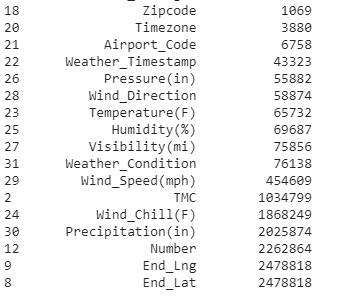
# Capstone Project - Car accident severity

Chris Han

**Introduction/Business problem**

Road accidents constitute a major problem in our societies around the world, also a significant source of death, injuries, property damage and a major cause of traffic jam. As a result, traffic safety has increasingly become a top issue today. As we all know a car accident cannot be happened single headed, there are varies of factor such as weather, driver’s conditions and ext will cause a car accident, and depend on different situation, there are different level of severity, Therefore, this study is mainly aimed to evaluate the most effective factors in severity of these accidents based on the limited data that I found online.

**Data Description**

* The dataset I chose for this study was US-Accidents: A Countrywide Traffic Accident Dataset which I found on Kaggle. This csv file covers around 3.5 million accident records from 49 states of the US, and this file has being collected continuously started from February 2016.
* This file has 49 different attributes for every single accident in record, such as temperature, humidity, and ext, some of the attributes are very useful for my study, so I believe this dataset will fit perfectly.
* For data cleaning, I printed out the missing value table, and clearly we can see that TMC, wind chill, precipitation, number, end lng and lat has a relatively large number of missing values, so for accuracy, I decided to drop those columns form the table.
* I also use the example dataset on coursera for more reference. Since I only have to use one or two attributes of it, so I did not make any data cleaning for this table.