**Data Wrangling Report Guide**

**Course Name:** Data Analytics   
**Date:** 14th February 2025  
**Submission Deadline:** 17th February 2025

**1. Dataset Overview**

* **Rows and Columns:** State the number of rows and columns in the dataset.
* **Features:** List all key columns, e.g., Order ID, Customer Name, Supplier Name, Order Amount, Shipment Cost, etc.

**2. Assumptions**

* Summarise the assumptions made prior to starting the data wrangling process (e.g., presence of missing values, possible inconsistencies, duplicates, outliers, misclassifications, etc.).

**3. Data Wrangling Process**

**3.1 Handling Missing Values**

* **Identification:** Identify columns with missing values (e.g., Order Amount, Shipment Cost, Customer Name).
* **Method:** Describe how missing values were handled (e.g., average for numerical values, “Unknown” for categorical).
* **Outcome:** State the number of missing values handled and the method applied.

**3.2 Detecting and Removing Duplicates**

* **Identification:** State how duplicates were detected (e.g., by Order ID).
* **Method:** Describe the process used to remove duplicates (e.g., Excel’s "Remove Duplicates" feature).
* **Outcome:** Document the number of duplicates removed.

**3.3 Resolving Inconsistencies**

* **Identification:** List any inconsistencies found (e.g., different payment status formats, inconsistent date formats).
* **Method:** Describe how inconsistencies were resolved (e.g., standardising payment statuses, reformatting dates).
* **Outcome:** State how many inconsistencies were corrected.

**3.4 Identifying and Handling Outliers and Errors**

* **Identification:** Describe the method used to identify outliers and errors (e.g., sorting numerical columns to spot extreme values).
* **Method:** Explain how outliers and errors were handled (e.g., correcting negative values, marking data as "Error").
* **Outcome:** Document how many outliers or errors were corrected.

**3.5 Correcting Misclassified Data**

* **Identification:** List misclassifications (e.g., Order Date after Delivery Date, invalid payment status entries).
* **Method:** Describe the correction process (e.g., swapping dates, changing invalid entries to “Pending”).
* **Outcome:** State the number of misclassifications corrected.

**4. Summary of Changes**

* Summarise all changes made during the wrangling process, e.g., missing values filled, duplicates removed, inconsistencies fixed, outliers handled, misclassified data corrected.

**5. Challenges Encountered**

* Briefly mention any challenges faced during the data wrangling process (e.g., difficulty in identifying errors, handling multiple missing values).

**6. Decisions Made**

* Summarise the key decisions made during the process (e.g., choice of method for handling missing values, standardisation approach).