1. **Data completeness and quality validation *(Performed in SQL and delivered in a .sql file)***

Used Sqlite3.

* Instead of downloading sqlite3 to the local system used the sqlite3 library in Python and performed the task in Google Collab.
* Created a database named ‘uk\_accidents.db’ in sqlite3 server. Added a table ‘uk\_accidents\_table’ and loaded it from a panda’s data frame.
* Primarily used PRAGMA table\_info to extract column information and checked for data types, missing values, etc.
* Also used a query to find duplicates based on Accident\_Index.

1. ***Data analysis (Performed in Python and delivered in a .html file)***

Performed a data analysis using pandas and plotted graphs with matplotlib, seaborn, and Folium.

* First loaded the data and concatenated the files to create a single dataframe.
* Data analysis is done by checking data types, length, various statistical values, etc
* Then removed duplicate records and also columns with a very high number of missing values. Then used data imputation in some categorical columns using mode().
* First created a pair plot with sns so the relationship between every column can be visualized.
* Then visualized the data using, barplot, pi plot, histogram, scatterplot, etc.

1. ***Interactive Visualisation (Performed in tool of choice, format should be readable independent of licenses)***

Microsoft PowerBI Desktop is used to create an interactive dashboard of the UK\_Accidents data.

Included 5 different graphs which can be changed by selecting the year and day of the week.