

# Data Cleansing Process

## Overview

This is a comprehensive and intensive data cleansing task. You will handle a large dataset and gain valuable learning experience in the process. Make use of the main channel for collaboration, creating working sessions, and learning from peers.

## Steps to Complete the Data Cleansing Process

1. **Review the Dataset**
  - Begin by examining the dataset to understand its structure. Look for any obvious errors or inconsistencies that may need attention.
2. **Identify Missing Data**
  - Use appropriate tools to detect missing values. Decide on the best method to handle these missing values, ensuring they are addressed properly.
3. **Correct Inconsistencies**
  - Identify and correct inconsistencies in the data. This includes:
    - Removing duplicate records
    - Ensuring data types are correct
    - Managing any outliers that could skew the analysis
4. **Standardize Data**
  - Make sure that all data formats are consistent across the dataset. This may include:
    - Aligning date formats
    - Standardizing categorical values
5. **Document Changes**
  - Keep a detailed record of all changes made during the data cleansing process, providing transparency and traceability.

## Completion Criteria: How to Know You're Done

You will know you're finished when:

- The dataset is thoroughly cleaned, free from missing values, duplicates, and inconsistencies.
- The data is well-organized, accurate, and ready for analysis.
- You have documented the entire data cleaning process, detailing:
  - The steps, techniques, and tools used
  - How each step contributed to improving data quality
- Your documentation demonstrates your ability to transform messy, unstructured data into a reliable, structured format.