.333333
10	2	1.000000
11	1	0.500000
	2	0.500000
12	1	1.000000



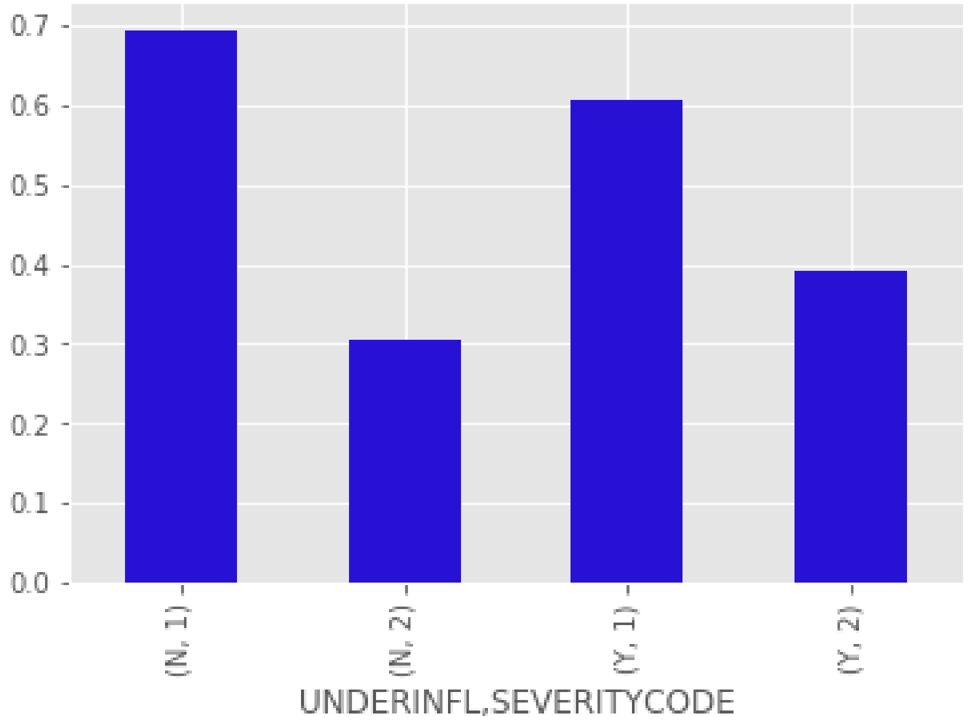
JUNCTIONTYPE - SEVERITYCODE

JUNCTIONTYPE	SEVERITYCODE	
At Intersection (but not related to intersection)	1	0.700243
	2	0.299757
At Intersection (intersection related)	1	0.563474
	2	0.436526
Driveway Junction	1	0.696264
	2	0.303736
Mid-Block (but intersection related)	1	0.678260
	2	0.321740
Mid-Block (not related to intersection)	1	0.782274
	2	0.217726
Ramp Junction	1	0.687500
	2	0.312500
Unknown	1	0.800000
	2	0.200000



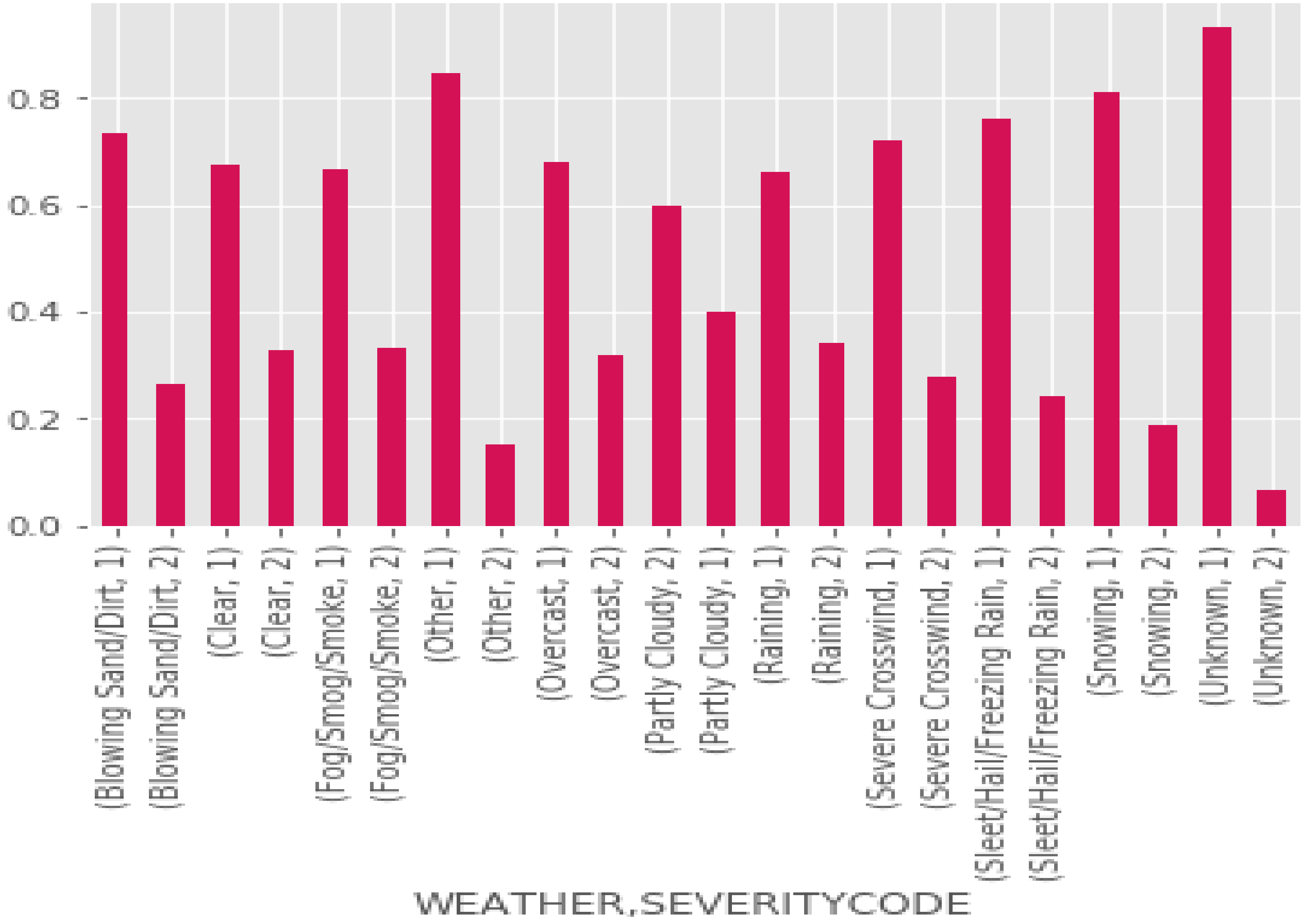
UNDERINFL - SEVERITYCODE

UNDERINFL	SEVERITYCODE	
N	1	0.694666
	2	0.305334
Y	1	0.607869
	2	0.392131



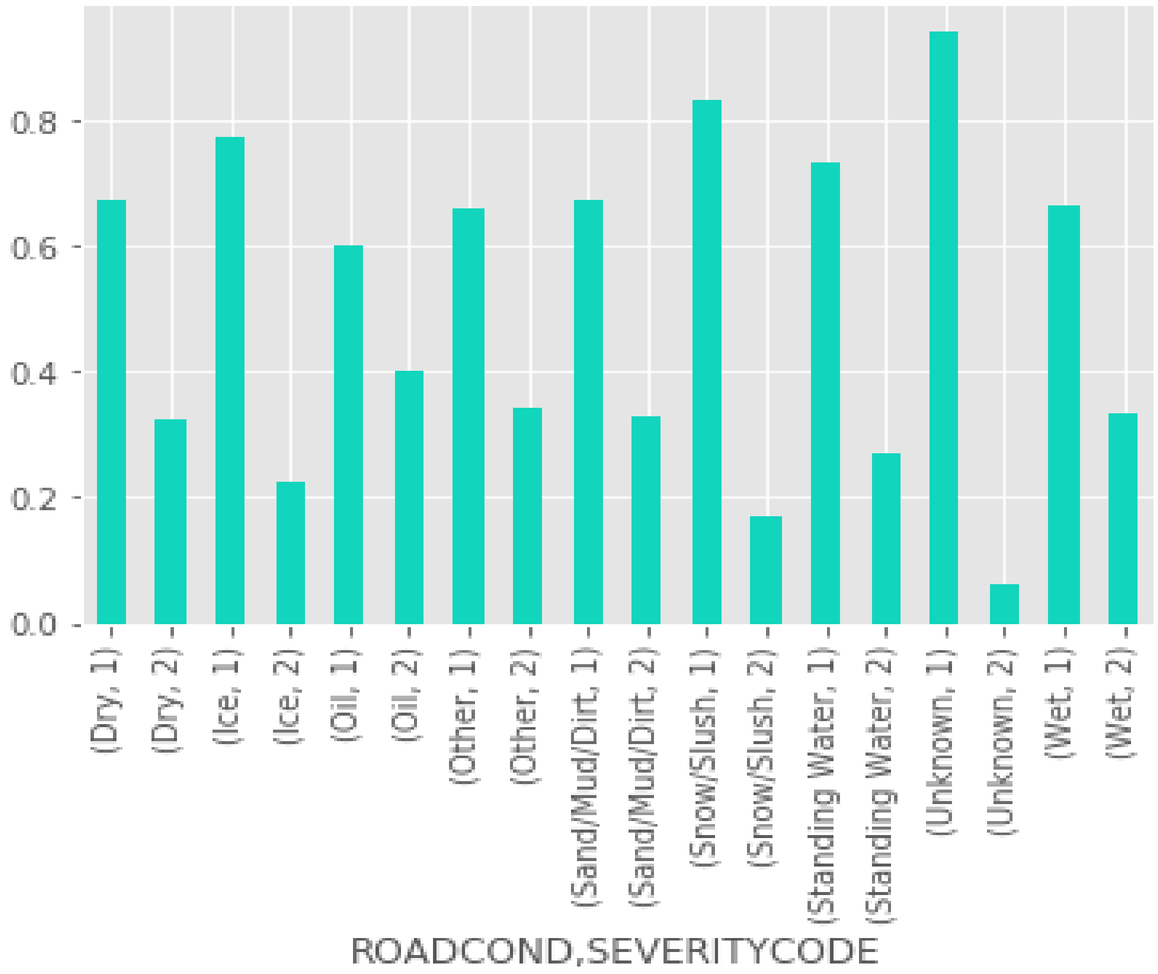
WEATHER - SEVERITYCODE

WEATHER	SEVERITYCODE	
Blowing Sand/Dirt	1	0.734694
	2	0.265306
Clear	1	0.673727
	2	0.326273
Fog/Smog/Smoke	1	0.665468
	2	0.334532
Other	1	0.847185
	2	0.152815
Overcast	1	0.681014
	2	0.318986
Partly Cloudy	2	0.600000
	1	0.400000
Raining	1	0.660468
	2	0.339532
Severe Crosswind	1	0.720000
	2	0.280000
Sleet/Hail/Freezing Rain	1	0.758929
	2	0.241071
Snowing	1	0.810443
	2	0.189557
Unknown	1	0.933746
	2	0.066254



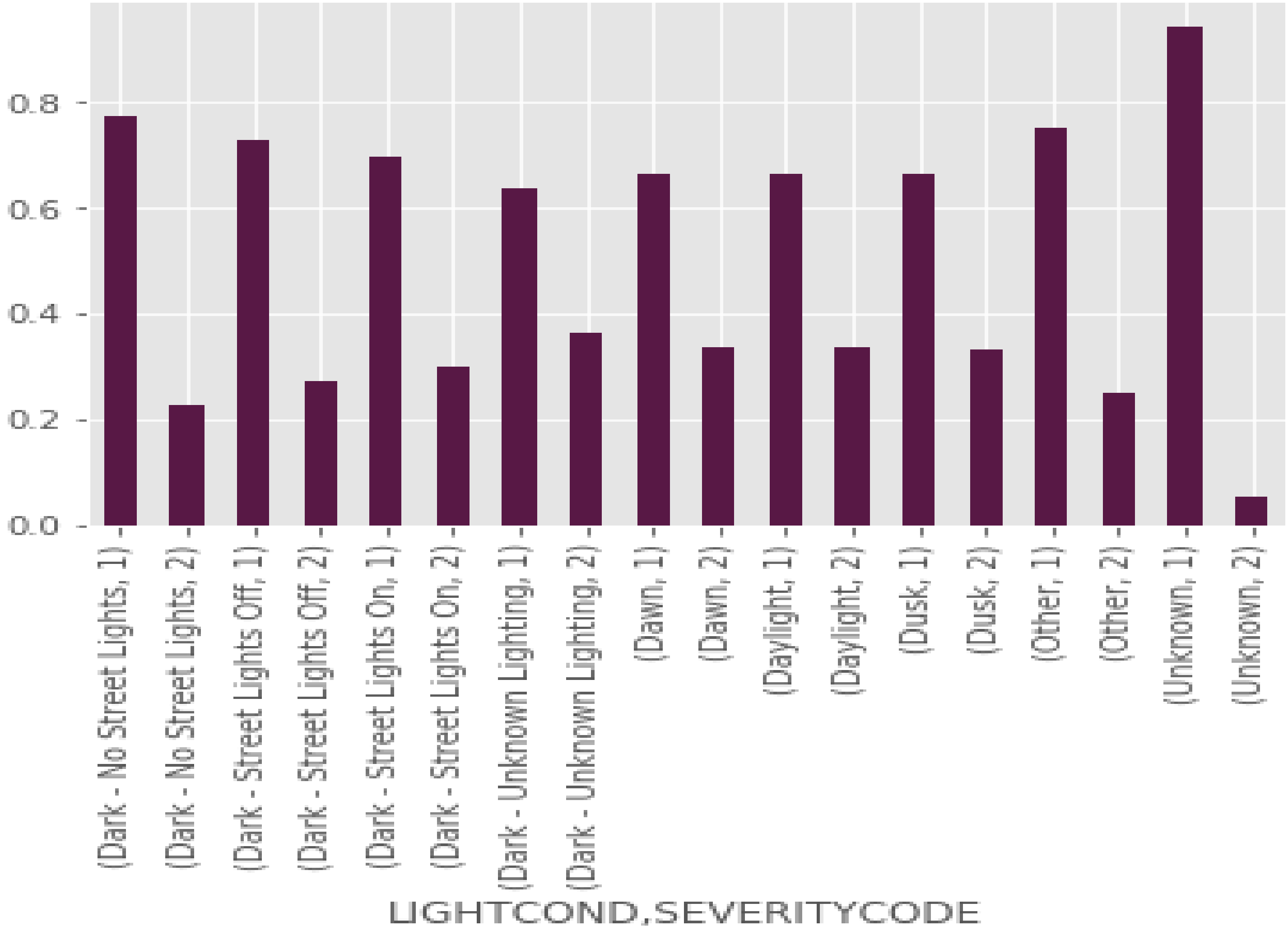
ROADCOND - SEVERITYCODE

ROADCOND	SEVERITYCODE	
Dry	1	0.674678
	2	0.325322
Ice	1	0.773152
	2	0.226848
Oil	1	0.600000
	2	0.400000
Other	1	0.658537
	2	0.341463
Sand/Mud/Dirt	1	0.671642
	2	0.328358
Snow/Slush	1	0.831288
	2	0.168712
Standing Water	1	0.731481
	2	0.268519
Unknown	1	0.938623
	2	0.061377
Wet	1	0.665382
	2	0.334618



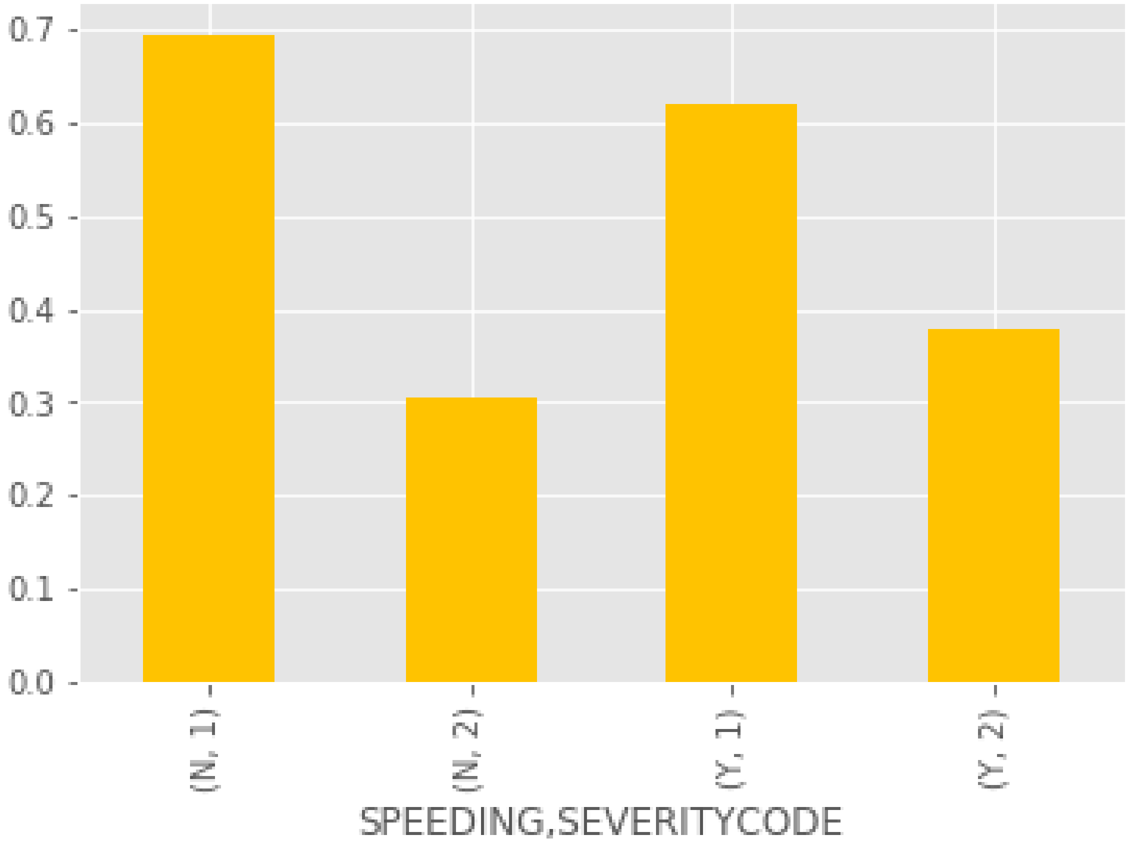
LIGHTCOND - SEVERITYCODE

LIGHTCOND	SEVERITYCODE	
Dark - No Street Lights	1	0.775496
	2	0.224504
Dark - Street Lights Off	1	0.729473
	2	0.270527
Dark - Street Lights On	1	0.698172
	2	0.301828
Dark - Unknown Lighting	1	0.636364
	2	0.363636
Dawn	1	0.666123
	2	0.333877
Daylight	1	0.663941
	2	0.336059
Dusk	1	0.666262
	2	0.333738
Other	1	0.752381
	2	0.247619
Unknown	1	0.944870
	2	0.055130



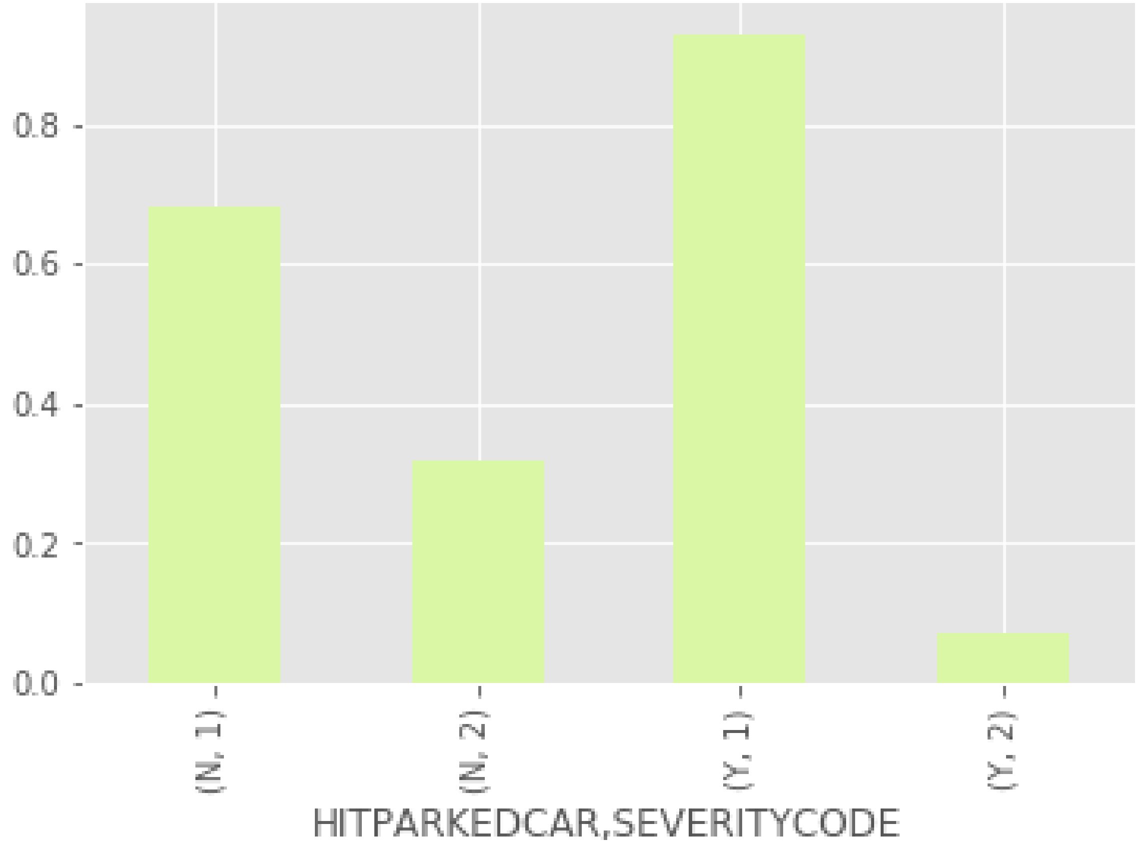
SPEEDING - SEVERITYCODE

SPEEDING	SEVERITYCODE	
N	1	0.694114
	2	0.305886
Y	1	0.620146
	2	0.379854



HITPARKEDCAR - SEVERITYCODE

HITPARKEDCAR	SEVERITYCODE	
N	1	0.682735
	2	0.317265
Y	1	0.928998
	2	0.071002



Predictive Modeling

Used following model/classifiers/algorithms for modeling

- ❑Logistic Regression
- ❑K-nearest neighbors
- ❑Random Forest Classifiers
- ❑Gaussian Naive Bayes Classifier
- ❑Gradient Boosting Classifier

Results

	Algorithm	Accuracy
0	Logistic Regression	0.710462
1	k-nearest neighbors	0.705651
2	Random Forest Classifier	0.738621
3	Gaussian Naive Bayes Classifier	0.698352
4	Gradient Boosting Classifier	0.740780

Conclusion

- In this project, I outlined the attributes as features that tend to affect the severity code of an incident such as address type, weather, light condition, road condition, just to mention a few.
- I developed different classification models to predict the severity code of an incident based on the selected features attributes provided.
- The Gradient Boosting Classifier proved to be the best model in making this prediction.
- This prediction will be helpful for residents as well traffic attendants and paramedics to predict the severity of incidents and plan in terms of providing medical attention and safety guidelines.