January 2025

*Data Cleaning in SQLite*

*Objective 1: Clean the tables order\_details\_dirty, orders\_dirty and pizzas\_dirty.*

Step 1: Open DB Browser for SQLite.

Step 2: ‘New Database’

Step 3: Create a file name for the database and save it to a preferred location.

Step 4: Close out the window to create a table.

Step 5: Import a table to clean via File > Import > Table from CSV file…

Step 6: Navigate to the location where the csv file is stored.

Step 7: Open the csv file.

Step 8: Ensure that ‘Column names in first line’ is checked.

Step 9: Click OK.

Tables will appear in ‘Database Structure’ under ‘Tables’.

Step 10: Right click on a table title to ‘Modify Table’.

Step 11: For each table in the database, under ‘Fields’, ensure that the Name and Type (data type) are correct. Further, after understanding the relationships between tables, designate certain fields as PK (primary key) and (by scrolling to the right) Foreign Key.

Step 12: In ‘Execute SQL’, use one tab to clean one table at a time.

Note: It is reasonable to switch around Steps 11 and 12.

*Result: 3 separate tabs with clean tables.*

The pizza\_types table was already clean as per the data source.

*Objective 2: Create a view to combine all four tables.*

Step 1: In ‘Execute SQL’, use one tab to create a view that joins the order\_details and orders tables. Do the same in another tab to join pizzas and pizza\_types.

Step 2: Ensure that the queries were executed correctly by selecting the Table under ‘Browse Data’ to examine the full table’s records.

Step 3: Create a total view in another tab to join the two views.

Step 4: Under ‘Database Structure’, right click on the total view and Export to CSV file to a preferred location.

Step 5: Open a blank report in Power BI.

Step 6: Get data (choose the Text/CSV option) from the location on the hard drive.

The original dirty dataset can be found at the bottom of this site: [link](https://www.wiley.com/en-us/How+to+Become+a+Data+Analyst%3A+My+Low-Cost%2C+No+Code+Roadmap+for+Breaking+into+Tech-p-9781394202232)

The clean dataset can be found here: [link](https://mavenanalytics.io/project/1412)