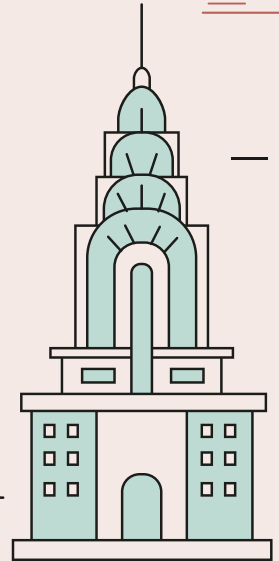
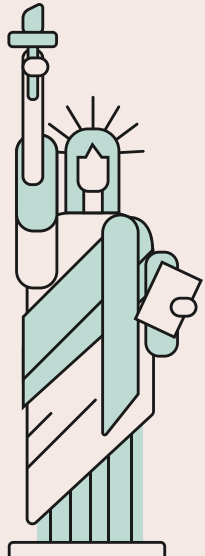


MTA Traffic Analysis for WTWY Street Teams

Presented by :
Manal AlQahtani
Nadia AlGhamdi

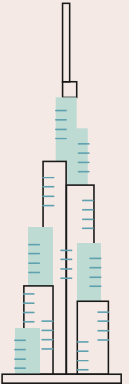
As Project 1 of SDAIA Data Science Bootcamp (T5)





Outline:

1. Motivation
2. Methodology
3. Results and Summary
4. Recommendations





Motivation

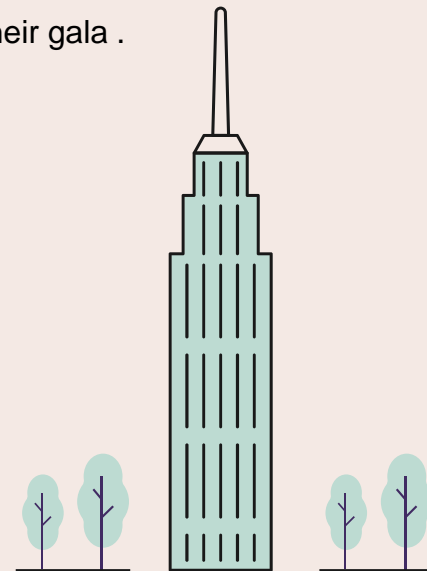
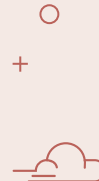
An imaginary client organization (WTWY) hosts summer gala, and want to deploy street teams on NYC subway stations to attract more attendance to sign up for their gala .



Goal

Using MTA data to provide WTWY with recommendation based on :

1. The top 10 busiest stations determined by daily average foot traffic .
2. The busiest stations by weekday .
3. The optimal time to deploy street teams on these stations .





1. Source the data :

- + - " MTA Turnstile " website [" 29 May 2021 - 27 August 2021 "]



2. Loading the data :

- 1- Creating local database using SQLite .
- 2- Delete duplicated columns names using DB browser .
- 3- Read it on Jupyter notebook using " Sqlalchemy " library and started exploring data .

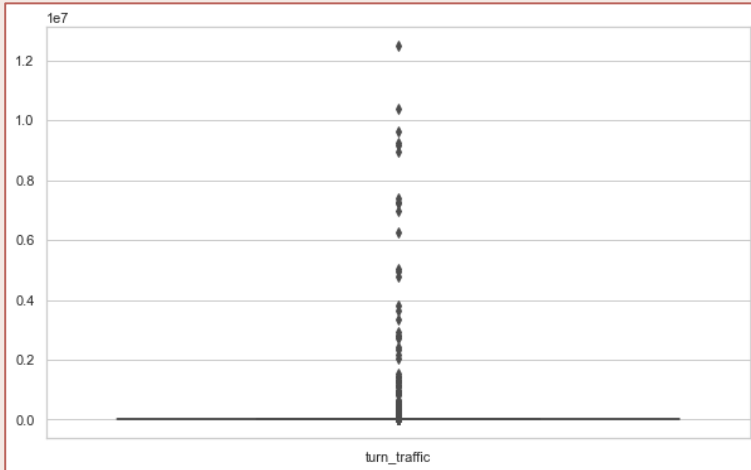
```
nadia — zsh — 80x24
Last login: Wed Sep 29 21:30:33 on ttys001
(base) nadia@Nadias-MacBook-Pro ~ % sqlite3 mta.db
SQLite version 3.35.4 2021-04-02 15:20:15
Enter ".help" for usage hints.
sqlite> create table data("C/A" TEXT,"UNIT" TEXT, "SCP" TEXT, "STATION" TEXT, "L
INENAME" TEXT, "DIVISION" TEXT, "DATE" TEXT,"TIME" INTEGER, "DESC" TEXT,"ENTRIES
" INTEGER, "EXITS" INTEGER);
sqlite> .mode csv
sqlite> .import /Users/nadia/Downloads/t1.csv data
sqlite> .import /Users/nadia/Downloads/t2.csv data
```



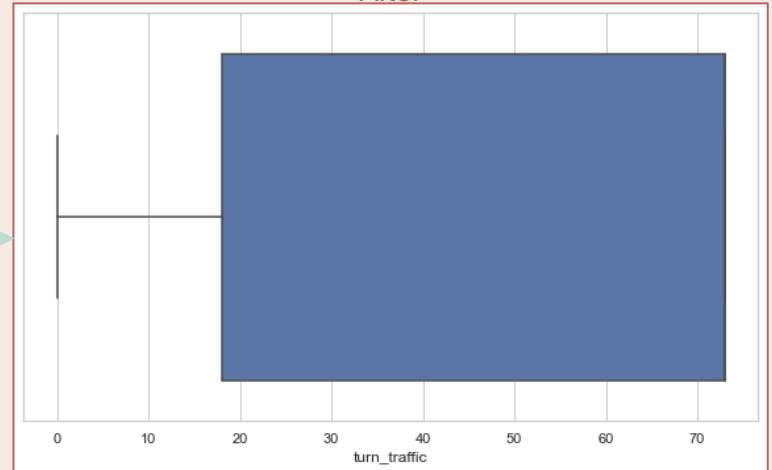
3. Cleaning the data :

- Checking for nulls, outliers and duplicates then drop them .
- Combining Date and time columns .
- Getting total turnstile traffic out of entries and exits .
- Drop unnecessary column .
- Add week days column .

Before

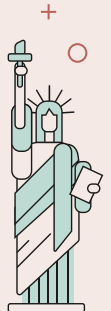
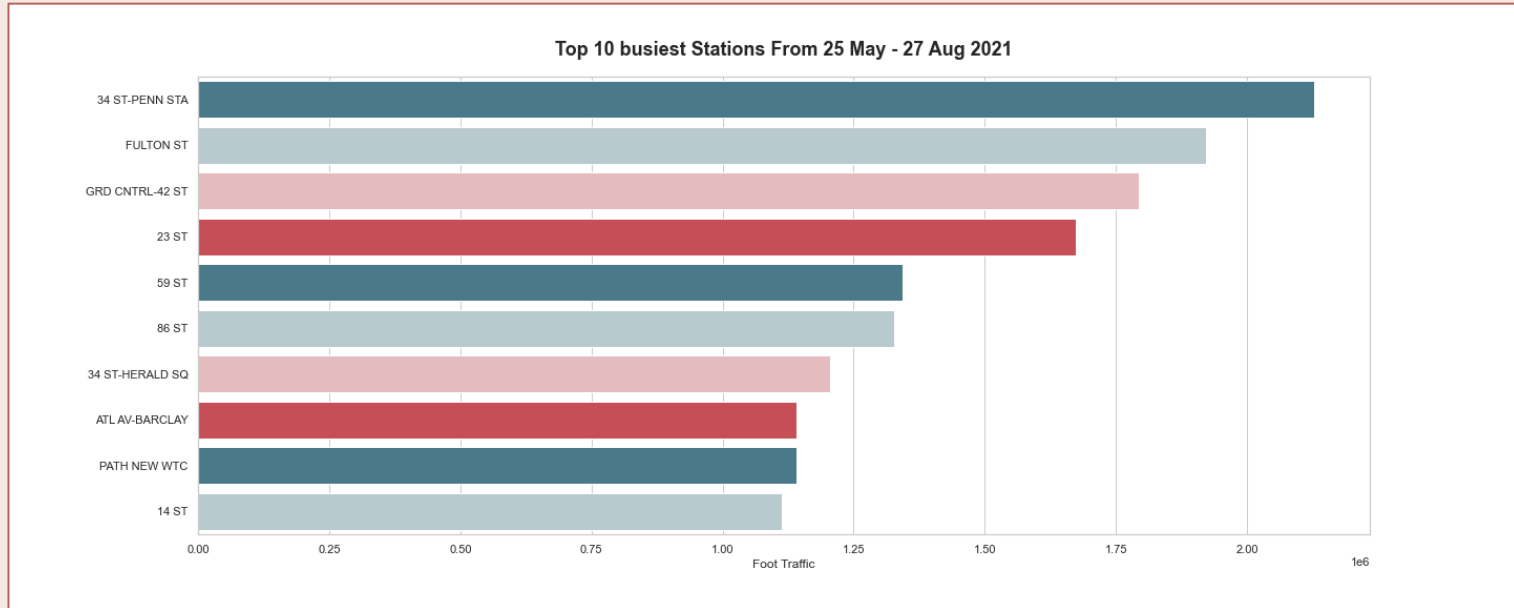


After

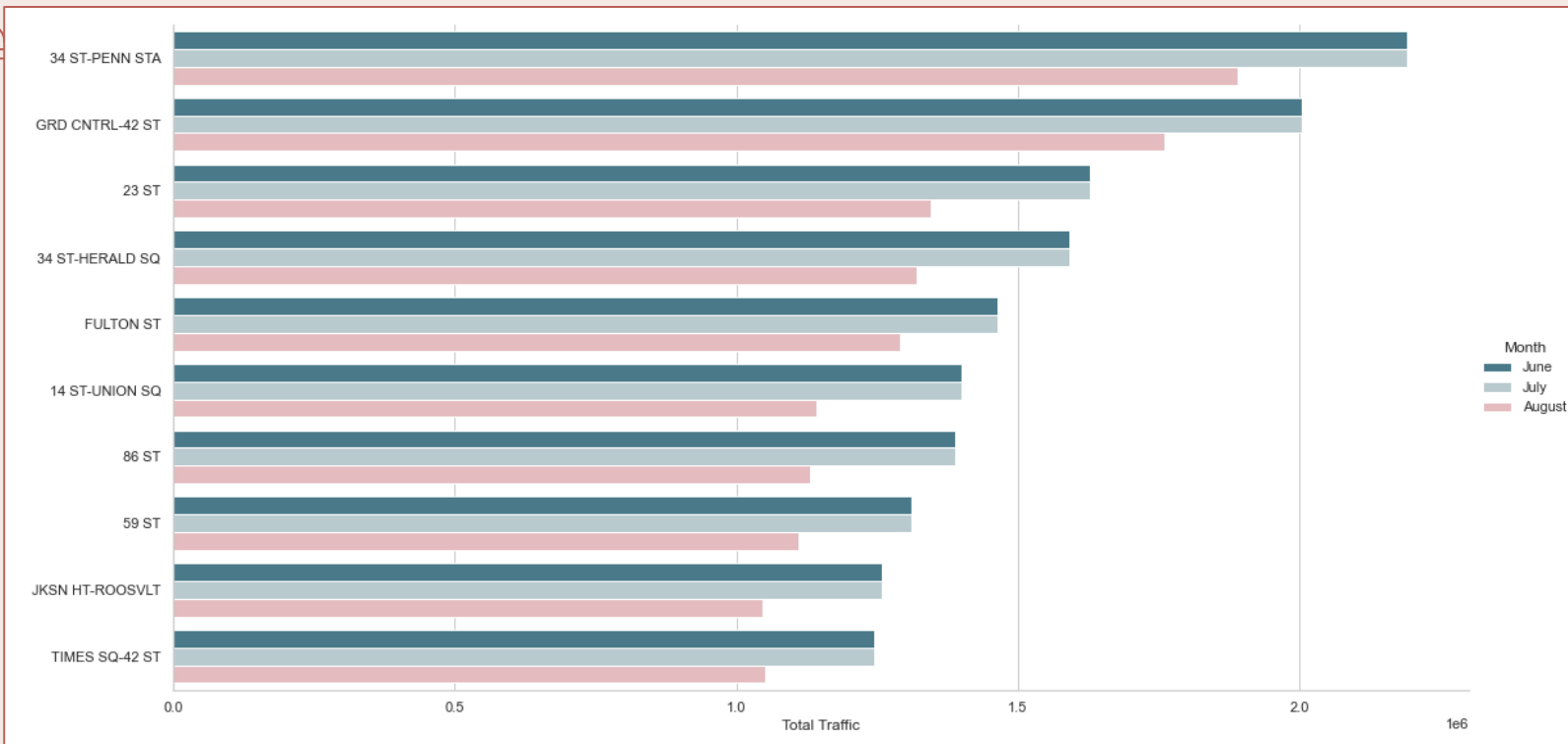


Results

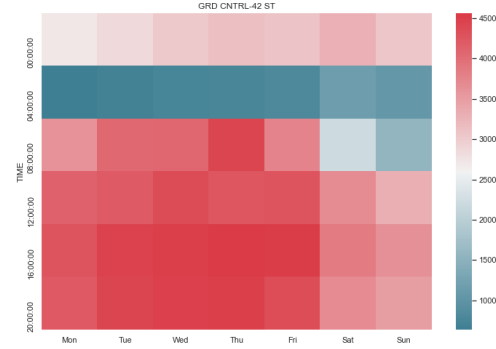
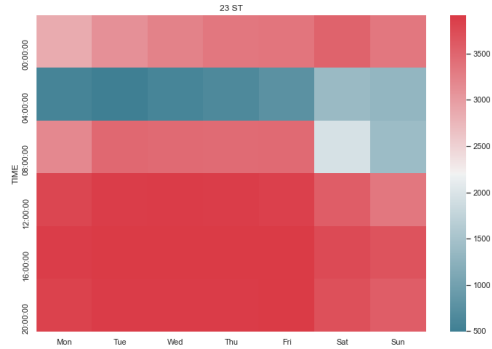
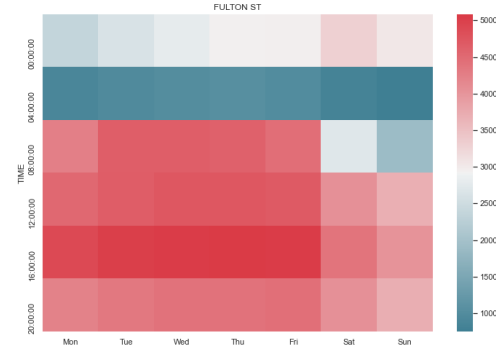
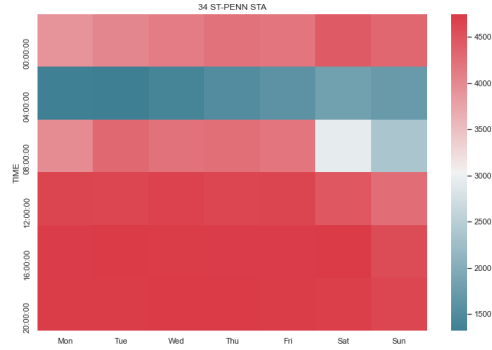
+ 1- The top busiest stations by daily average



2- The busiest stations for each month



3- Busiest Hours for top 4 stations in terms of traffic

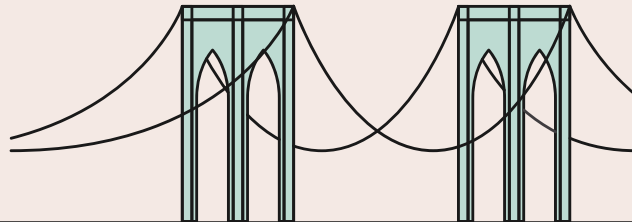


Recommendations

I recommend targeting stations that have a high traffic to maximize the number of sign-ups people to attend the gala. Based on this criteria :

- **34 St-Penn STA** from Monday through Friday between (8:00 am – 8:00 pm) & Saturday between (12:00 pm – 8:00 pm) .
- **FULTON ST** from Monday through Friday between (12:00 pm – 4:00 pm) .
- **23 ST** from Monday through Friday between (8:00 am – 8:00 pm) .
- **GRD CNTRL-42 ST** from Monday through Friday from (12:00 pm – 8:00 pm) & Thursday from (4:00 am – 8:00 am) .





Thank you

for your kind attention

