

## Data Acquisition

### Data Source

There is a set of datasets about car accidents and related severity for Greater Manchester for 2010 to 2019 available for public use. There are three files and a data guide downloaded from the UK's Open Data platform (<https://data.gov.uk/dataset/25170a92-0736-4090-baea-bf6add82d118/gm-road-casualty-accidents-full-stats19-data>). They refer to accident, vehicle and casualty records for Greater Manchester between 2010 and 2019. The data stored in the datasets was processed at the completion of the end of year process:

- STATS19AccData20102019.csv – contains accident records for each of the accidents. It includes severity, accident index, date of the accident, time of the accident, road condition, weather condition and other types of condition, BNG coordinate and so on.
- STATS19CasData20102019.csv – contains casualty records that linked to vehicle records. It contains accident index, year, number of casualties, casualty type, sex, age group and so on. The casualty can be either the vehicle struck the casualty or the casualty was a driver or a passenger in a vehicle.
- STATS19VehData20102019.csv – contains vehicle records that link to each accident in the accident file. It contains accident index, year, vehicle type, vehicle involved in the accident, sex of the driver, age group of the driver and so on.
- GMRoadSTATS19DataGuide.xlsx – data guide of the data files.

All features in the datasets are coded which is good for machine learning in order to predicting the severity under certain conditions. There is a form STATS19 describes the meaning of the codes.

### Licence

The datasets are licenced to use under the UK's Open Government Licence. The licence details located at <https://www.nationalarchives.gov.uk/doc/open-government-licence/version/3/>.