Summary of main.ipynb (Tabular ML on Segmentation Features)

Purpose

This notebook performs a classical machine-learning analysis using quantitative features derived from FSL FAST segmentation files (the *_fseg.txt outputs) in the OASIS-1 dataset. The goal is to classify subjects (e.g., Control vs Demented) using simple, interpretable features rather than deep learning.

What Was Done
---
1) Data parsing – Located segmentation summary files (*_fseg.txt) for each subject and parsed CSF, GM, and WM volumes.
2) Feature engineering – Built a tabular dataset per subject including CSF, GM, WM, GM/WM ratio, and total brain volume; then merged with metadata labels.
3) Modeling – Trained classical ML models (Logistic Regression, SVM, Random Forest), with class balancing and repeated stratified cross-validation for stability.
4) Evaluation – Reported Accuracy and ROC-AUC; plotted confusion matrices.

Outcome / Results

• A clean tabular feature matrix per subject based on tissue volumes (no MRI pixels used here).
• Stable performance estimates via repeated stratified cross-validation.
• Clear baseline showing how far simple global volumetrics can go on this dataset.