# Coursera Capstone Project

## Intro / business problem

Target audience: Local authorities in UK

Problem Statement: predict likelihood of severe accidents on roads on a specific day/time given the weather conditions.

This could be used to model the current risk and then the reduced risk if temporary measures are put in place (i.e. create a prediction with lower speed limits), therefore allowing the most effective measures to be put in place to prevent accidents from occurring.

## Data

Data source(s): [UK Department for Transport – Road Safety Data: 2018 dataset](https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data)

I am planning to use the following columns from the “[Road Safety Data – Accidents 2018](http://data.dft.gov.uk/road-accidents-safety-data/dftRoadSafetyData_Accidents_2018.csv)” Dataset:

* Severity of accident
* Police Force
* Road class
* Road Type
* Speed limit
* Day of week
* Time of day \*there may be crossover with lighting conditions, however this will vary throughout the year and the time of day is important to capture timing of rush hour traffic
* Date – extract month \*this may be dropped during the data exploration phase if there is high crossover with light conditions/weather
* Light Conditions
* Weather Conditions
* Road Surface Conditions (e.g. ice, standing water)
* Urban/Rural area

This data has been chosen as it includes sufficient information about the time and conditions leading to an accident along with the severity. I am interested to use the day of the week/time to see if the frequency/severity of accidents are more influenced by the day/time (i.e. rush hour on weekdays) or by the conditions of the road/weather.

I have included the police force in the dataset initially because of the target audience and use case. Originally, I was only going to differentiate by road class and urban or rural areas. However, if I find the police force has a high relation to the severity of accident during the data exploration phase I will include it in the final model.

## Methodology

### Exploratory Data Analysis

### Inferential Statistical testing

### Machine learnings used

## Results

## Discussion

## Conclusion

## Expansion Ideas

Expansion ideas:

* predict who the casualty is - the driver/passenger/pedestrian
* include some of the driver data, so that drivers could look at the personal risk of their journey allowing for the route taken & vehicle type etc

Additional info from vehicle/casualty datasets of interest:

* Junction Detail
* Age of driver
* Vehicle Type
* Age of vehicle
* Engine capacity
* Journey purpose
* Take into account drink driving & speeding
* Casualty age
* Casualty sex
* Casualty category (driver/passenger/pedestrian)
* Casualty severity