

Operational guidelines

Below are the phases of the project in chronological order with the related files used:

1. Acquisition and Preprocessing phase:

This part was developed through the creation of the notebook "Acquisition and Preprocessing of the kaggle dataset.ipynb".

The Python code receives as input the datasets present in the "Initial Dataset" folder And it returns the Kent_2022_data.csv dataset in the "Acquisition Dataset" folder.

2. Use of API:

It was performed through the "API Phase - Retrieving weather conditions data.ipynb" notebook containing the code relating to the use of the API which allowed us to obtain the twelve datasets within the "API Datasets" folder.

3. Integration phase:

The notebook for this step is "API merge and Integration.ipynb".

At the end of the process .csv files are created inside the RDBMS folder which were later used to create the Accidents.db database with the help of the software "DB Browser(SQLite)". Inside the folder there is also the project file relating to the creation.

4. Data Quality:

It was carried out using the notebooks present in the "Data Quality and Profiling" folder which return the metrics and graphs relating to this phase.