**Objective**

The goal of this assessment is to evaluate your ability to develop an **SSIS package** that extracts, transforms, and loads (ETL) complaint data from a CSV file into a **SQL Server database**, ensuring data quality, consistency, and integrity throughout the process.

You are expected to demonstrate proficiency in:

* File handling
* Data transformation (including validation and cleaning)
* Error handling
* Loading data into SQL Server tables
* Following best practices in SSIS design

**Source Data**

You will be provided with a **CSV file** containing complaint data from the Consumer Financial Protection Bureau (CFPB). The dataset includes the following fields:

| **Field Name** | **Data Type** | **Notes** |
| --- | --- | --- |
| Date received | DateTime | Format: MM/DD/YY |
| Product | Text | Categorical |
| Sub-product | Text | Categorical, optional |
| Issue | Text | Categorical |
| Sub-issue | Text | Categorical, optional |
| Consumer complaint narrative | Text | Optional, large text |
| Company public response | Text | Optional |
| Company | Text | Categorical |
| State | Text | US state abbreviation |
| ZIP code | Text | 5-digit, may include leading 0 |
| Tags | Text | Optional |
| Consumer consent provided? | Text | Optional |
| Submitted via | Text | Categorical |
| Date sent to company | DateTime | Format: MM/DD/YY |
| Company response to consumer | Text | Categorical |
| Timely response? | Text | yes/no |
| Consumer disputed? | Text | yes/no/N/A |
| Complaint ID | Integer | Unique key |

**Task Requirements**

**1. Create SQL Server Destination Table**

* Name the table: CFPB\_Complaints
* Define Complaint ID as the **primary key**
* Use appropriate data types (e.g., NVARCHAR, DATETIME, INT, VARCHAR)

**2. Develop SSIS Package**

Create an SSIS package that performs the following:

**a. Extract**

* Source: CSV file
* Delimiter: Comma

**b. Transform**

Perform the following transformations:

| **Transformation** | **Description** |
| --- | --- |
| Date parsing | Convert Date received and Date sent to company from MM/DD/YY to DATETIME(DD/MM/YY). |
| Trim strings | Remove leading/trailing whitespace from all text fields. |
| Empty to NULL | Replace blank strings with NULL for optional fields (Sub-issue, Narrative, Tags, etc.) |
| ZIP Code formatting | Preserve leading zeros by treating ZIP code as VARCHAR(10) |
| Normalize booleans | Convert yes/no fields to lowercase, and handle blanks as NULL |
| Validate Complaint ID | Ensure uniqueness and valid integers |
| Handle special characters | Ensure no invalid or problematic characters in text columns (quotes, newlines) |

**c. Load**

* Target: SQL Server table CFPB\_Complaints
* Ensure **no duplicate Complaint IDs**
* Use **data flow task with appropriate error handling**

**Additional Requirements**

* Implement error redirection or logging for invalid rows.
* (Optional ): Create **dimension tables** for Product, Sub-product, Issue, State, and Company using Lookup/Slow Changing Dimension transformations.