


PREDICTING ACCIDENT SEVERITY



PREDICTING ACCIDENT SEVERITY IS IMPORTANT FOR ALL PARTIES INVOLVED

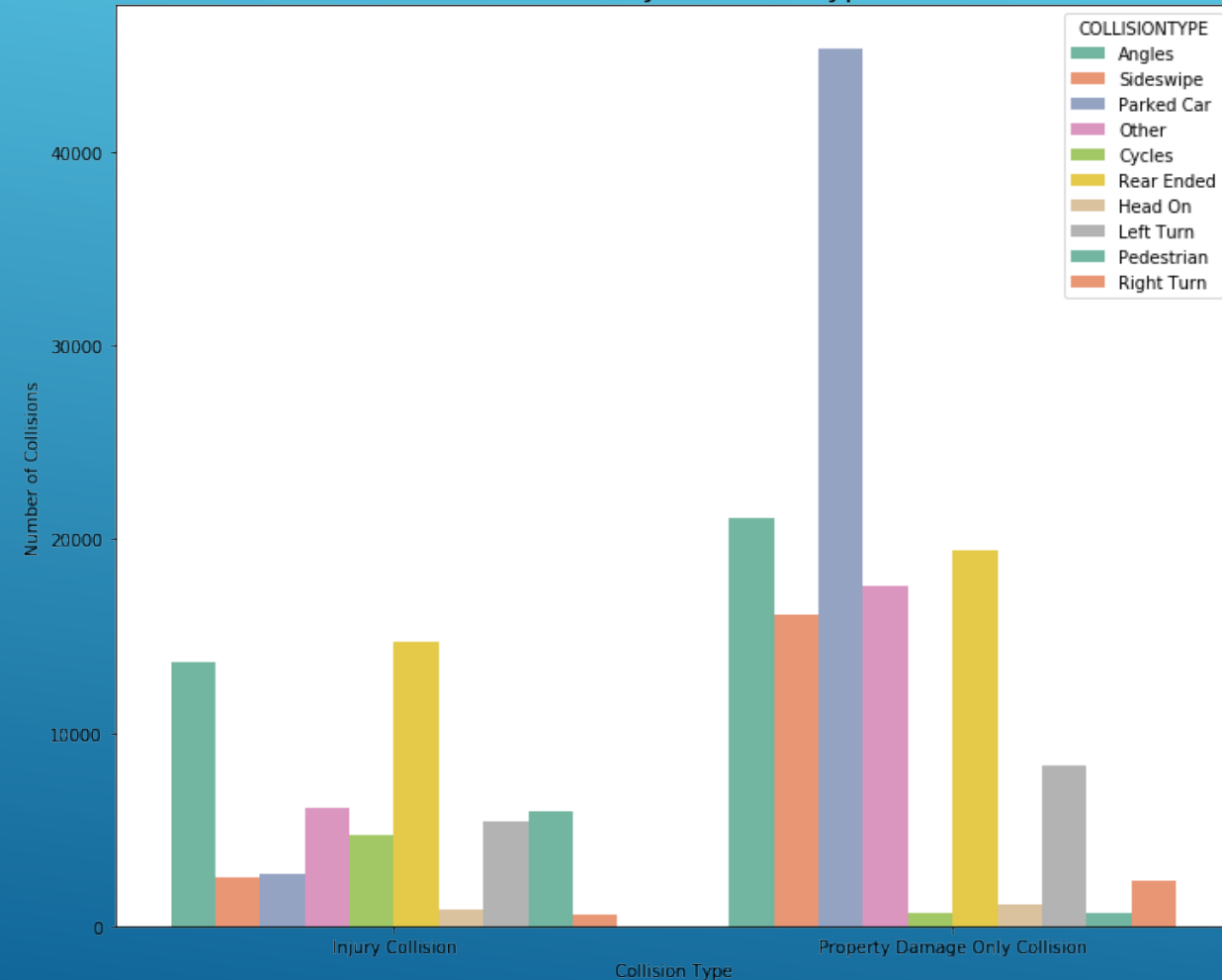
- ▶ Car accidents cause more injuries and deaths than any other type of personal injury
 - ▶ Value in having tools and means to provide drivers with a warning, given the weather and road conditions, about the possibility of getting into a car accident & severity
 - ▶ Helps avoid accidents
 - ▶ Allows transport, security, and emergency services to analyze crashes and dispatch appropriate response swiftly
- 
- A series of three parallel white diagonal lines in the bottom right corner of the slide, pointing towards the top right.

DATA ACQUISITION AND CLEANING

- ▶ Collision data from Seattle Police Department and recorded by Traffic Records
 - ▶ In total, 194,673 rows with 38 attributes in the raw dataset
 - ▶ Helps avoid accidents
 - ▶ Filled in missing/ambiguous values for columns of interest and converted all qualitative values into quantitative ones
 - ▶ Rows containing NaN values were dropped
- 
- Several white lines of varying lengths and orientations are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.

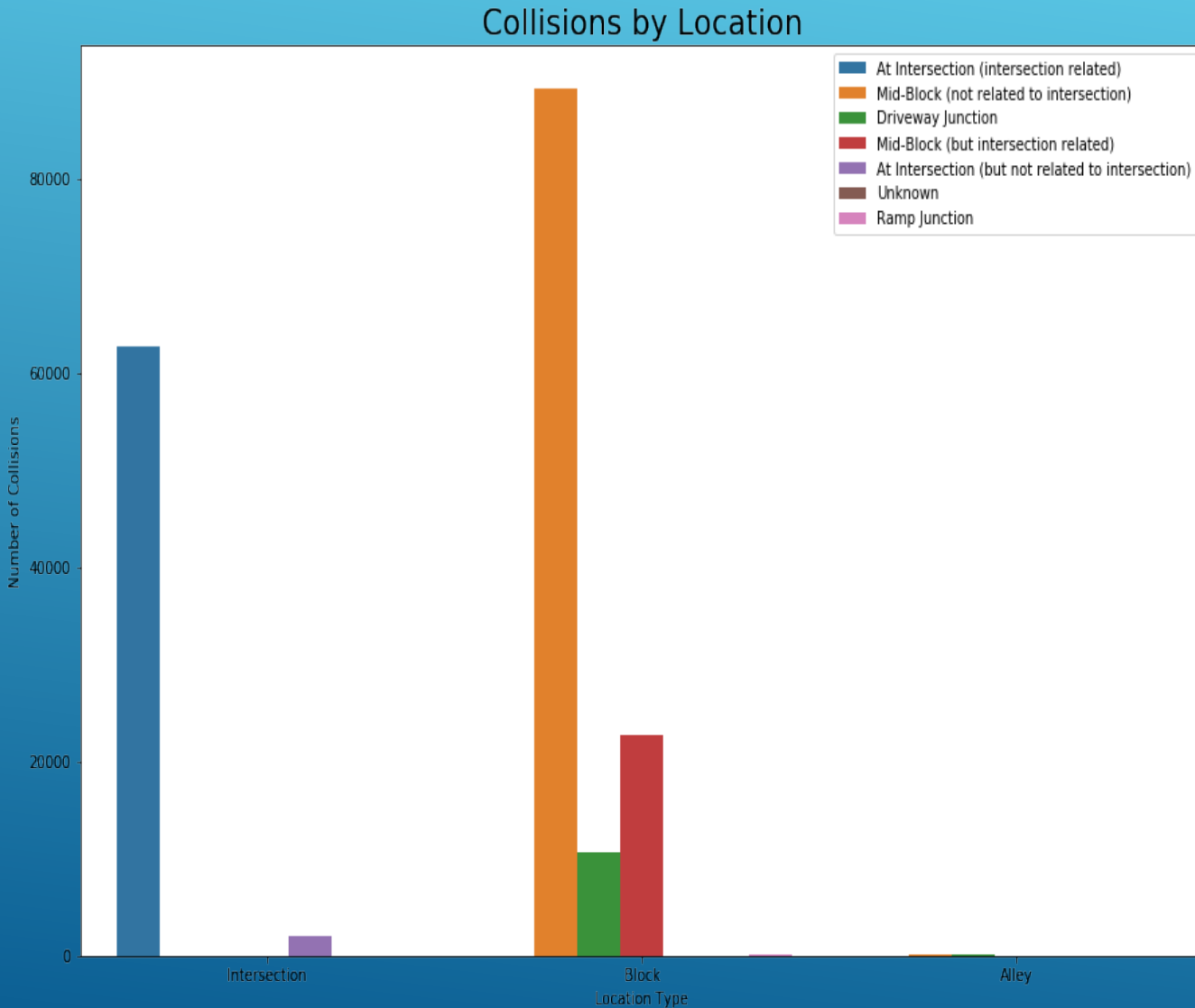
COLLISION TYPE AND NUMBER OF COLLISIONS

Accidents by Collision Type



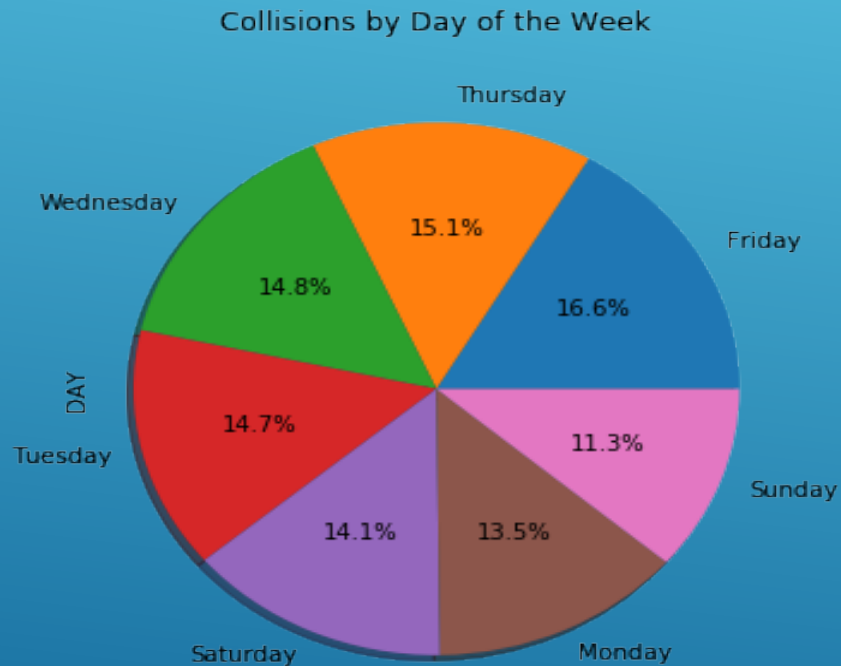
- ▶ Injury Collisions → Most accidents due to cars getting hit on the rear end/at an angle
- ▶ Property Damage → Majority involve hitting a parked car

LOCATION TYPE AND NUMBER OF COLLISIONS



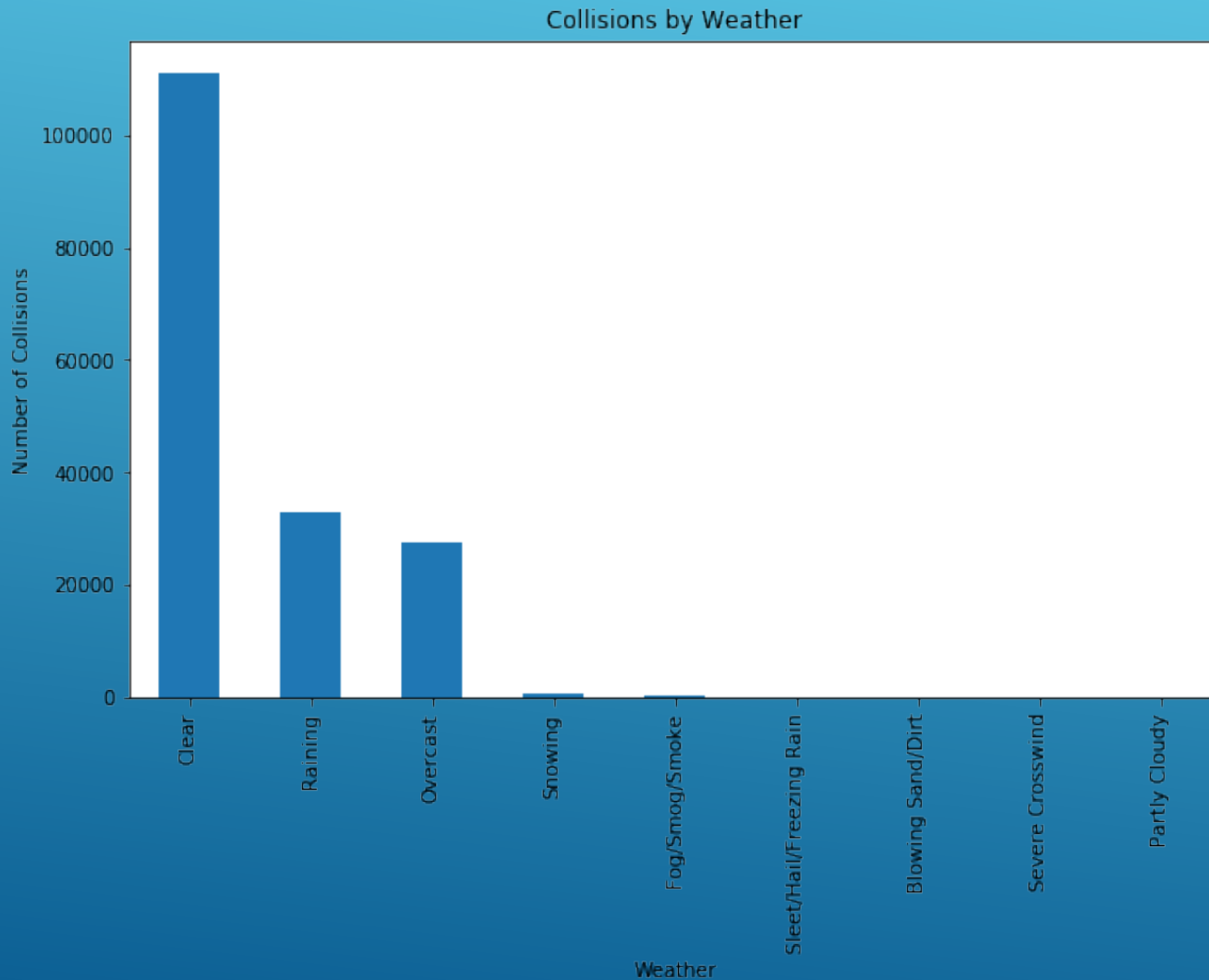
- ▶ Most block-related collisions happen mid-block
- ▶ Very little collisions occur in an alley

DAY OF THE WEEK AND NUMBER OF COLLISIONS



- Most accidents occur on a weekday as opposed to the weekend

WEATHER AND NUMBER OF COLLISIONS



- ▶ Most collisions occurred on a clear day
- ▶ Weather may not play an important factor when predicting collisions/collision severity

INFERENCEAL STATISTICAL TESTING

	SEVERITYCODE	Intersection	Alley	Block
SEVERITYCODE	1.000	0.199	-0.026	-0.185
Intersection	0.199	1.000	-0.044	-0.970
Alley	-0.026	-0.044	1.000	-0.085
Block	-0.185	-0.970	-0.085	1.000

- ▶ Weather & Location had no strong correlation with collision severity
- ▶ Most severe injuries involved pedestrians and cyclists

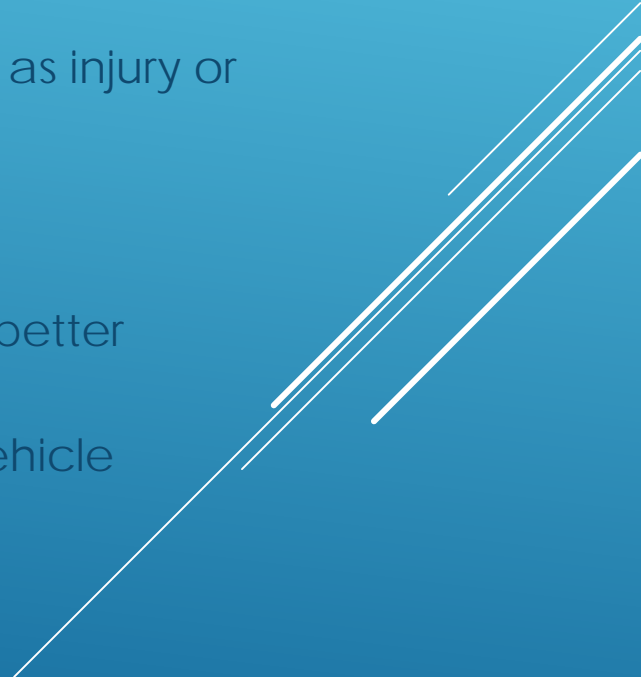
	SEVERITYCODE	PERSONCOUNT	PEDCOUNT	PEDCYLCOUNT	VEHCOUNT
SEVERITYCODE	1.000	0.131	0.246	0.214	-0.055
PERSONCOUNT	0.131	1.000	-0.023	-0.039	0.381
PEDCOUNT	0.246	-0.023	1.000	-0.017	-0.261
PEDCYLCOUNT	0.214	-0.039	-0.017	1.000	-0.254
VEHCOUNT	-0.055	0.381	-0.261	-0.254	1.000

DECISION TREE AND LOGISTIC REGRESSION MODELING


Evaluation	Decision Tree	Logistic Regression
Jaccard Index	0.73	0.73
Precision	0.71	0.72
Recall	0.99	0.97
F1 Score	0.66	0.68

- ▶ Needed to normalize due to values not being in the same range
- ▶ Split the data set into a train set and test set
 - ▶ Test Size = 20%
 - ▶ Random State = 3

DISCUSSION

- ▶ Based on our analysis of the data, we have some key observations:
 - ▶ Most accidents occur during weekdays at intersections
 - ▶ Weather conditions do not play a significant role in accidents
 - ▶ Road and lighting conditions have a weak correlation with accidents
 - ▶ Between blocks, maximum accidents occur at mid-blocks
 - ▶ In collision accidents, maximum damage is done to parked cars
 - ▶ Analyzed different machine learning classification methods to classify accidents as injury or collision accidents
 - ▶ Logistic Regression model offered maximum accuracy
 - ▶ It correctly predicted 72% as injury collisions.
 - ▶ Data could be used by enable transport, security, and emergency agencies to better understand car accidents and take preventative measures to prevent them.
 - ▶ Also aimed at every-day commuters, whether they are pedestrians, cyclists or vehicle owners to be more careful at intersections to prevent an accident.
- 
- A series of three parallel white diagonal lines are positioned in the bottom right corner of the slide, extending from the bottom edge towards the right edge.

CONCLUSION

- ▶ Able to achieve an accuracy of 72% using Logistic Regression Classifier
 - ▶ Utilization of this work could narrow down to the location where maximum accidents occur and most affected
 - ▶ There is very less importance of human and weather factors in causing an accident
 - ▶ Prediction could be improved by capturing real-time data during traffic incidents
- 
- Several white lines of varying lengths and thicknesses are positioned in the bottom right corner of the slide, creating a modern, abstract graphic element.