Seattle City Traffic Accident Severity Analysis

How Traffic Accident Severity is Affected by Weather and Road Conditions

By: Sarah Fallon

Introduction to business problem

Problem Background

In 2007, Seattle city drivers were ranked the 5th worst in the nation.

Seattle City Police Department (SCPD) are now under pressure to improve their rating.

Plan of Action

SCPD hired a Data Scientist to review the traffic accident data.

They want a model to help predict traffic accident severity.

Stakeholder Goals

The goal of the SCPD is to be able to use a model to predict days/times that may be predisposed to more severe car accidents.

They want to create public warnings to reduce these accidents.

Data Understanding

All Seattle City traffic accident records
(2014 - present)

The data contained 37 attributes about each traffic accident record, including the following information:

- Collision address
- Date and time
- Type of intersection
- Severity of collision
- Number of people/objects
- Driver state (distracted, under the influence, speeding)
- Weather
- Road condition
- Light condition

Data challenges deep-dive

Challenge 1	Challenge 2	Challenge 3

Expand dataset

Initially, the data received only contained accidents of severity 1 (property damage) and 2 (minor injury).

Data scientists re-pulled the data for a complete cross-section.

Unbalanced data

Within the new dataset, the majority (68%) of accidents were Severity 1 while only .1% were the highest severity (Level 4, fatal).

The data had to be balanced for analysis.

Inconclusive data

There was some incomplete data within the dependent (accident severity) and independent (weather, road condition) variables.

These had to be removed for model training.

Methodology of Analysis

Initial analysis and machine learning models

Initial Analysis:

- Most accidents occurred on clear weather days and dry road conditions
- Correlation between weather and accidents may be minimal
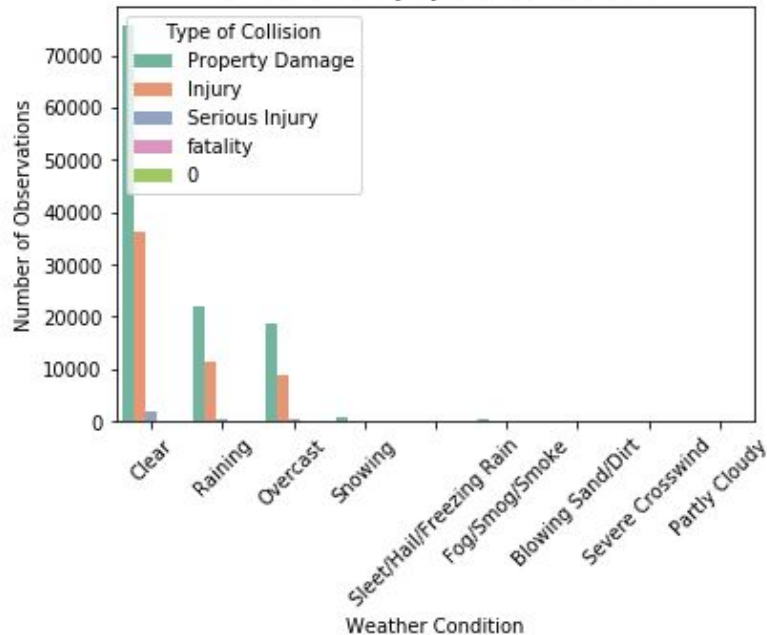
Machine Learning Models:

- 3 types of analysis were completed:
  - Correlation coefficient
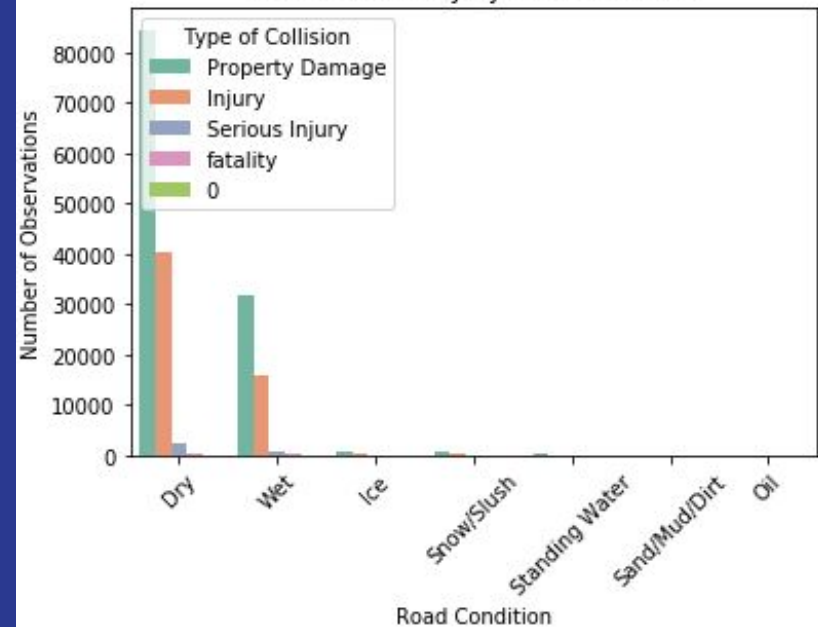 - K-nearest Neighbor
  - Decision Tree

Accident Severity
By weather and road condition

Accident Severity by Weather Condition

Accident Severity by Road Condition

Accuracy of Models

K-NN and decision tree model cccuracy

The model accuracy was calculated for the K-NN and decision tree models.

Both were noted to be similarly accurate, while the K-NN was slightly more effective at predictions based on weather and road condition.

Machine Learning Model	Accuracy
K-Nearest Neighbor	58.9%
Decision Tree	58.5%

Conclusion and Recommendations

Conclusion

Weather and road condition

Based on these two independent variables identified by the SCPD, the model created will only marginally help predict accident severity.

Other Observations

Clear days and dry roads

Through review of the data, most accidents occur on clear weather and dry road days.

Drivers may already be exercising caution for inclement weather or avoiding roads.

Recommendations

Further analysis for resource deployment

SCPD could explore the day, time, intersection, people/objects involved, crosswalk involvement, lane of accident, etc. to help better deploy resources to accidents and save lives.

Thank you!

Any questions?