**Project: SSN and Driver’s License Number Validation**

**Dataset**

* Total rows: ~1000–1200 (small enough to practice but not overwhelming).
* Columns (dirty with missing values, spaces, duplicates, wrong casing, etc.):
  1. **SSN** (Social Security Number)
  2. **DLN** (Driver’s License Number)

**Data Cleaning & Preprocessing**

1. **Missing Data**
   * Some rows will have missing SSN or DLN.
   * You’ll decide whether to drop or impute (for practice, drop is fine).
2. **Duplicates**
   * Introduce a few duplicate SSNs and DLNs → remove them.
3. **Spaces & Case**
   * Some values padded with leading/trailing spaces.
   * Convert everything to proper format (SSN as XXX-XX-XXXX, DLN uppercase).

**Validation Rules**

**Social Security Number (SSN)**

Format: AAA-GG-SSSS

* Exactly 9 digits (with or without dashes).
* Cannot start with 000, 666, or 9.
* Group GG cannot be 00.
* Serial SSSS cannot be 0000.
* Example Valid: 123-45-6789
* Example Invalid: 000-12-3456, 666-45-6789, 12345678.

**Driver’s License Number (DLN) (take New York style for practice)**

Format: 1 letter + 7 digits (e.g., A1234567)

* Exactly 8 characters.
* First character must be alphabetic (A–Z).
* Next 7 characters must be numeric.
* Example Valid: K1234567
* Example Invalid: 12345678, AA234567.

**Categorisation**

* **Valid** if all format rules are satisfied.
* **Invalid** if format broken, incomplete, or contains special characters.

**Summary Report**

* Total valid SSNs
* Total invalid SSNs
* Total valid DLNs
* Total invalid DLNs
* Total missing/incomplete records
* Write the validation results and missing or incomplete records to different sheets of single Excel file