➢ Problem:

The United States of America has the highest rates of homeless people, as the number of homeless people in New York reached 78,676 in 2018 ,New York City aspires to realize its vision that by 2024, there will be no homeless people.

Shelter is a company that shelters homeless people and provides them with permanent housing, contributes to be part of achieving the vision and help collect donations from people to contribute to reducing homelessness.

A shelter company want to start collecting donations by placing cash donation boxes at the entrances and exits to the crowded New York City subway stations .

➢ Question/Need:

What are the names of the stations that are crowded? Where are the residential areas located? and commercial areas?

The mission is to use subway data from the New York City Public Transportation Authority (MTA) to help the shelter company raise as many donations as possible. The data retrieved through the online MTA portal is used to extract a list of the most frequently used and crowded stations in the morning, noon, and evening to explore residential areas and attract the kindness of nearby residents to donate, as well as commercial areas, gathering people at stations and the presence of a large number of homeless in them.

➢ Description data:

C/A: Control Area.

UNIT: Remote Unit for a station.

SCP: Subunit Channel Position represents a specific address for a device.

STATION: Represents the station name the device is located at.

LINENAME: Represents all train lines that can be boarded at this station Normally lines are represented by one character.

DIVISION: Represents the Line originally the station belonged to BMT, IRT, or IND.

DATE: Represents the date (MM-DD-YY).

TIME: Represents the time (hh:mm:ss) for a scheduled audit event

DESc: Represent the "REGULAR" scheduled audit event (Normally occurs every 4 hours).

ENTRIES: The cumulative entry register value for a device.

EXIST: The cumulative exit register value for a device.

➢ Tools:

Pandas with python language for clean and explore data and Matplotlib library for data visualization.

Shahad Almubki