Transform Crimes_Boston:

load_csv_to_dataframe:

- Purpose: Loads data from a CSV file into a pandas DataFrame.
- Why: This is the first step to read and manipulate the data stored in the CSV file.

project_columns:

- Purpose: Selects specific columns from a DataFrame.
- Why: It's used to focus only on the relevant columns needed for further analysis, reducing memory usage and processing time.

convert_shooting_to_boolean:

- Purpose: Converts the values in the 'SHOOTING' column to boolean (True for 'Y', False otherwise).
- Why: Standardizes the representation of shooting incidents for easier analysis and comparison.

clean_crime_data:

- Purpose: Cleans the crime dataset by removing rows with missing or invalid location information.
- Why: Ensures data integrity and accuracy by eliminating problematic rows that may affect analysis results.

new_csv:

- Purpose: Saves a DataFrame to a new CSV file with specified columns.
- Why: Creates new CSV files containing relevant subsets of data for easier storage and sharing.

remove_slccolumn:

- Purpose: Removes specified columns from a DataFrame.
- Why: Eliminates unnecessary columns to focus only on relevant data, reducing memory usage and improving clarity.

read_weather_html_file:

- Purpose: Reads an HTML file and extracts a DataFrame representing a table within it.
- Why: Extracts weather data from HTML files for further analysis and integration with crime data.

rename_columns:

- Purpose: Renames columns in a DataFrame based on a provided dictionary.
- Why: Standardizes column names for consistency and clarity in data analysis.

join_dataframes:

- Purpose: Joins two DataFrames based on specified columns and method.
- Why: Integrates weather data with crime data to enable combined analysis and insights to see after if there is any relation between the weather and crimes.

select_interesting_rows:

- Purpose: Selects rows from a DataFrame based on specified conditions.
- Why: Filters the data to focus only on shooting incidents, allowing for specific analysis on these incidents.

save_dataframe_to_csv:

- Purpose: Saves a DataFrame to a CSV file.
- Why: Persists the transformed data to a CSV file for further use and sharing.

add_primary_key:

- Purpose: Adds a new column with unique values acting as a primary key to the DataFrame.
- Why: Facilitates identification and referencing of individual records in the DataFrame, enhancing data management and analysis.

Transform Shootings :

load_csv_to_dataframe:

- Purpose: Loads data from a CSV file into a pandas DataFrame.
- Why: This function is necessary to read the data from the input CSV file and prepare it for further processing.

rename_columns:

- Purpose: Renames specified columns in the DataFrame.
- Why: It standardizes the column names to improve clarity and consistency in data analysis.

convert_victims_to_boolean:

- Purpose: Converts values in the 'multiple_victims' column to boolean (True for 'T', False for 'F').
- Why: Standardizes the representation of multiple victims for easier analysis and comparison.

convert_Shooting_to_boolean:

- Purpose: Converts values in the 'Shooting_type' column to boolean (True for 'FATAL', False for 'NON-FATAL').
- Why: Standardizes the representation of shooting types for easier analysis and comparison.

select_interesting_rows:

- Purpose: Selects rows from the DataFrame where 'multiple_victims' column equals 1 and 'gender' column equals 'Female'.
- Why: Filters the data to focus only on shooting incidents committed by females with multiple victims, allowing for specific analysis on these incidents.

replace_nan_with_unknown:

- Purpose: Replaces NaN values in specified columns with 'unknown'.
- Why: Ensures that missing values are handled consistently and uniformly, improving data quality and analysis.

save_dataframe_to_csv:

- Purpose: Saves the DataFrame to a CSV file.
- Why: Persists the transformed data to a CSV file for further use and sharing.

add_primary_key:

- Purpose: Adds a new column with unique values acting as a primary key to the DataFrame.
- Why: Facilitates identification and referencing of individual records in the DataFrame, enhancing data management and analysis.