

```

# Crop Data Cleaning without external libraries
# Works in Programiz Online Compiler (no pandas or sklearn)

# -----
# SAMPLE DATA (You can replace this with your own)
# -----
data = [
    {"Crop": "Wheat", "Rainfall": 500, "Temperature": 25, "Yield": 3.2},
    {"Crop": "Rice", "Rainfall": None, "Temperature": 30, "Yield": 4.0},
    {"Crop": "Maize", "Rainfall": 400, "Temperature": 28, "Yield": None},
    {"Crop": "Wheat", "Rainfall": 500, "Temperature": 25, "Yield": 3.2}, # duplicate
    {"Crop": "Barley", "Rainfall": 350, "Temperature": 20, "Yield": 2.5}
]

# -----
# STEP 1: REMOVE DUPLICATES
# -----
unique_data = []
for row in data:
    if row not in unique_data:
        unique_data.append(row)

# -----
# STEP 2: HANDLE MISSING VALUES
# Replace None with average (for numeric) or "Unknown" (for text)
# -----
def average(column):
    values = [row[column] for row in unique_data if isinstance(row[column], (int, float))]
    return sum(values) / len(values) if values else 0

avg_rainfall = average("Rainfall")
avg_yield = average("Yield")

for row in unique_data:
    if row["Rainfall"] is None:
        row["Rainfall"] = avg_rainfall
    if row["Yield"] is None:
        row["Yield"] = avg_yield
    if row["Crop"] is None:
        row["Crop"] = "Unknown"

# -----
# STEP 3: NORMALIZE NUMERIC COLUMNS (0–1 range)
# -----

```

```
def normalize(column):
    values = [row[column] for row in unique_data]
    min_val, max_val = min(values), max(values)
    for row in unique_data:
        row[column] = (row[column] - min_val) / (max_val - min_val) if max_val != min_val else 0
```

```
normalize("Rainfall")
normalize("Temperature")
normalize("Yield")
```

```
# -----
# STEP 4: ENCODE CROP NAMES INTO NUMBERS
# -----
```

```
crop_map = {}
counter = 0
for row in unique_data:
    crop = row["Crop"]
    if crop not in crop_map:
        crop_map[crop] = counter
        counter += 1
    row["Crop"] = crop_map[crop]
```

```
# -----
# STEP 5: DISPLAY CLEANED DATA
# -----
```

```
print("✅ Cleaned Crop Data:\n")
for row in unique_data:
    print(row)
```