**MEL 2024 IRS SEASONAL STAFF PRACTICAL INTERVIEW**

**Instructions:**

1. **About the dataset:**

The dataset uses **IRS-Number** column as the unique ID. If **Sprayed** column is 0, then we can allow **IRS-Number** to duplicate hence **Re-visit** column should have a value of 1. If **Sprayed** column value is 0, the **Reason for No Spray** column must have a value. **SOP Code** is a unique ID of a data collector.

1. **About the test:**

Use the dataset provided in a csv file to answer the questions below. Save the provided csv dataset file as an Excel workbook using the **candidate code** provided (**DO NOT WRITE YOUR NAME ANYWHERE**). You must use the correct code, otherwise, your workings will be misplaced. The first sheet to hold the dataset provided, name this sheet “**Dataset**”. Answer each question below on the specific sheets within the workbook created. Label each question’s sheet using the corresponding question numbers (alphabets a – j).

**The test is strictly 2 hours, starting from 8:00 am to 10:00 am. Any test sent beyond 2 hours will be discarded.**

**Questions:**

1. Identify data issues in the dataset that make it untidy. [2 marks]
2. How would you clean the identified data issues in **a.** above? [2 marks]
3. Extract and summarize in a table the number of duplicate records by sub-county by county. [5 marks]
4. Use the most suitable chart in Excel to represent the status of spray by sub-county per county. {**Hint**: spay status is recorded in the "**Sprayed**" column} [3 marks]
5. Generate a summary showing the percentage of the number sleeping under nets against the total population by sub-county per county. [5 marks]
6. Generate a summary showing the percentage of rooms sprayed against the rooms found by sub-county per county. [4 marks]
7. Use the most suitable chart in Excel to represent the reason for no spray by sub-county per county. [4 marks]
8. Generate a list showing the number of events (records) per SOP Code per day grouped by sub-county per county. [4 marks]
9. What factors could lead to the varying number of events (records) per SOP Code per day as observed in **h.** above? [3 marks]
10. What verification criteria would you put in place to ensure that the data collected is of good quality. [3 marks]