We wanted to link UK road accidents to weather in order to identify factors that may contribute to the likelihood of an accident occurring. This would allow insurance companies to assess risk factors that may contribute to an increase in incidents mainly dependent upon weather, but also the age band of the driver, the gender of the driver and other observable road conditions. We therefore needed data pertaining to incidents themselves, as well as weather information, from the U.K. All files are metric.

Data sets used are:

* This data set contained details of each tracked incident in the UK, including potential hazards and the casualties resulting from the incident. We used both files. <https://www.kaggle.com/tsiaras/uk-road-safety-accidents-and-vehicles?select=Accident_Information.csv>
* The Met Office Weather data includes summarized, monthly data covering high and low temperatures (in Celsius), total rainfall (mm), total sunshine (hours) and total days with air frost (af), which can predict levels of slipperiness from ice on the road. <https://www.kaggle.com/josephw20/uk-met-office-weather-data>

We created three tables in the database. The Vehicles tables contains information about the vehicle type and the details of the accident, including the gender of the driver. The Accident table includes the date, number of casualties, location and physical characteristics of the location of the incident. The Weather table contains the year, month, temperature, rainfall, sunshine and weather station location.

The regions covered in the datasets were comprehensive, so we limited our analysis to Greater London by using the below Local Authority Districts for observations in the Accident and Vehicles tables to:

* City of London
* Hammersmith and Fulham
* London Airport Heathrow
* Kensington and Chelsea
* Kingston upon Thames

The Weather table was limited to Heathrow for the local weather station.

Each table can be linked via Primary Keys. The Vehicles table can be linked to Accident table by the Accident\_Index in each one. Weather can be linked to Accident via either the Year or Month. See schema information below for more information.

Vehicles and Accident data required cleansing to remove rows that were missing information. The missing data was referenced in a uniform manner, so any rows with missing date were dropped by specifying “Data missing or out of range” to be removed. There were also a number of columns that were not used in the tables, and such, were dropped from the data.

A column was added to Weather and populated with only the month data. Additionally, Weather date data needed to be converted to integers.

The month column was added to Accident in order to provide a key between the data sets.

The time period covered was also extensive so we limited the years included to 2005 until 2017.