Car Accident Severity

Meet Shah

October 3, 2020

1. Introduction

1.1 Background, Problem & Interest

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. Over-speeding accounted for 64.4% of the persons killed. India, ranks 1st in the number of road accident deaths across the 199 countries reported in the World Road Statistics, 2018 followed by China and US. 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 percent of total road network, accounted for 30.2 per cent of total road accidents and 35.7 per cent of deaths in 2018. State Highways which account for 2.97% of the road length accounted for 25.2 percent and 26.8 percent of accidents and deaths, respectively. As India having a major problem of water-logging during monsoon, this analysis can help BMC and the Road Department can predict the weather and take precautions in advance so that accidents especially during rainfalls can be reduced to quite a bit.

I would like to know if we can reduce this severity and avoid future accidents by analyzing and applying various Classification Machine Learning Algorithms on the data provided by IBM. This study will hopefully reveal what, if any, measures we can take as individuals and municipalities to make travel in India safer especially under harsh weather conditions.

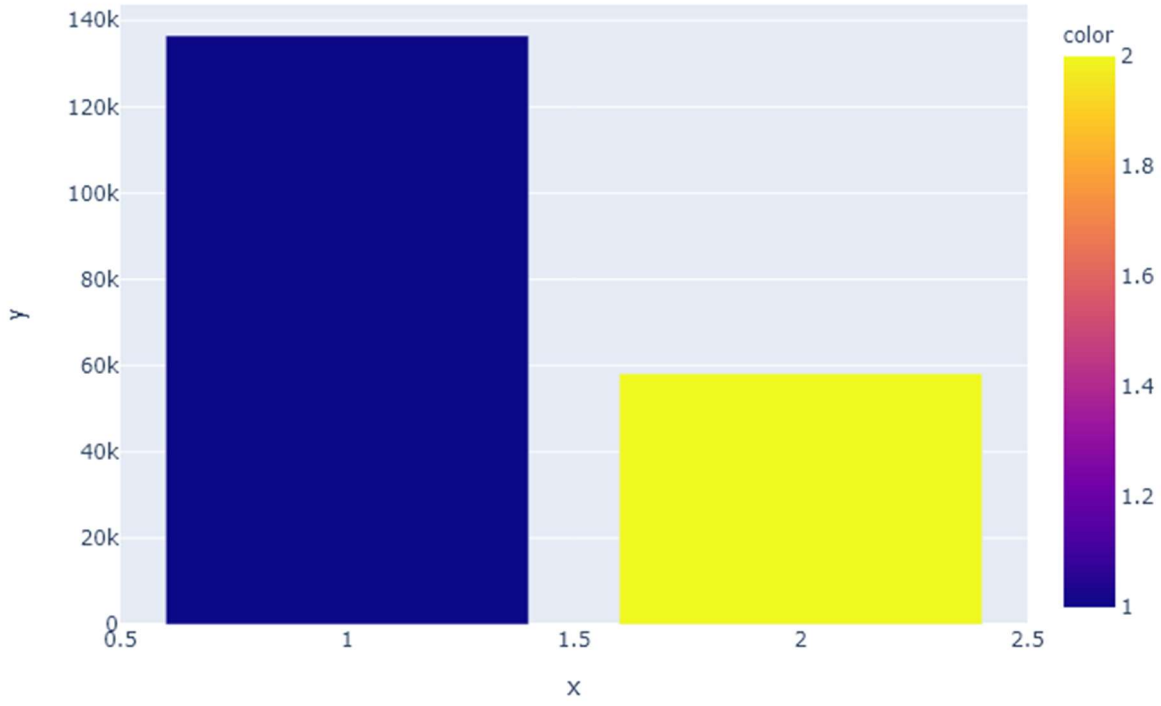
2. Exploratory Data Analysis

2.1 Calculation of Severity (Target/Dependent Variable)

Severity counts with other variables is not a feature we see in any dataset and hence is needed to be calculated. One easiest way to both calculate and understand severity with other variables is to visualize it. That way we can make conclusions regarding our data. This also helps in judging if either a certain variable is fit for our purpose or should we drop it.

2.2 Severity count of Level 1 & 2

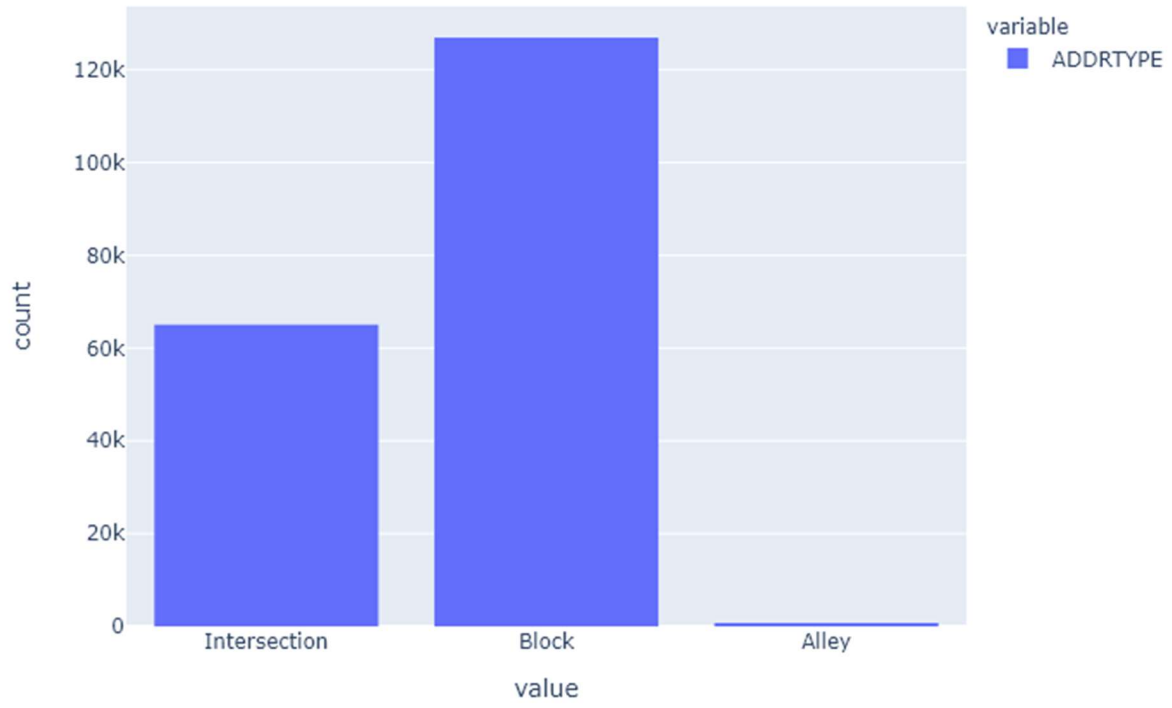
It is widely known that as the level of severity increases, so does the chances of a casualty. Below figure shows the count of each severity.



This 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 will do this count and their relationship with one another until we come to a hard conclusion.

2.3 Accident Location Count

Here we will see the count of each location where accidents took place In order to understand which location(s) are a hotspot for accidents. This will make us understand if there is something wrong (like road construction, traffic lights, pedestrian area etc.) in that area. We could pass on this information to the respective department so they could investigate this matter.

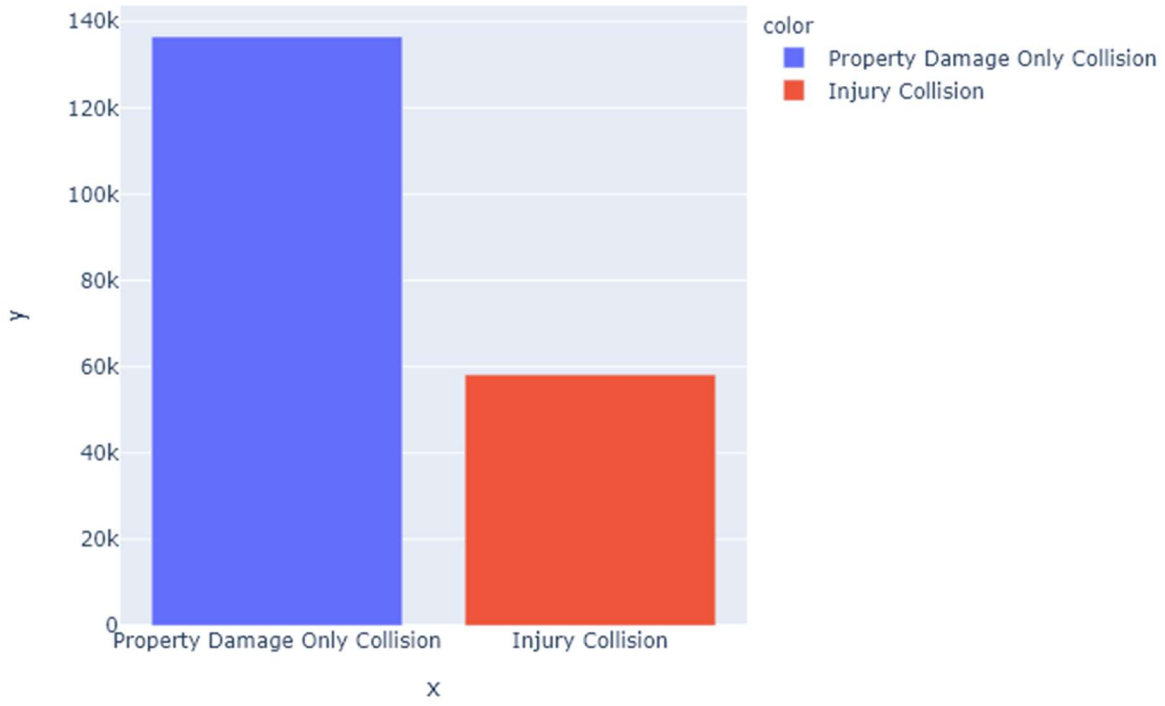


As seen here, most of the accidents occur at blocks and intersections. This is because people on these roads must be in a rush and are driving so fast that they fail to notice another vehicle coming right at them. Alleys, on the other and have significantly less accidents. This is because not many vehicles go

into alleys because most of the alleys are way too narrow for a big vehicle to go into.

2.4 Count of every Accident Description

We will now count each accident description in order to find a hidden pattern. This will let us know if certain type of accidents occurs more than other types. We know the description of these accidents but if we know which one occurs more then maybe we can stop those from happening or at least reduce those numbers.

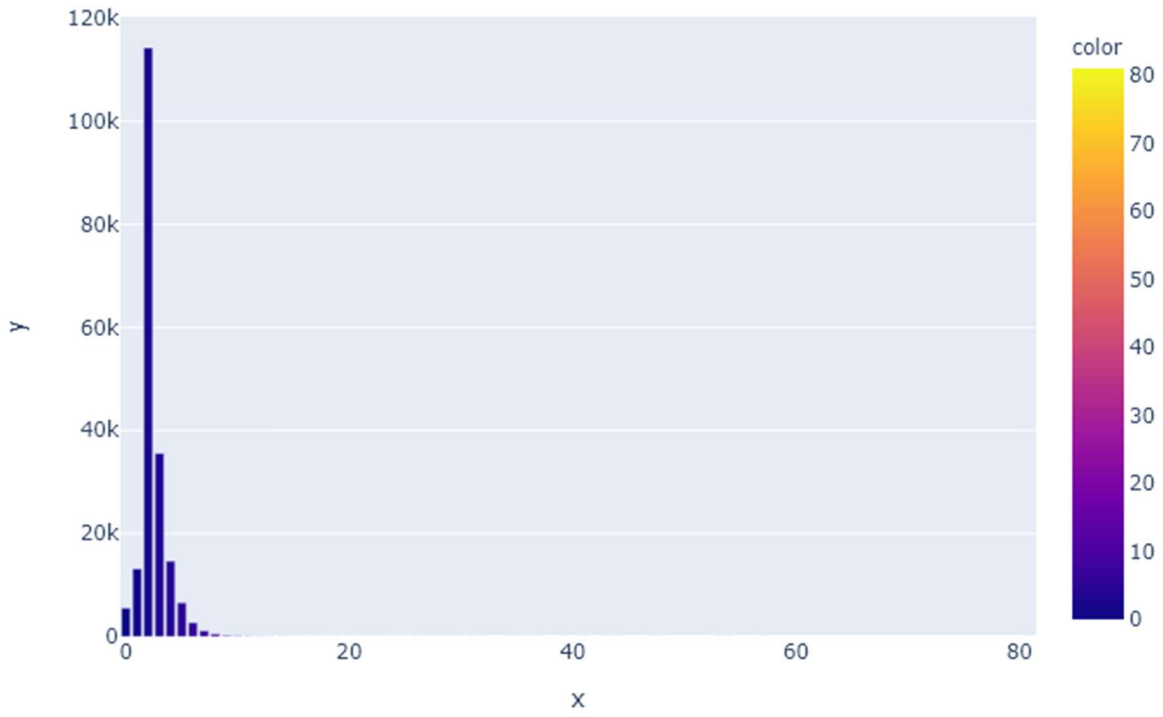


Hmm… So, this looks like only single vehicles were involved in accidents where they crashed with any property on the road. It can be water pumps on

the walking area of the road or a traffic light or worse any building. We will get to know this further down when we visualize many more aspects of this data.

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.

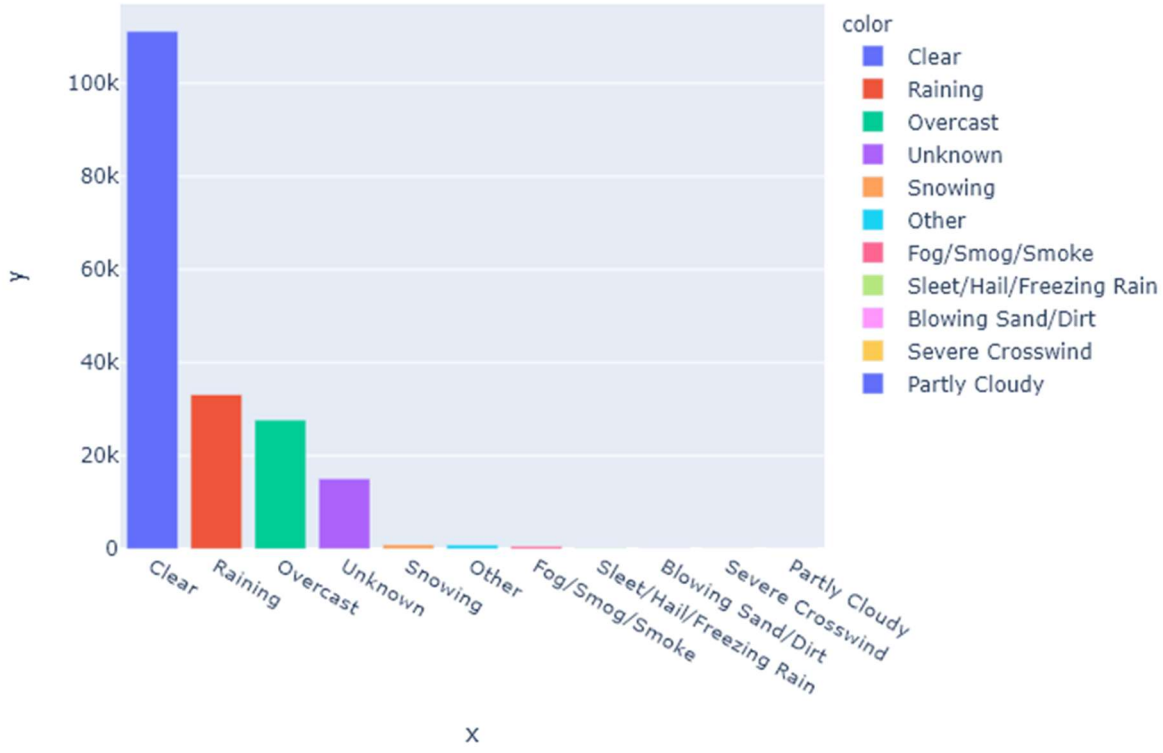


This is exactly what I suspected. Many accidents had more than 1 person involved. Most of them had 2 and 3 persons involved. We can assume that there are two people in their vehicles and the third one involved might be a

pedestrian walking by. Next, we will predict the weather to see if it must play any role in these accidents.

2.6 Visualizing Weather

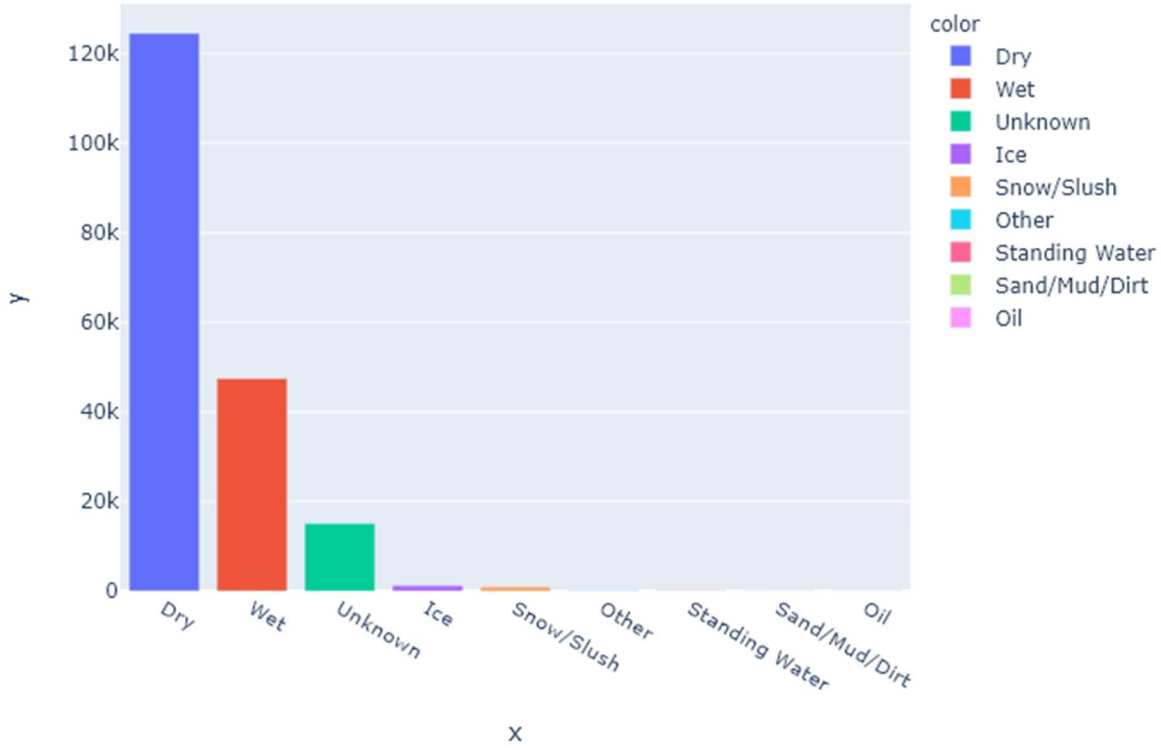
Now, we will visualize weather conditions for the day that accidents occurred and see if any hidden pattern pops up.



Now this is strange. Most of the accidents that have occurred had a clear weather on that day. We still cannot conclude that weather has no role to play until we look at Road & Light conditions, respectively.

2.7 Road Condition

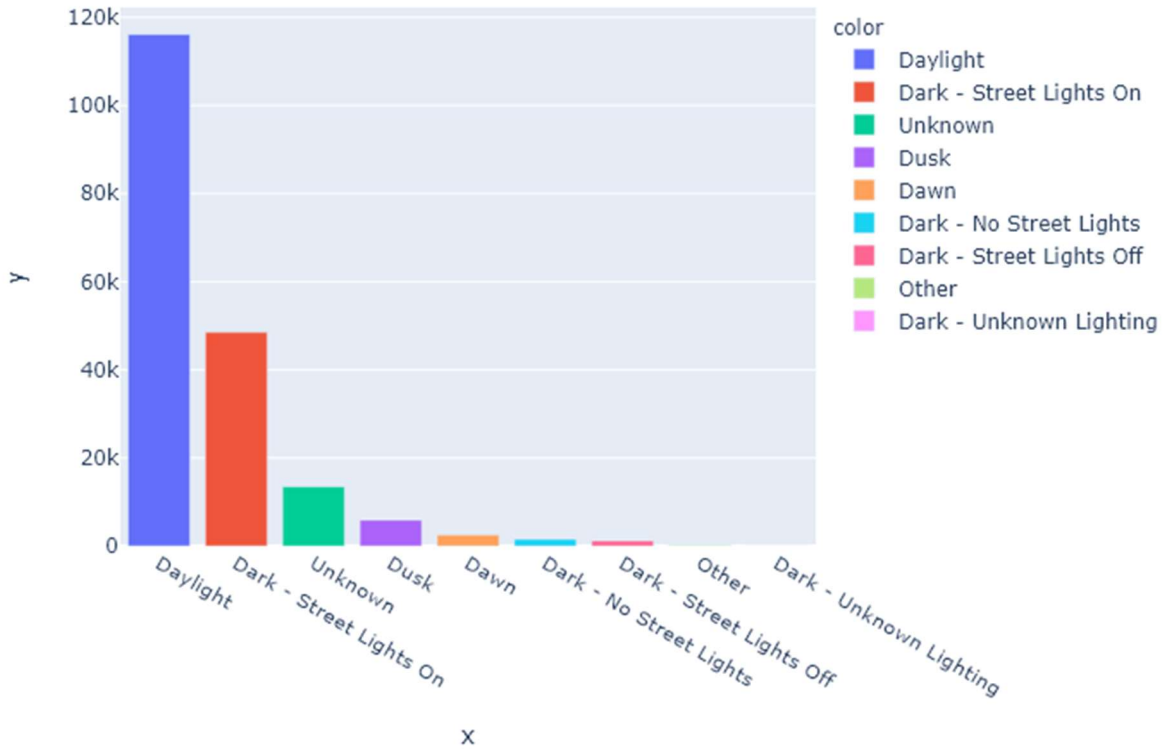
Let us have a visual of the road condition on the day of those respective accidents.



As expected, road was dry during most of the accidents. This means that roads absolutely did not have anything to do with those accidents. This really was a helpful visual for us. Whenever an accident occurs, first intuition is that it was because of the rain that the roads got wet and two-wheelers slipped. We now know that there is something else that is making these accidents happen or are the cause of major accidents. Next, we will have a look at the Light conditions and see if it has anything to do with these accidents.

2.8 Light Conditions

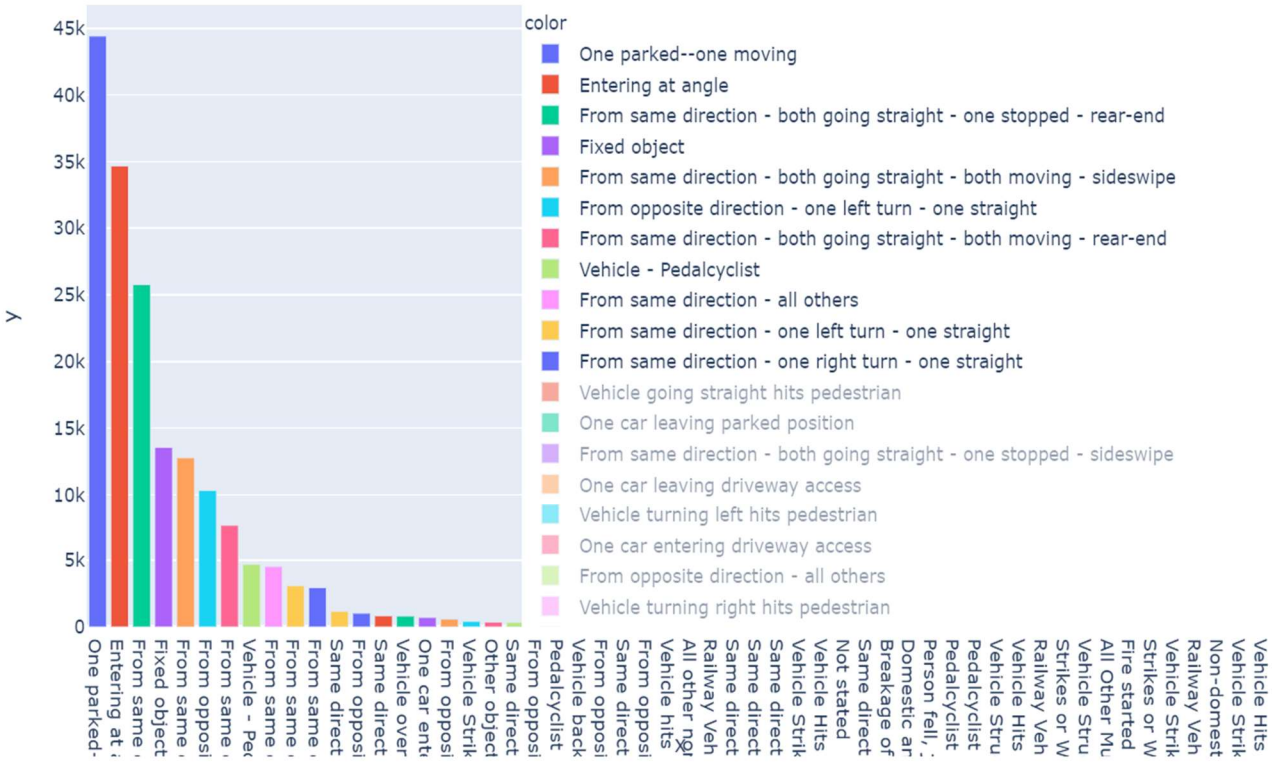
Let us see if light has got anything to do with these accidents.



Aaaaannndddd, NO. Even the light has got nothing to do with people crashing at each other. Then why are accidents happening?? Maybe it is the driving of people that must be improved? Or maybe it is the roads that need maintenance work or maybe the traffic lights are not in sync with each other? Whatever that is we need to improve it reduce the number of accidents that are occurring. Also, we will see the actual reasons in my Results section. But before that, let us have a detailed description of all the collisions.

2.9 Collision Description

Now the final visualization. Let us see which collision types are famous and most occurring.



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

Below we will the result section where everything will fall into place. After the result section, we will investigate our conclusion of this problem.

3.0 Result

Following are the observations we have concluded after visualizing our data:

1. Number of accidents with Severity 1 is greater than that of Severity 2. Severity 1 has total a of 124.258k fatalities while that of Severity 2 is of 55.809k.

2. More accidents occur at Blocks compared to Intersections. Number of accidents at occurred at Block are 117.085k while that of Intersections are 62.982k.

3. As seen in point No.1, it is good to see that most collisions caused only property damage like roads, vehicles etc. rather than causing Injuries. The

numbers are also the same - Property Damage = 124.258k & Injury Collisions = 55.809k

4. Top 3 accidents have occurred when:

1 - Cars were parked and not moving. Total of 42.886k Fatalities.

2 - At road angles. Probably occurred when one or more person(s) failed to notice another vehicle coming out from the other side of the road. Total of 34.353k Fatalities.

3 - At Rear Ends. This one occurs mostly when a person tries to overtake another vehicle in front of them.Total of 32.778k Fatalities.

5. Maximum 2 to 3 Persons were involved in a particular accident. No of accidents with 2 persons - 104.408k. No of accidents with 3 persons - 34.356k.

4.0 Conclusion

Most surprising thing to see is that most accidents have occurred when one of the two cars involved in an accident was parked and still. Also, most accidents have occurred in broad Daylight when the weather conditions were good. This is probably because of the roads. The roads are not well maintained and must be crooked or bumped here and there. From my observation, it is the Roads that need maintenance although we will come to our conclusion only after applying our Machine Learning Models on this data.