

Data Cleaning



Data cleaning is the process of identifying, correcting, and removing errors, inconsistencies, and inaccuracies in datasets. It involves a series of steps such as data validation, data transformation, data normalization, and data enrichment to ensure that the data is accurate, complete, and consistent. The purpose of data cleaning is to improve the quality and reliability of data, making it more suitable for analysis, reporting, and decision-making. Data cleaning is a critical step in data preprocessing and is often performed before data analysis or machine learning algorithms are applied to the data. **Record the entire tasks in a macro and save the excel file.**

Assignment-3

Task 1

Where **Name** is missing fill the value in place by extracting the name from the **Email-id** column.

Task 2

Fill the Missing **Age** value with the **mean** of the **Age** column.

Task 3

Clean and Format the **Phone number** column and it should show like **(91-xxxxxxx)**

Task 4

Where **Name** and **Email-Id** is missing the delete the entire row.

Task 5

Highlight the **Top 10 Employees** with highest salary in the dataset.

Task 6

Where **Name** is there but no **Email-Id** then fill the blank with **(support@unmessenger.com)**

Task 7

In the **Department column** make the data proper with the useful functions and rename the wrong spelled words in the department column.

Task 8

Where **Address** is not there fill it with **(Address not Available)**

Submission: The Entire assignment should be submitted by (Sunday, 11/02/2024). You can save the excel file as Macro enabled Excel file. Upload the excel file in a repository in your GitHub Account.