PREDICTING TRAFFIC ACCIDENT SEVERITY

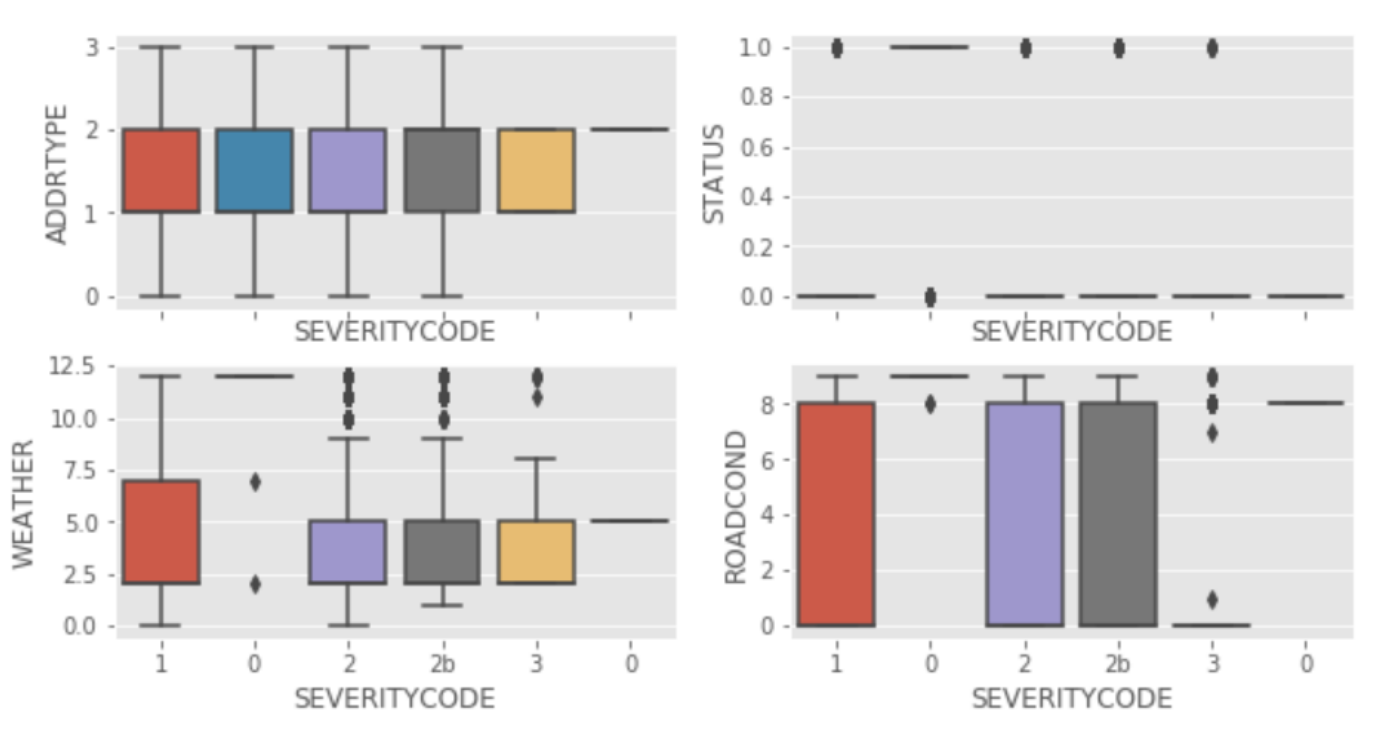
OBJECTIVE

The objective of the project is to use a dataset with conditions and the severity of the reported occurrence of car accidents in a city to predict the severity label which describes the fatality of an accident given the conditions. The conditions such as weather, light, speeding, inattention and user information is used for training and forecasting the severity and occurrence of accident.

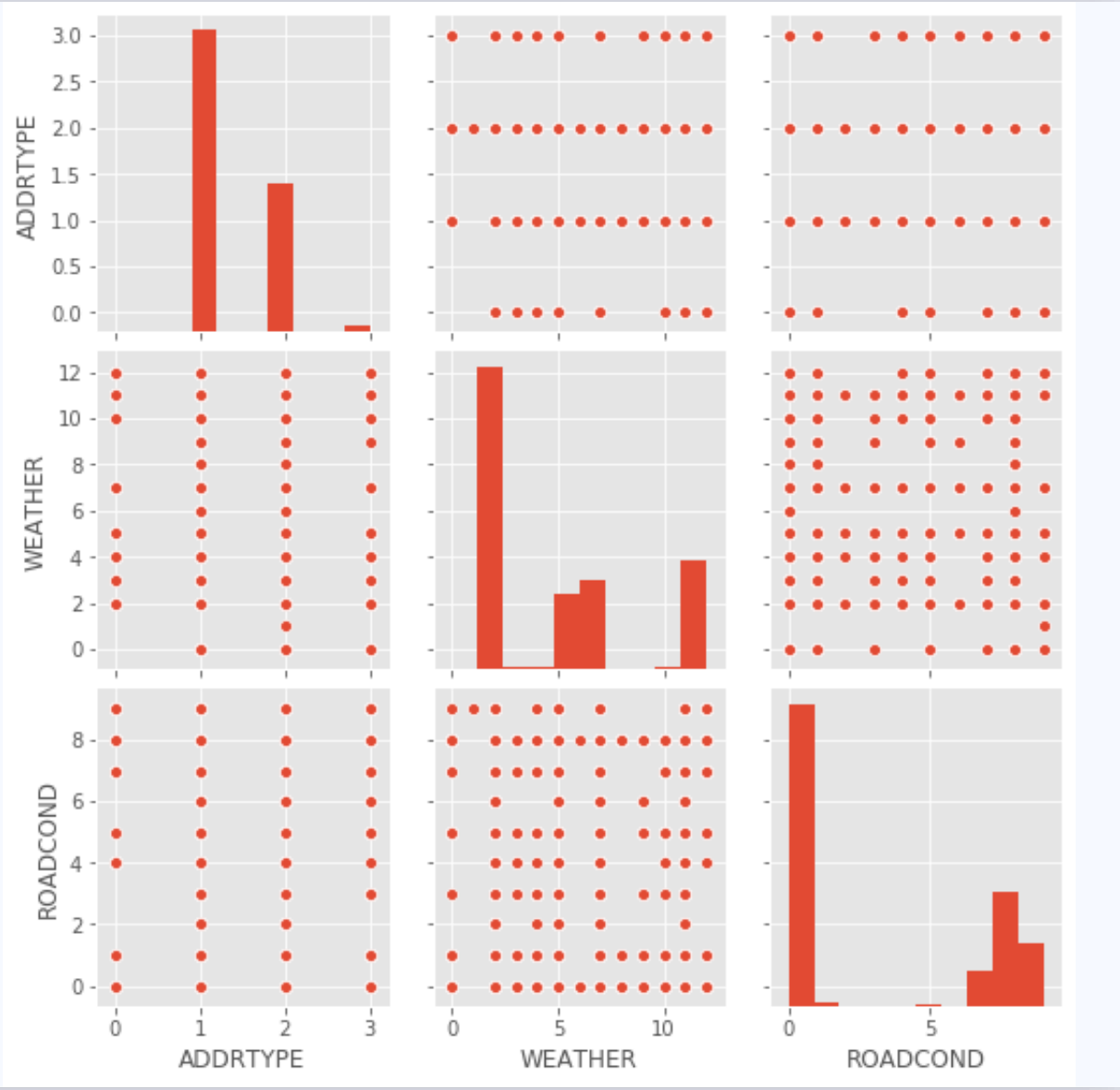
DATA

	X	Y	OBJECTID	INCKEY	COLDETKEY	REPORTNO	STATUS	ADDRTYPE	INTKEY	LOCATION	EXCEPTRSNCODE	EXCEPTRSND
0	-122.320757	47.609408	1	328476	329976	EA08706	Matched	Block	NaN	BROADWAY BETWEEN E COLUMBIA ST AND BOYLSTON AVE		
1	-122.319561	47.662221	2	328142	329642	EA06882	Matched	Block	NaN	8TH AVE NE BETWEEN NE 45TH E ST AND NE 47TH ST		
2	-122.327525	47.604393	3	20700	20700	1181833	Unmatched	Block	NaN	JAMES ST BETWEEN 6TH AVE AND 7TH AVE	NaN	

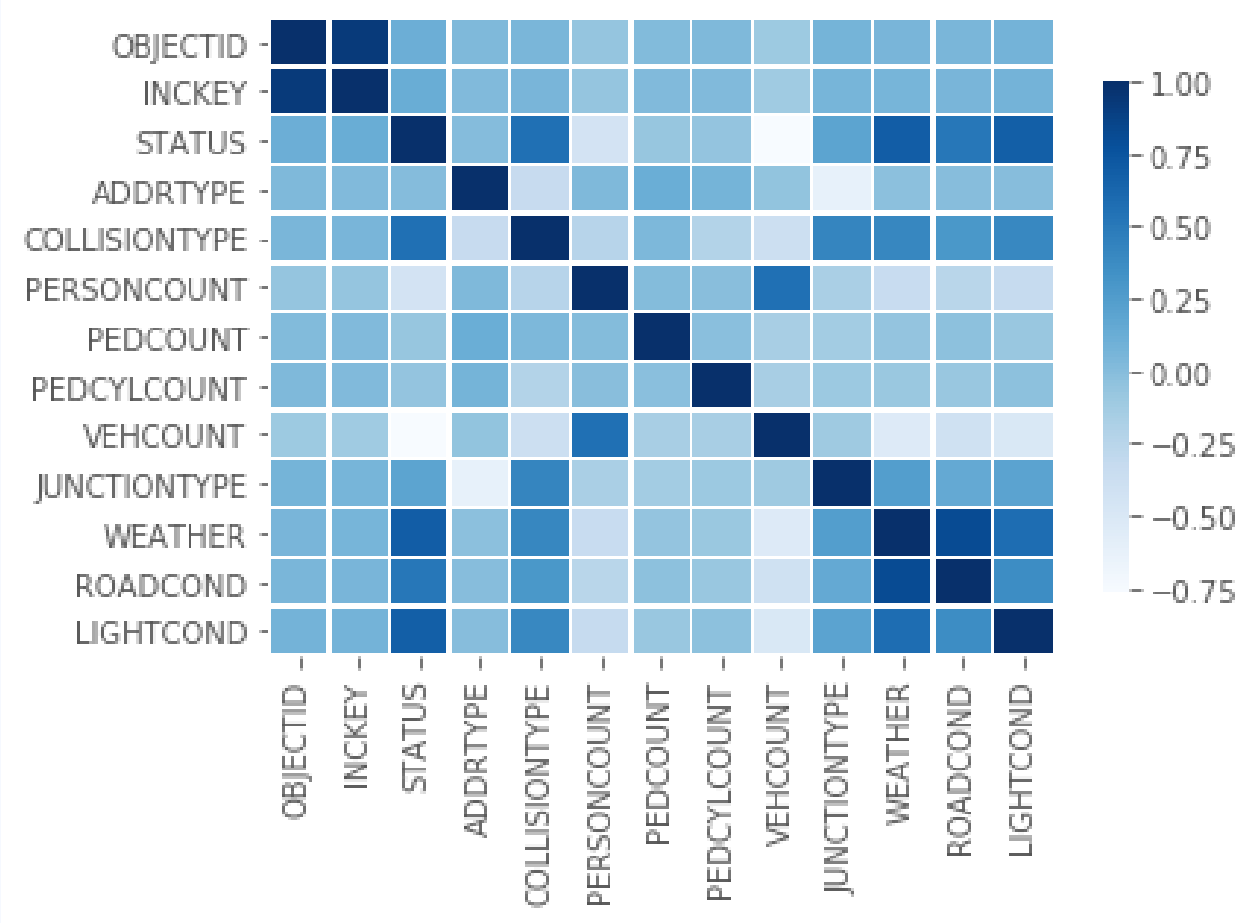
EXPLORATORY ANALYSIS
BOX PLOT



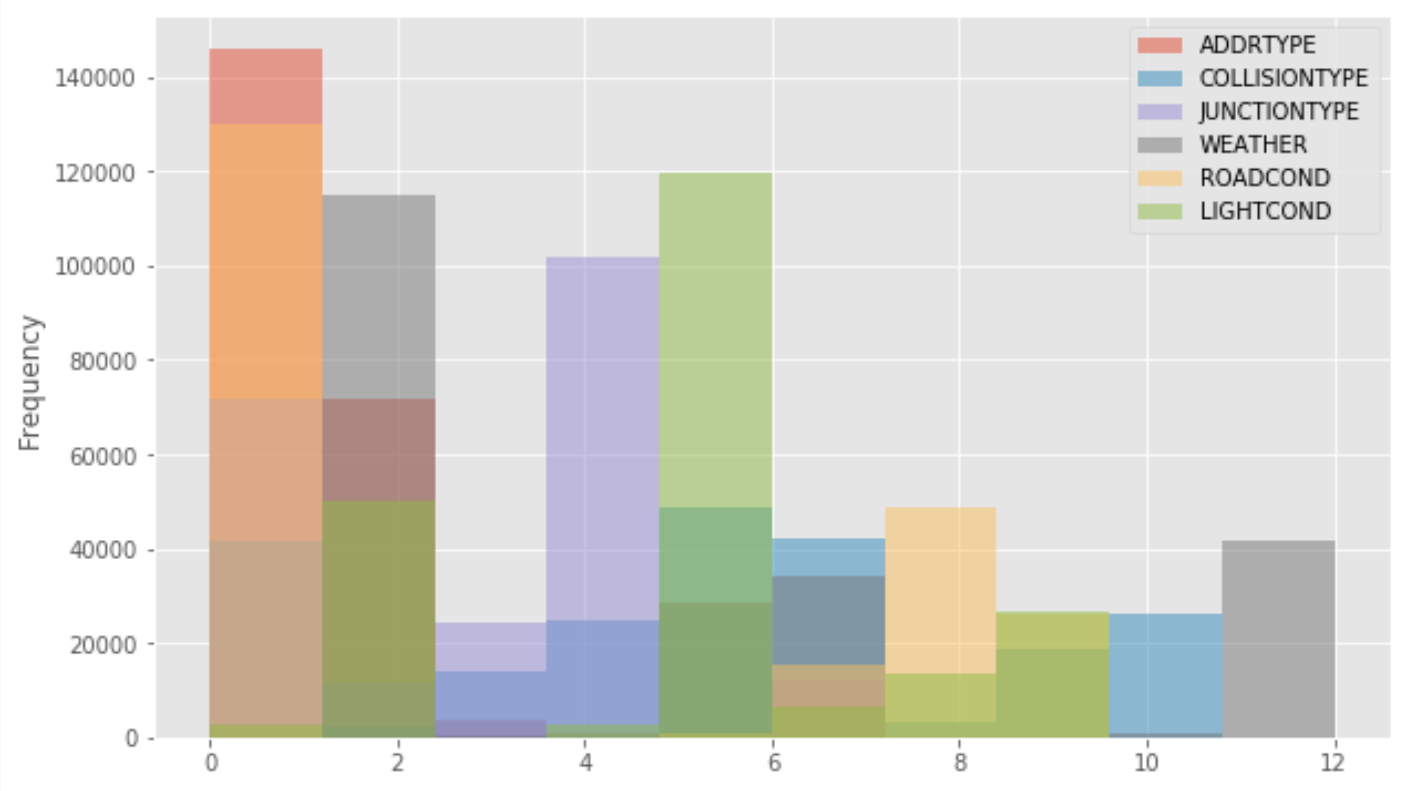
PAIR PLOT



COVARIANCE MATRIX



HISTOGRAM


ADDRTYPE
COLLISIONTYPE
JUNCTIONTYPE
WEATHER
ROADCOND
LIGHTCOND
Frequency
140000
120000
100000
80000
60000
40000
20000
0
0
2
4
6
8
10
12

CLASSIFICATION ALGORITHM SCORES

Classifier	Score	F1 Score
KNN	0.68	0.81
Decision Tree	0.75	0.84
Logistic Regression	0.69	0.82
Naïve Bayes	0.70	0.82
Random Forrest	0.71	0.80

CONCLUSION

the best classifier of this problem is Decision Tree. It's gotten best score and more true positive values.