Question: Make observations and write down the problems that you see with the data like missing values, duplicates, date formats, data conversions, special characters (if any to be identified) and inconsistencies. (For this implementation you can skip the cleaning part as its not covered)

Missing Values: There are apparent gaps in certain columns, notably in CATEGORY2 and CATEGORY3. These missing values could potentially affect the accuracy of any subsequent analysis and thus necessitate attention.

Data Types and Formats:

Date Formats: The CREATION DATE field follows a MM/DD/YYYY format. Depending on the analysis, transforming it into a standard YYYY-MM-DD format might enhance consistency and clarity.

Time Formats: The CREATION TIME field employs a 12-hour AM/PM format. Standardizing it to a 24-hour format could streamline analysis processes and ensure uniformity.

Numeric Fields: Fields like ZIP CODE and PARCEL ID NO likely contain numeric data. It's crucial to verify their consistency across the dataset to maintain data integrity.

Duplicates: Identifying and addressing duplicate records is essential to prevent distortions in analysis outcomes.

Inconsistencies and Special Characters:

Textual data may exhibit inconsistencies in capitalization or the presence of special characters, particularly in fields such as SOURCE, DEPARTMENT, and WORK GROUP. Normalizing these inconsistencies is vital for accurate analysis.

Special characters, if present, may require handling, especially within text fields, to maintain data integrity.

Data Conversions: Certain fields may necessitate conversion to facilitate more insightful analysis. For instance, transforming LATITUDE and LONGITUDE into a geospatial format could enhance mapping capabilities.

Range and Validity Checks: Ensuring that data, including dates, times, and numeric fields, falls within expected ranges and adheres to appropriate formats is crucial for maintaining data quality and reliability.

To address these issues effectively, the following cleaning steps could be planned:

Handle Missing Values: Begin by investigating the reasons behind missing data and determine whether to fill them with a default value, employ interpolation techniques, or consider removing the affected rows or columns altogether.

Standardize Date and Time Formats: Normalize the formats of dates and times to ensure consistency across the dataset. This involves converting them into a uniform format for ease of analysis.

Remove Duplicates: Identify and eliminate any duplicate records present in the dataset to prevent skewing analysis results and ensure data integrity.

Normalize Textual Data: Standardize textual data by converting them to a consistent case, such as upper or lower case, and consider removing or replacing special characters as necessary to maintain uniformity and clarity.

Data Type Conversions: Verify that the data types assigned to each column accurately reflect their content. For example, consider converting ZIP codes to strings if they contain leading zeros to preserve their integrity.

Validation Checks: Implement thorough checks to validate the range and validity of data, particularly for dates, times, and numeric fields. This process ensures that the dataset conforms to expected standards and enhances its reliability for analysis purposes.