

# PAN Number Validation Project

## Objective

Clean and validate a dataset containing the Permanent Account Numbers (PAN) of Indian nationals. The goal is to ensure every PAN number follows the official format and to categorize each as **Valid** or **Invalid**. The dataset is provided in a separate Excel file: **PAN Number Validation Dataset.xlsx**.

---

## Instructions

### 1. Data Cleaning and Pre-processing

- **Handle Missing Data:** Identify PAN numbers with missing values and decide whether to remove the rows or impute values, depending on the context.
  - **Remove Duplicates:** Check for duplicate PAN numbers and delete any duplicates.
  - **Trim Spaces:** Remove any leading or trailing spaces from the PAN numbers.
  - **Correct Letter Case:** Ensure that all PAN numbers are in uppercase letters.
- 

### 2. PAN Format Validation

A valid PAN number must meet all of the following conditions:

1. **Length:** Exactly 10 characters.
2. **Pattern:** Format is AAAAA1234A
  - First **five** characters: Uppercase alphabets.
  - Next **four** characters: Digits.
  - Last **one** character: Uppercase alphabet.
3. **Alphabet Rules:**
  - No two adjacent alphabetic characters can be the same (e.g., **AABCD** is invalid; **AXBCD** is valid).
  - The first five alphabets cannot form a continuous sequence (e.g., **ABCDE**, **BCDEF** are invalid; **ABCDX** is valid).
4. **Numeric Rules:**
  - No two adjacent digits can be the same (e.g., **1123** is invalid; **1923** is valid).
  - The four digits cannot form a continuous sequence (e.g., **1234**, **2345** are invalid).

**Example of a valid PAN:** AHGVE1276F

---

### 3. Categorization

Classify each PAN number as:

- **Valid PAN:** Meets all the above criteria.
  - **Invalid PAN:** Fails to meet the format, is incomplete, or contains non-alphanumeric characters.
- 

#### 4. Summary Report

Provide a final report containing:

- Total records processed.
- Total valid PANs.
- Total invalid PANs.
- Total missing or incomplete PANs (if applicable).