Project

Module
Data Processing

Module Code
DE43003FP

Duration
15 hours

Title:

Apply data integration and transformation techniques to process a given dataset for visualisation

Tools, Equipment and Materials:

- Personal Computer with Internet access
- Data Processing Software

Instructions & Guidelines:

1. Submissions should be made via the myConnexion by the stated deadline.

2. No marks will be awarded, if the work is copied or you have allowed others to copy your work.

   Submission:

3. Submission should include the following items:

   - Microsoft Excel file / PowerBI file

Project

Module
Data Processing

Module Code
DE43003FP

Duration
15 hours

Task
Requirements

Dataset:
A research study has been conducted on the Tree Seedling and its survival base on the sunlight availability.
Your task is to perform simple data processing to prepare the dataset for analysis.

Using Microsoft Excel to perform these tasks below:

Assign data types

1. Assign the correct data types for “Species”, “Time” “Phenolics” and “Lignin” columns.

Perform data manipulation columns into tables

2. Apply conditional formatting to highlight the "Time" column cells with a value of 1, indicating a seedling's death, using a different color.

3. Using the "Alive" field, create a new column called "Survival_Status" that indicates "Survived" if the seedling is alive ("X" in the "Alive" field) and "Dead" if not.

4. Create a new column called "Time_Category" using nested IF statements to categorize the "Time" variable into three groups: "Early," "Mid," and "Late".

Replace row values in the dataset.

5. Display a new column named using VLOOKUP to match the "Adult" column with the corresponding species in the "Species" column. Use the given Lookup Table.

Identify variables for aggregation

6. Create a second pivot table to calculate the sum values of “Phenolics” and “Lignin”
another

7. Build a ~~second~~ pivot table to average "NSC" values for each species.

Project

Module

Data Processing

Perform combining data into required tables

8. In another spreadsheet, create a new table with these following columns:
a) Plot
b) PlantDate
c) Species
d) Adult_Species
e) Survived_Status

survival

Within the table, it includes seedlings with "Survived_Status" categorized as "Survived".

Create visualisation for the dataset

9. Create a pivot table vertical bar chart to show the count of plot compared against species.

10. Create a pivot table vertical bar chart to show the soil compared against count of harvest.

Complete a workflow diagram

11. Review and complete the workflow process in the dataset.

~Thank You~

Module Code

DE43003FP

Duration

15 hours