**Data Management I**

**Goal-B :**

Sales Data Analysis dataset from Kaggle is a cleaned dataset. This dataset contains sales data for a variety of products. It includes columns like Order ID, Product,Quantity Ordered, Price Each, Order Date, Purchase, Address, Month, Sales, City, Hour.

**Dimensions considered while making the dataset raw**

**1)Conformity:** This can potentially introduce non-unique data, as case-sensitivity can impact record uniqueness.  
 Action taken : values changed to lower case for CITY column.

2) **Validity** : randomly replaced about 30% of the "City" entries with "error”.

3) **Completeness :** made approximately 40% of the "Sales" column empty.

4) **Uniqueness** : Duplicated one product column to create redundancy, which affects the uniqueness.

5) added "RandomText" to 10% of the rows.

**Summary of Changes**

Introduced Duplicates: Reduces uniqueness.

Replaced with NULL: Affects completeness.

Added Characters: Impacts accuracy and conformity.

Mixed Data Types: Disrupts consistency.

I have attached the flow and datasets to support my performance of uncleaning a cleaned based on the dimensions.