**Day 8 — 30 June 2025 (Monday)**

**Topic: Advanced Pandas — Data Cleaning & Transformation**

Today’s session focused on *advanced data preprocessing techniques* using Pandas. We explored dataset inspection in depth — identifying null values, inconsistent formatting, special characters, and duplicate entries. The trainer demonstrated how raw datasets from emails require cleansing before passing to ML models. We learned column manipulation, label encoding basics, and data export. A cleaned dataset improves model accuracy and ensures reliable training. This session strengthened my ability to prepare raw email text data before feeding it into ML pipelines.

**Key Code Practiced:**

import pandas as pd

df = pd.read\_csv("emails.csv")

df.dropna(inplace=True) # remove null entries

df.drop\_duplicates(inplace=True)

df['label'] = df['label'].map({'ham':0, 'spam':1}) # convert text labels to numbers

print(df.head())