

Algorithm for Clustering and Grouping Safe Zones (For Analysis can be monthly or weekly):

Input:

- The input is a DataFrame (safe_zones_df) that contains:
 - coordinates: A column with latitude and longitude values stored as a string.
 - last_visited: A column representing the date when each safe zone was last visited.

Data Preprocessing:

1. Split the coordinates column into two separate columns: one for latitude (lat) and one for longitude (lon).
2. Ensure that no invalid or missing values exist after splitting the coordinates.

Clustering (Spatial Grouping):

1. Use the K-Means clustering algorithm to group the safe zones based on their latitude and longitude values.
2. The default is to create **a minimum of 3 clusters**, but this can be adjusted. If you want, you can increase the number of clusters (e.g., to 4, 5, or more) by changing the n_clusters parameter.
3. After clustering, store the cluster labels in a new column called cluster to indicate which group each safe zone belongs to.

Grouping by Date:

1. Convert the last_visited column to a valid date format.
2. Extract only the date portion (ignoring the time) and store it in a new column called visited_date.
3. Group the safe zones based on this visited_date to see how many safe zones were visited on different days.
4. Store the grouped data (including the coordinates, last visited date, latitude, and longitude) in a dictionary where the keys are the visit dates and the values are the corresponding safe zone data for each date.

Output:

1. A DataFrame with an added cluster column that shows which cluster each safe zone belongs to.
2. A dictionary where the keys are visit dates (visited_date) and the values are DataFrames containing the safe zone data for each day.

FLOWCHART

