Problem Description

Traffic accidents are a common occurrence of everyday life. These can cause material losses, personal injury, emotional distress, traffic disruption and unfortunately in some cases death. In many countries and cities there is an increased amount of cars on the road. Various reasons can be attributed to this rise in traffic:

- Increased population
- Lack or unreliable public transport
- Low cost of cars
- Increased disposable incomes

Due to the increase of cars on the road and potential accidents, it is essential to understand the causes of accident severity which would be useful for various bodies such as police departments, hospitals, insurance companies, transport companies, among many others.

These stakeholders will benefit of these predictions by being able utilise their resources more efficiently. For example hospitals would be able to have staff available only in times when conditions are higher for more severe accidents and reduce staff when it is lower. This would be a similar situation for police departments.

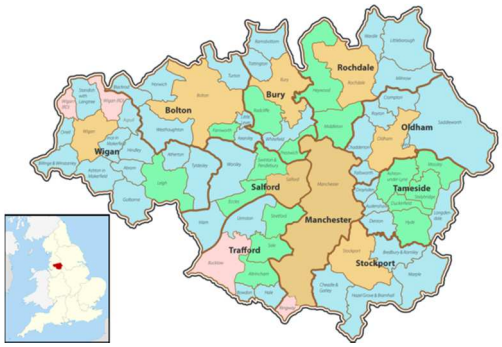
Insurance companies would be able to understand when and why high severity accident happens. For example if road condition is a large factor, they might incentivise customers to have a more stringent tyre changes or use snow tyres.

Transport companies like taxis might avoid certain days or conditions when a high number of severe accidents happen.

The objective of the capstone project is to predict the severity of a traffic accident in Greater Manchester, England. Using data science and machine learning techniques, this project will analyse accident data from 2018 to understand the factors that affect the severity of an accident.

Data Description and Approach

The data being used has been gathered by the UK Department of Transport and it includes all traffic accidents during 2018 in the Greater Manchester area. Greater Manchester is a large metropolitan area in the north of England with an approximate population of 2.8 million and includes 10 boroughs: Bolton, Bury, Oldham, Rochdale, Stockport, Tameside, Trafford, Wigan and the cities of Manchester and Salford.



The dataset is publicly available on the following site: https://data.gov.uk/dataset/cb7ae6f0-4be6-4935-9277-47e5ce24a11f/road-safety-data

The data includes more than 122,000 data points and more than 30 attributes such as:

- Accident data: location, number of vehicles involved, date
- Environmental data: light conditions, weather conditions
- Others: local authority, special conditions, police attendance

The dataset uses 3 different attributes to identify the severity of the accidents – Fatal (1), Serious (2) and Slight (3).

To solve the problem, firstly the plan is to understand the available data and the available attributes. Next, look for trends on which attributes are the most relevant to use. Afterwards, ensure the data is ready for modelling including data balancing, filling missing data and generally cleaning the dataset. Since the aim of the analysis is to predict and categorise accident severity, I will use a model using supervised machine learning techniques. Lastly, evaluate the model to ensure the business objectives are achieved using metrics such as recall, precision, and F1-score.