# Weekly Report (Week 1–3)

Course: Applied Data Science with AI

Semester: BSSE 7th

Week #: 1–3 Combined Progress

Student Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Roll Number: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Project Title: Image-based Waste Classification

## 1. Reading Summary

Week 1:  
- Installed Python environment (Google Colab).  
- Learned basics of Jupyter/Colab notebooks.  
- Understood course structure and selected project (based on roll no = 9 → Waste Classification).

Week 2:  
- Studied Pandas and NumPy basics for data cleaning.  
- Learned about handling missing values, duplicates, and outliers.

Week 3:  
- Studied visualization libraries (Matplotlib & Seaborn).  
- Learned how to plot histograms, bar charts, and show images with labels.

Key Learnings:  
1. How to set up Colab + GitHub workflow.  
2. How to prepare and clean raw datasets before modeling.  
3. How visualization helps in understanding dataset patterns.

Reflection:  
These readings directly connect with my project because cleaning images and checking distribution is critical for training a correct waste classification model.

## 2. Classroom Task Documentation

- Loaded dataset from Kaggle into Colab.  
- Organized into TRAIN and TEST folders.  
- Created bar charts of class distribution.  
- Displayed sample images from each class (O = Organic, R = Recyclable).  
  
📸 (Add screenshots of Colab outputs: bar chart, sample images)

## 3. Weekly Assignment Submission

Assignment Title: Data Cleaning + Visualization of Waste Dataset

Steps Taken:  
1. Uploaded Kaggle dataset (archive.zip) into Colab via Google Drive.  
2. Unzipped and explored folder structure.  
3. Applied duplicate detection and removed 326 duplicate images.  
4. Resized all images to 224x224 and saved cleaned dataset.  
5. Plotted class distribution graph and displayed random samples from each class.

Output:  
- Clean dataset: /content/processed\_data/  
- Classes: O (Organic), R (Recyclable)  
- Visualization showed training set slightly imbalanced but usable.

Challenges Faced:  
- Understanding dataset folder structure (nested DATASET → TRAIN/TEST).  
- Handling large zip file upload on Colab.  
- Confusion about class names (O and R) — later clarified.

GitHub Link: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## 4. Project Progress Milestone

Week 1: Project selected + dataset downloaded.  
Week 2: Dataset cleaned (resized, duplicates removed).  
Week 3: EDA completed (class distribution, sample images).

Next Week’s Goal:  
Perform basic statistics and correlation analysis on dataset features.

## 5. Self-Evaluation

☑️ I completed all tasks on time.

## 6. Questions for Instructor

- Is it okay to use image-based statistics (like pixel intensity) for correlation in Week 4?  
- Should we balance the dataset further if classes are unequal?