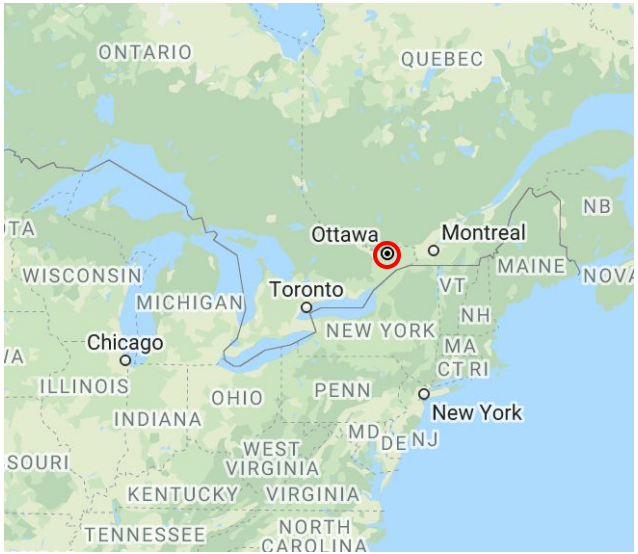
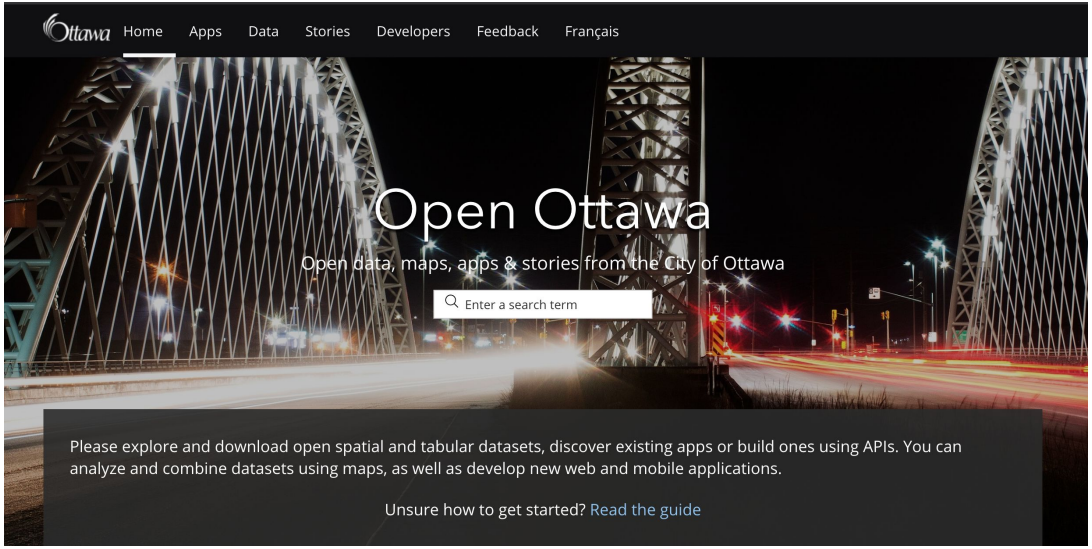
Predicting Car Collision Outcomes Based on Situational Classification Data

Julia Tsaltas
Feb 22, 2021
GA Data Science Capstone

The Situation





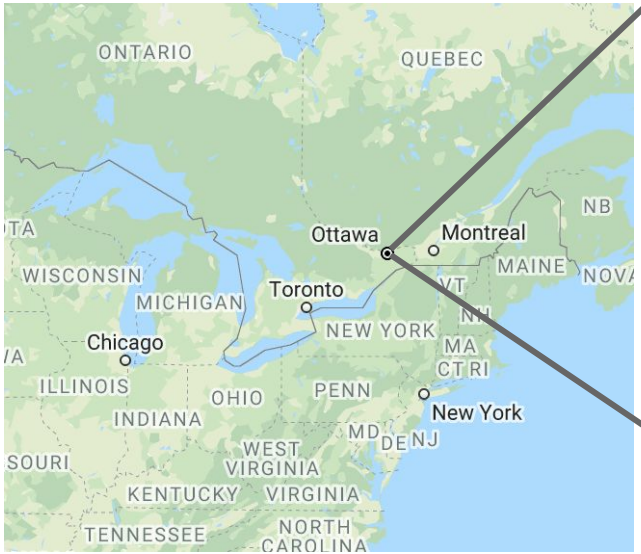
Infrastructure

Parks & Recreation

Transportation

Image Sources: Google Maps 2021, https://open.ottawa.ca/

The Situation



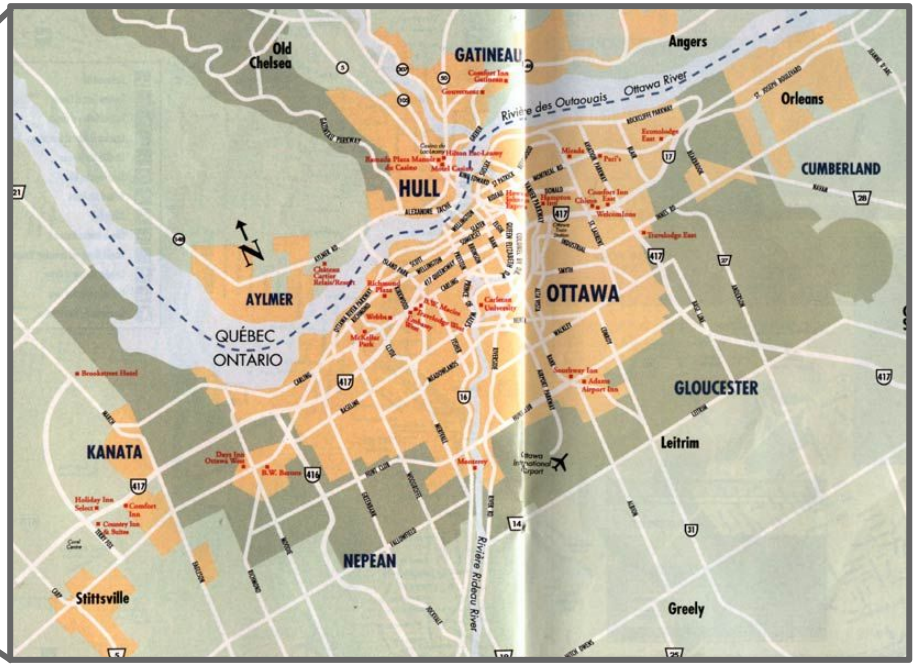
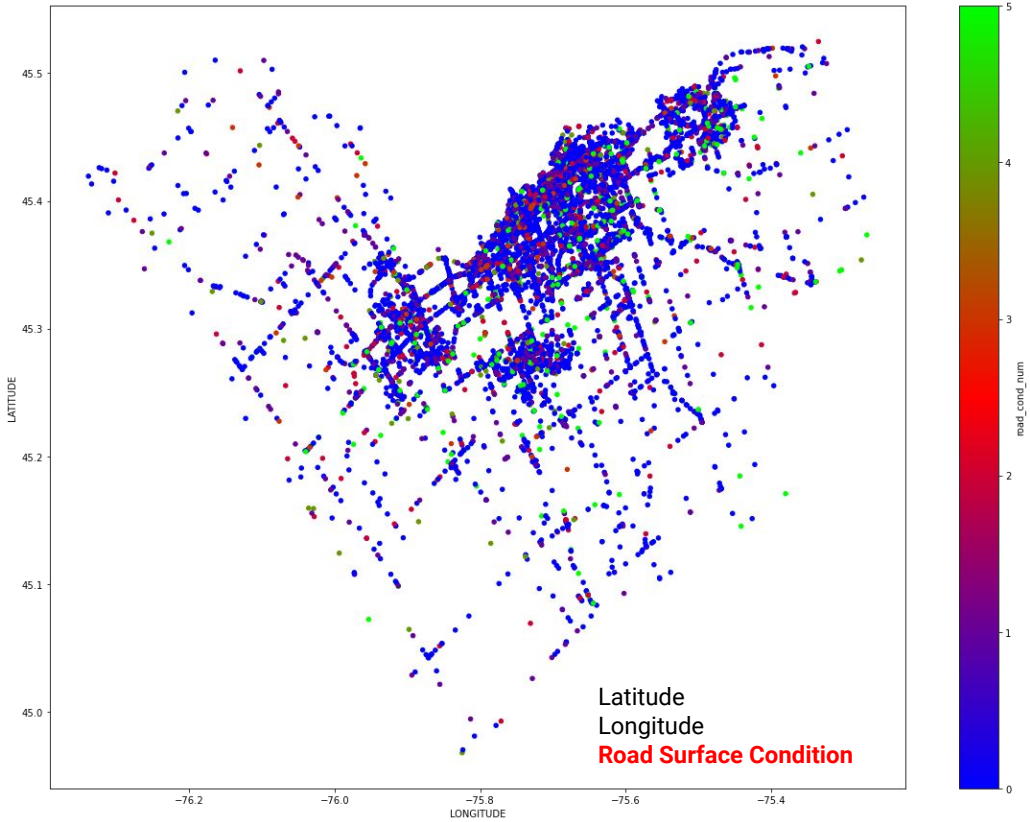


Image Source: https://www.4data.ca/ottawa/largerMaps/regionalMap.jpg

The Situation

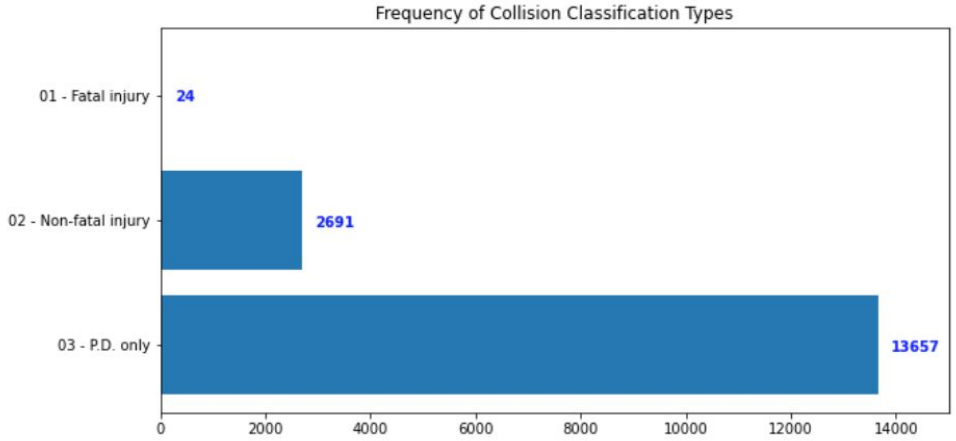
Hue	Road Condition	Count
Bright Green	Ice	719
Dark Green	Packed Snow	580
Orange	Slush	557
Red	Loose Snow	997
Purple	Wet	2788
Blue	Dry	10711



API Source: https://open.ottawa.ca/datasets/2019-tabular-transportation-collision-data

The Outcomes

CLASSIFICATION OF ACCIDENT

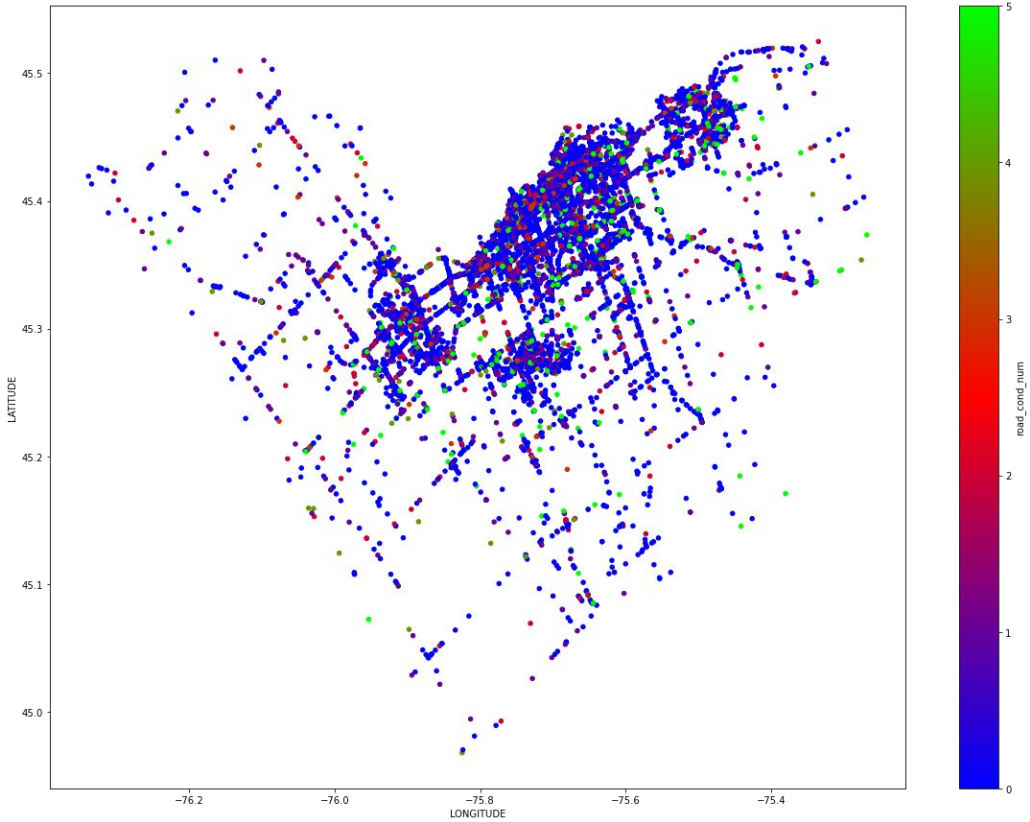



```
01 - Property Damage    13657
02 - Injury              2715
Name: Classification_Of_Accident, dtype: int64
```

Combine injury and fatality categories into one category to create a binary outcome.

The Setup

- ID#
- Accident Date
- Accident Time
- Location
- Geo ID
- Accident Location
- Classification of Accident — y
- Initial Impact Type
- Environment Condition — X
- Light — X
- Road Surface Condition — X
- Traffic Control — X
- Traffic Control Condition
- Longitude
- Latitude

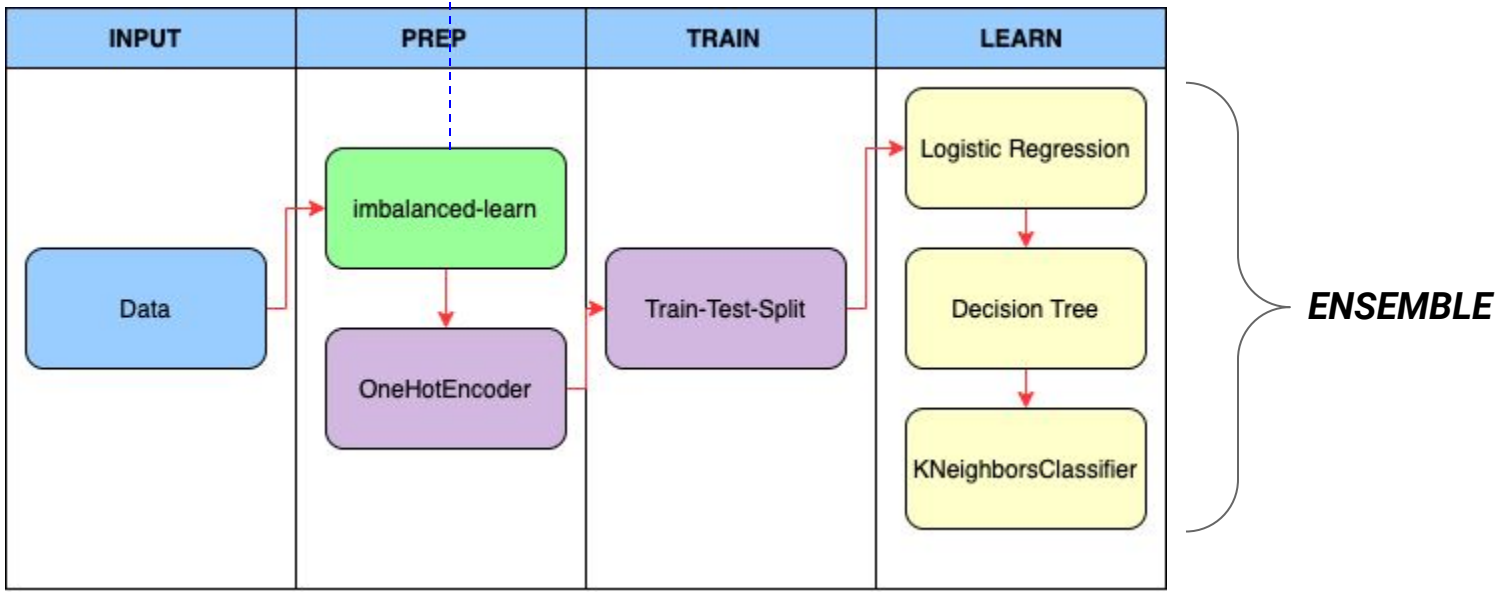


API Source: https://open.ottawa.ca/datasets/2019-tabular-transportation-collision-data

The Model

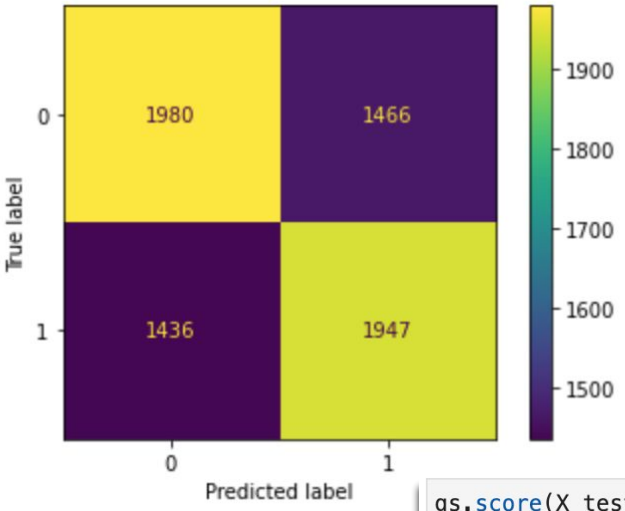
```
01 - Property Damage    13657
02 - Injury              2715
Name: Classification_Of_Accident, dtype: int64
```

```
01 - Property Damage    13657
02 - Injury             13657
Name: Classification_Of_Accident, dtype: int64
```



The Model Accuracy

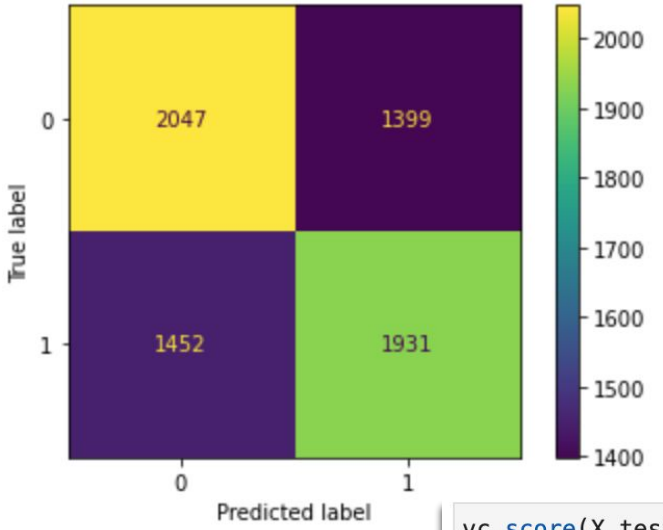
LOG REG ONLY




```
gs.score(X_test, y_test)
```

```
0.5750475911553669
```

ENSEMBLE (LR, DT, KNC)




```
vc.score(X_test, y_test)
```

```
0.5825157416898521
```

Ensemble method using Log Reg, Decision Tree, and KNN Classification improved model score.

The Future

- Explore model results to determine which input criteria predict a more serious collision outcome.
- Reduce scope to specific intersections or streets to determine safety trends of smaller area over some number of years.