Project proposal

**Client company**: Buna coffee(imaginary coffee shop)
**Question**: The company has multiple coffee shops located in multiple train stations and after the post covid pickups, they are opening a few of those shops. My analysis will be done on the recent three-month period and it will be done to study which train stations have more traffic flow during this period; which days of the week and what times of the day have the highest traffic. This data will help the coffee shop to know which stores have more demand and helps them with effective resource and labor allocation.
**Data to be used**: MTA turnstile data from Dec 19, 2020, to Mar 20, 2020
The MTA Turnstile has 11 fields;
C/A = Control Area
UNIT = Remote Unit for a station (R051)
SCP = Subunit Channel Position represents an specific address for a device (02-00-00)
STATION = Represents the station name the device is located at
LINENAME = Represents all train lines that can be boarded at this station
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 comulative entry register value for a device
EXIST = The cumulative exit register value for a device

Since my analysis is going to be focused on the traffic flow to find out which stations has the highest traffic, at what time and day there will be maximum traffic ; I would be focusing on the STATION, TIME, ENTRIES AND EXITS fields. Therefor the table that I am planing to create will contain the four field.