

# Determining the Traffic Accidents in the UK.

Author: Tra Thomas  
Date: March 8, 2019

## ***What is the problem you want to solve?***

How has changing traffic flow impacted accidents? The United Kingdom, like many countries, have had a series of traffic problems and road conditions driven by several different factors like weather conditions, road construction, Rural to Urban area populations, etc. I will predict the accident rates over time. I will identify what might improve accident rates. I will determine what weather conditions produced the most accidents. I will plot interactive maps of changing time, population, speed limit and accident severity to see how they correlate and differ. I will determine the busiest roads in the UK. I will identify infrastructure needs, failings and successes. I will also determine how Rural and Urban areas differed. I will determine which areas never change and why during a given period.

## ***Who is your client and why do they care about this problem? In other words, what will your client do or decide based on your analysis that they wouldn't have done otherwise?***

My clients are the UK government, companies within the automotive industry and the UK Department of Transportation. By identifying changes in the traffic flow and how it impacts accidents, each client can use the data collected to help guide their decision making on how to prevent accidents in the area and track the percent of change in traffic flow over a given period to monitor improvements. By identifying accident rates over time, weather conditions impact on accidents, and changing time, population, speed limit, and accident severity; companies within the automotive industry will be able to use the data to identify the biggest challenges in the UK market to help improve their development of safer vehicles, self-driving vehicles, and vehicles that are adaptable in various weather conditions. The UK government and the UK Department of Transport can determine the areas within the data set where accidents occur most and determine a course of action to improve those areas.

## ***What data are you using? How will you acquire the data?***

The data used for this project comes from the UK government Open Government License on all data and the UK Department of Transportation. The data is available on [UK Gov. License](#) and [UK Traffic](#) and includes Accident type, accident severity, Date, number of casualties, day of week, time, road type, speed limit, light and weather conditions, road surface conditions, Urban or Rural area indicator, location and year. I will use information from 2005 - 2014. There are four datasets; three containing accidents data. The other dataset tracks how much traffic there was on all major roads. These datasets can be found on CSV files.

## ***Briefly outline how you'll solve this problem. Your approach may change later, but this is a good first step to get you thinking about a method and solution.***

Part of the data wrangling process will be combining the 3 data sets. After this is done, bad data (like the variables not being used from the original data set) will need to be found and removed.

To explore the data, I'll use a combination of inferential statistics and data visualization on variables like location, accident types, the amount of accidents within each area, accident severity, the date, etc. To examine each factor to solve business problems. I will code this data to display the frequency of accidents in the UK, locations and other metrics to visualize the accidents from 2000-2016. This could include mapping locations and using time series.

I'll then apply a similar approach to track the weather conditions of the accidents to determine how this metric affected the accidents occurring, as well as light conditions, road types, and junction details.

After identifying trends and significant variables, I'll use ML to build a model to predict an accident based on speed limit, road type, location, weather conditions, etc.

***What are your deliverables? Typically, this includes code, a paper, or a slide deck.***

My deliverables will be a jupyter notebook and slide deck published to my GitHub account. Note will include a report of my findings and related python code.