Instructions for Handling Tabular Data

1. Import Necessary Libraries:
   * Ensure that all required libraries, such as pandas for data manipulation and SDV for synthetic data generation, are installed.
2. Load the Header and Data File:
   * Load the header file separately, which contains the column names.
   * Load the data file (which does not have headers) containing the original data.
   * Assign the header columns to the data to form a combined DataFrame.
3. Apply Custom Functions to Randomize Sensitive Columns:
   * Identify which columns in the DataFrame contain sensitive information.
   * Apply custom functions to these columns to randomize or anonymize the data as required, without affecting the overall structure.
4. Prepare Data for SDV Model:
   * Ensure the DataFrame is properly formatted before using the SDV model, which includes correcting data types and handling missing values appropriately.
5. Train the SDV Model:
   * Train the SDV model to learn the patterns and distributions of the original data.
6. Generate Synthetic Data:
   * After training, use the SDV model to generate synthetic data that mimics the characteristics of the original data.
7. Save Synthetic Data to the Final DataFrame:
   * Finally, DataFrame is only the synthetic data generated by the model.
8. Save the Final DataFrame to a Text File Without Headers:
   * Export the final DataFrame (containing only synthetic data) to a text file. Make sure the header row is not included in the output file.

Instructions for Handling Positional Files with Synthetic Data Generation

1. Import Necessary Libraries

Ensure required libraries, such as `pandas` for data manipulation and `SDV` for synthetic data generation, are installed.

2. Define File Paths

Specify paths for the layout file, data file, header layout file, and the metadata storage.

3. Create Output File Path

Construct the output file path by appending `\_syn` to the original input file name to indicate it contains synthetic data.

4. Load Layout and Data Files

Load the layout from the CSV file and the data from the text file. **Note that the data does not contain a header row.**

5. Process Records

Process the data to separate CE records (starting with "CD") and process the main sample data based on the defined layout. For each line, extract data according to the specified character lengths for each column.

6. Generate Synthetic Data

Pass the processed DataFrame to the SDV model. Use existing metadata if available; otherwise, create new metadata from the DataFrame.

7. Evaluate Synthetic Data

After generating synthetic data, evaluate its quality against the original data using diagnostic tools provided by SDV.

8. Build Page Trailer

Calculate the record count and sums for relevant columns (available in detailed records) and construct the trailer record with this information.

9. Write Output File

Write the synthetic header, synthetic data, CE records, and the trailer to the output file, ensuring that no column header row is included.

10. Main Function Execution

Implement the main function to orchestrate loading data, generating synthetic data, and writing the output file, while also handling any optional parameters as needed.