

## Case Study: Data Cleaning & Analysis for Antibiotics

### Objective:

- To properly clean the data regarding the inconsistency in Level 5: active constituents.
- You can use any tools you choose to do some of the necessary analysis.

### Background:

You are given a Word and Excel document named “ProjectLyft\_Casestudy” in this project. The Word document contains the instructions and questions, while the Excel file includes the dirty data.

While working with the Excel file, you will be expected to pay attention to detail, especially in the Level 5: Active Constituent column. This column contains various misspelt representations of various active constituents. Cleaning must be done to achieve the correct results. Next, you will be expected to combine the three tables by using lookup functions or joining the worksheets using the columns “Brand,” as it is common on the three tables. After cleaning and joining the worksheets together, you can then proceed to answer the questions below.

### Core Deliverables:

Submit a Word document with the answers to the expected questions and an Excel file containing the cleaned data. The files must be named accordingly.

**CaseStudy\_firstname\_lastname** where “firstname” and “lastname” denote your first name and surname, respectively.

### Questions:

1. Which ATC level 5 has the highest sales value market share in 2021?
2. Which ATC level 5 has the lowest sales volume market share?
3. Which are the top 10 brands with the highest % sales volume market share?
4. What is the total sales volume of the Active Constituents J01AA02 Doxycycline in 2020?
5. What is the total sales value of the Active Constituents J01CA04 Amoxicillin in 2021?
6. Which Brand has the highest Sales volume in the year 2021?
7. Which Brand has the lowest Sales value in the year 2022?

8. Which Level 4 Code: Drug Class has the highest Sales value across the four years?
9. Which Level 3 Code: Pharmacological Segment has the highest Sales volume in 2020?
10. What tool did you use for this data cleaning and analysis?

### **Grading**

You will be graded on the clean Excel sheet, and the answers will be supplied in a Word document.