# Week 1 – Data Cleaning Report

## 1. Objective

Prepare a clean and reliable logistics dataset for SLA analysis by fixing data issues and creating meaningful columns.

## 2. Dataset Understanding

The dataset contains shipment-level data including carrier, shipping mode, origin, destination, delivery timelines, and cost.

## 3. Data Cleaning Steps

- Data Type Fixing:

Converted cost, delay, and shipment values into proper numeric formats. Binary columns kept as 0/1.

- Handling Negative Values:

Negative delays (early delivery) retained. Created new column "Delay Date Only" where only delays (>0) are considered.

- Derived Columns:

Created SLA Status column:

SLA Breached / On-Time Shipment

- Duplicate Check:

No major duplicates found based on shipment_id.

- Outlier Review:

Extreme values retained as they represent real business scenarios.

- Logical Validation:

No negative costs, delay logic consistent, binary flags correct.

## 4. Key Calculations Prepared

SLA Breach Rate, On-Time Rate, Average Delay, Cost metrics.

## 5. Final Dataset Status

Cleaned, validated, and ready for dashboard creation.

6. Summary

Week 1 focused on cleaning, validation, and preparing data for analysis.