

# Car Accident Severity

## Introduction

In this part will be highlighting this Capstone Project to Car accident severity leveraging Cross-industry standard process for data mining phases. It is an initial phase to understand the project's objective from the business or application perspective. Then, you need to translate this knowledge into a machine learning problem with a preliminary plan to achieve the objectives.

Car accidents is one of the leading deaths causes around the world and according to the World Health Organization, road traffic injuries caused an estimated 1.35 million deaths worldwide in the year 2016. Also, car accidents have number of consequences, ranging from minor property damages to the more severe loss of human lives.

There are many studies on the factors causing traffic accidents. Also, it is very important is to determine what affect the level of severity of accidents. Additionally, it is also imperative to understand what causes severe accidents so that targets, priorities strategies can be developed to reduce high severity accident occurrences first, as an efficient use of limited resources.

Having such insights, governments would be interested in accurate predictions of the severity of an accident and reduce high severity accidents by eliminating different factors which could be mitigatable and un-mitigatable.

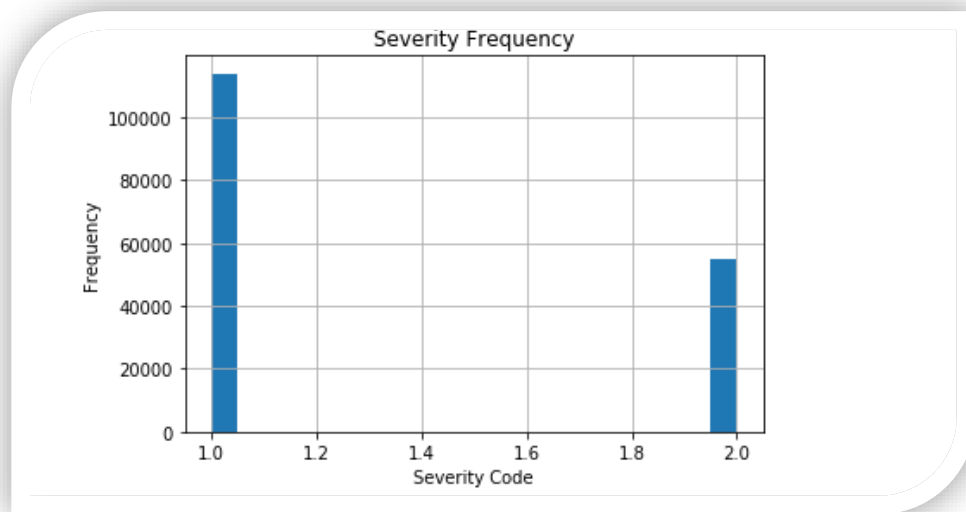
## Data Understanding

Given dataset needs re-processing and cleaning in order to support machine learning model to be executed. There are number of observations of the dataset where the total number of records is 194673 with 38 columns. Additionally, there are variations in each feature. Also, there are unknown values. Therefore, data cleansing has been applied for Unknown values.

## Methodology

Once the dataset was cleaned-up, number graphs were plotted to help which severity was commonly to arise and give insights.

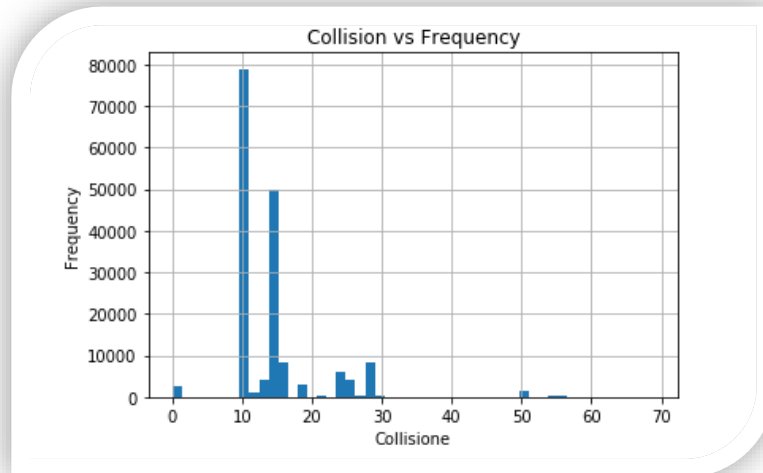
As you can see that the severity code 1 which is corresponding to property damage where severity code 2 is corresponding to injury



Severity code:

SEVERITYCODE	Text, 100	A code that corresponds to the severity of the collision: <ul style="list-style-type: none"><li>• 3—fatality</li><li>• 2b—serious injury</li><li>• 2—injury</li><li>• 1—prop damage</li><li>• 0—unknown</li></ul>
--------------	-----------	---

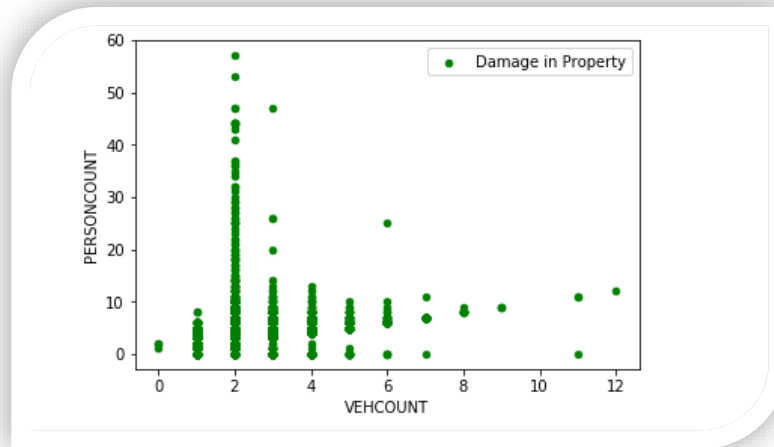
In the second view, it is from Collision and frequency perspective, but it is showing that Collision with code 10 corresponding to Entering At Angle and code 11 which is From Same Direction -Both Going Straight-Both Moving- Sideswipe. They were higher than the others.



### State Collision Code Dictionary

Code	Description
0	Vehicle Going Straight Hits Pedestrian
1	Vehicle Turning Right Hits Pedestrian
2	Vehicle Turning Left Hits Pedestrian
3	Vehicle Backing Hits Pedestrian
4	Vehicle Hits Pedestrian - All Other Actions
5	Vehicle Hits Pedestrian - Actions Not Stated
10	Entering At Angle
11	From Same Direction -Both Going Straight-Both Moving- Sideswipe
12	From Same Direction -Both Going Straight-One Stopped- Sideswipe
13	From Same Direction - Both Going Straight - Both Moving - Rear End

Additionally, we have taken in account Severity code of Damage in property and Collision involving number of injuries for people and cars. We noticed that most Damage in property involves two or more cars and persons.



From prediction perspective, we leverage one of the models which k-nearest neighbors algorithm (k-NN). However, it is practical to test the accuracy for the value of K to recognize which one would give high accuracy value.

## Results:

It has been noticed that K value of 9 was resulted as highest accuracy value. Therefore, we predict the value  $\hat{y}$  and produced 12 values out 20 in comparison to the actual values.

## Discussion:

Based on the results, incorporating additional variable to predict intended variable which Car Accident Severity, would be higher. We could also apply other machine learning supervised models.

## Conclusion:

In this review, the relationships between number of persons injured, damaged cars, type of Collision and Severity. Classification was KNN leveraged to predict Car Accident Severity