Quality Check (QC) Report – ShipEase Logistic Data Analysis

1. Starting the QC Process

When I first received the logistics dataset, my main goal was to make sure the data was reliable before jumping into analysis. Since the dataset came from seven different CSV files, I expected some inconsistencies and I found quite a few.

So before touching any insights, I began with data validation, cleaning and structure checks.

2. Dataset Exploration

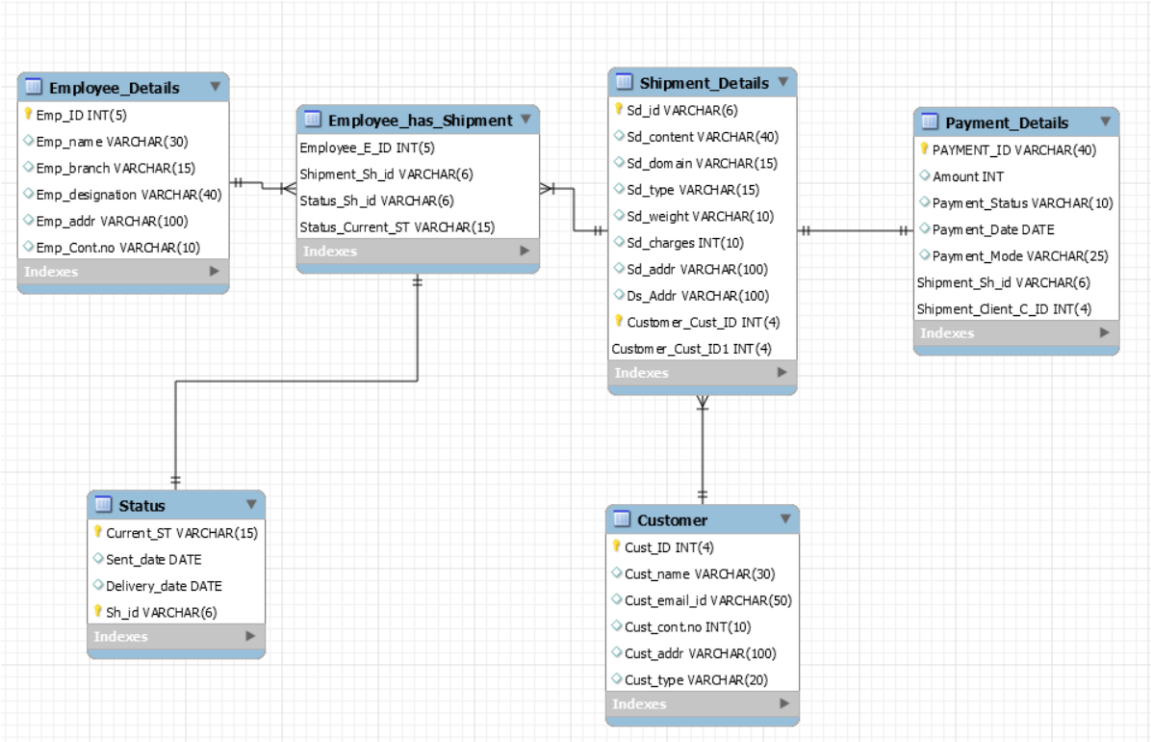

Table Definitions:

- Employee_Details Table: Contains the information of the employees.
- Membership Table: Contains membership details of customers or clients.
- Customer Table: Contains the information of customers or clients.
- Payment_Details Table: Contains the payment details.
- Shipment_Details Table: Contains shipment details.
- Status table: Contains delivery status details.
- Shipment Table: Relationship table between employee and shipment.

Data Quality Framework

Quality Dimension	Customer Table	Shipment Details	Status Table	Payment Details	Membership Table
Accuracy	No issues found	No issues except weight stored as VARCHAR	Wrong / impossible dates (month > 12, day > 31)	Incorrect/ Unclean payment status text entries (e.g., NA)	No issues
Completeness	No missing values	No missing shipment IDs	Missing or empty Delivery/Sent Dates	Missing Payment Dates	Missing End Dates in some memb
Consistency	Formats consistent	Weight stored as text instead of numeric	Date formats inconsistent	Payment Date TEXT not DATE	Date formats inconsistent
Timeliness	No issues	No issues	Invalid timeline entries	No issues	No issues
Relevancy	No irrelevant columns	No irrelevant columns	No irrelevant fields	No irrelevant fields	No irrelevant fields
Uniqueness	Customer_ID unique	Shipment ID unique	SH_ID repeated (expected)	Payment ID unique	Membership ID unique
Validity	Email/contact valid	Weight contains non-numeric value	Delivery < Sent (invalid)	Invalid Payment Dates	No issues

3. How I Identified Data Quality Issues (QC Findings)

Incorrect / Impossible Dates

• Checked months > 12, 30/31 Feb.

• Detected invalid delivery timelines using STR_TO_DATE(), MONTH(), DAY().

Missing & Empty Values

• Searched for '', NA, 0000-00-00, NULL.

• Revealed missing Delivery, Sent, Payment and Membership End Dates.

Inconsistent Formats

• Detected mixed MM/DD/YYYY and TEXT formats.

• Payment Date not stored as DATE.

Data Type Issues

• Weight stored as VARCHAR.

• All date columns stored as TEXT.

• Inconsistent payment status text.

Logical Validity Checks

• Found Delivery < Sent using DATEDIFF().

• Incorrect membership durations detected.

Uniqueness Checks

• Verified primary key uniqueness across all tables.

4. How I Resolved All Issues (Step-by-Step Fixes)

Issue 1: Wrong / Impossible Dates

• Identified invalid months using SUBSTRING().

• Checked Feb 30/31.

• Converted dates using STR_TO_DATE().

Issue 2: Missing or Empty Dates

• Converted blanks and placeholders to NULL.

• Cleaned Sent_Date, Payment_Date, Membership End_Date.

Issue 3: Date Format Inconsistency

• Standardized all dates using STR_TO_DATE().

• Converted column types to DATE.

Issue 4: Invalid Delivery Timeline

• Detected Delivery < Sent.

• Added Delivery_Flag column.

• Flagged invalid rows.

Issue 5: Payment Date Stored as TEXT

• Converted text to DATE.

• Updated column data type.

Issue 6: Payment Status Inconsistency

• Standardized empty/NA statuses.

Issue 7: Shipment Weight as VARCHAR

• Validated for numeric consistency.

Issue 8: Missing Membership End Dates

• Used COALESCE() for ongoing memberships.

Issue 9: Data Enrichment Through Views

• Created Payment_Summary view.

• Created PaymentNotDone view.

5. Final Summary of QC Resolution

Issue	Solved Through
Invalid dates	Validation + STR_TO_DATE()
Missing dates	Converted to NULL
Wrong date formats	Standardized + DATE conversion
Delivery timeline errors	Flagging invalid rows
TEXT data types	Converted to DATE / numeric
Payment inconsistencies	Standardized values
Membership gaps	Proper NULL handling
Weight format	Validated numeric consistency
Multi-table inconsistencies	Views + relational cleanup