



TASK

Data Analysis – Data Cleaning

Model-Answer Approach

Visit our website

Auto-graded Task 1

The approach uses various data-cleaning techniques to standardise the country names and date formats within the DataFrame. It utilises lowercase conversion, whitespace removal, and fuzzy matching for country name standardisation, ensuring data consistency. The model answer uses custom minimum ratios for the fuzzy matching, because the minimum ratio of 90 (used in the example notebook provided) is not always the best fit for all cases. Some of the **strings_to_match** values require a lower or higher ratio to be matched correctly. The solution is fitting for the task because it allows for the standardisation of the country names, even when the country names are hard to match. Common pitfalls that students might encounter are not considering that the minimum ratio of 90 is not always the best fit for all cases.