# Capstone Project - Car accident severity

1. **Business Problem:**

With increase in population, and the buying capacity, people are using more and more of private means of transport instead of public transport. This results in two huge problems which most of the countries are facing.

1. **Increase in number of accidents.**
2. Pollution

In this project, we will be focused on point “a”. Once an accident occurs, the most important goal is to minimize the loss of human lives and to make sure we decrease the probability of it happening again. The most important factor to achieve this is “**if we can predict the severity of an accident**?”

Now, for every problem we can have either **Reactive** approach or **Proactive** approach. Using reactive approach, we can analyze the historical data and make improvements on accident-prone sites. Proactive approach is if you predict the severity of an accident, and then try to prevent the accident from happening.

In this project, we will go with the Proactive approach to increase road safety. This can help many departments such as healthcare authorities, road transport authorities. Some of the benefits using this model will be :

1. Government can take proactive approach and provide ambulance near the sites where severity is very high. This will help in saving lives and reducing healthcare cost involved.
2. Road transport authorities can use this model and improve the road conditions, give warnings etc to prevent accidents.
3. Government can use this model to make better traffic management systems.
4. **Data:**

The most important requirement in solving any business/ data science problem is the **Data.** We will be using a labeled dataset provided by Coursera. Our dataset consists of 36 features, a target variable “accident-severity”, and 194673 observations. Each accident event can be identified by a unique key. The data is spanned across years 2004 to 2020.

Accident severity can have values unknown, property damage(1), injuries(2).

Some of the import features in the dataset are :

1. the address, junction type, location
2. Collision types : Different type of collisions such as angled, parked car, right turn etc
3. Features such as count of people, pedestrians, vehicles involved in the collision.
4. Features such as total injuries, serious injuries, fatalities.
5. Date, time of the accident
6. If collision caused was because of not paying attention.
7. Driver was under influence of any alcohol or drug.
8. Conditions such as weather, road, light conditions.
9. If parked car was hit or not.
10. Was the car speeding or not.

As you can clearly see, our dataset has lot of features and it contains both categorical and continuous features. We will surely be doing analysis such as correlation, PCA and pick up the most important features and which are really important for our business problem. There are some blank observations also in data, but we can handle that in our data preparation stage.