Accident Severity Analysis

GREATER MANCHESTER, UK

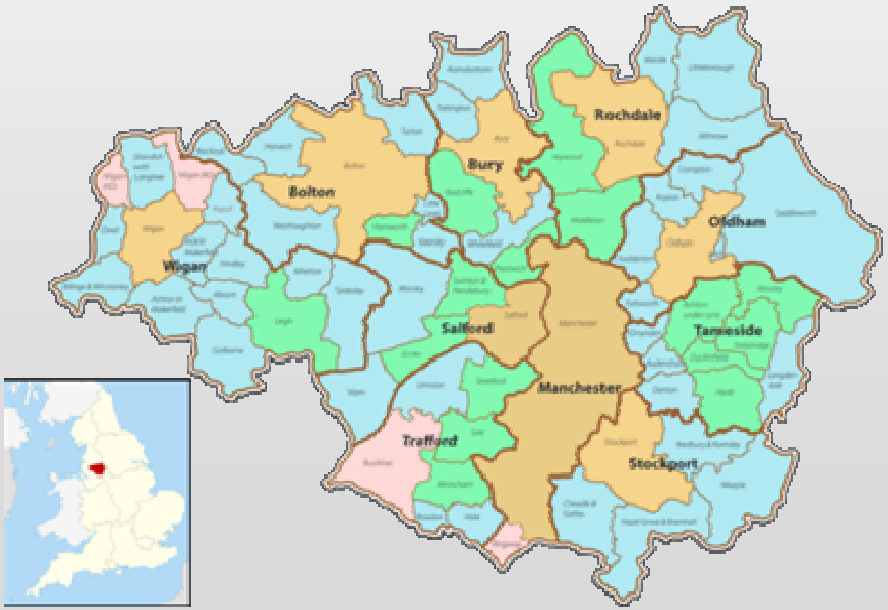
Table of Contents

Introduction

- 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.
- The objective of the capstone project is to predict the severity of a traffic accident in Greater Manchester, UK.
- 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

- 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.



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, road 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).

Methodology

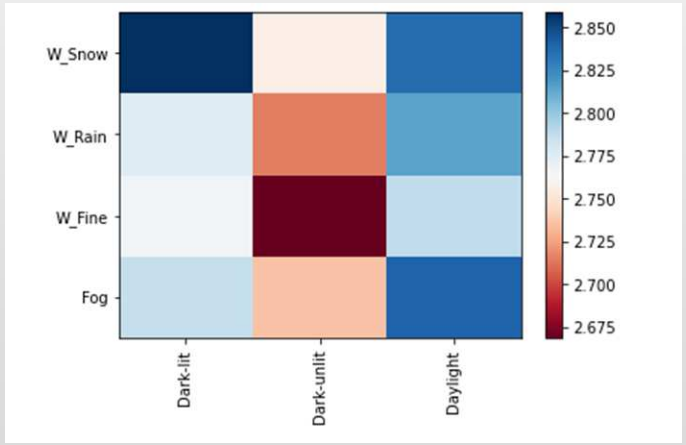
- Total data includes 32 attributes and 122,635 rows
- Data Cleaning
  - Remove unnecessary data:
    - Attributes related to the police or authority: Police_Force, Local_Authority_(District), Did_Police_Officer_Attend_Scene_of_Accident
    - Attribute too specific to be relevant to the study: Special_Conditions_at_Site, Pedestrian_Crossing-Human_Control, Location_Easting_OSGR, Location_Northing_OSGR, 1st_Road_Class, 1st_Road_Number
  - Remove any unknown, missing or other data.
  - After the clean-up total dataframe became 116,806 so less than 5% data loss.

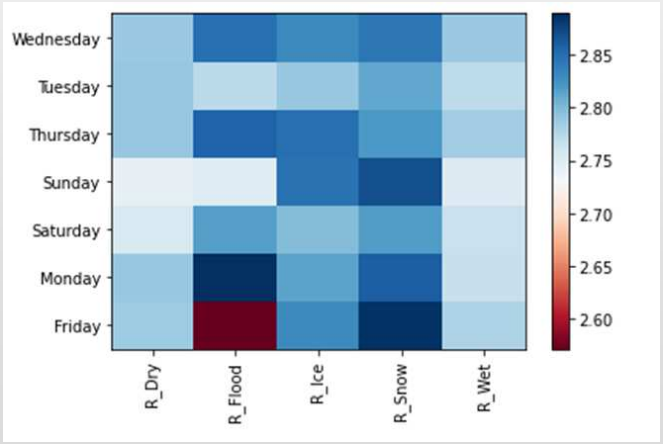
Methodology

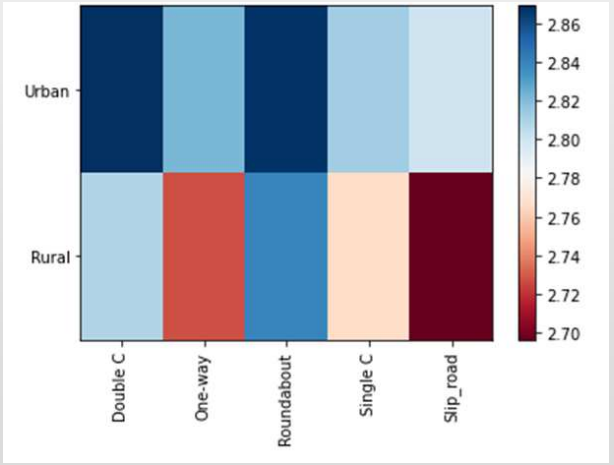
- Analysis
  - After this clean-up, the study ended up with 6 relevant attributes and the predicted one:
    - Day of Week: Monday through Sunday
    - Road Type: One way, single carriageway, double carriageway, slip road or roundabout
    - Light Conditions: Daylight, dark-lit and dark-unlit
    - Urban or Rural Area
    - Road Surface Conditions: Dry, Wet, Snow, Ice or Flood
    - Weather Conditions: Fine, Rain, Snow or Fog

Methodology

- Analysis
  - Categorical Variables
  - Correlation between the variables using heatmaps







Methodology

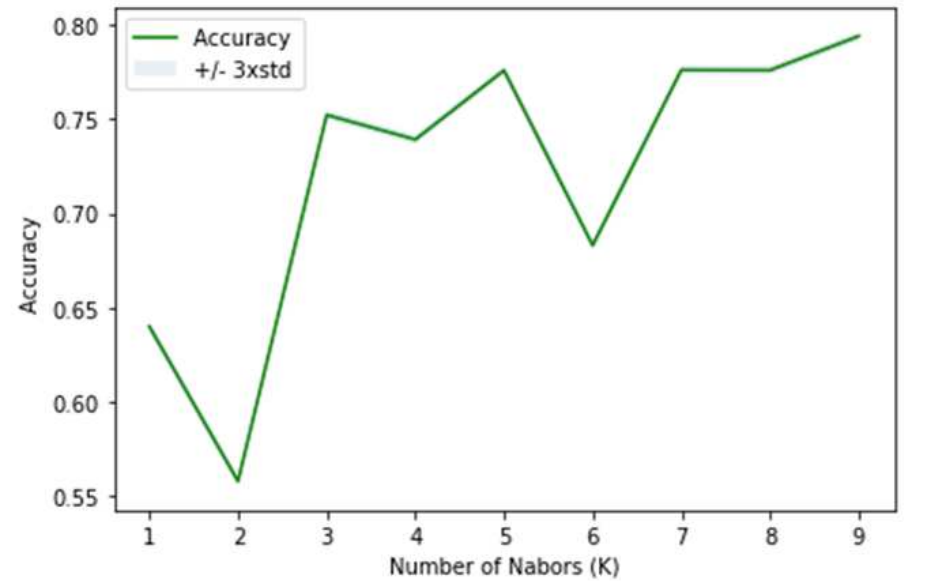
- Pearson Correlation

```
The DW Pearson Correlation Coeff is 0.0040940400006618656  with a P-value of = 0.16175105216748345
The RT Pearson Correlation Coeff is -0.046740972246208476  with a P-value of = 1.6739455486280756e-57
The LC Pearson Correlation Coeff is -0.053607268023001156  with a P-value of = 4.40769664260657e-75
The UR Pearson Correlation Coeff is -0.089915939344425 96  with a P-value of = 3.29891835955637e-208
The RS Pearson Correlation Coeff is 0.0048818699599547 48  with a P-value of = 0.09522374237112245
The WC Pearson Correlation Coeff is 0.011521538409716666  with a P-value of = 8.22347012056342e-05
```

  - Removed 2 of the variables due to a very low correlation
    - Day of Week
    - Road Surface Condition

Methodology

- Split data into Test and Train
- Prediction Modelling
  - K-Nearest Neighbour (k=9)
  - Decision Tree
  - Logistic Regression
  - Support Vector Machines
- Consistent 79% predictable results



Results

- Accidents in unlit darkness are more serious than in lit darkness. Further analysis can be made to see where these accidents are more common to possibly invest in public lighting.
- Accidents are more severe in winter due to the darkness, so maybe temporary staff might be needed in health departments.
- Accidents are more severe in wet conditions, so might be relevant for insurance companies to incentivise the purchase and use of wet tread tyres. Another option would be to review the speed limits in some of these zones.

Results

- This report and analysis would be interesting for various stakeholders in accident managementssuch as:
  - Insurance companies
  - Hospitals
  - Police departments
  - Road designers
  - Government councils

Conclusion

- In conclusion, the Capstone allowed me to go through the process of identifying a problem, accessing publicly available data and have a good understanding of it. Afterwards cleaning and preparing it for analysis and looking for relations between the variables which can give us a good prediction. Lastly looked at different modelling for accurate prediction