EX2- PERFORM EXPLORATORY DATA ANALYSIS (EDA) WITH DATASETS LIKE EMAIL DATA SET

**✅ DATASET LINK: Email Spam Detection Dataset**

You can use the classic email spam dataset from UCI:

**📥DatasetURL:**  
<https://www.kaggle.com/datasets/uciml/sms-spam-collection-dataset>

**✅ LAB EXERCISE ALGORITHM (Step-by-Step EDA)**

Algorithm: EDA on Email (SMS) Spam Dataset

1. Import necessary libraries
2. Load the dataset
3. Understand dataset structure (head, info, describe)
4. Check for null/missing values
5. Analyze class distribution (spam vs ham)
6. Visualize data distribution using plots
7. Perform text-specific analysis
   * Message length
   * Word frequency
   * Word clouds for spam and ham
8. Check correlations (if any)
9. Conclude insights from data

**📘 Expected Output:**

* Distribution plot of spam vs ham
* Histogram of message lengths
* Word clouds highlighting most frequent words in spam and ham
* Statistical insights: average message length, outliers