

Car Accident Severity

Purbayan Chatterjee
October 4, 2020

1. Introduction

1.1 Background, Problem & Interest

Road accidents are fatal all over the world. Not only do they claim lives, they also leave the serious injuries behind or may be an unforgettable scar. These accidents are increasing year by year claiming more lives than ever and leaving so many wounded or paralysed.

Over-speeding accounted for about 64.4% of the persons killed world-wide (World Road Statistics, 2018) where India ranked 1st followed by China and the United States.

Road accidents in India claimed over 1.5 lakh lives in 2018. The ministry of road transport and highways issued a report on Road accidents in India in 2018, which showed that road accidents last year increased by 0.46% as compared to 2017.

A total of 4,67,044 road accidents have been reported by States and Union Territories (UTs) in the calendar year 2018, claiming 1,51,417 lives and causing injuries to 4,69,418 persons.

As per the WHO Global Report on Road Safety 2018, India accounts for almost 11% of the accident related deaths in the World.

National Highways which comprise of 1.94% of total road network, accounted for 30.2% of total road accidents and 35.7% of deaths in 2018.

State Highways which account for 2.97% of the road length accounted for 25.2% of accidents and 26.8% of deaths.

As India has a major problem of waterlogging during monsoon, this analysis can help BMC and the Road Department to predict the weather and take precautions in advance to reduce accidents.

My project works on various Machine Learning Algorithms to predict and thereby avoid/reduce future accidents.

Again, this is just a humble attempt to try and figure out measures to make commute safer especially in rough environmental and road conditions.

2. Data Analysis

2.1 Calculation of Severity

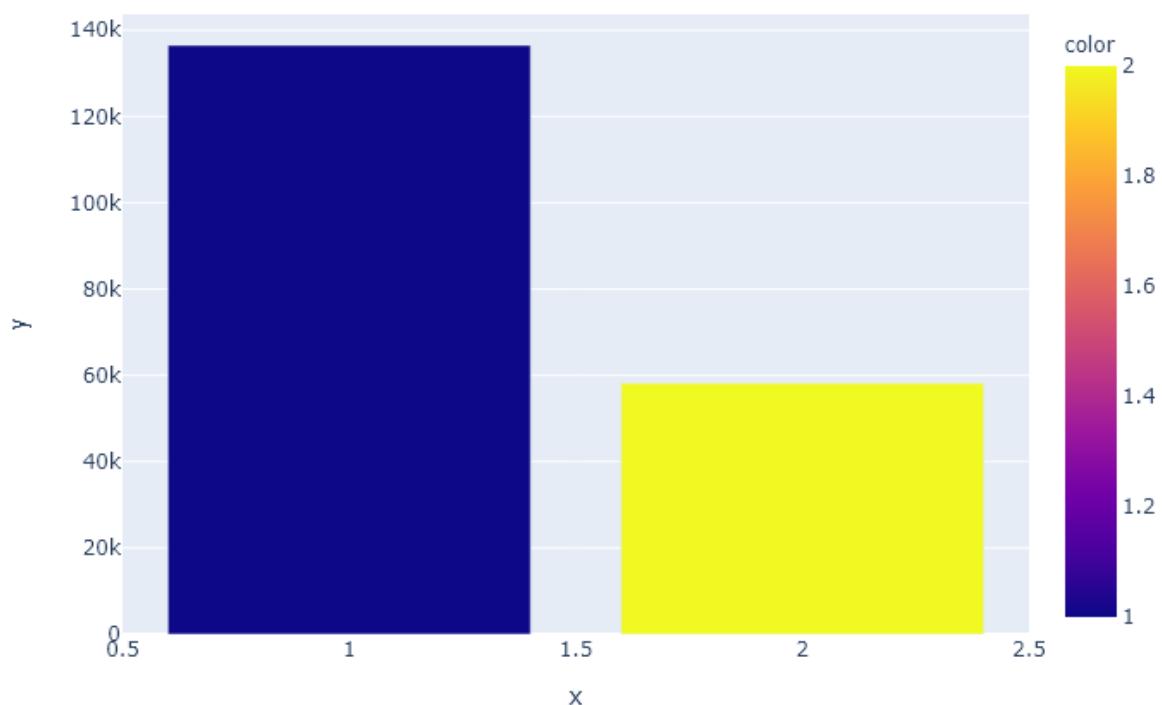
Easiest way to calculate severity is to visualize it. It helps us to make conclusions and draw inferences from the provided dataset or any data sample we have collected.

It also helps us to judge whether to keep certain parameters or to drop them.

2.2 Severity count

There are basically 2 levels of severity- Level 1 & Level 2.

With increase in severity level chances of casualty also increases.



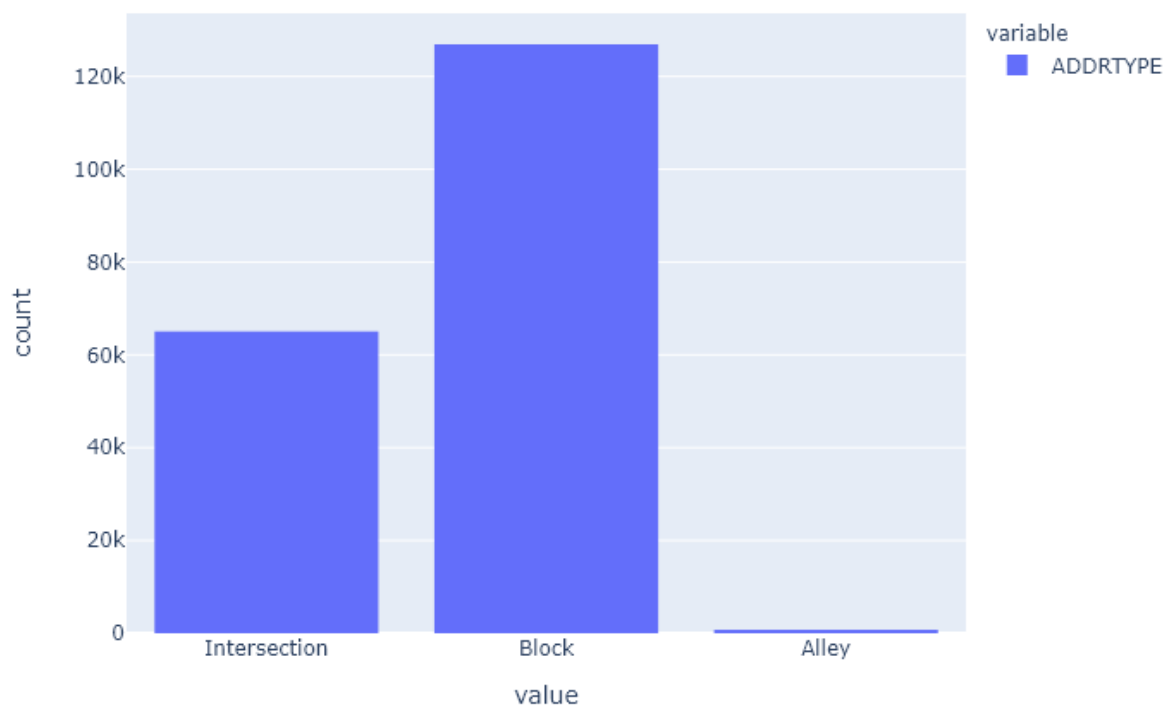
The figure shows that most of the accidents have occurred with a severity level of 1. But this does not mean that we can ignore severity of level 2.

We need to work on both levels and find their relationships with one another until we come to a proper conclusion.

2.3 Accident Location Count

Count of each location where accident has taken place is acutely visualized. This will help us to understand the factors that are going wrong like construction of road, walk-ways traffic lights, etc.

The information so obtained can be forwarded to the respective departments so that they could investigate the matter.



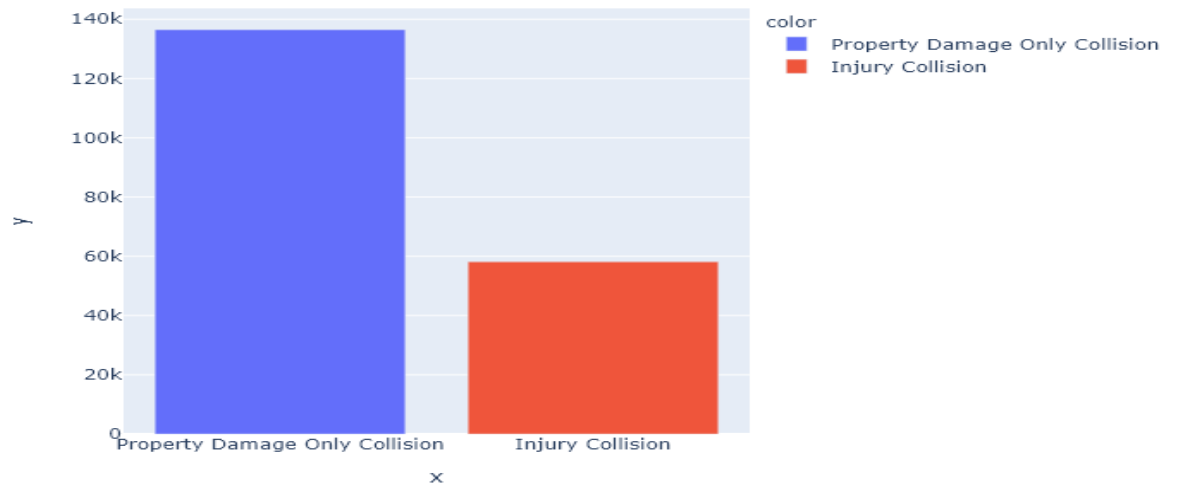
As seen in the graph, most of the accidents occur at blocks and then at intersections. This is because people on these roads must be in a rush and are driving so fast that they fail to notice another vehicle coming right at them.

Alleys, on the other hand have significantly less accidents.

This is because not many vehicles go into alleys because most of the alleys are way too narrow for a big vehicle to go into. Also, maybe due to the fact drivers become cautious while inside an alley as there are no proper rules there.

2.4 Count of every Accident Description

We will now count each accident description in order to find a hidden pattern. This will help us understand if a certain type of accident occurs more than other types.

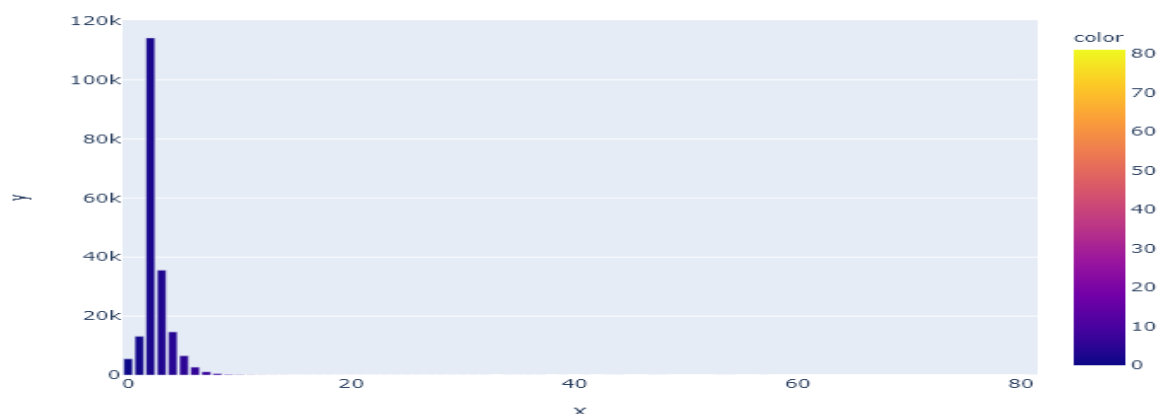


According to the visualization, property damage due to accidents is more compared to injuries.

So, this looks like only single vehicles were involved in accidents where they crashed with any property on the road (water pumps, traffic lights, shops or buildings).

2.5 No of People

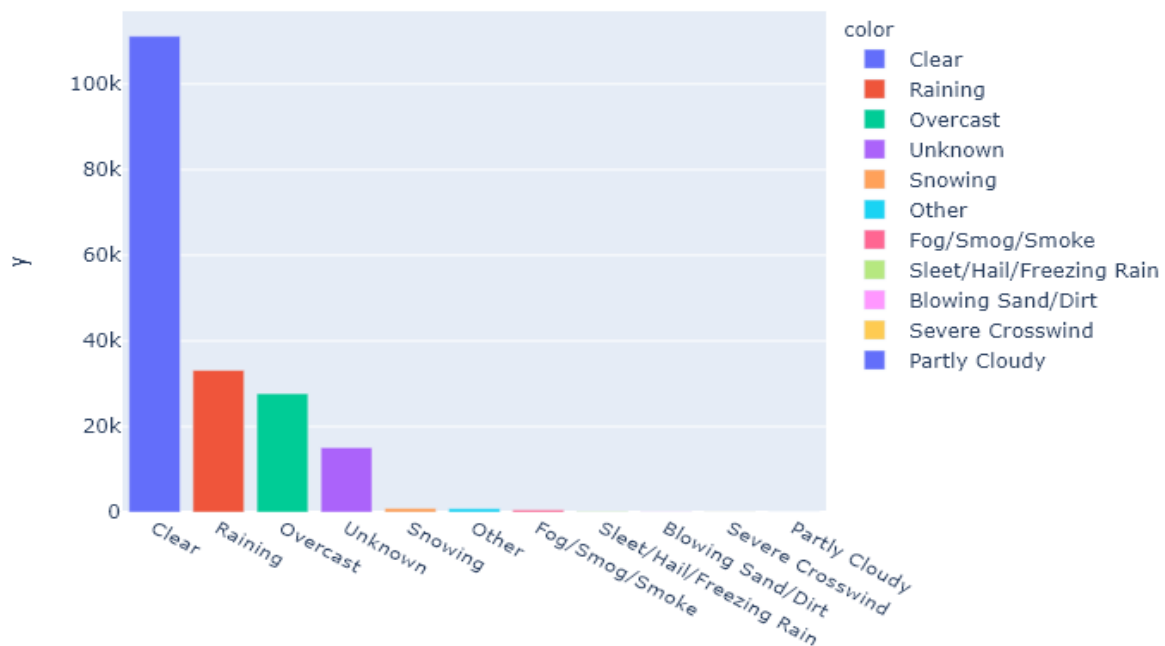
Now let us see what number of people get involved in a single accident and which is the highest in them. This will let us know if the accidents involved only had a driver or many others with them.



The figure tells that many accidents had more than 1 person involved. Most of them had 2 and 3 persons involved. We can assume that there are two people in their vehicles and the third one involved might be a pedestrian walking by.

2.6 Visualizing Weather

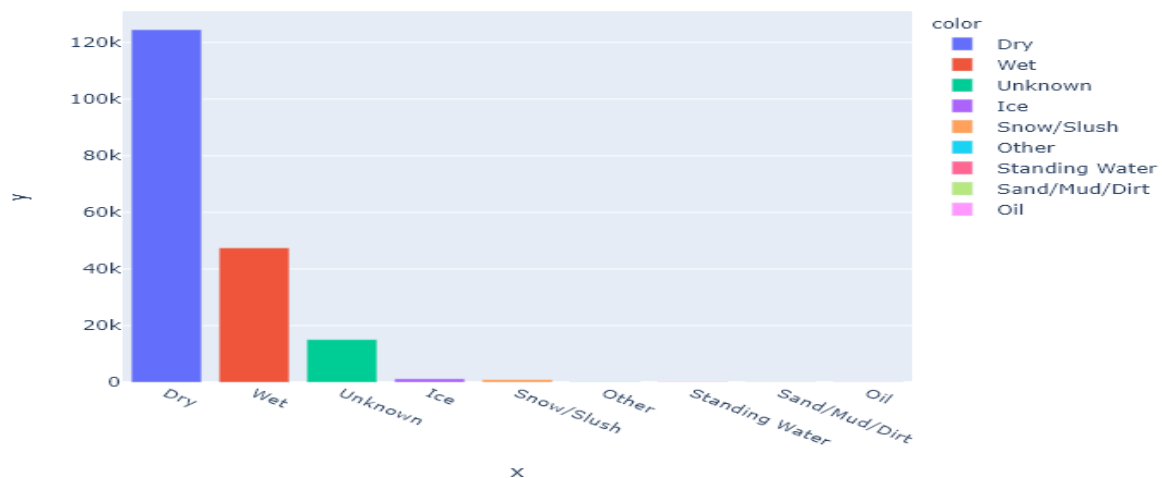
Visualizing the weather conditions may open up a new horizon.



Most of the accidents that have occurred had a clear weather on that day. However, we still can't conclude that weather has no role to play until we look at other conditions (Road & Light).

2.7 Road Conditions

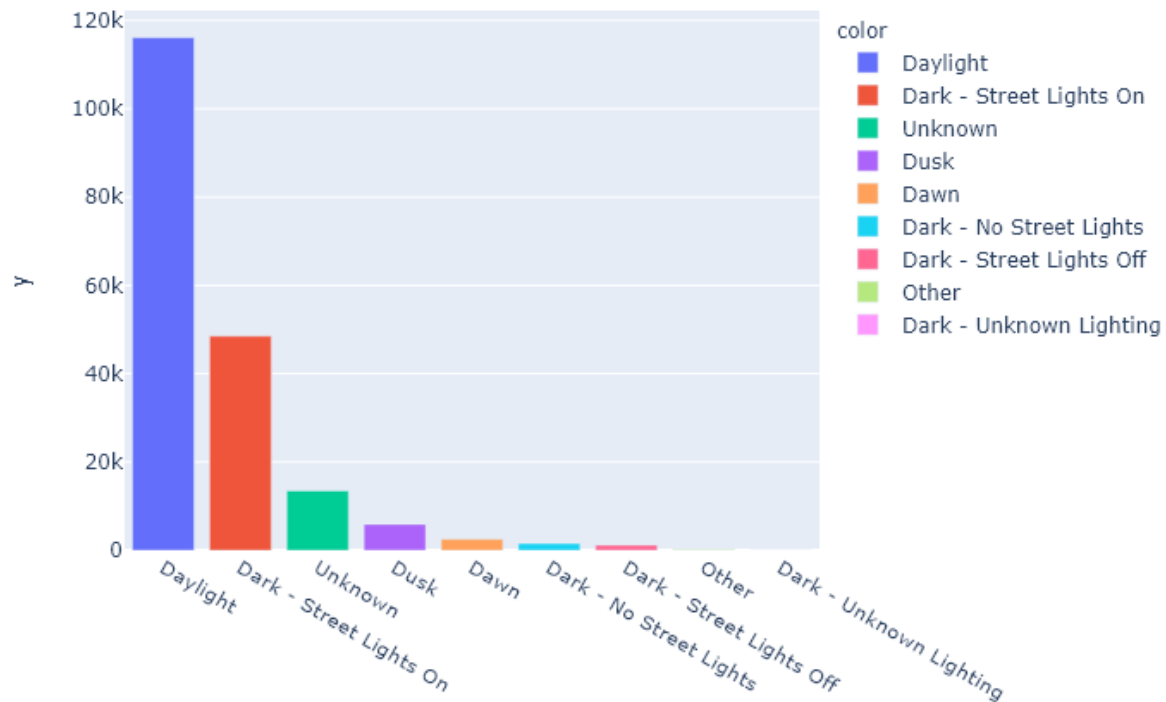
Let us have a visual of the road condition on the day of those respective accidents.



Roads were dry during most of the accidents. This means that roads absolutely did not have anything to do with these accidents. We now know that there is something else in the making these accidents.

2.8 Light Conditions

Let us see if light has got anything to do with these accidents.



Even the light has got nothing to do with these accidents. Maybe, the skillset of the drivers needs to be improved or the roads need maintenance work. Let us now, have a detailed discussion about all the collisions.

2.9 Collision Description

Let us observe which type of collisions are most frequent.



Most of accidents seem to occur when one of the cars was parked and still while the other was moving. This really seems to be a case of trash-driving.

3. Result

The following observations can be concluded after visualization of data.

1. Number of accidents with Severity 1 is greater than that with Severity 2.
2. More accidents occur at Blocks compared to Intersections or Alleys.
3. Most accidents cause only property damage rather than causing injuries.
4. Most accidents occurred when cars were not moving.
5. Most of the accidents occurred in broad daylight

All these succumb to the conclusion that most probably the roads weren't well maintained. Therefore, I believe the roads need proper maintenance, although we need to apply Machine Learning to come to any proper conclusions.

4. Conclusion

After training the dataset through various Machine Learning algorithms, we can conclude:

1. While 'parked car' category results in few injuries, 'number of vehicles involved' tells another story. When the number of vehicles is three or more, the chance of injury is nearly 50% and if the collision involves only 1 vehicle, the chance for injury far outweighs property damage only.
Remember, this is the feature that is most predictive for the model.
2. Taking this information and pairing it with the feature, 'number of pedestrians involved', clears the fact that any collision with a pedestrian is likely to cause injury. Chance of injury also goes up when the number of people involved goes up.
3. Collisions between midnight and 1 am are frequent and result in many injuries. Collisions also tend to climb throughout the day culminating at 5 pm.

Therefore, as individuals, we can be more aware of ours and others' driving habits in certain conditions, such as Friday at 5 pm, or at midnight when the bars close down. Downtown and highways are also areas where injury tends to occur.

As a city, implementing no turn on red and letting pedestrians go first may be ways to improve injury rates. Lowering speed limits in the areas where injuries occur more often may also be another way.