**PAN Number Validation Project**  
**Data Cleaning and Validation**

**Objective**  
The task involves cleaning and validating a dataset of Permanent Account Numbers (PAN) for Indian nationals. The aim is to ensure that every PAN number conforms to the official format and is classified as either **Valid** or **Invalid**. The dataset is provided in a separate Excel file.

**Instructions**

**1. Data Cleaning and Preprocessing**

* **Identify and handle missing data:** PAN numbers may have missing values. Handle them appropriately, either by removing the rows or imputing values, depending on the context.
* **Check for duplicates:** Ensure there are no duplicate PAN numbers. Remove any duplicates if found.
* **Remove leading/trailing spaces:** Eliminate any extra spaces before or after the PAN numbers.
* **Correct letter case:** Ensure all PAN numbers are in uppercase. Convert any lowercase letters to uppercase.

**2. PAN Format Validation**  
A valid PAN number must meet the following criteria:

* It must be exactly **10 characters long**.
* The format should follow: **AAAAA1234A**
* The first **five characters must be uppercase letters**.

**PAN Format Rules**

**1. Alphabetic Part (First 5 Characters)**

* Adjacent letters cannot be the same (e.g., AABCD is invalid, AXBCD is valid).
* All five letters cannot form a sequential series (e.g., ABCDE, BCDEF are invalid, ABCDX is valid).

**2. Numeric Part (Next 4 Characters)**

* All four characters must be digits.
* Adjacent digits cannot be the same (e.g., 1123 is invalid, 1923 is valid).
* All four digits cannot form a sequence (e.g., 1234, 2345 are invalid).

**3. Last Character**

* Must be an uppercase letter.

**Example of a Valid PAN:** AHGVE1276F

**Categorisation**

* **Valid PAN:** Matches all the above format rules.
* **Invalid PAN:** Does not match the correct format, is incomplete, or contains any non-alphanumeric characters.

**Tasks**

1. Validate all PAN numbers based on the rules above.
2. Categorize PAN numbers into two groups:
   * Valid PAN
   * Invalid PAN
3. Create a summary report including:
   * Total records processed
   * Total valid PANs
   * Total invalid PANs
   * Total missing or incomplete PANs (if any)

NOTE

You can use **either SQL or Python** to perform the data cleaning and PAN validation tasks for this project.