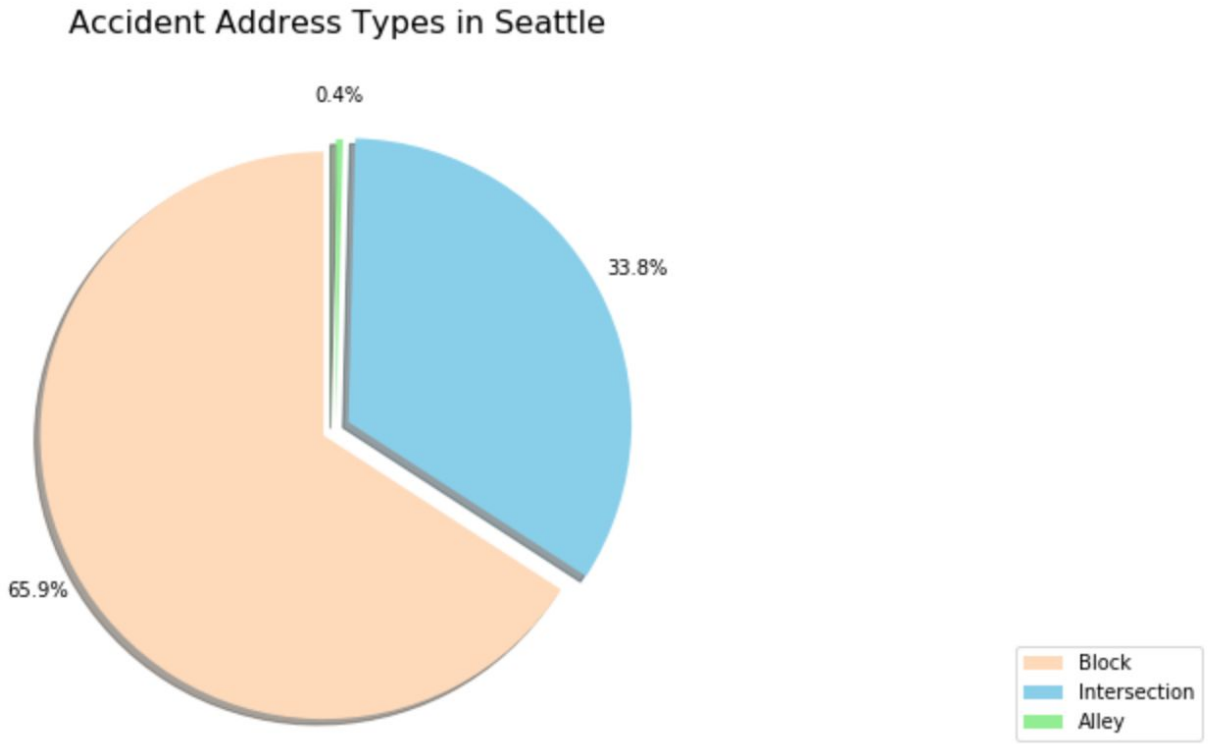
Capstone Project for Data Science - Car Accident Severity in Seattle

By: Camelia Melissa Mallek

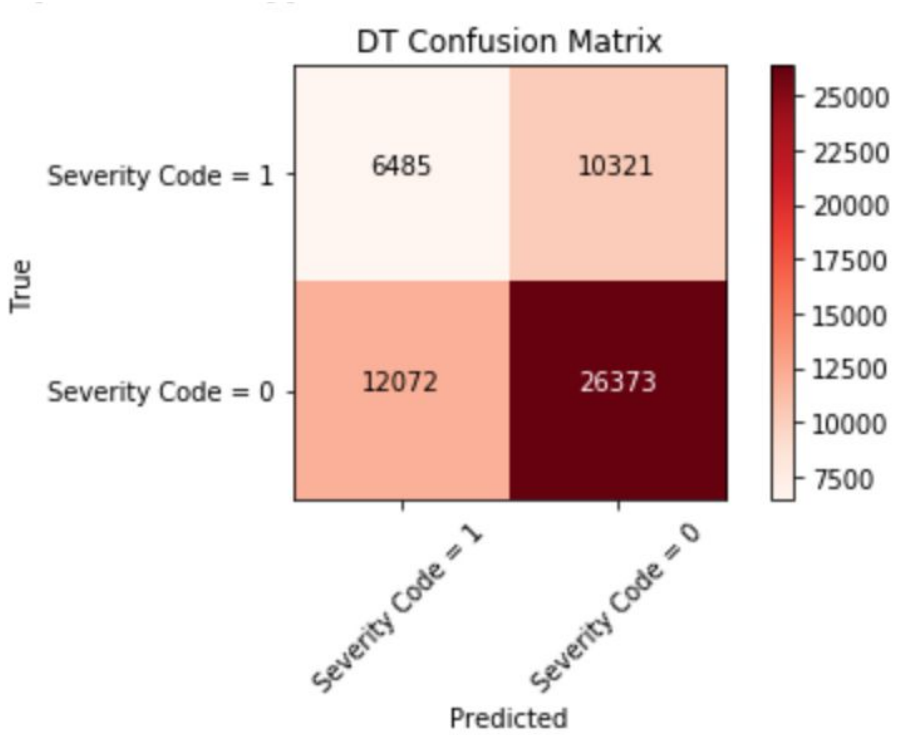
Where do the majority of accidents occur?



Accuracy Results for Decision Tree Model

	precision	recall	f1-score	support
0	0.69	0.72	0.70	36694
1	0.39	0.35	0.37	18557
accuracy			0.59	55251
macro avg	0.54	0.53	0.53	55251
weighted avg	0.59	0.59	0.59	55251

Decision Tree Model - Confusion Matrix

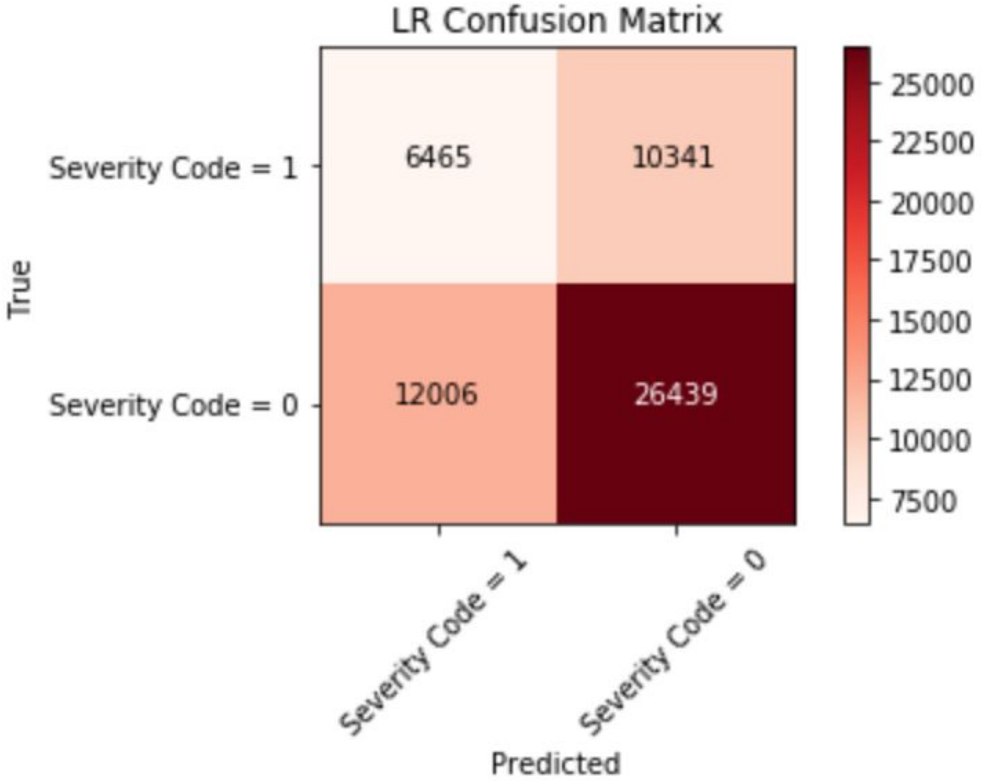


Accuracy Results for Logistic Regression Model

Accuracy score of logistic regression model: 0.595536732366835

	precision	recall	f1-score	support
0	0.72	0.69	0.70	38445
1	0.35	0.38	0.37	16806
accuracy			0.60	55251
macro avg	0.53	0.54	0.53	55251
weighted avg	0.61	0.60	0.60	55251

Logistic Regression Model - Confusion Matrix

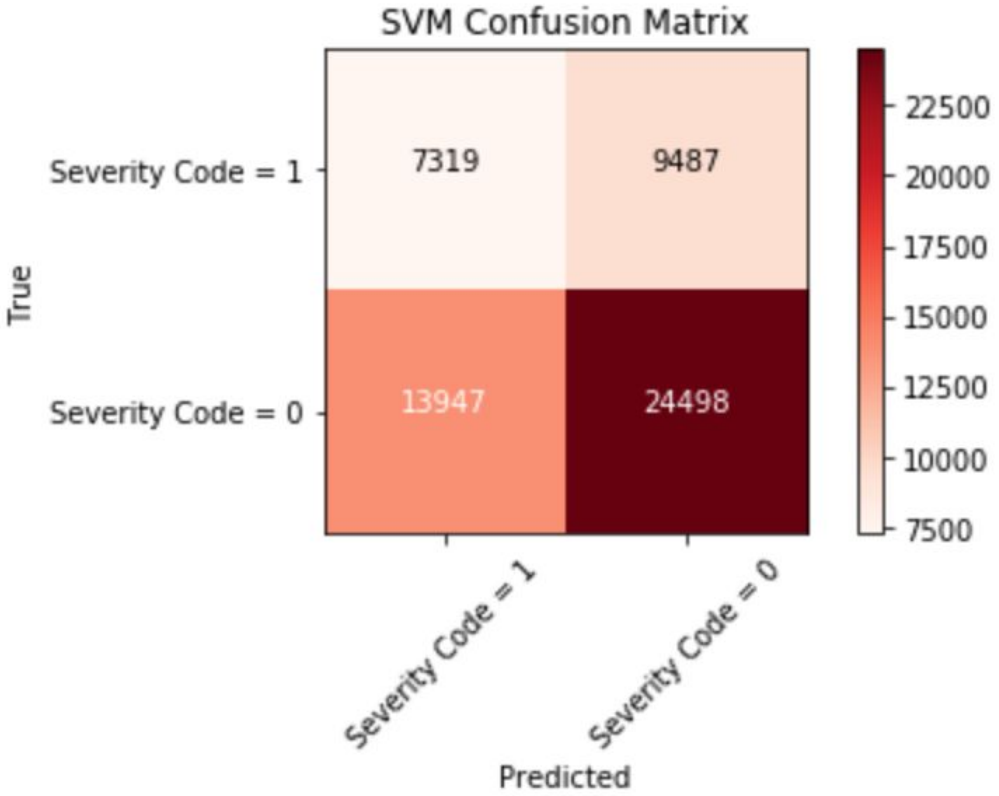


Accuracy Results for Support Vector Machine Model

Accuracy of SVM model: 0.5758628803098587

	precision	recall	f1-score	support
0	0.72	0.64	0.68	38445
1	0.34	0.44	0.38	16806
accuracy			0.58	55251
macro avg	0.53	0.54	0.53	55251
weighted avg	0.61	0.58	0.59	55251

Support Vector Machine Model - Confusion Matrix



Conclusion

- The main takeaways from this project is that all of the machine learning models that were conducted had similar results and value can be derived from each of them
- The bulk of accidents occur in a particular address type (block) and this should be further considered as an indicator for which areas to avoid during certain driving conditions
- Future analysis should include different machine learning methods in order to further determine if there is an even more effective model type for this kind of data analysis