# Data Cleaning Project

In this project, I performed data cleaning on a raw patient visit dataset using SQL Server Management Studio 2021. The original data contained various quality issues such as mixed text casing, missing values, inconsistent status labels, and possible duplicates.

## Steps Taken:

1. Imported the dataset (PatientVisits\_Dirty.csv) into SQL Server using manual table creation.

2. Created a new cleaned table called `patientVisistsCleaned` using SQL queries.

3. Standardized text formatting for Department and Doctor names using SQL string functions.

4. Normalized `Status` values (e.g., empty → 'Unknown', lowercase → title case).

5. Converted `Charges` column to proper FLOAT type and handled invalid or blank entries.

6. Identified missing values in `Charges` and `Status` fields using COUNT queries.

7. Checked for duplicate entries based on `PatientID` and `VisitDate` combinations.

8. Exported the cleaned data to Excel for use in reporting or Power BI.

## Outcome:

The cleaned table is now consistent, free of structural errors, and ready for visualization or analysis. This process demonstrated practical SQL skills in data profiling, transformation, and integrity checking.