IBM Data Science Capstone Project

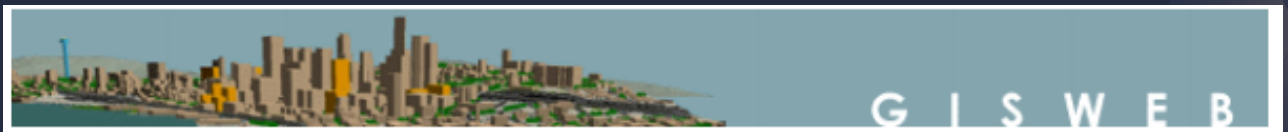
Predict the Accident severity for vehicle GPS

Introduction

- Increasing demand for GPS navigation
- GPS to predict accident severity and find safest route
- Use machine learning to build prediction model

Data

- Collision data in Seattle
- Target: Severity code
- Independent: Location, road condition, weather, junction type, etc.



ArcGIS Metadata Form

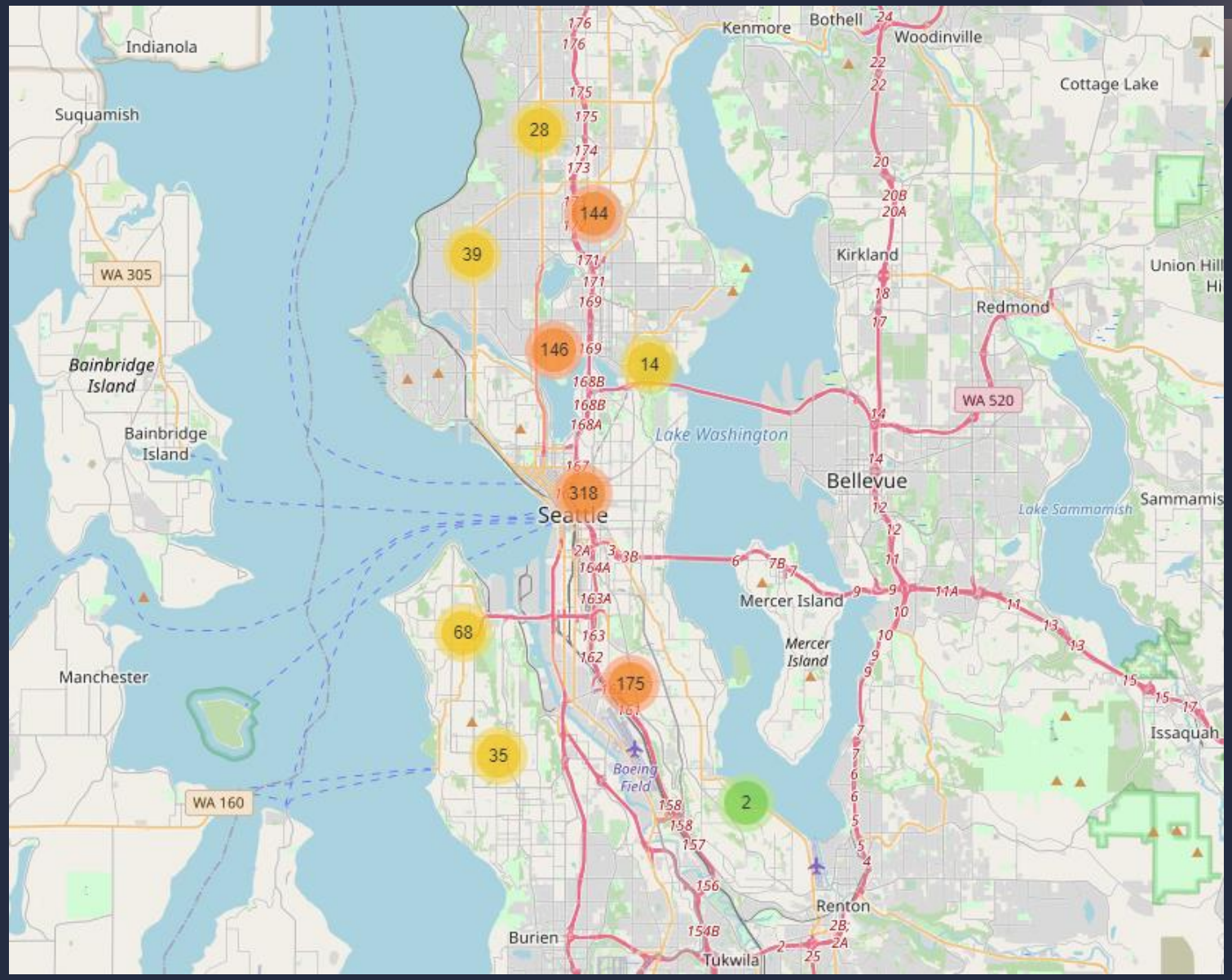
Collisions—All Years

Data Set Summary

Data Set Basics	
Title	Collisions—All Years
Abstract	All collisions provided by SPD and recorded by Traffic Records.
Description	This includes all types of collisions. Collisions will display at the intersection or mid-block of a segment. Timeframe: 2004 to Present.
Supplemental Information	
Update Frequency	Weekly
Keyword(s)	SDOT, Seattle, Transportation, Accidents, Bicycle, Car, Collisions, Pedestrian, Traffic, Vehicle
Contact Information	
Contact Organization	SDOT Traffic Management Division, Traffic Records Group
Contact Person	SDOT GIS Analyst
Contact Email	DOT_IT_GIS@seattle.gov

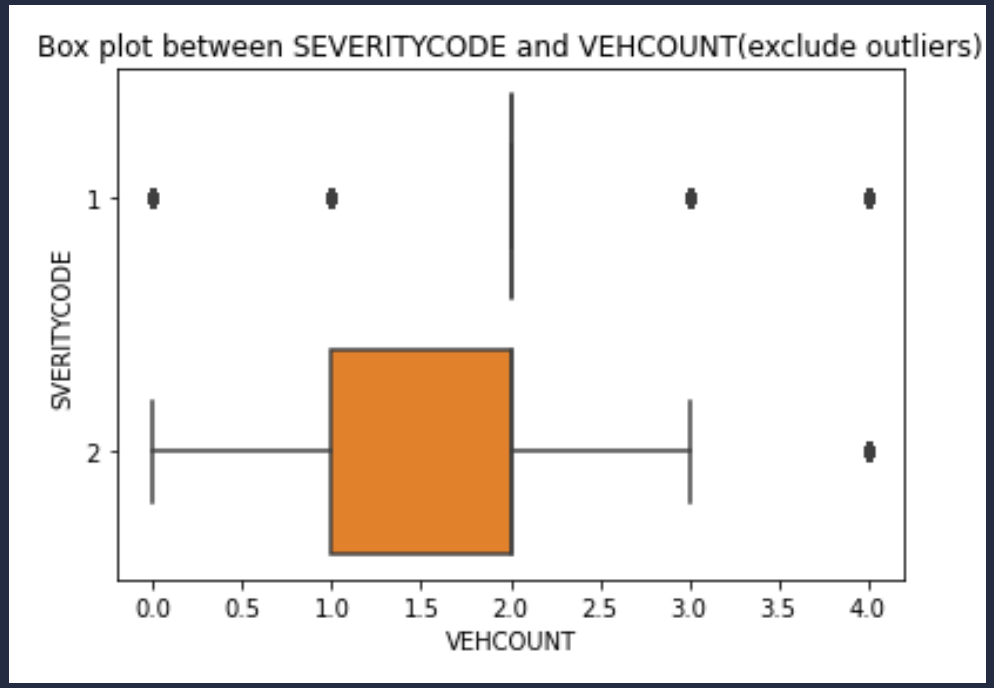
Data Seattle collisions

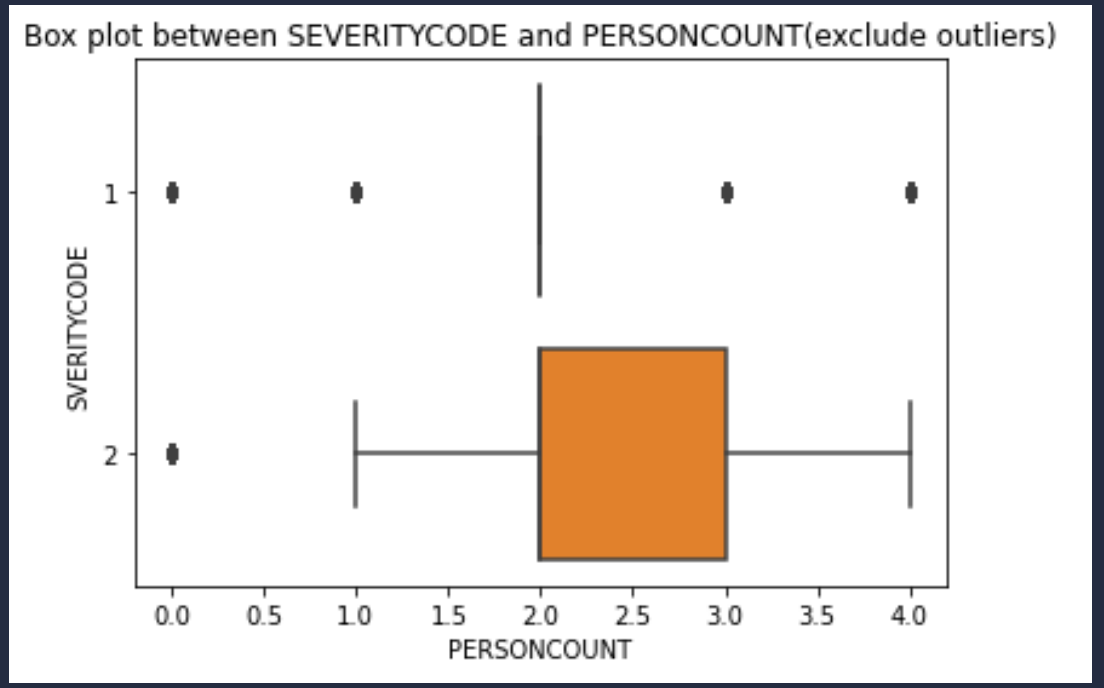
- Frequent collisions along the highway



Methodology
Dataset preparation

	SEVERITYCODE	PERSONCOUNT	VEHCOUNT	ROADCOND	WEATHER	JUNCTIONTYPE	SPEEDING	LIGHTCOND
0	2	2	2	Wet	Overcast	At Intersection (intersection related)	NaN	Daylight
1	1	2	2	Wet	Raining	Mid-Block (not related to intersection)	NaN	Dark - Street Lights On
2	1	4	3	Dry	Overcast	Mid-Block (not related to intersection)	NaN	Daylight
3	1	3	3	Dry	Clear	Mid-Block (not related to intersection)	NaN	Daylight
4	2	2	2	Wet	Raining	At Intersection (intersection related)	NaN	Daylight



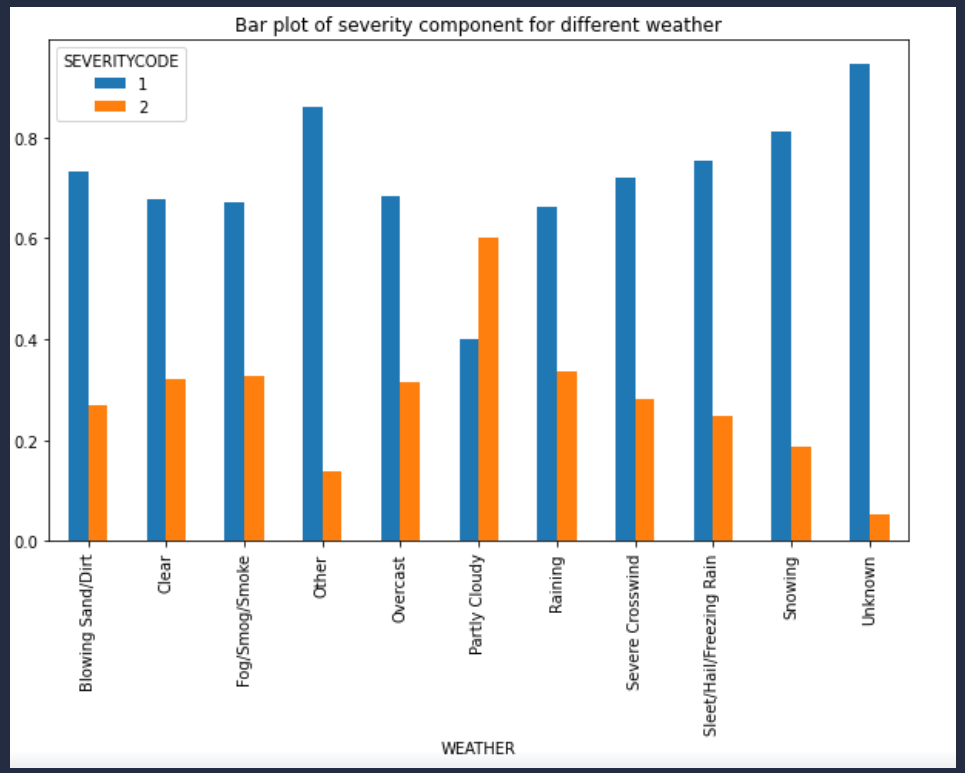


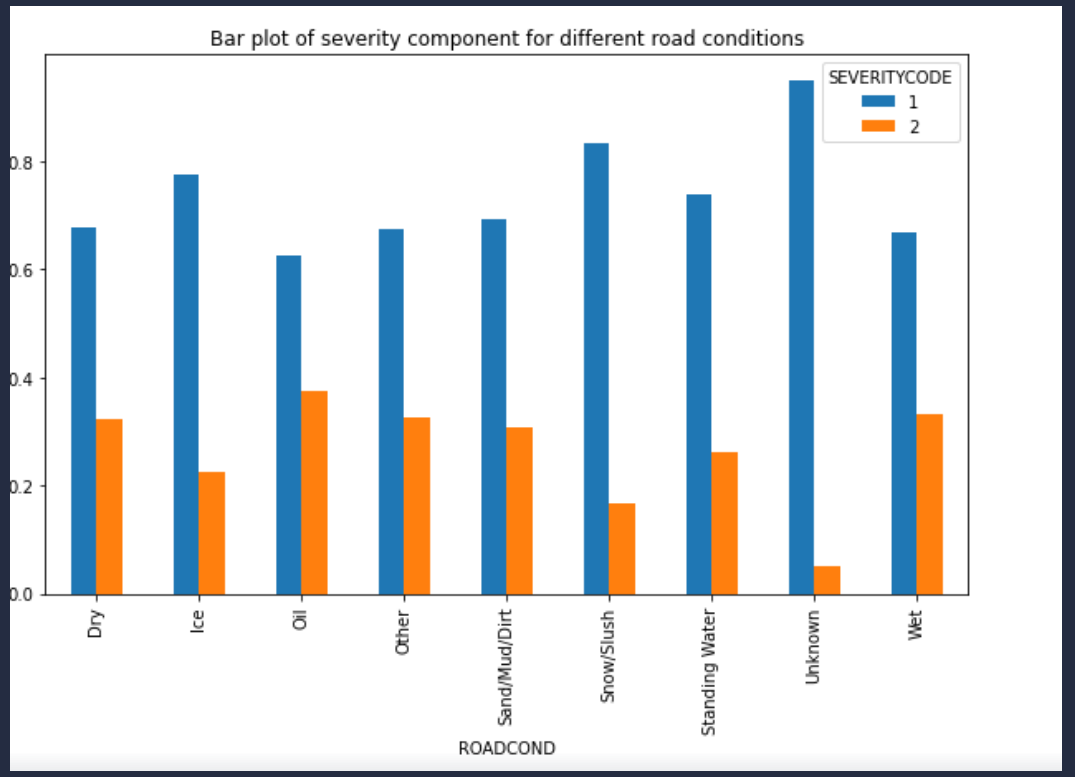
Methodology
Data analysis

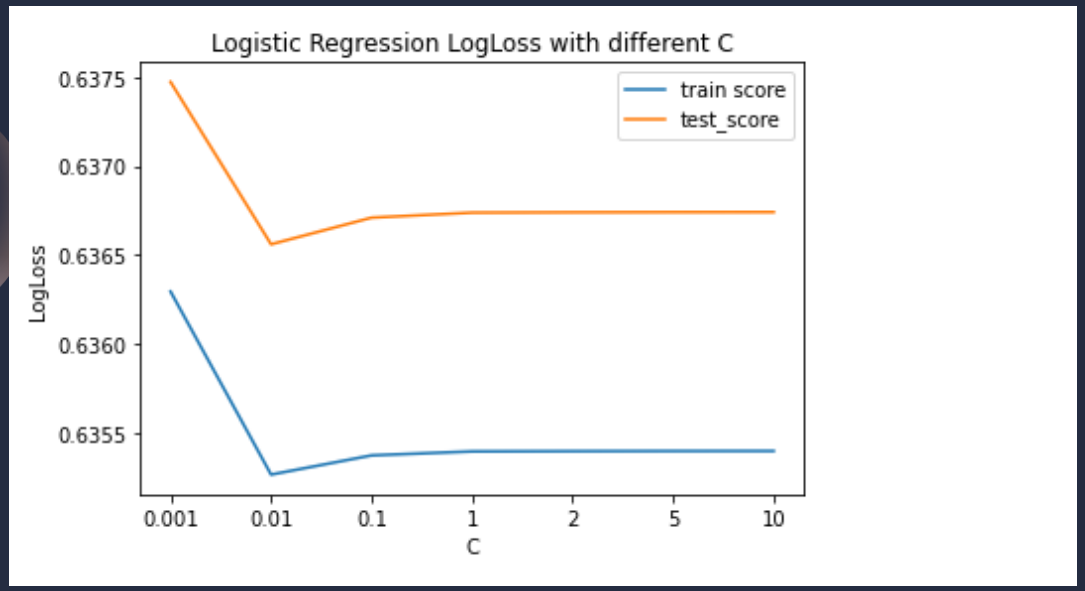
- Analysis numerical independent variables: vehicle count, person count

Methodology
Data analysis

- Deal with categorical variables

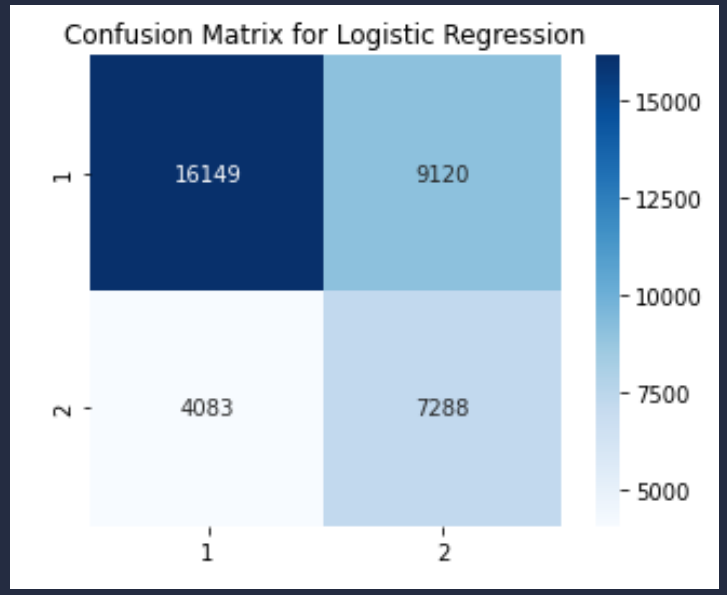


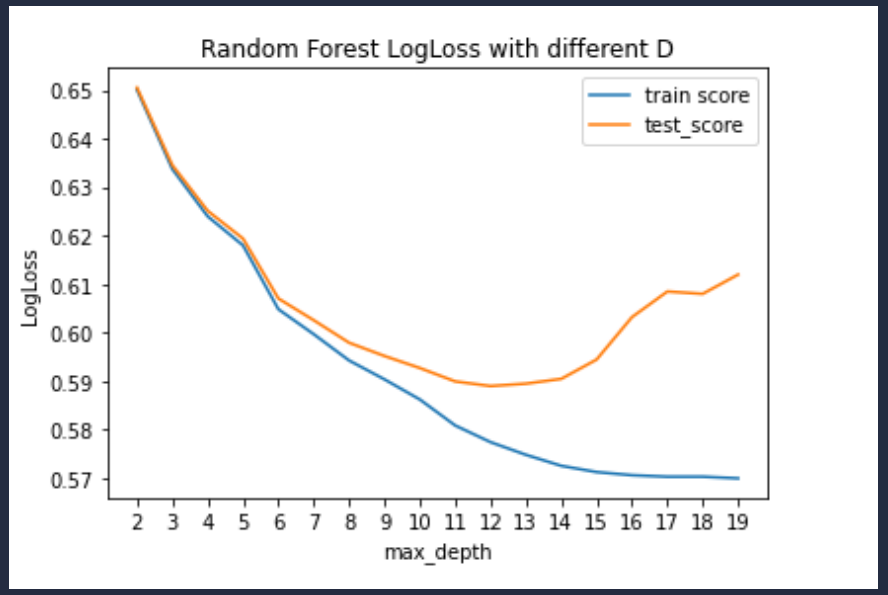


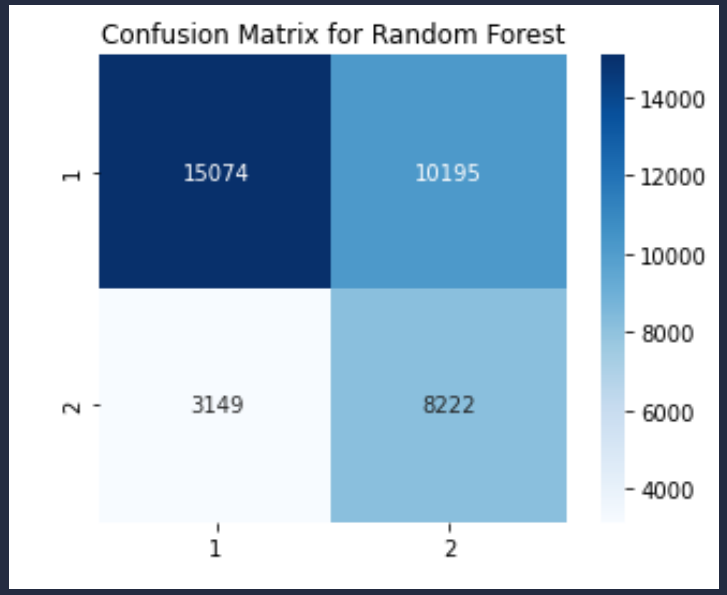


Model and Evaluation

Logistic Regression







Model and Evaluation

Random Forest

Conclusion

- Developed machine learning models to predict the accident severity and the corresponding possibility
- More potentials to be explored further
- Consider weights on type II errors