Capstone Project
Car Accident Severity

Business Problem

Traffic collisions are in first place based on the number of deaths and injuries

Accidents occur for many reasons, including both technological and human factors

Is there a mechanism to predict the possibility of a car accident and its severity?

Goal :

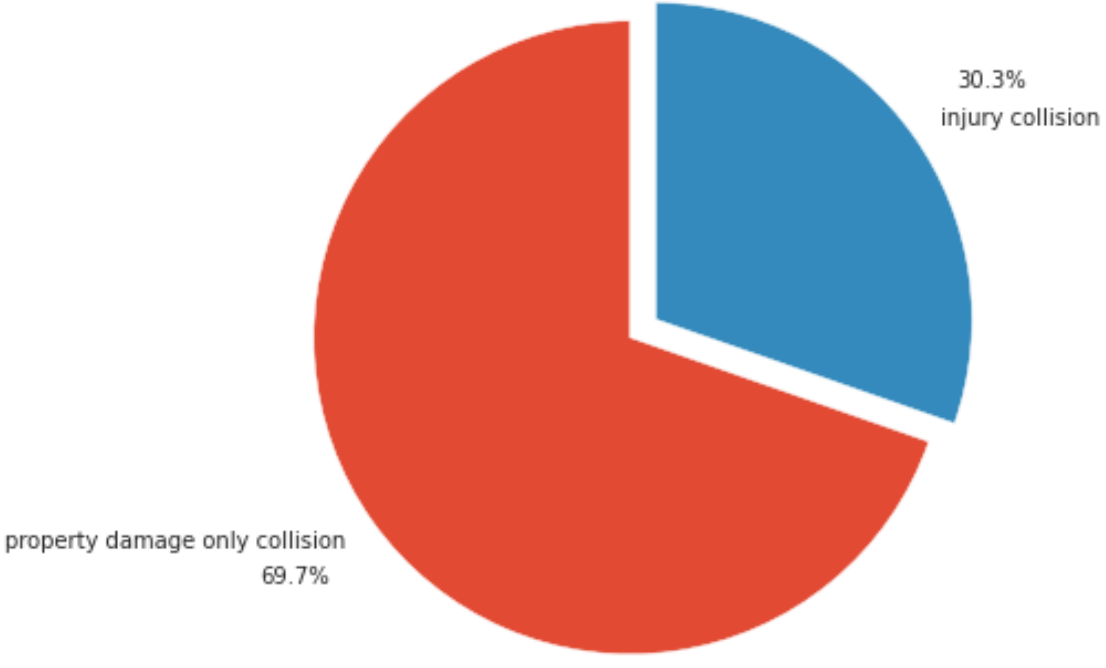
- identify and analyze the factors that cause traffic collisions
- create a model that will predict the severity of car accidents

Data

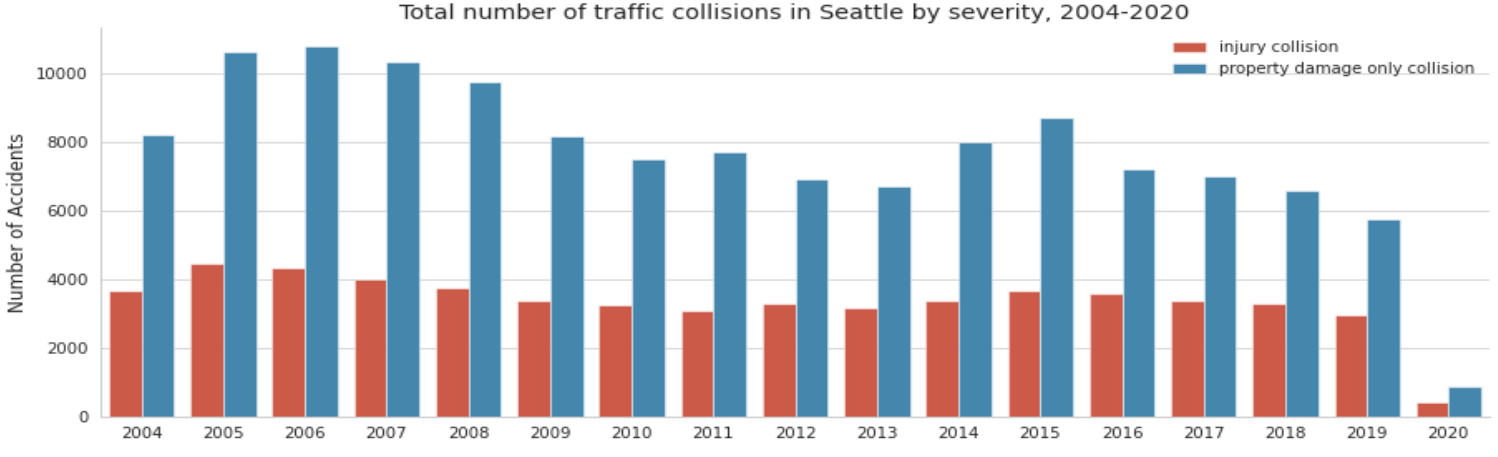
- Dataframe with accident severity of all collision types in Seattle, 2004-2020, available from IBM
- Total: 194,673 entries and 37 features
- After cleaning: 187,524 entries and 9 features

- Target variable: SEVERITYCODE
- Independent variables:
  - 'WEATHER' - A description of the weather conditions during the time of the collision;
  - 'ROADCOND' - The condition of the road during the collision;
  - 'LIGHTCOND' - The light conditions during the collision.

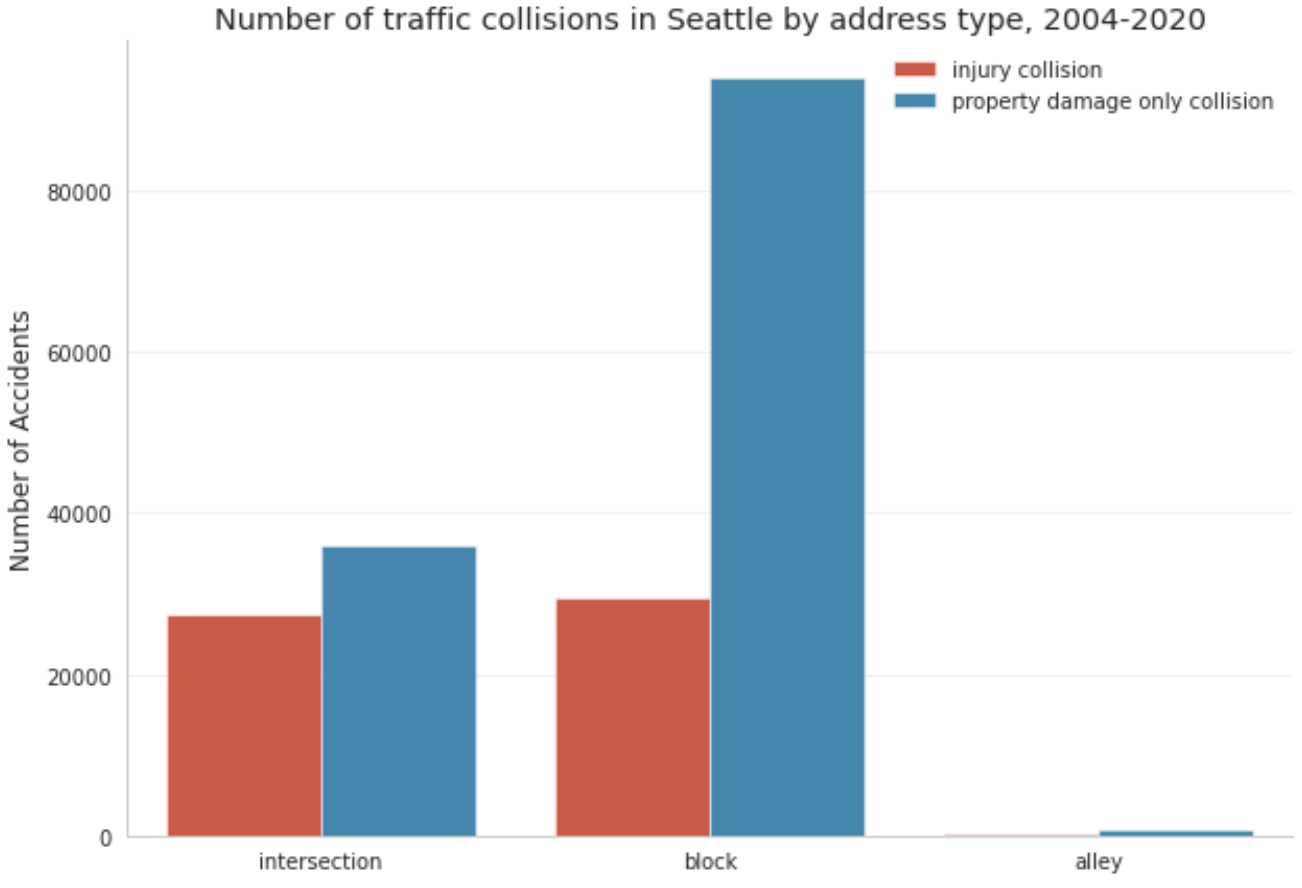
Accident severity distribution



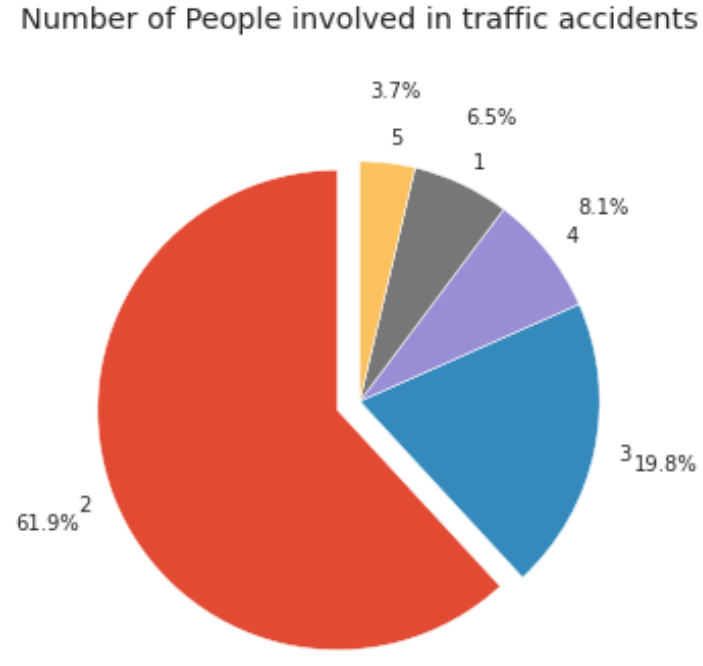
Annual amount of traffic accidents

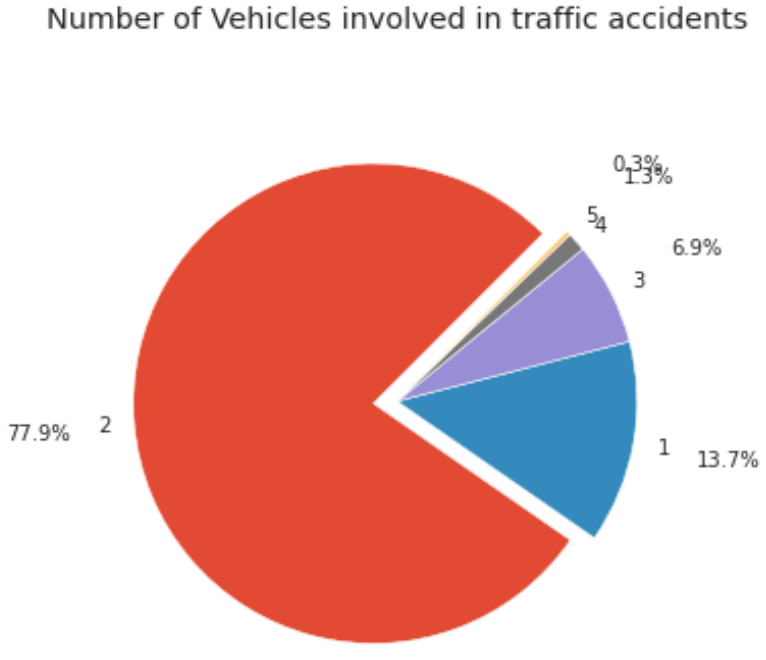


Address type of traffic accidents

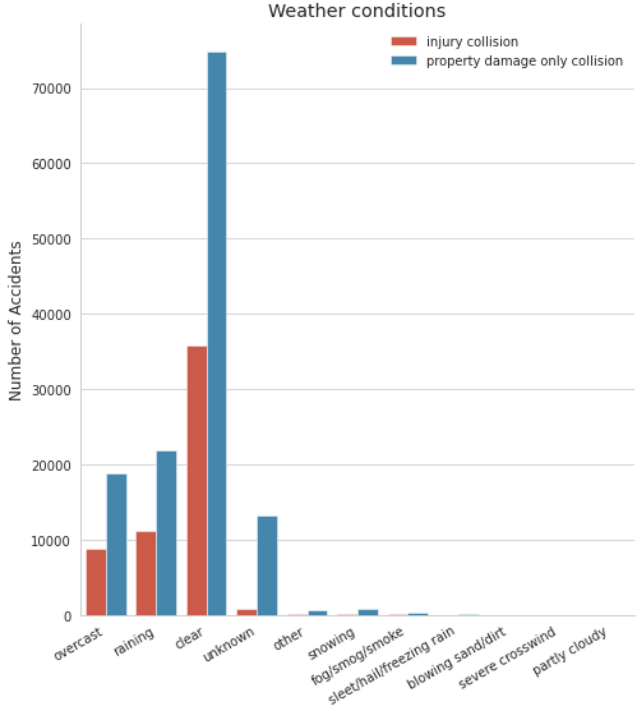


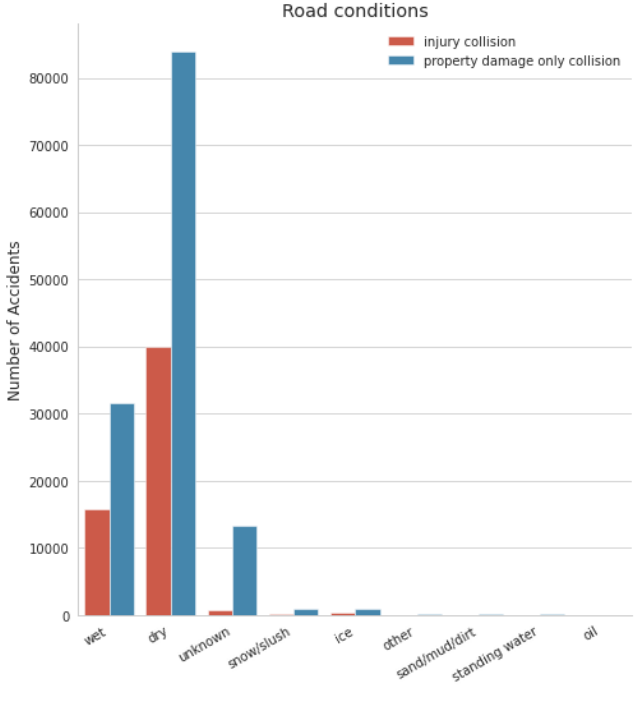
Number of People and Vehicles involved in traffic accidents

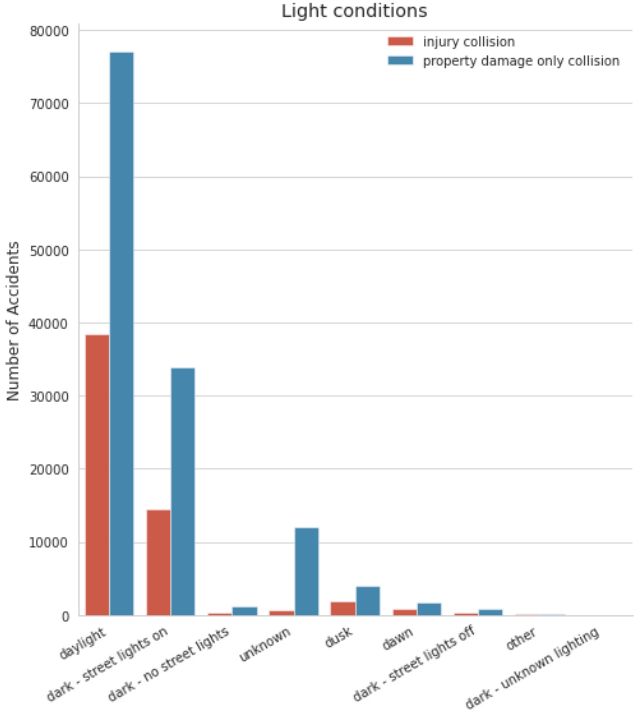




Weather, Road and Light conditions and Accident Severity







Classification algorithms

Algorithm	Jaccard	F1-score	Precision
Logistic Regression	0.6979	0.5737	0.4871
KNN	0.6673	0.6051	0.5958
Decision Tree	0.6979	0.5737	0.4871

Jaccard’s score between 66.7% and 69.7%

F1-score is between 57.3% and 60.5%

Precision is between 48.7% and 59.5%

Conclusions

- 70% of accidents resulted in property damage while other 30% involved injuries.
- Most of collisions happened either at the block or at the intersection.
- In almost 62% of cases 2 people get in a car accident.
- In nearly 78% 2 cars are affected by the collision.
- Traffic accidents usually happen during daytime with clear weather and dry road condition.
- All three algorithms showed similar results, while Logistic regression and Decision Tree showed better results in evaluating model accuracy.