

Car Accident Severity

A data analysis project by Waleed
Muhammad





Aim of The Project

- Understand factors that influence accident severity
 - Road conditions, Weather, Time of day, etc
- Accidents that result in injuries can be reduced
- This data can be used across heavily populated cities
 - Local governments can use this
 - Organizations can use this

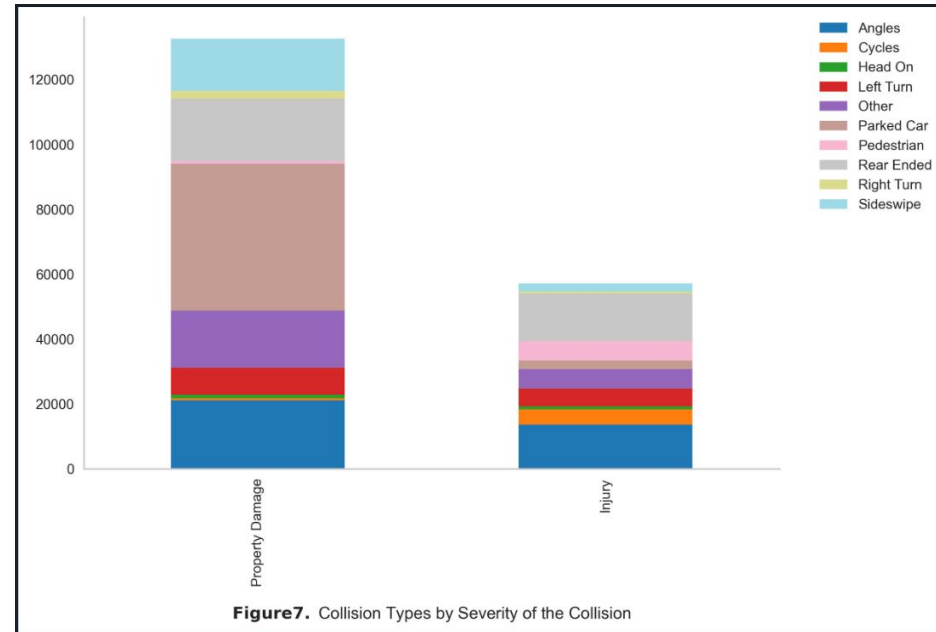


Data Acquisition & Cleaning

- Data was acquired via public car collision data.
- In total, there are 194,673 rows and 38 features in the raw dataset.
- The cleaned data has 54 properties

Be aware of pedestrians, cyclists, and parked cars

- Most property damage accidents are due to parked cars.
- Accidents involving cyclists and pedestrians result in injury.





Road Conditions

Dry road conditions result in the most injury and property damage related accidents. Do not underestimate this fact.

It is also important to be aware when it rains and the roads are wet.

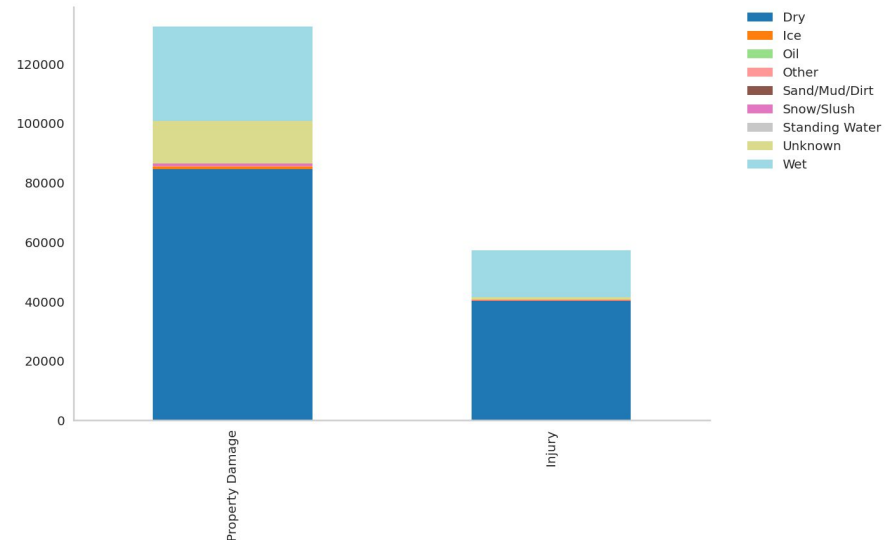


Figure4. Road Conditions by Severity of the Collision



Take caution at intersections

People should be very aware when driving passed Intersections.

These lead to a significant amount of property and personal injury.

It is important to note parked cars are surprisingly involved in many accidents as well.

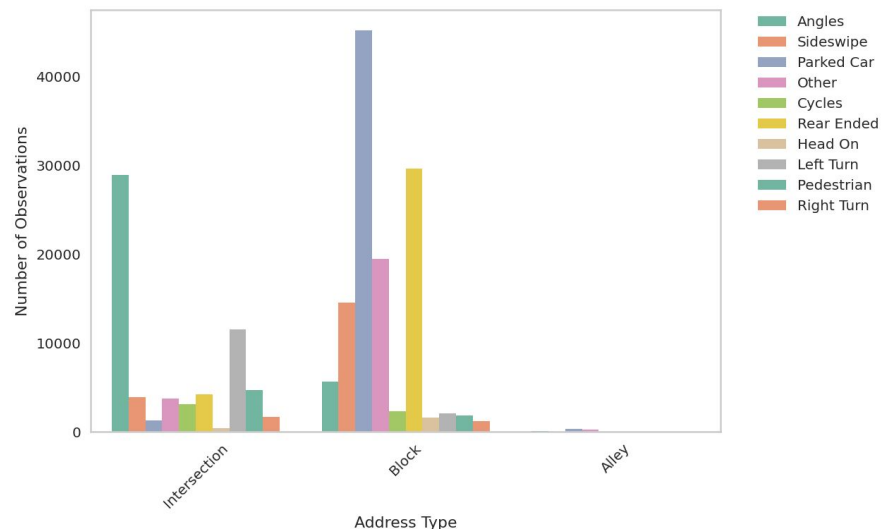


Figure6. Collision Types by Address Type

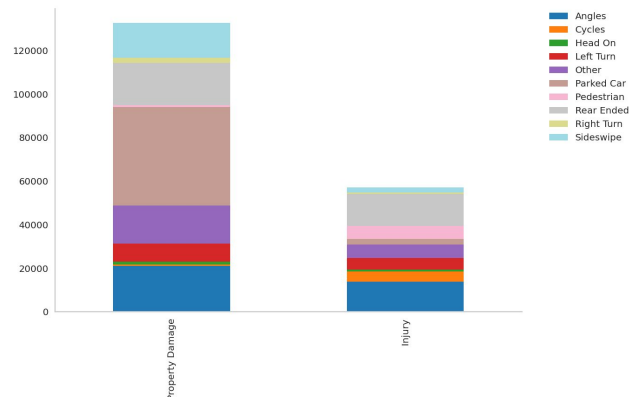


Figure7. Collision Types by Severity of the Collision



Be careful on Friday at 5pm

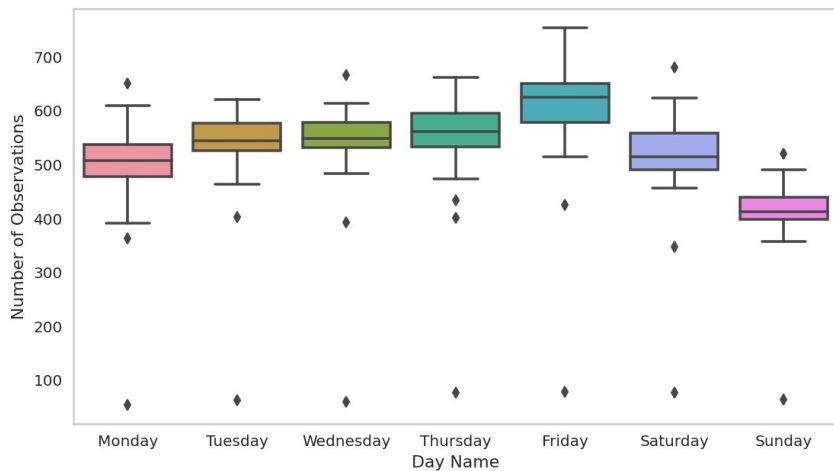


Figure10. Number of Collisions by Day Name

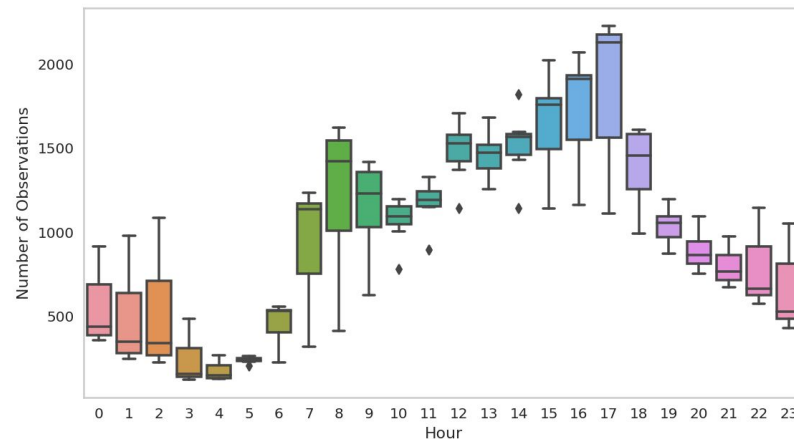


Figure11. Number of Collisions by Hour



Best Classification Models

According to the initial situation, we can say that the recall value to accidents resulting in injury increases.

	Precision	Recall	F1 Score
Property Damage	.74	.98	.84
Injury	.80	.21	.34
Accuracy			.75
Macro Avg	.77	.59	.59
Weighted Avg	.76	.79	.69

Table 1. Logistic Regression results with class with class weight param


	Precision	Recall	F1 Score
Property Damage	.78	.80	.79
Injury	.50	.46	.48
Accuracy			.70
Macro Avg	.64	.63	.63
Weighted Avg	.69	.70	.70

Table 5. Random Forest results with undersampling (Tuned Model)

Best Classification Models

	Precision	Recall	F1 Score
Property Damage	.74	.99	.85
Injury	.85	.20	.33
Accuracy			.75
Macro Avg	.80	.59	.59
Weighted Avg	.78	.75	.69

Table 7. Gradient boosting results with undersampling



According to the initial situation, we can say that the recall value to accidents resulting in injury increases quite well.



Conclusion

- Built a model to evaluate features that correlate to car accidents
- All the important features relating to car accident severity were identified
- Best practices were recommended to drivers and interested organizations