

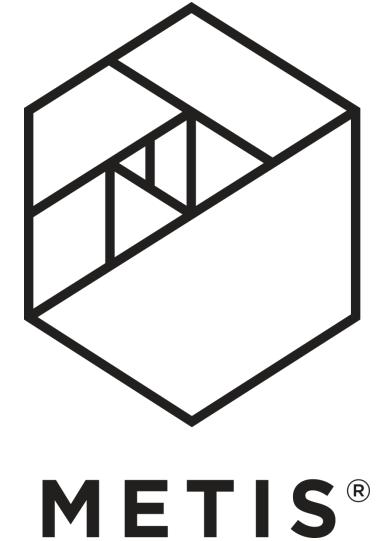
# MTA Stations Traffic Analysis

## For MTA Advertising

project for Metis EDA Bootcamp

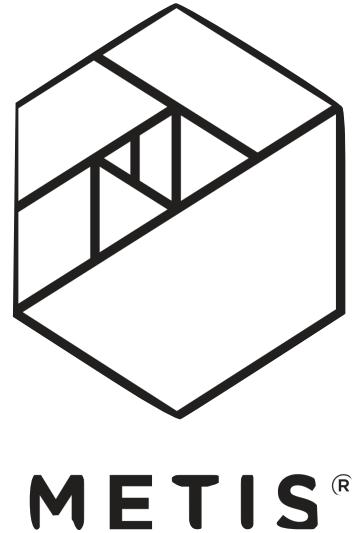
by Krystian Krystkowiak, 2022

# Introduction



- MTA Advertising would like to introduce new experimental advertising panels at their stations.
- First 10 systems will be installed in September 2022, to reach Christmas traffic in 2022.
- GOAL: To identify 10 stations that could fully show potential of new advertising medium during September - December 2022 period.

# Methodology

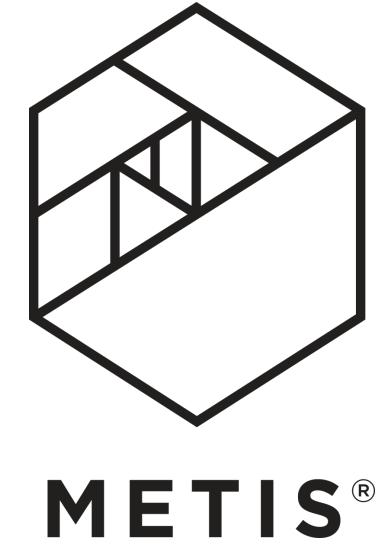


- I used online MTA turnstile data imported from SQL using SQLAlchemy.

C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS	
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	08/31/2019	00:00:00	REGULAR	7183242	2433142
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	08/31/2019	04:00:00	REGULAR	7183258	2433149
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	08/31/2019	08:00:00	REGULAR	7183278	2433176
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	08/31/2019	12:00:00	REGULAR	7183393	2433262
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	08/31/2019	16:00:00	REGULAR	7183572	2433312

- I considered 3 periods: Sept-Dec 2021, to analyse last years traffic. Jan-May 2022, for latest trend. Sept-Dec 2019, as traffic data from before pandemic traffic drop may add more context.
- I focused on entries, as it is group spending more time at the station, therefore longer exposed to advertisement.

# Methodology

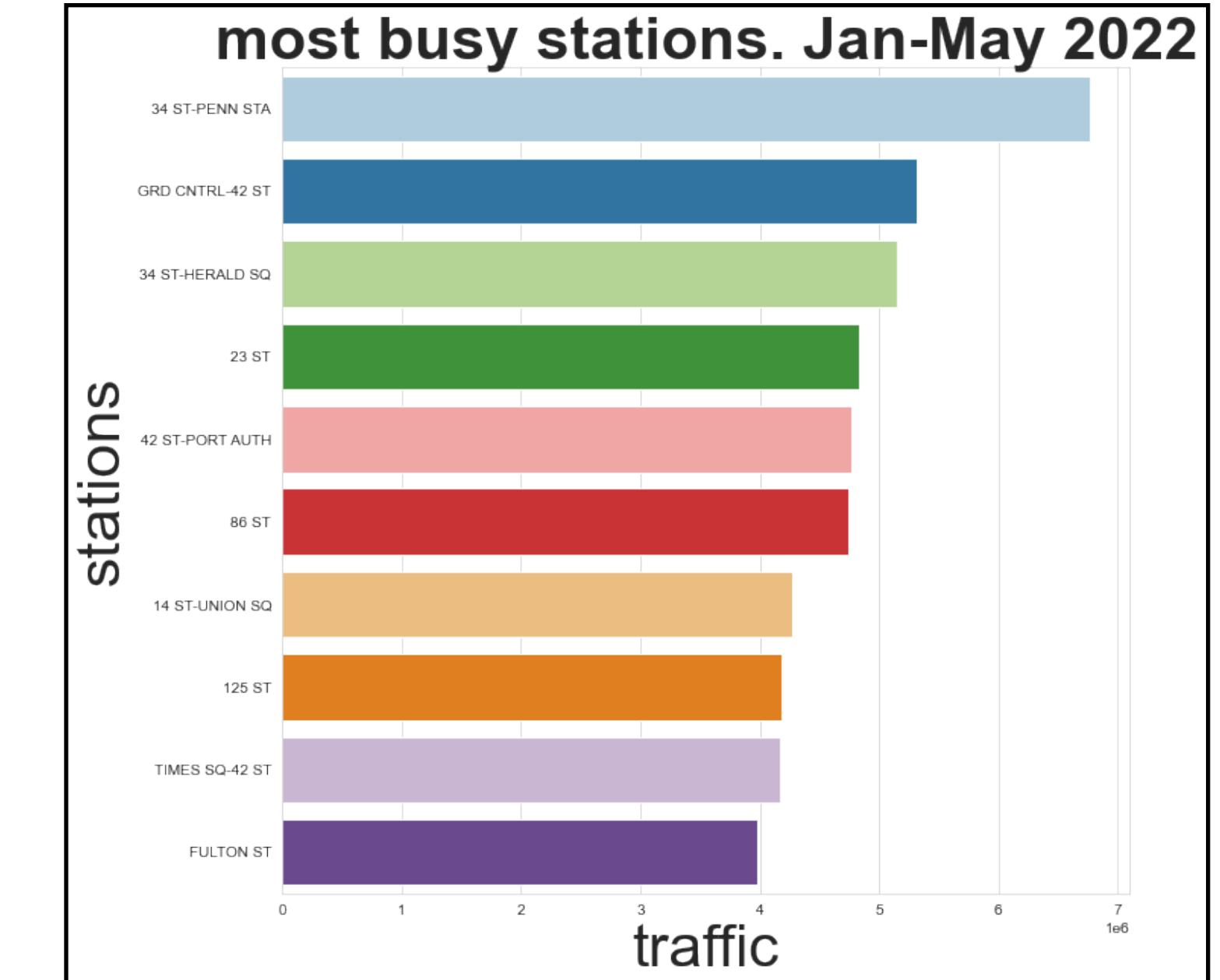
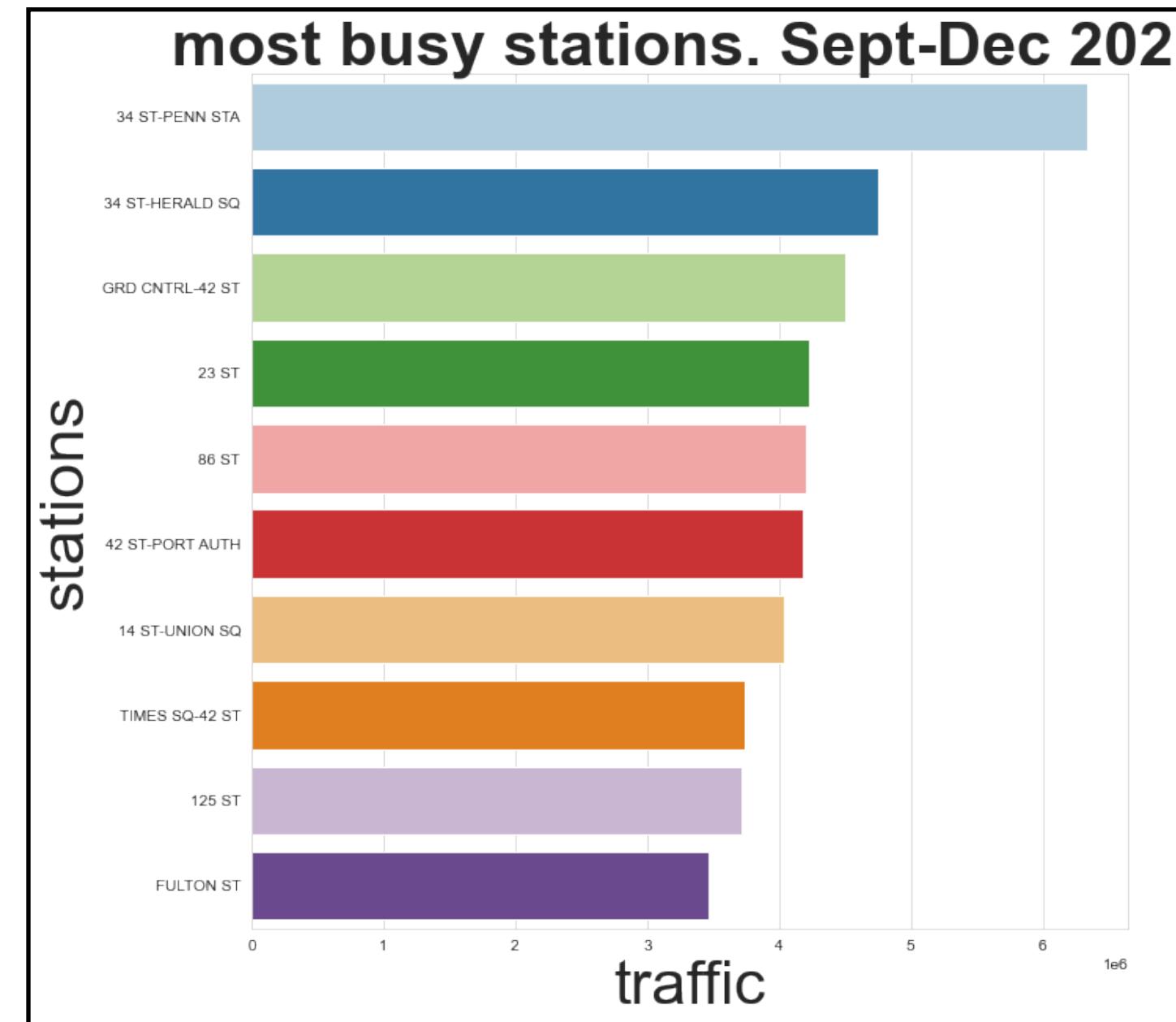
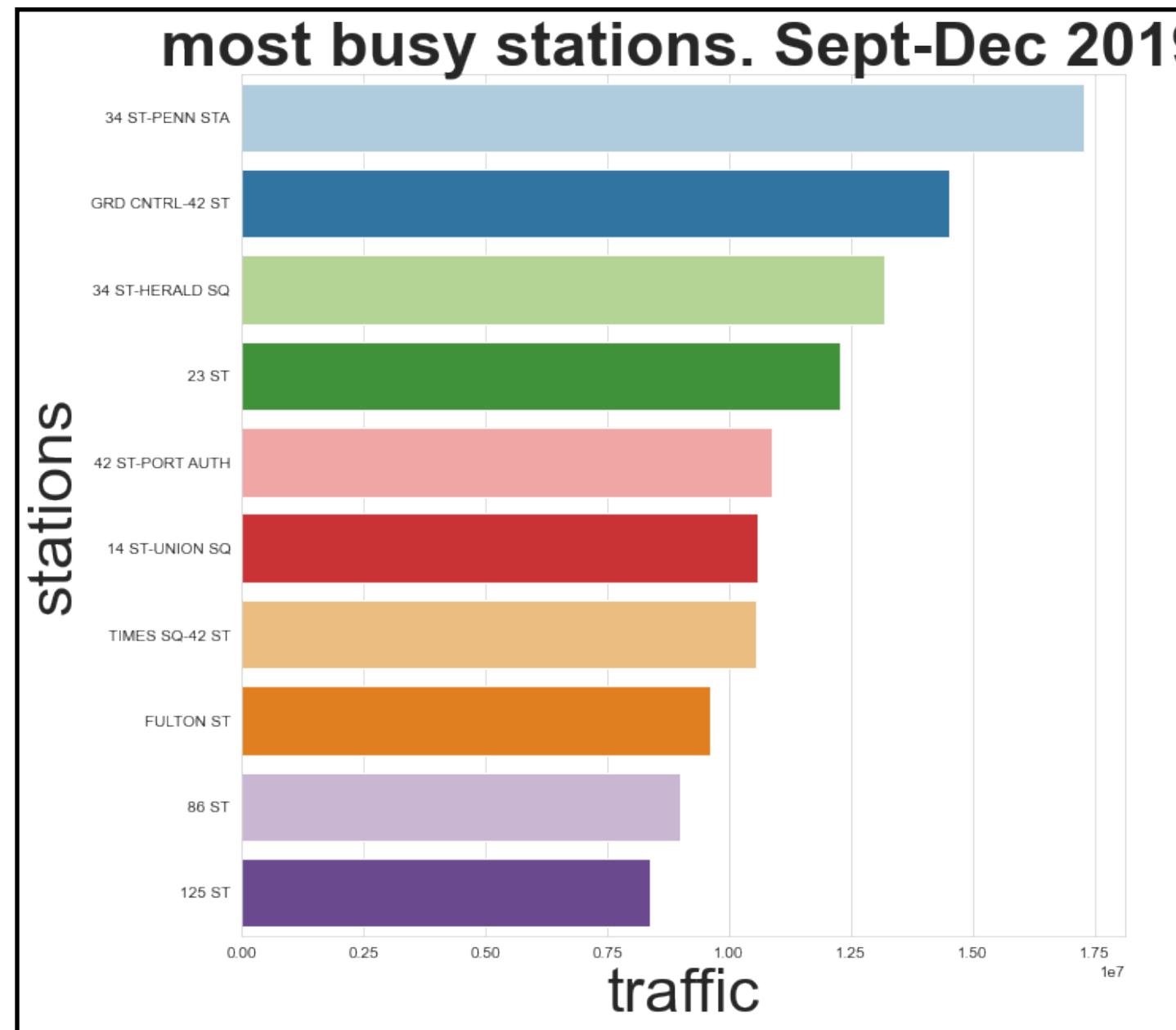
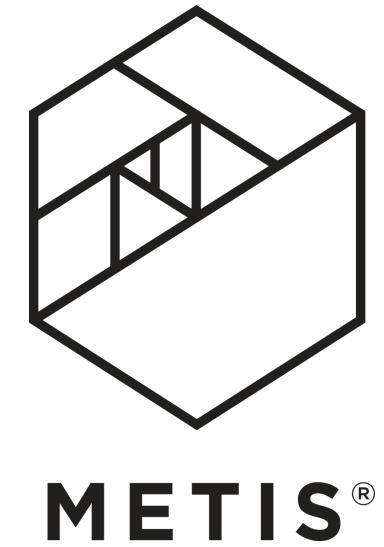


- I cleaned the data of errors and duplicates.
- My processing was based on daily number of entries per turnstile.
- I calculated it by finding the difference between maximum day record and maximum record day before.

```
def get_daily_counts(row, max_counter):  
    counter = row["ENTRIES"] - row["PREV_ENTRIES"]  
    if counter < 0:  
        counter = -counter  
    if counter > max_counter:  
        counter = min(row["ENTRIES"], row["PREV_ENTRIES"])  
    if counter > max_counter:  
        return 0  
    return counter  
  
turnstile_daily["DAILY_ENTRIES"] = turnstile_daily.apply(get_daily_counts, axis=1, max_counter=50000)
```

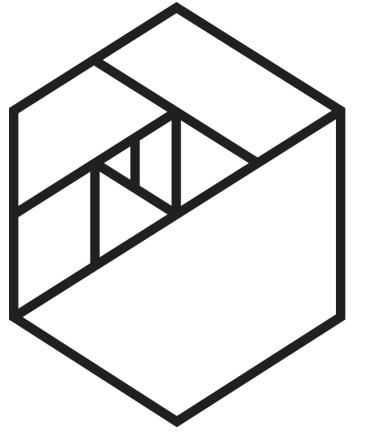
- Reason for negative counts could be reset of the turnstile counter.

# Results

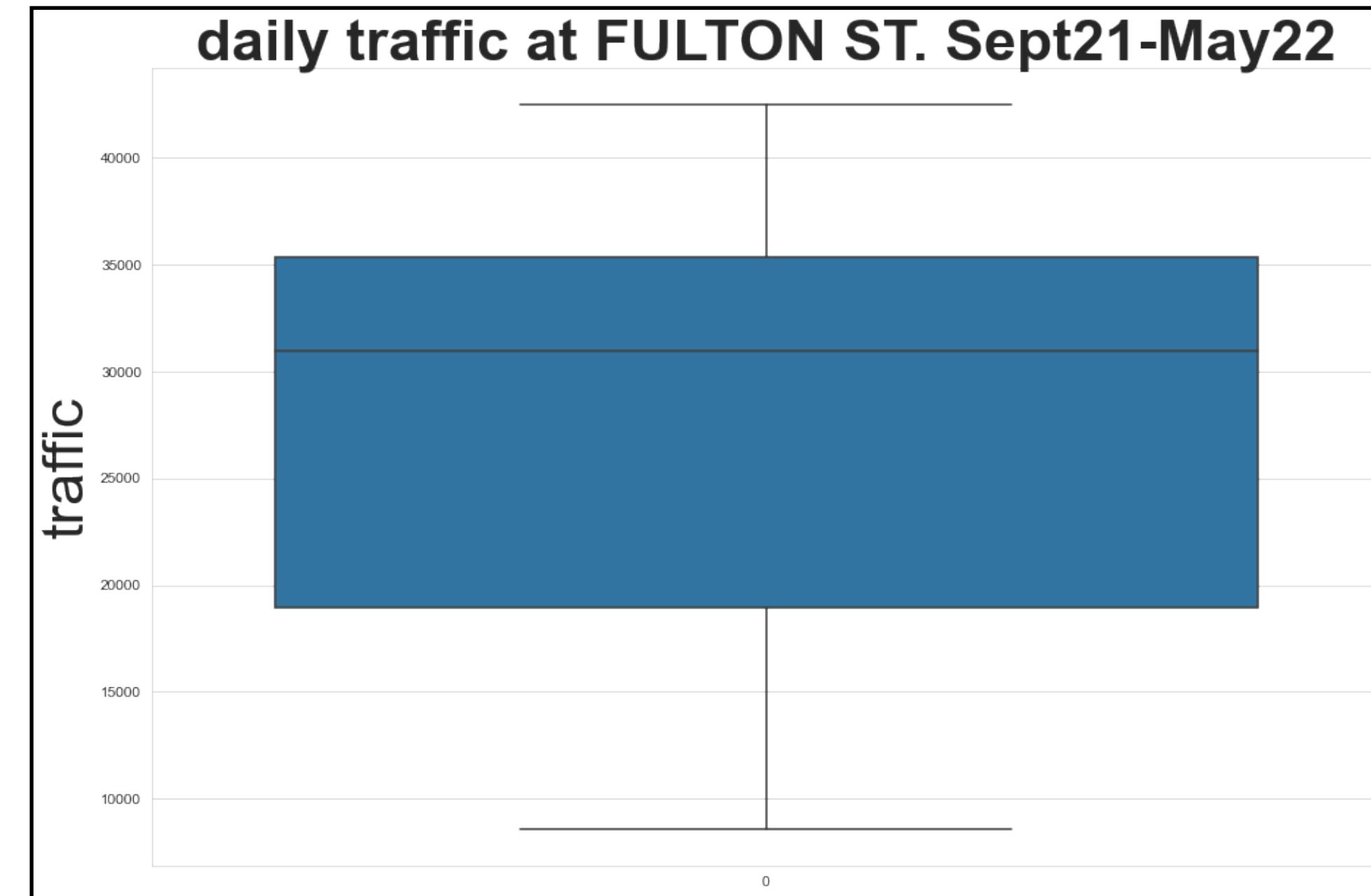
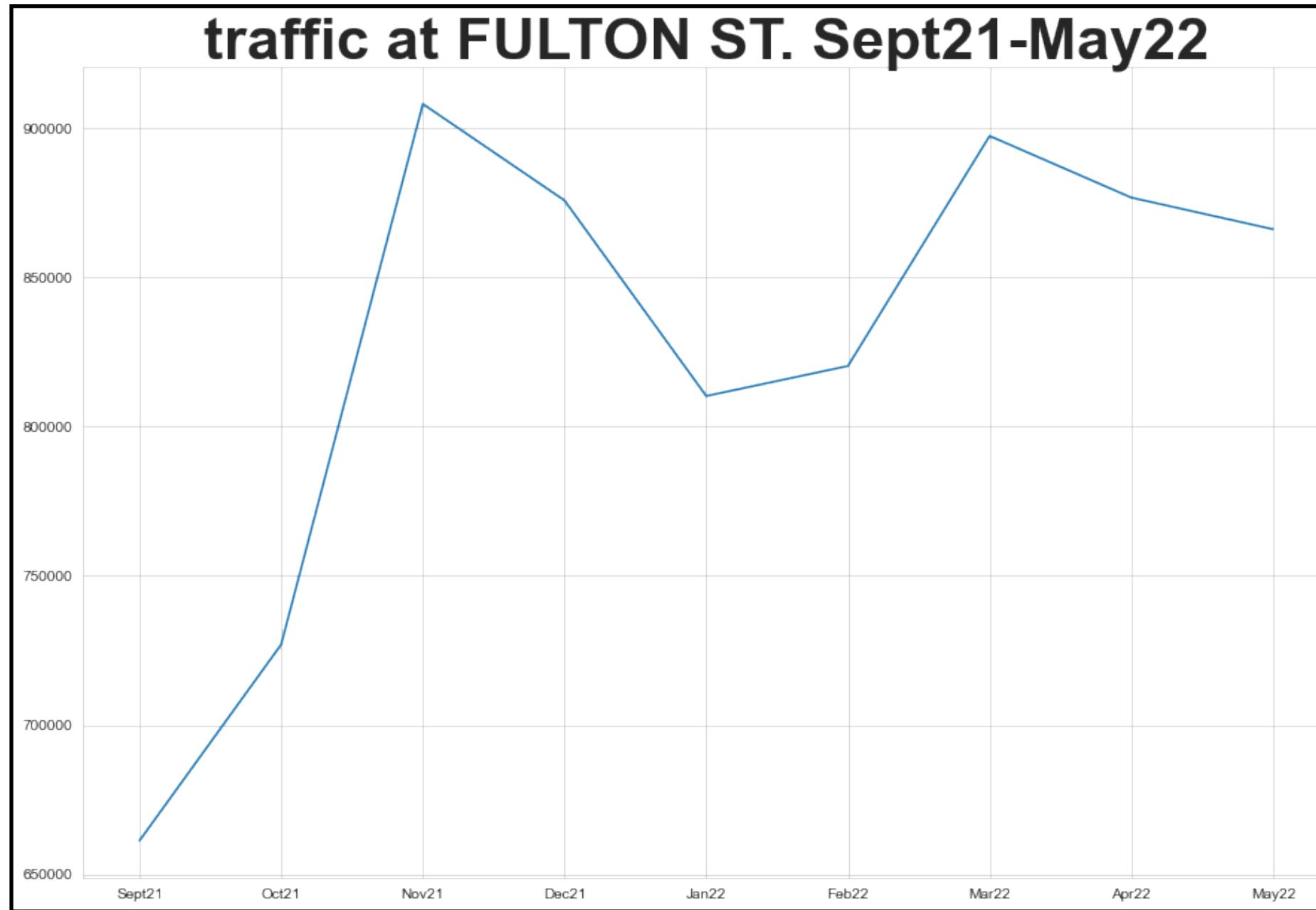


- Most busy stations in specified periods.
- All 3 list contain the same stations. 34 ST-PENN STA is significantly busier than other stations.

# Results

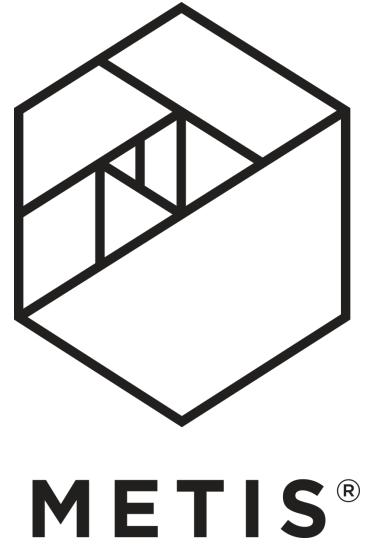


METIS®



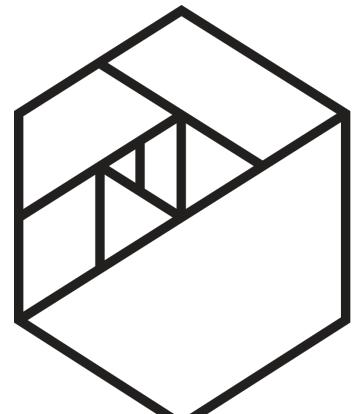
- Monthly traffic and daily traffic at FULTON ST, bottom of top 10 station list.
- Trend shows that traffic is stable and most probably will rise.

# Conclusions



- Identified stations may fully show potential of new advertising medium during September - December 2022. I would recommend installation of new advertising panels at all these locations.
- In all 3 analysed periods - before pandemic, end of the 2021 and first 5 months of 2022 - the same 10 MTA stations had most entries, indicating that the most people were waiting for the transport, making them most exposed to advertisement.
- MTA traffic dropped significantly after pandemic and is slowly recovering.

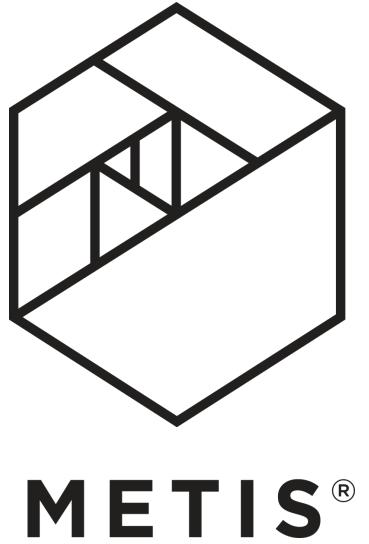
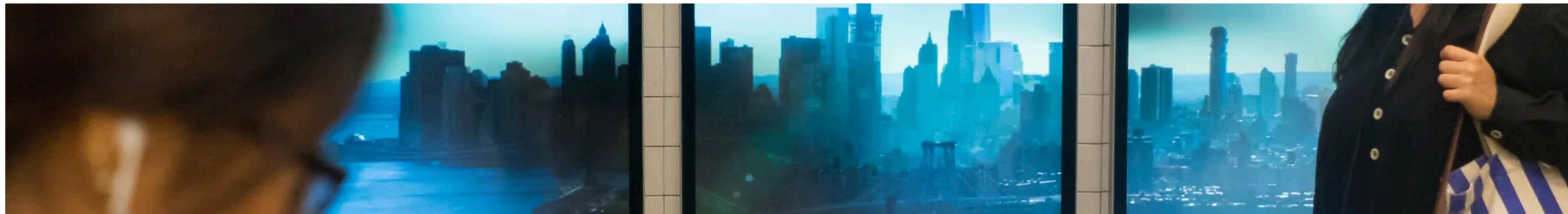
# Future Work



METIS®

- I would take closer look at traffic during pandemic periods. That could give insight into connections essential for the city.
- I would analyse exits counters, for deeper understanding of the NY traffic.
- I would create map visualisation of the most busy stations and get more information about population of those areas.





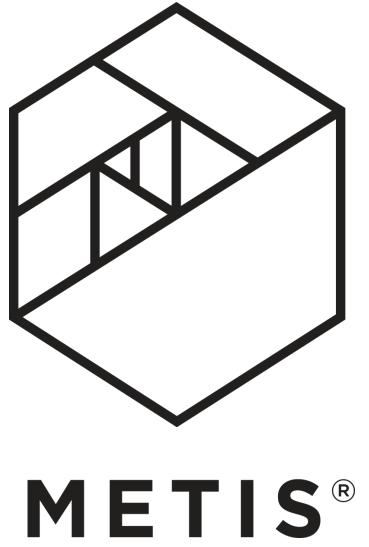
# Thank you!

## Questions?

project for Metis EDA Bootcamp

by Krystian Krystkowiak, 2022

# Sources



- Title photo: [shutterstock.com](https://www.shutterstock.com)
- Last photo: [www.nymag.com](https://www.nymag.com)
- MTA map: MTA