

Dissertation / Project / Project Work Title:
**Predictive Modeling and Data Quality Assurance
for EDI Order Processing in Retail Supply Chain**

Course No.: S1-25_DSECLZG628T

Course Title: Dissertation

Dissertation / Project /Project Work Done by:

Student Name: Ajitabh Tiwari

BITS ID: 2020FC04613

Degree: M. Tech. (Data Science & Engineering)

**Research Area: Supply Chain Management,
Predictive Modeling, and Data Quality.**

Dissertation / Project Work carried out at:

7-Eleven Global Solution Center, Bangalore



**BIRLA INSTITUTE OF TECHNOLOGY &
SCIENCE, PILANI VIDYA VIHAR, PILANI,
RAJASTHAN – 333031
(DEC / 2025)**

Abstract

This report document is the initial phase of a dissertation aimed at improving the reliability of Electronic Data Interchange (EDI)-based order processing in retail supply chains.

The study focuses on analysing how data quality issues in inbound purchase orders (EDI 850) contribute to order failures within Oracle E-Business Suite Order Management.

A preliminary framework that integrates rule-based data quality validation with basic predictive modelling is proposed to assess failure risk at an early stage of processing. Synthetic EDI datasets are used to demonstrate the feasibility of the approach, and initial observations reveal a strong correlation between poor data quality and order failure.

The outcomes of this phase establish a foundation for advanced machine learning models and dashboard-based analytical insights to be developed in the final phase of the dissertation.

Ajitabh Tiwari

Signature of the Student

Name: Ajitabh Tiwari

Date: 16/December/2025

Place: Bengaluru

Rajasekhar Kandula

Signature of the Supervisor

Name: Raja Sekhar Kandula

Date: 16/December/2025

Place: Bengaluru

Contents

1.	Introduction	4
1.1	Problem Statement	
2.	Literature Review.....	5
3.	Research and Methodology.....	6
3.1	Overview of Data Quality	
3.2	Data Quality Matrices	
3.3	Weight Assignment for Matrices	
3.4	Data Quality Score Calculation	
3.5	Example Data Quality Score Calculation	
3.6	Data Quality Score Interpretation	
4.	Analysis and Design	10
4.1	EDI Order Processing Flow	
5.	Future-Plan.....	11
●	Figure 1 : Oracle Standard : EDI Order Processing.....	6
●	Figure 2: System Architecture : EDI Order Processing (Proposed).....	10
●	Figure 3: Data Quality Score process flow.....	7
●	Table 1: Data Quality Matrices	8
●	Table 2: Data Quality Matrices Weights	8
●	Table 3: Data Quality Score Bands	10

1. INTRODUCTION

1.1 Problem Statement

- Retail supply chains increasingly depend on **Electronic Data Interchange (EDI)** to automate the exchange of purchase orders and reduce manual processing between trading partners.
- In **Oracle E-Business Suite (EBS)**-based order management environments, inbound EDI purchase orders (EDI 850) are staged and processed through predefined interface tables and validation programs before sales order creation.
- Despite automation, a substantial proportion of inbound EDI orders fail during processing due to **data quality issues**, including missing mandatory attributes, incorrect data formats, invalid reference data, and non-compliance with partner-specific business rules.
- Such data quality issues result in **order rejections, repeated reprocessing, and manual intervention**, leading to delays in order fulfilment and increased operational workload for support teams.
- Existing validation mechanisms in Oracle EBS are primarily **rule-based and reactive**, identifying errors only after they occur, with limited capability to assess the overall reliability of an incoming transaction.
- The lack of a **quantitative measure of data quality** makes it difficult to prioritize high-risk orders or proactively intervene before order creation failures.
- Furthermore, current systems provide minimal insight into the **likelihood of order failure** or the potential operational impact associated with poor data quality.
- There is therefore a clear need for a **proactive and data-driven approach** that can evaluate the quality of inbound EDI orders at an early stage and estimate the risk of processing failure.
- This study addresses the identified problem by proposing a framework that integrates **data quality assurance with predictive analysis**, enabling early identification of high-risk orders and improving the reliability and efficiency of EDI order processing in retail supply chains.

2. LITERATURE REVIEW

Electronic Data Interchange (EDI) has long been adopted in retail supply chains to enable automated and standardized exchange of transactional documents between trading partners. Prior studies highlight that while EDI improves processing speed and reduces manual effort, its effectiveness is highly dependent on the quality of incoming data. In enterprise resource planning (ERP) environments such as Oracle E-Business Suite, data quality issues in EDI purchase orders frequently result in interface failures, manual rework, and delayed order fulfilment.

Research on data quality management emphasizes the importance of validating accuracy, completeness, consistency, and conformity of transactional data before core system processing. Traditional approaches largely rely on rule-based validation mechanisms that detect errors after data ingestion. Although effective for identifying known issues, these methods are reactive in nature and provide limited insight into the overall reliability of a transaction or the likelihood of downstream failure.

Recent literature in supply chain analytics suggests that predictive modeling techniques can enhance operational decision-making by identifying risk patterns in transactional data. Machine learning methods have been applied to predict exceptions, delays, and failures in various supply chain processes; however, their application in EDI-based order processing remains limited. Most existing studies focus on post-processing analysis rather than early-stage failure prevention.

Studies integrating data quality metrics with predictive analytics indicate that quantitative data quality indicators can serve as strong predictors of process outcomes. By converting validation results into numerical features, organizations can move beyond binary pass-fail checks and adopt a risk-based processing approach. This shift enables proactive intervention, targeted manual review, and improved system efficiency.

Despite these advancements, there is a noticeable gap in literature addressing the combined use of data quality assurance and predictive modeling within Oracle ERP-driven EDI order processing. This project addresses this gap by proposing a framework that integrates rule-based data quality validation with predictive failure estimation to improve the reliability and efficiency of retail supply chain order processing.

2.1 Standard EDI Process Flow:

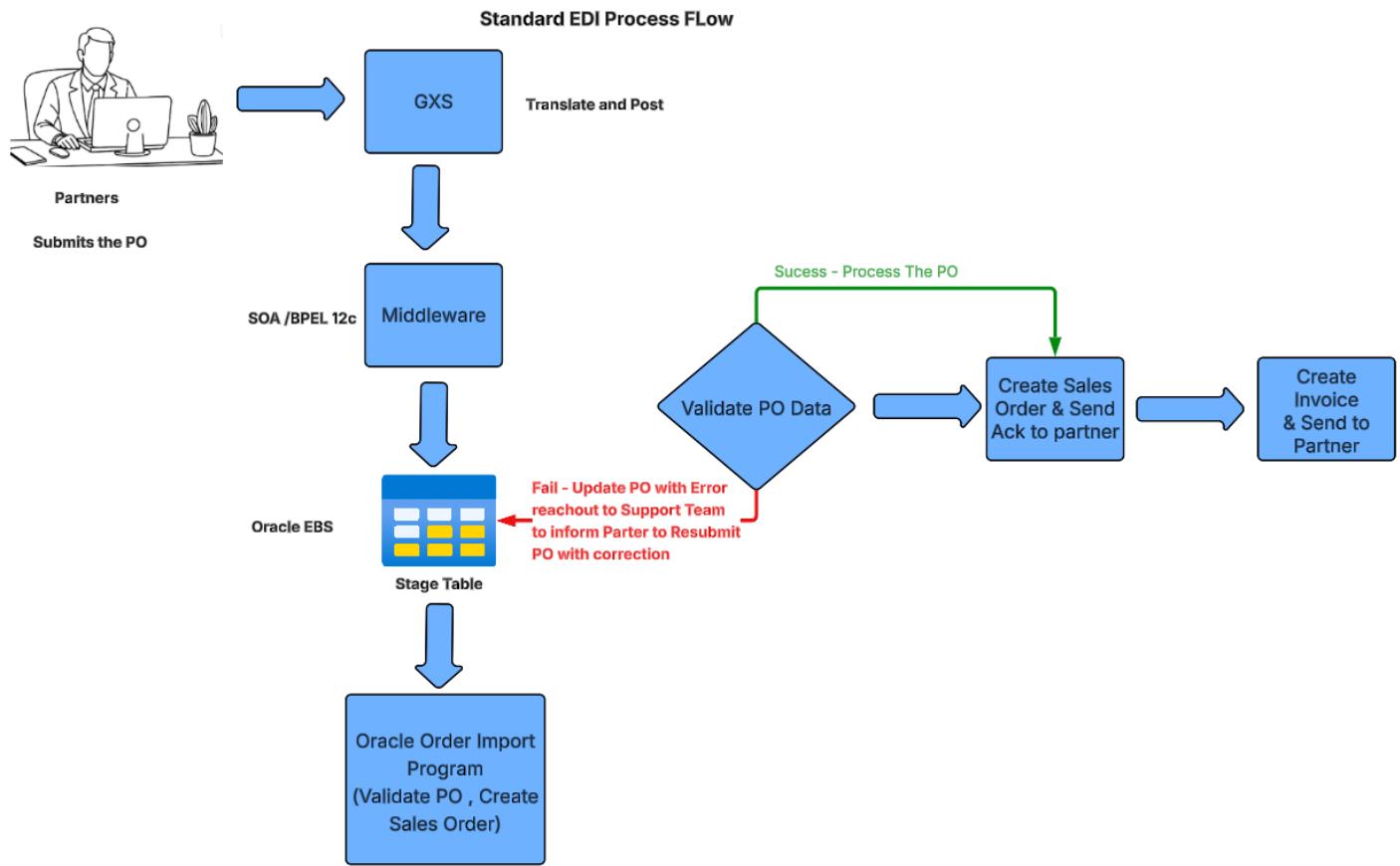


Figure 1 : Oracle Standard : EDI Order Processing

3. PROPOSED ARCHITECTURE

The proposed architecture is designed to integrate data quality assurance into the inbound EDI order processing flow of Oracle E-Business Suite. Incoming purchase orders in EDI 850 format are first staged in Oracle Order Management interface tables, which act as a buffer between external partner data and core order tables. A Data Quality Gate is introduced at the interface layer to perform structured validations on the staged data. These validations assess mandatory fields, data formats, reference data consistency, and partner-specific business rules. The results of these validations are aggregated to generate data quality metrics and a composite Data Quality score.

Based on the computed score, orders can be categorized as low-risk or high-risk before order creation. This architecture enables proactive identification of problematic transactions and forms the basis for further predictive analysis in the final phase of the project.

3.1 EDI order Processing Flow (Proposed):

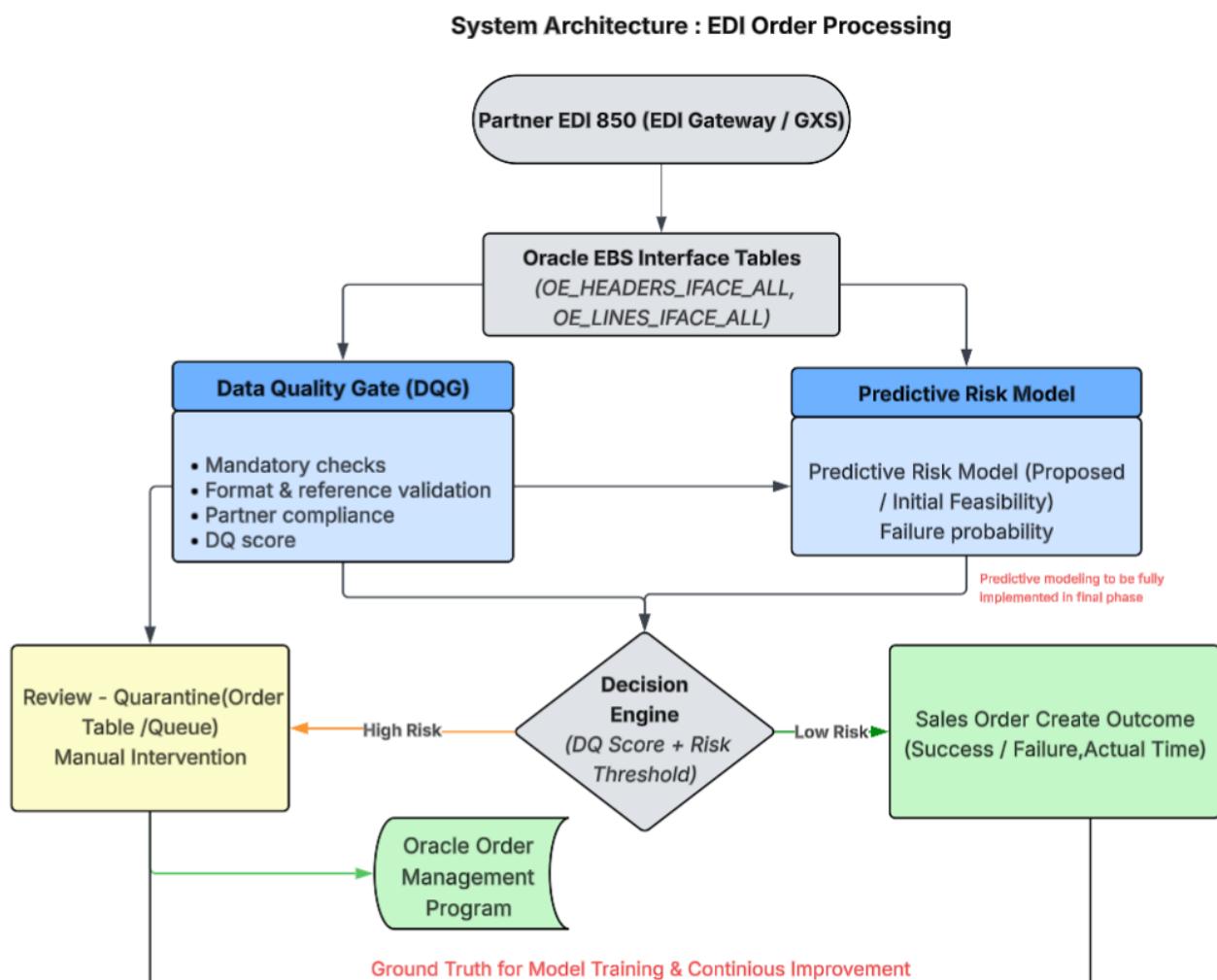


Figure 2 : System Architecture : EDI Order Processing

4.RESEARCH METHODOLOGY (INITIAL IMPLEMENTATION)

4.1 DATA QUALITY FRAMEWORK

The research methodology adopted in this study follows a design-oriented and experimental approach, focusing on improving the reliability of EDI-based order processing in retail supply chains through data quality assurance and predictive modeling. The methodology is structured to ensure practical relevance while maintaining academic rigor and data confidentiality.

The study begins with an analysis of the existing EDI order processing workflow in Oracle E-Business Suite, particularly the inbound processing of purchase orders (EDI 850) through interface tables. Based on this analysis, common causes of order failures related to data quality issues are identified, including missing mandatory fields, invalid reference data, format inconsistencies, and partner-specific rule violations.

Due to the confidential nature of enterprise EDI and ERP data, the research utilizes a synthetically generated dataset that replicates the structure, attributes, and behavioural patterns of real-world EDI purchase orders. The synthetic dataset is designed to simulate realistic scenarios by deliberately introducing data quality issues and varying partner behaviour, while ensuring no production or proprietary data is used.

A rule-based data quality framework is then applied to the dataset. These rules mirror Oracle interface validations and are used to generate data quality metrics such as missing mandatory field count, invalid reference count, and partner rule violations. A Data Quality score is computed using a weighted penalty approach to represent the overall reliability of each transaction.0o

Subsequently, the derived data quality metrics and scores are used as inputs for predictive modeling. Classification models are trained to estimate the likelihood of order failure based on historical patterns in data quality behaviour. The models are evaluated using standard performance measures to assess their effectiveness in identifying high-risk orders.

4.1.1 Overview of Data Quality –

The Data Quality (DQ) score is used to quantify the reliability of an incoming EDI purchase order during the pre-processing stage. It is derived by consolidating the outcomes of multiple rule-based data quality checks performed on data staged in Oracle interface tables. By translating individual validation results into a single numerical value on a scale of 0 to 100, the DQ score enables consistent assessment of data integrity, where higher scores reflect more reliable transactions and reduced risk of downstream processing failures.

4.1.2 Data Quality Score process flow -

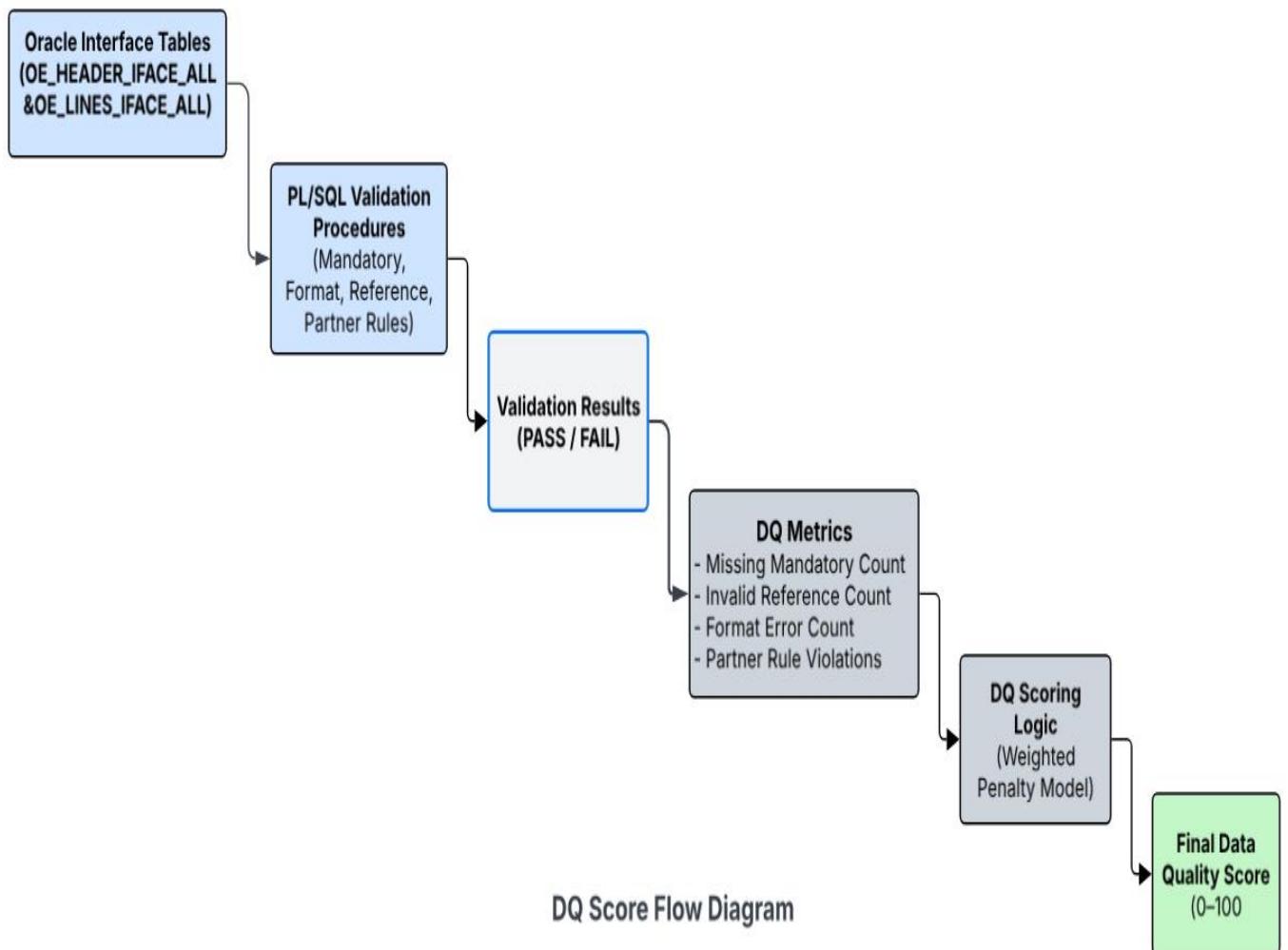


Figure 3 : Data Quality Score process flow

4.1.3 Data Quality Metrics

The DQ score is derived from multiple data quality metrics generated after PL/SQL validations.

Metric Name	Description	Source of Validation
missing_mandatory_count	Number of missing mandatory fields	Mandatory field validation
invalid_reference_count	Invalid item, customer, or location	Master data checks
format_error_count	Invalid date or numeric format	Format validation
partner_ruleViolation_count	Partner-specific rule failures	Custom business rules
warning_count	Non-critical validation warnings	Advisory checks

Table 1: Data Quality Metrics

4.1.4 Weight Assignment for Metrics

Each metric is assigned a weight based on its severity and impact on order processing.

Metric	Weight (Penalty Points)	Severity Level
Missing mandatory field	15	Critical
Invalid reference data	20	Critical
Format error	5	Medium
Partner rule violation	10	Medium
Warning	2	Low

Table 2: DQ Metric Weights

4.1.5 DQ Score Calculation Formula

The DQ score calculated using a weighted penalty model.

Formula

$$\text{DQ Score} = 100 - \sum_{i=1}^n (M_i \times W_i)$$

Where:

- M_i = Count of data quality metric i
- W_i = Weight assigned to metric i

The final DQ score is bounded between 0 and 100.

4.1.6. Example DQ Score Calculation

Sample Validation Results

Metric	Count
Missing mandatory fields	2
Invalid reference data	1
Format errors	0
Partner rule violations	1
Warnings	2

Calculation :

$$\text{DQ Score} = 100 - (2 \times 15) - (1 \times 20) - (1 \times 10) - (2 \times 2)$$

$$\text{DQ Score} = 100 - (30 + 20 + 10 + 4) = 36$$

4.1.7 DQ Score Interpretation

DQ Score Range	Quality Band	Interpretation
80 – 100	High (Green)	Order is reliable and low risk
50 – 79	Medium (Amber)	Order requires attention
Below 50	Low (Red)	Order is high risk

Table 3: DQ Score Bands

4.2 Predictive Risk Model

The validated data and DQ metrics are then passed to a predictive risk model, which estimates the probability of order failure. The model uses historical transaction patterns and data quality indicators to classify orders as low-risk or high-risk.

Predictive modeling to be fully implemented in final phase.

5. PRELIMINARY RESULTS

Preliminary analysis of the synthetic dataset indicates a clear relationship between data quality scores and order processing outcomes. Orders with higher data quality scores exhibit a significantly lower likelihood of failure, while orders with lower scores are more prone to validation and processing issues.

The results confirm that aggregating validation outcomes into a structured scoring mechanism provides meaningful insight into transaction reliability. These findings justify the use of data quality metrics as input features for predictive modeling in the final phase of the project.

6. PLAN FOR FINAL PHASE

The final phase of the dissertation will extend the current framework by incorporating predictive analytics to enhance decision-making in EDI order processing. Advanced classification models will be developed using data quality metrics and historical outcomes to estimate order failure probability and enable risk-based processing. In addition, interactive dashboards will be implemented to visualize order status, data quality trends, and risk distribution for operational insight. The framework will also be expanded to estimate end-to-end processing time for inbound EDI purchase orders, as processing duration is influenced by data quality, order complexity, and manual intervention. Key process timestamps such as EDI receipt, interface loading, and order completion will be used to derive processing time, while data quality scores and order attributes will serve as explanatory variables. A regression-based modeling approach will be explored to support processing time estimation, SLA forecasting, and operational planning. The final phase will conclude with a detailed evaluation of model performance and business impact, demonstrating the practical applicability of the proposed solution.

Phases	Start Date-End Date	Work to be done	Status
Dissertation Outline	05 Nov 2025 – 21 NOV 2025	Literature Review and prepare Dissertation Outline	COMPLETED
Design & Development	22 Nov 2025 – 31 Dec 2025	Design & Development Activity	COMPLETED
Testing	31 Dec 2015 – 31 Jan 2026	Software Testing, User Evaluation & Conclusion	PENDING
Dissertation Review	15 Feb 2026- 15 March 2026	Submit Dissertation to Supervisor & Additional Examiner for review and feedback	PENDING

Submission	15 Mar 2026- 15 APR 2026	Final Review and submission of Dissertation	PENDING
------------	--------------------------	---	----------------

Conclusion -

This mid-semester report establishes a strong foundation for improving EDI order processing reliability through data quality assurance. By introducing a structured validation and scoring framework, the study demonstrates how data quality can be quantified and analysed at an early stage of processing. The initial results support the feasibility of extending this approach with predictive modeling and analytical dashboards in the final phase of the dissertation.