

Predicting The Severity of Automobile Collisions

Matthew Levins

Making Our Roadways Safer

- The overall goal is to provide safer roadways for everyone who uses it
 - Includes Drivers and Pedestrians

Three Steps:

1. Understand the factors that lead to severe collisions
2. Be able to predict the severity of a potential collision based on the current factors at play
3. Plan, Design and Control the roadways based on these conclusions to make the roadways safer

The Data

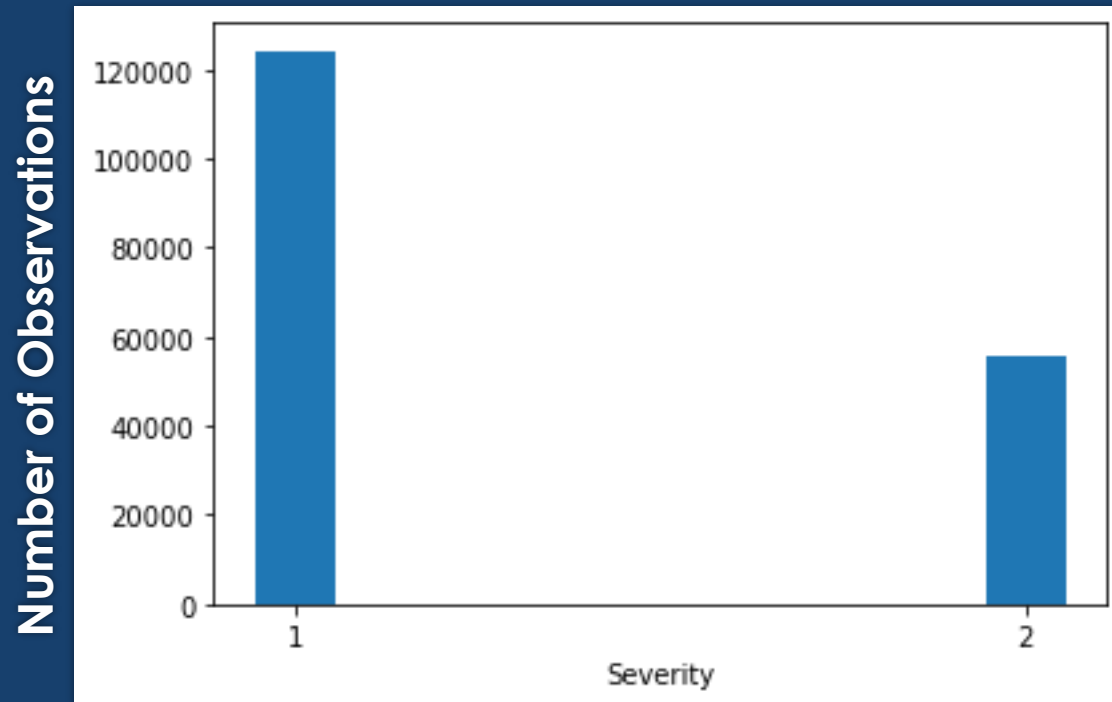
- The data used for investigation was Seattle's collision data
 - "Collisions – All Years" by SDOT
 - From seattle.gov

Data Information

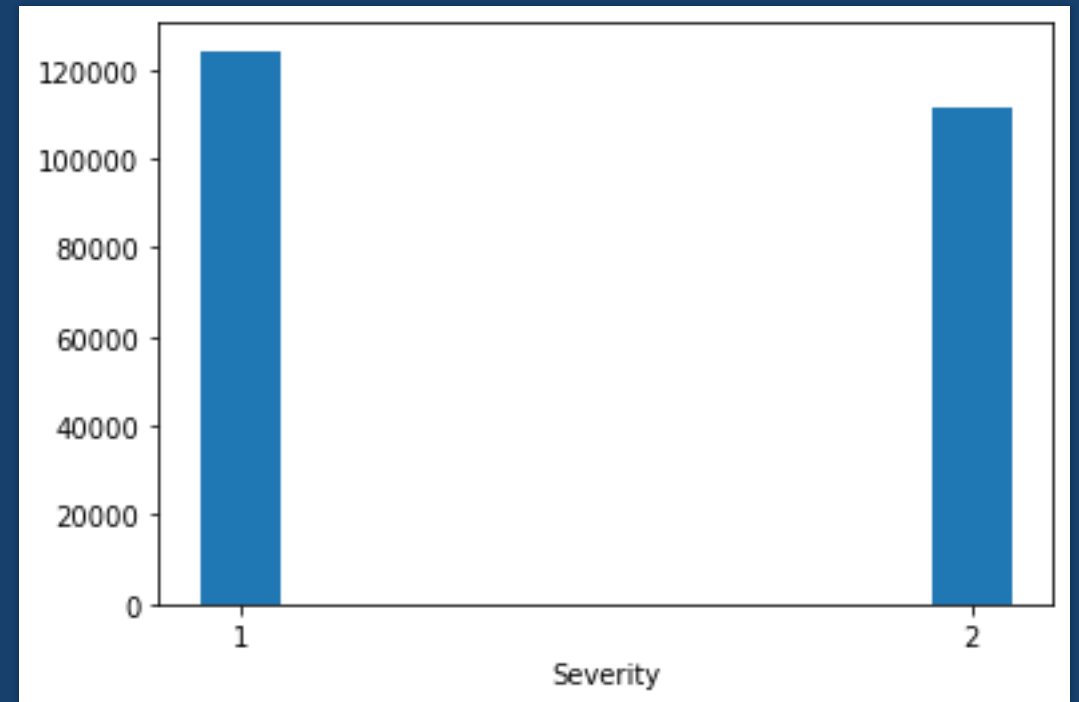
- Date Range: 2004 to Present (2020)
- 180,068 cleaned observations
- 21 Chosen Features (20 Predictors, 1 Label)

Exploratory Data Analysis

Balancing The Data



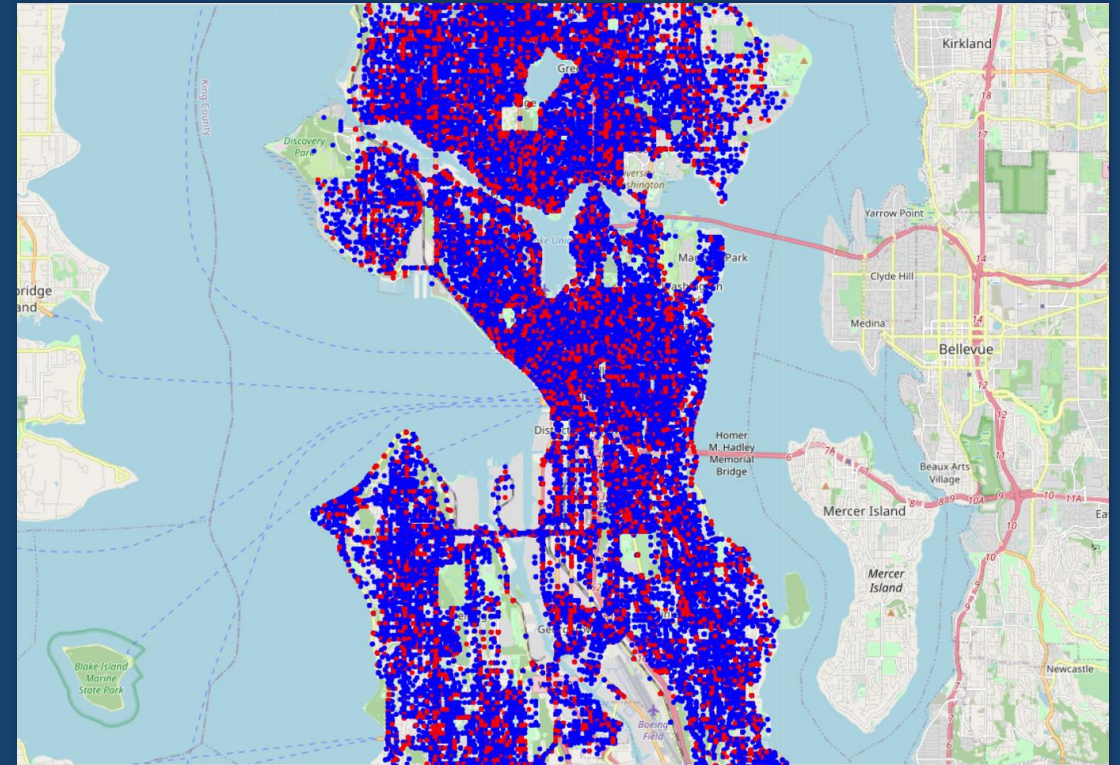
Original



Balanced

Geographical Data

- Blue – Level 1 Severity
- Red – Level 2 Severity
- No predominant clustering of Level 2 collisions
- Geographical location is not an effective predictor of collision severity



Severity vs X

Junction Type

- Intersections = More Severe (60%)

Alcohol Impairment

- Impaired = More Severe (55%)

Weather Conditions

- Harsh Weather = Less Severe (54 – 68%)

Speeding

- Collision from Speeding = More Severe (55%)

Road Conditions

- Predictable Harsh Conditions = Less Severe (58 – 71%)
- Unpredictable Harsh Conditions = More Severe (52 – 58%)

Light Conditions

- Dark Conditions = Less Severe (53 – 63%)

Pedestrian ROW

- Pedestrian Involvement = More Severe (18x More Likely)

Severity vs X

Time of Day

- Darker Times = Less Severe (52 – 57%)

Day of the Week

- Weekends = Less Severe (54 – 57%)

SEVERITYCODE	1	2
DAYOFTHEWEEK		
Thursday	0.514614	0.485386
Tuesday	0.518030	0.481970
Wednesday	0.519133	0.480867
Monday	0.519947	0.480053
Friday	0.531125	0.468875
Saturday	0.536265	0.463735
Sunday	0.555040	0.444960

SEVERITYCODE	1	2
TIMEOFDAY		
Afternoon	0.498278	0.501722
Morning	0.525797	0.474203
Evening	0.527289	0.472711
Early Morning	0.569526	0.430474

Predictive Modeling

Data Set (One Hot Encoded)

Id	Attribute	Count
0	SEVERITYCODE	235877
1	INTERSECTIONRELATED	235877
2	BLOCKNOTINTERSECTIONRELATED	235877
3	UNDERINFL	235877
4	Raining	235877
5	Fog/Smog/Smoke	235877
6	Blowing Sand/Dirt	235877
7	Sleet/Hail/Freezing Rain	235877
8	Snowing	235877
9	Oil	235877
10	Sand/Mud/Dirt	235877
11	Wet	235877
12	Standing Water	235877
13	Ice	235877
14	Snow/Slush	235877
15	DarkNoLights	235877
16	PEDROWNOTGRNT	235877
17	SPEEDING	235877
18	WEEKEND	235877
19	Afternoon	235877
20	Early Morning	235877

Predictive Modeling

Training Notes

- 70% Training, 30% Testing
- Hyperparameters tuned iteratively for each model type
- KNN had the highest computation time
- Accuracy scores based on testing set

Results

Model	Accuracy
Decision Tree	61.6%
KNN	61.2%
Logistic Regression	61.4%
Logistic Regression (Probability Prediction)	64.8%

Predictive Modeling

Training Notes

- 70% Training, 30% Testing
- Hyperparameters tuned iteratively for each model type
- KNN had the highest computation time
- Accuracy scores based on testing set
- Principal Component Analysis was conducted for second iteration of training

Results

Model	Accuracy
Decision Tree	61.6%
KNN	61.2%
Logistic Regression	61.4%
Logistic Regression (Probability Prediction)	64.8%

Model	Accuracy
PCA Decision Tree	61.5%
PCA Logistic Regression	60.8%
PCA Logistic Regression (Probability Prediction)	66.4%

Conclusions

- Results from EDA demonstrate that current drivers take extra precautions in unfavourable conditions shown in the lesser severity collisions
- More caution is needed for drivers' interactions with pedestrians

Next Steps:

- Instruction on ROW rules
- More defined pedestrian routes
- Predictive models need further investigation but offer preliminary benefits if used as a probability predictor