

## Data Cleaning and Preprocessing

### SUMMARY

To clean and prepare the **Medical Appointment No-Shows** dataset by handling null values, duplicates, text inconsistencies, and encoding issues using **Microsoft Excel**.

---

#### Cleaning Steps (Using Excel)

- Checked and handled missing values using filters and "Go To Special → Blanks".
  - Removed duplicate entries using the **Remove Duplicates** feature.
  - Fixed encoding issues in the Neighbourhood column (e.g., SÃfO → SÃO) using manual correction and UTF-8 interpretation.
  - Corrected typos and formatting issues in neighbourhood names.
  - Trimmed extra spaces and standardized text fields using TRIM(), UPPER(), and PROPER().
  - Converted date columns (ScheduledDay, AppointmentDay) to a uniform dd-mm-yyyy format.
  - Checked and fixed data types (e.g., Age as number, binary values for medical indicators).
  - Renamed column headers for consistency (e.g., No-show → no\_show).
-