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 UK. All files are metric.

Data sets used are:

* UK Road Safety: Traffic Accidents and Vehicles 2005 – 2017. 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>
* UK MET Office Weather Data. 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 table 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 as this is most relevant to the districts mentioned above.

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.

Weather month and year 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 date column was in YYYY-DD-MM format and month was extracted to provide the date.

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

**Schema**

The vehicle and accident tables connect via the primary key of Accident\_Index. The Weather table connects to the accident table via year and month. This can be seen in the schema diagram below:

Graphical user interface, text, application

Description automatically generated

For each table in each category, the result types are listed below to help the user know what parameters are included in the database to help with any specific searches.

**Vehicle Table**

|  |  |
| --- | --- |
| **Category** | **Result type** |
| Accident\_Index | Unique letter number combination |
| Age\_Band\_Of\_Driver | 0-5, 16-20, 21-25, 26-35, 36-45, 46-55,56-65, 66-75, Over 75 |
| Age\_of\_Vehicle | Integer |
| Hit\_Object\_in\_Carriageway | Bridge (roof), Other object, Any animal (except ridden horse), None, Road works, Central island of roundabout, Previous accident, Parked vehicle, Bollard or refuge, Bridge (side), Open door of vehicle, Kerb |
| Hit\_Object\_off\_Carriageway | Bus stop or bus shelter, None, Lamp post, Wall or fence, Submerged in water, Tree Near/Offside crash barrier, Road sign or traffic signal, Telegraph or electricity pole, Entered ditch, Other permanent object, Central crash barrier |
| Make | Too many entries to list, contact JJA for electronic list |
| Model | Too many entries to list, contact JJA for electronic list |
| Sex\_of\_Driver | Female, Male, Not Known |
| Skidding\_and\_Overturning | Jackknifed, Jackknifed and Overturned, None, Overturned, Skidded, Skidded and Overturned |
| Towing\_and\_Articulation | Articulated, Caravan, Double or Multiple Trailer, No tow/articulation, No tow |
| Vehicle\_Manoeuvre | Changing lane to left, Changing lane to right, Going ahead left-hand bend, Going ahead other, Going ahead right-hand bend, Moving off, Overtaking- nearside, Overtaking moving vehicle – offside, Overtaking static vehicle – offside, Parked, Reversing, Slowing or stopping, Turning left, Turning right, U-turn, Waiting to go – held up, Waiting to turn left, Waiting to turn right |
| Vehicle\_Type | Goods 7.5 tonnes mgw and over, Car, Goods over 3.5t. and under 7.5t, Other vehicle, Agricultural vehicle,Motorcycle 125cc and under, Minibus (8 - 16 passenger seats), Van / Goods 3.5 tonnes mgw or under, Bus or coach (17 or more pass seats), Mobility scooter, Electric motorcycle, Motorcycle over 125cc and up to 500cc, Motorcycle over 500cc, Goods vehicle - unknown weight, Taxi/Private hire car, Motorcycle - unknown cc, Motorcycle 50cc and under |
| Was\_Vehicle\_Left\_Hand\_Drive | No, Yes |
| Year | YYYY |

**Accident Table**

|  |  |
| --- | --- |
| **Category** | **Result type** |
| Accident\_Index | Unique letter number combination |
| Date | YYYY-MM-DD |
| Day\_of\_Week | Monday, Tuesday, Wednesday, Thursday, Friday, Saturday, Sunday |
| Light\_Conditions | Daylight, Darkness - lights unlit, Darkness - lighting unknown, Darkness - no lighting, Darkness - lights lit |
| Local\_Authority\_(District) | Hammersmith and Fulham, City of London, Kensington and Chelsea, London Airport (Heathrow) |
| Number\_of\_Casualties | Integer |
| Number\_of\_Vehicles | Integer |
| Pedestrian\_Crossing-Human\_Control | Integer |
| Pedestrian\_Crossing-Physical\_Facilities | Integer |
| Road\_Surface\_Conditions | Dry, Flood over 3cm. deep, Frost or ice, Snow, Wet or damp |
| Road\_Type | Dual carriageway, One way street, Roundabout, Single carriageway, Slip road, Unknown |
| Special\_Conditions\_at\_Site | Auto signal part defective, Auto traffic signal – out, Mud, None, Oil or diesel, Road sign or marking defective or obscured, Road surface defective, Roadworks |
| Time | HH:MM |
| Year | YYYY |
| Month | MM |

**Weather Table**

|  |  |
| --- | --- |
| **Category** | **Result type** |
| Year | YYYY |
| Month | MM |
| Tmax | Float |
| Tmin | Float |
| Af | Integer |
| Rain | Float |
| Sun | Float |
| Station | Heathrow |
| Year | YYYY |
| Month | MM |

From this combined database, using the SQL query function, many different categories can be compared against each other to investigate trends. Please see below for examples of different queries that can be produced.

**Top 5 Car makes involved in the highest number of accidents from 2005 to 2017**

*SQL query:*

select v.make,count(0) from accident a inner join

vehicles v on (a."Accident\_Index" = v."Accident\_Index")

group by v.make order by 2 desc limit(5);

*Result:*

Table

Description automatically generated

**Top 5 car makes involved in the least number of accidents from 2005 - 2017**

*SQL query:*

select v.make,count(0) num\_of\_accidents from accident a inner join

vehicles v on (a."Accident\_Index" = v."Accident\_Index")

group by v.make order by 2 limit(5);

*Result*

Table

Description automatically generated

**Summary of total number of accidents for each month (2005 to 2017) with associated average temperature ranges and total rainfall**

*SQL query*

select a."Year", a."month", count(0) num\_of\_accidents, w.tmax, w.tmin,w.rain from accident a

inner join weather w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year", a."month", w.tmax, w.tmin,w.rain order by 1,2;

*Results (first 10 results displayed only)*

Table

Description automatically generated

**Summary of total number of accidents for each year (2005 to 2107) with average temperatures and total rainfall**

*SQL Query*

select a."Year", count(0) num\_of\_accidents, round(avg(w.tmax)::numeric,1) avg\_max\_temp,

round(avg(w.tmin)::numeric,1) avg\_min\_temp,

round(sum(w.rain)::numeric,1) total\_rain from accident a

inner join weather w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year" order by 1;

*Results*

Table

Description automatically generated

**Total number of accidents grouped by month for years 2005 – 2017 with average temperature and total rainfall data**

*SQL Query*

select a."month", count(0) num\_of\_accidents, round(avg(w.tmax)::numeric,1) ave\_max\_temp, round(avg(w.tmin)::numeric,1) avg\_min\_tmp,

round(avg(w.rain)::numeric,1) avg\_rain\_fall, round(sum(w.rain)::numeric,1) total\_rain from accident a

inner join weather w on (a."Year" = w."Year" and a."month" = w."month")

group by a."month" order by 1;

*Results*

Table

Description automatically generated

**Top 5 month and year combination with the most accidents with temperature and rainfall information**

*SQL Query*

select a."Year", a."month", count(0), w.tmax, w.tmin,w.rain from accident a

inner join weather w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year", a."month", w.tmax, w.tmin,w.rain order by 3 desc limit(5);

*Results*

Table

Description automatically generated

**Top 5 month and year combination with the least accidents with temperature and rainfall information**

*SQL Query*

select a."Year", a."month", count(0), w.tmax, w.tmin,w.rain from accident a

inner join weather w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year", a."month", w.tmax, w.tmin,w.rain order by 3 limit(5);

*Results*

Table

Description automatically generated

**Month and Year with the greatest rain fall**

*SQL Query*

with greatest\_rainfall\_month\_year as

(select w.\* from weather w where rain in

(select max(w.rain)

from weather w inner join accident a

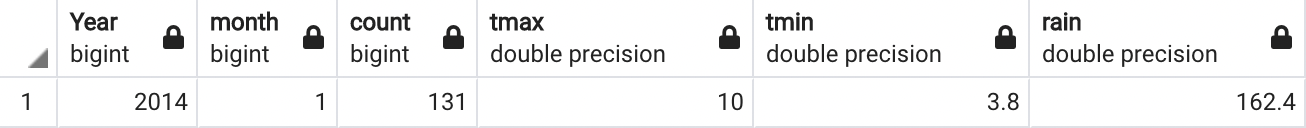
on (a."Year" = w."Year" and a."month" = w."month")))

select a."Year", a."month", count(0), w.tmax, w.tmin,w.rain from accident a

inner join greatest\_rainfall\_month\_year w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year", a."month", w.tmax, w.tmin,w.rain;

*Results*



**Month and Year with the lowest temperature**

*SQL Query*

with minimum\_temp\_month\_year as

(select w.\* from weather w where w.tmin in

(select min(w.tmin)

from weather w inner join accident a

on (a."Year" = w."Year" and a."month" = w."month")))

select a."Year", a."month", count(0), w.tmax, w.tmin,w.rain from accident a

inner join minimum\_temp\_month\_year w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year", a."month", w.tmax, w.tmin,w.rain;

*Results*



**Month and Year with the least rainfall**

*SQL Query*

with least\_rainfall\_month\_year as

(select w.\* from weather w where rain in

(select min(w.rain)

from weather w inner join accident a

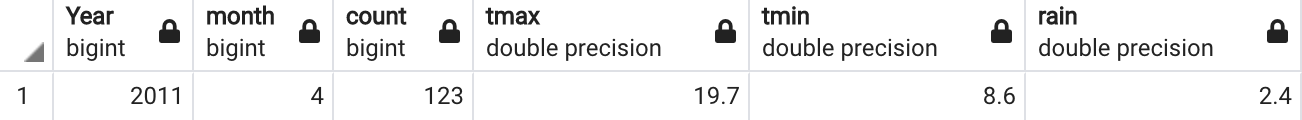
on (a."Year" = w."Year" and a."month" = w."month")))

select a."Year", a."month", count(0), w.tmax, w.tmin,w.rain from accident a

inner join least\_rainfall\_month\_year w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year", a."month", w.tmax, w.tmin,w.rain;

*Results*



**Month and Year with the maximum temperature**

*SQL Query*

with maximum\_temp\_month\_year as

(select w.\* from weather w where w.tmax in

(select max(w.tmax)

from weather w inner join accident a

on (a."Year" = w."Year" and a."month" = w."month")))

select a."Year", a."month", count(0), w.tmax, w.tmin,w.rain from accident a

inner join maximum\_temp\_month\_year w on (a."Year" = w."Year" and a."month" = w."month")

group by a."Year", a."month", w.tmax, w.tmin,w.rain;

*Results*



**Sex of driver vs number of accidents**

*SQL Query*

select v."Sex\_of\_Driver",count(0) num\_of\_accidents from accident a inner join

vehicles v on (a."Accident\_Index" = v."Accident\_Index")

group by v."Sex\_of\_Driver";

*Results*

Table

Description automatically generated

**Age Band of Driver vs number of accidents**

*SQL Query*

select v."Age\_Band\_of\_Driver",count(0) num\_of\_accidents from accident a inner join

vehicles v on (a."Accident\_Index" = v."Accident\_Index")

group by v."Age\_Band\_of\_Driver";

*Results*

Table

Description automatically generated

**Weather season vs number of accidents**

*SQL Query*

select season,count(0) number\_of\_accidents from (

select case when month in (12,1,2) then 'Winter'

when month between 3 and 5 then 'Spring'

when month between 6 and 8 then 'Summer'

when month between 9 and 11 then 'Autumn'

end season

from accident a) a group by season order by season;

*Results*

Table

Description automatically generated

**Day\_of\_Week vs number of accidents**

*SQL Query*

select a."Day\_of\_Week",count(0) num\_of\_accidents from accident a

group by a."Day\_of\_Week" order by 2 desc;

*Results*

Table

Description automatically generated