# Data Cleaning Project Summary – SQL Server Management Studio (SSMS)

This document summarizes my hands-on data cleaning project completed using SQL Server Management Studio (SSMS). The project involved working with a healthcare dataset containing patient visit records. All data preparation and validation were performed using SQL queries, without using Excel or external tools.

## Activities Performed:

* 1. Explored the raw patient visits dataset to identify key data quality issues.
* 2. Used SQL queries to detect missing values in important fields such as 'cleanedCharges'.
* Example: SELECT COUNT(\*) AS missingCharges FROM [dbo].[PatientVisitsCleaned] WHERE cleanedCharges IS NULL;
* 3. Flagged and counted records with unclear 'Status' values labeled as 'Unknown'.
* Example: SELECT COUNT(\*) AS unknownStatus FROM [dbo].[PatientVisitsCleaned] WHERE Status = 'Unknown';
* 4. Identified duplicate patient visit entries by grouping on PatientID and VisitDate.
* Example:
* SELECT PatientID, VisitDate, COUNT(\*) AS DuplicateCount
* FROM [dbo].[PatientVisitsCleaned]
* GROUP BY PatientID, VisitDate
* HAVING COUNT(\*) > 1;
* 5. Applied SQL functions like COALESCE, REPLACE, and TRIM to clean and standardize the dataset.
* 6. Ensured the final dataset was structured, consistent, and ready for analysis or reporting.

Screenshots of the queries and results have been captured as proof of the work completed. This project strengthened my ability to clean and prepare data using SQL and deepened my understanding of practical data quality workflows.