

➤ 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.

Shelter Company is a company that shelters the homeless and provides them with permanent housing, as this company collects donations from people to contribute to reducing homelessness.

A shelter company want to start collecting donations every month of the year for one day only by placing cash donation boxes at the entrances to the crowded New York City subway stations and also ads with QR code on the walls of train in the most important lines for a quick donation.

➤ Question/Need:

What day of the month should the company focus on? What are the most important stations that are often crowded? What are the names of the most important train lines?

The task 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. Data retrieved via the MTA's online portal is used to extract a list of the top train stations with the most passengers and which day of the month the Shelter company must focus on it to put donation boxes in and also the names of important train lines to add posters to donate .

➤ **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.