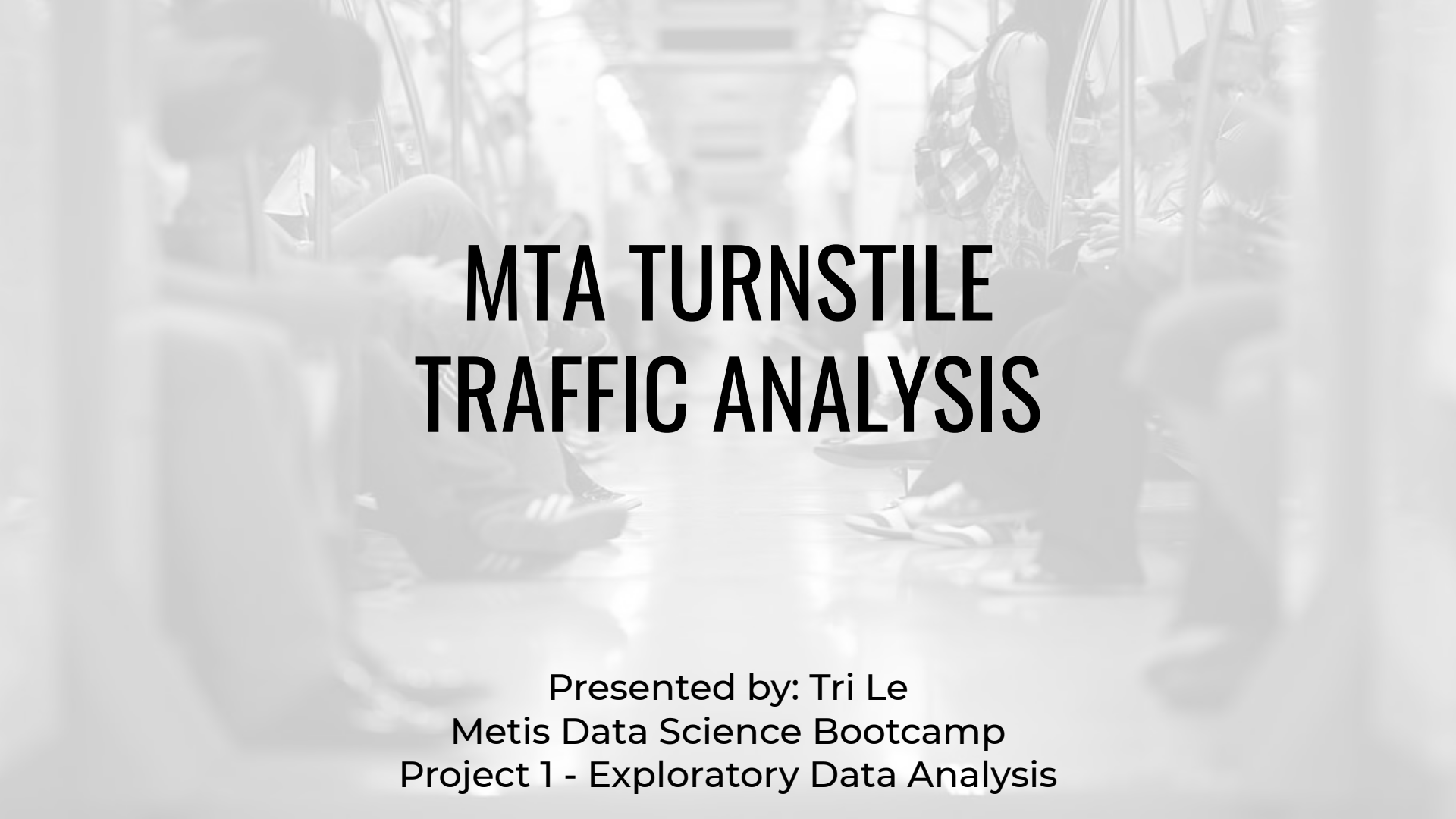


MTA TURNSTILE TRAFFIC ANALYSIS

Presented by: Tri Le
Metis Data Science Bootcamp
Project 1 - Exploratory Data Analysis

Introduction

Background and Motivation:

Bagels & Brew is a café that features breakfast items, with considerations to open a new location in New York City. Planned for spring of 2022, the owners are hoping commuting traffic will resume close to pre-pandemic levels. More importantly, Bagels & Brew is seeking to make an informed decision about ideal locations to set up shop, with particular focus on weekday morning hours as it closely aligns with the purpose of their business.

Objective:

Determine which locations in NYC that experience daily high-volume traffic

Goals:

Recommend ideal business locations to Bagels & Brew based on our findings

Methodology
Approach

Why NYC MTA Data?

- Publicly available
- Detailed turnstile information
- Entry/Exit counts, dates, timestamps, and station names/locations

Which months, days, and hours are relevant to our analysis?

- Spring, pre-pandemic — JAN-MAR, 2019
- Peak morning business hours — 6AM-10AM, MON-FRI

Objectives (Revisited)

- Which MTA stations experience the most traffic, during peak morning weekday hours?
- Are some weekday mornings busier than others?

Methodology
Tools & Data

Languages:

- Python
- SQL

Libraries:

- SQLITE and SQLAlchemy — Data importing and querying
- Pandas and NumPy — Data manipulation
- Matplotlib and Seaborn — Data visualization

A Closer look at the Data obtained from the MTA:

- JAN-MAR, 2019 (Pre-pandemic)
- 13 Weeks of data
- Turnstile counts, cumulatively captured every 4 hours

- Rows: 2,639,743
- Stations: 378
- Turnstiles: 4,906

	STATION	DATE	TIME	ENTRIES	EXITS	TURNSTILE	DAY	HOUR	DAILY_ENTRIES	DAILY_EXITS	TOTAL_TRAFFIC
0	1 AV	01/01/2020	07:00:00	15313672	17120097	H007-R248-00-00-00	Wednesday	7	2140.0	2897.0	5037.0
1	1 AV	01/01/2020	07:00:00	60963028	38072778	H007-R248-00-00-01	Wednesday	7	2397.0	1397.0	3794.0
2	1 AV	01/01/2020	07:00:00	370821703	387998731	H007-R248-00-03-00	Wednesday	7	813.0	1139.0	1952.0
3	1 AV	01/01/2020	07:00:00	2572792	1111648	H007-R248-00-03-01	Wednesday	7	692.0	401.0	1093.0
4	1 AV	01/01/2020	07:00:00	6582098	554503	H007-R248-00-03-02	Wednesday	7	914.0	139.0	1053.0

Methodology
Metrics

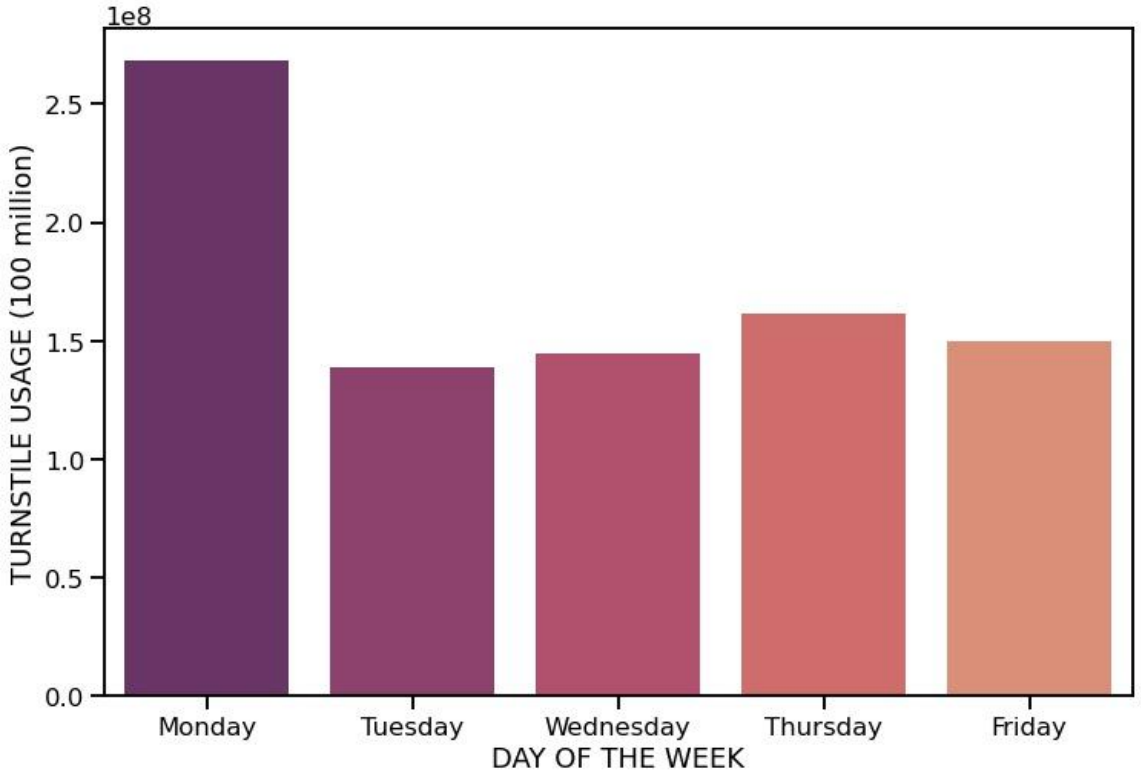
Station	Represents the station name the turnstile is located
Turnstile	Represents a individual turnstile identifier *(combination of C/A, UNIT, SCP*)*
Day	Represents the day of the week
Hour	Represents the hour as a single numerical digit
Daily_Entries	Represents the Entries calculated during a 4-hour period
Daily_Exits	Represents the Exits calculated during a 4-hour period
Total_Traffic	Represents sum of Daily_Entries and Daily_Exits during a 4 hour period

	STATION	DATE	TIME	ENTRIES	EXITS	TURNSTILE	DAY	HOUR	DAILY_ENTRIES	DAILY_EXITS	TOTAL_TRAFFIC
0	1 AV	01/01/2020	07:00:00	15313672	17120097	H007-R248-00-00-00	Wednesday	7	2140.0	2897.0	5037.0
1	1 AV	01/01/2020	07:00:00	60963028	38072778	H007-R248-00-00-01	Wednesday	7	2397.0	1397.0	3794.0
2	1 AV	01/01/2020	07:00:00	370821703	387998731	H007-R248-00-03-00	Wednesday	7	813.0	1139.0	1952.0
3	1 AV	01/01/2020	07:00:00	2572792	1111648	H007-R248-00-03-01	Wednesday	7	692.0	401.0	1093.0
4	1 AV	01/01/2020	07:00:00	6582098	554503	H007-R248-00-03-02	Wednesday	7	914.0	139.0	1053.0

**See Appendix*

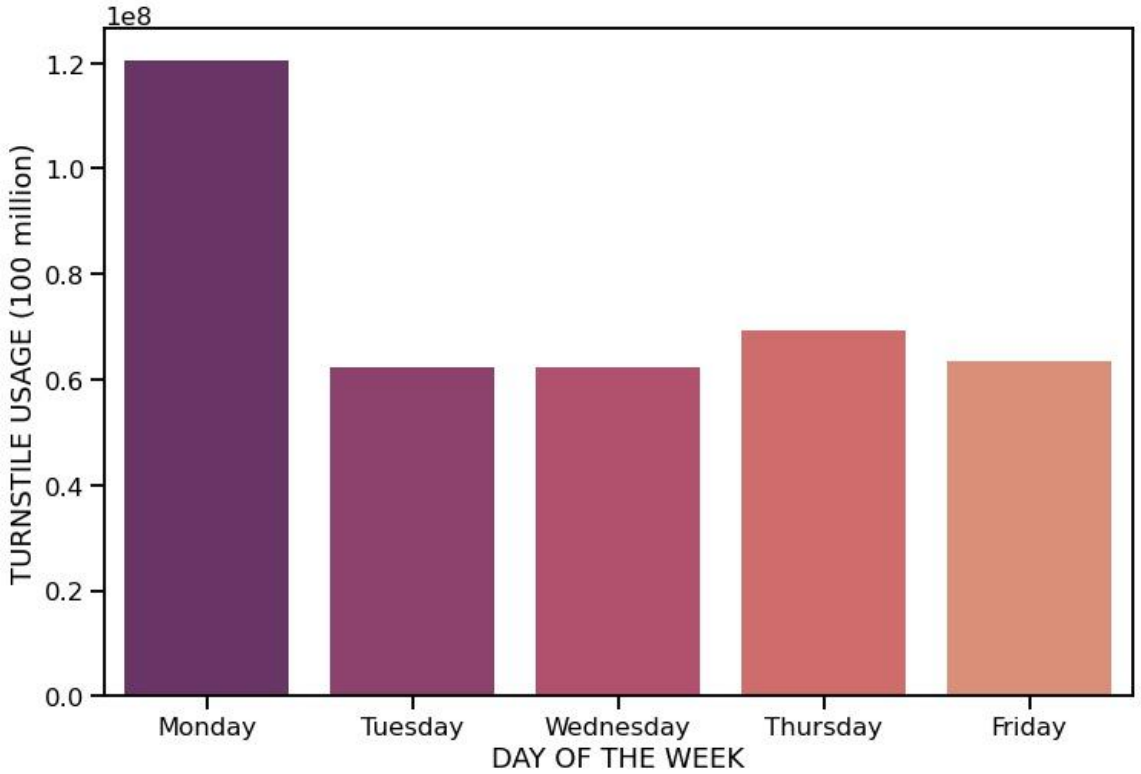
Results

Turnstile Traffic (Total) by Weekday (4AM-12PM)



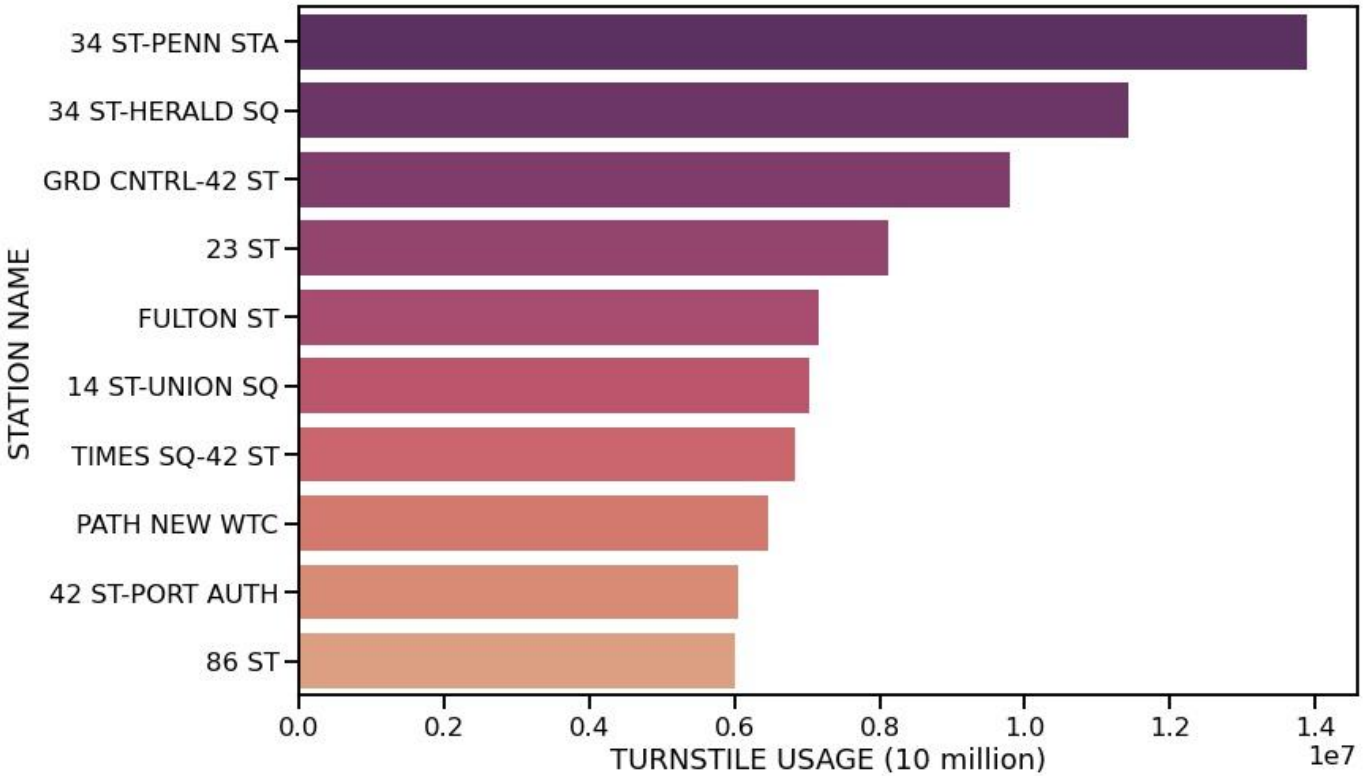
Results

Turnstile Traffic (Exits Only) by Weekday (4AM-12PM)



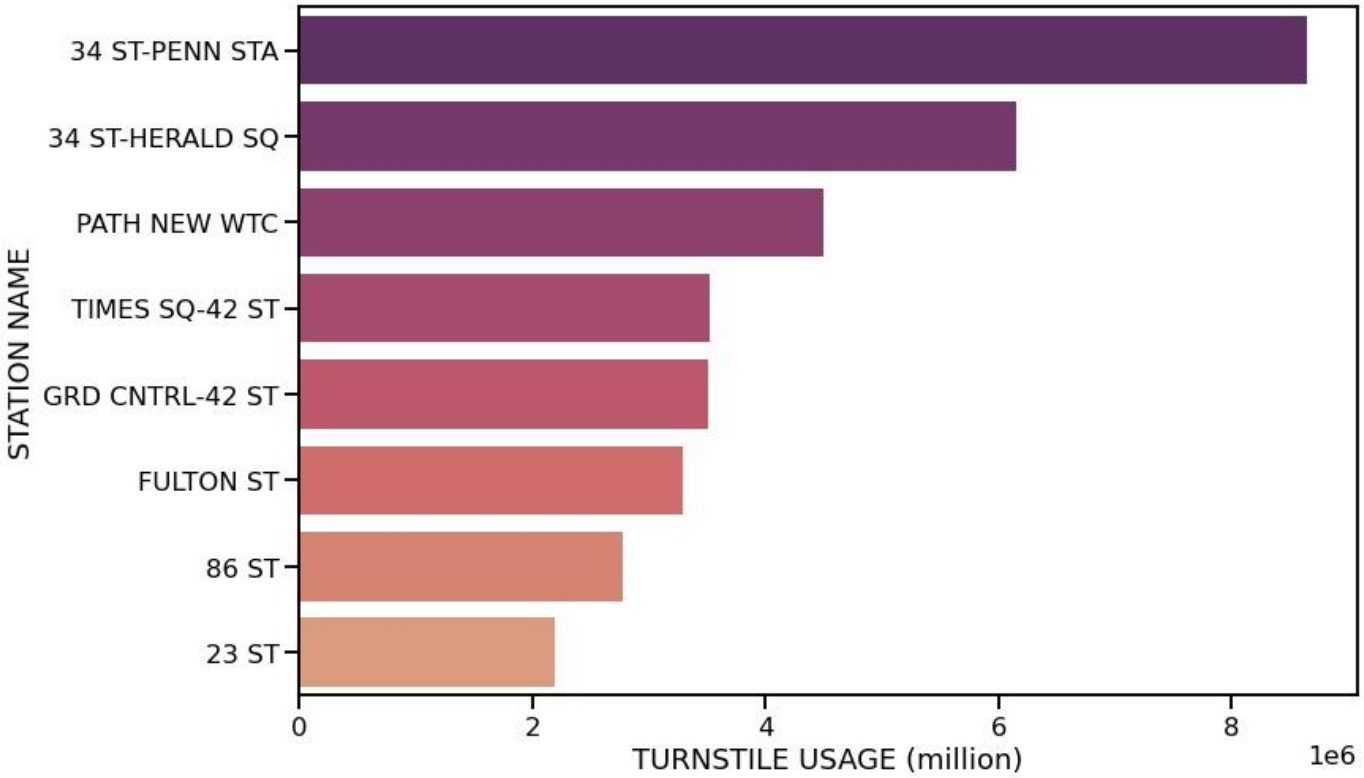
Results

Stations with Most Exit Traffic (Weekdays, 4AM-12PM)



Results

Stations with Most Exit Traffic (Weekdays, 6AM-10AM)



Conclusions
Recommendations

34 St - Penn Station

34 St | Herald Square

World Trade Center

Times Square | 42 St

Grand Central | 42 St

Future Considerations

Potential Further Analysis:

- Deep-dive into traffic by the hour
- Examine other day parts, and weekends
- Explore other seasons
- Geo-mapping of population density
- Geo-mapping of competitors

- Offer dynamic pricing or specials
- Optimize business hours
- Anticipate demand and customer traffic
- Identify neighborhood “hot spots”
- Avoid market saturation

Thank You!
Questions?

Appendix

MTA DataFrame Sample Prior to Manipulation and Analysis

	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
0	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/23/2019	00:00:00	REGULAR	6989774	2370411
1	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/23/2019	04:00:00	REGULAR	6989795	2370413
2	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/23/2019	08:00:00	REGULAR	6989813	2370436
3	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/23/2019	12:00:00	REGULAR	6989924	2370512
4	A002	R051	02-00-00	59 ST	NQR456W	BMT	03/23/2019	16:00:00	REGULAR	6990200	2370573

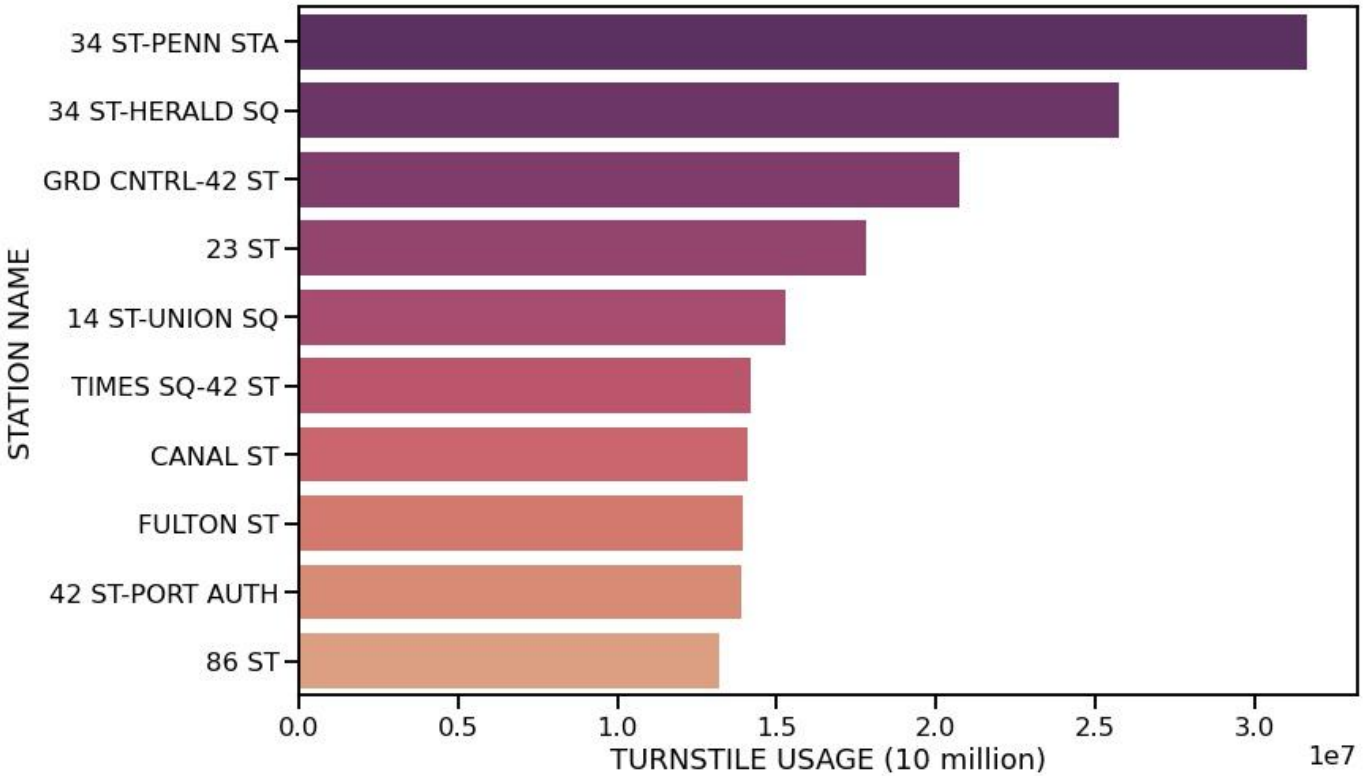
Appendix

MTA Field Name Definitions

FIELD NAME	DESCRIPTION
C/A	Control Area (A002)
UNIT	Remote Unit for a station (R051)
SCP	Subunit Channel Position represents a specific address for a device (02-00-00)
STATION	Represents the station name the device is located at (59 ST)
LINENAME	Represents all train lines that can be boarded at this station (NQR456W)
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	Represents the type of audit event (“REGULAR” or “RECOVR_AUD”). Scheduled events occur every 4 hours
ENTRIES	The cumulative entry register value for a device
EXITS	The cumulative exit register value for a device

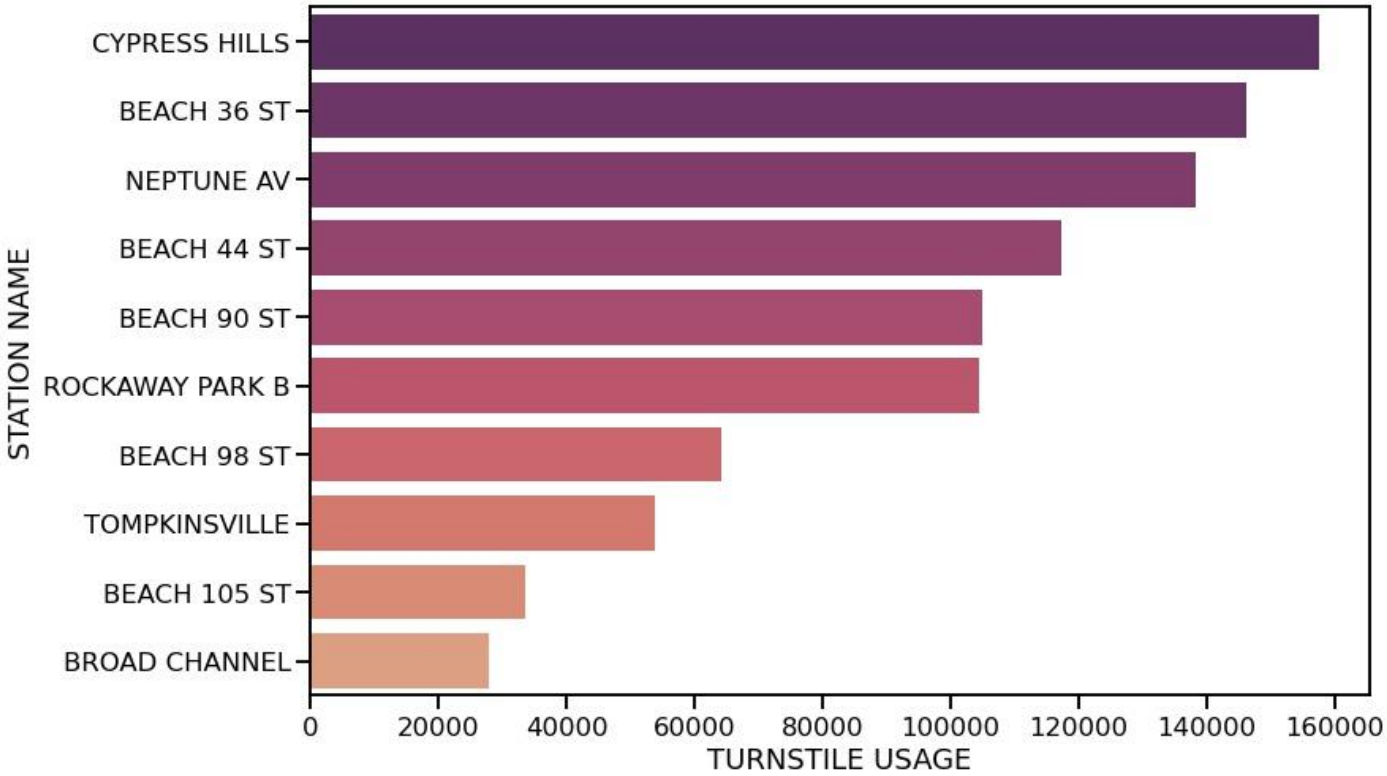
Appendix

Busiest stations during weekday morning total traffic (4AM-12PM)



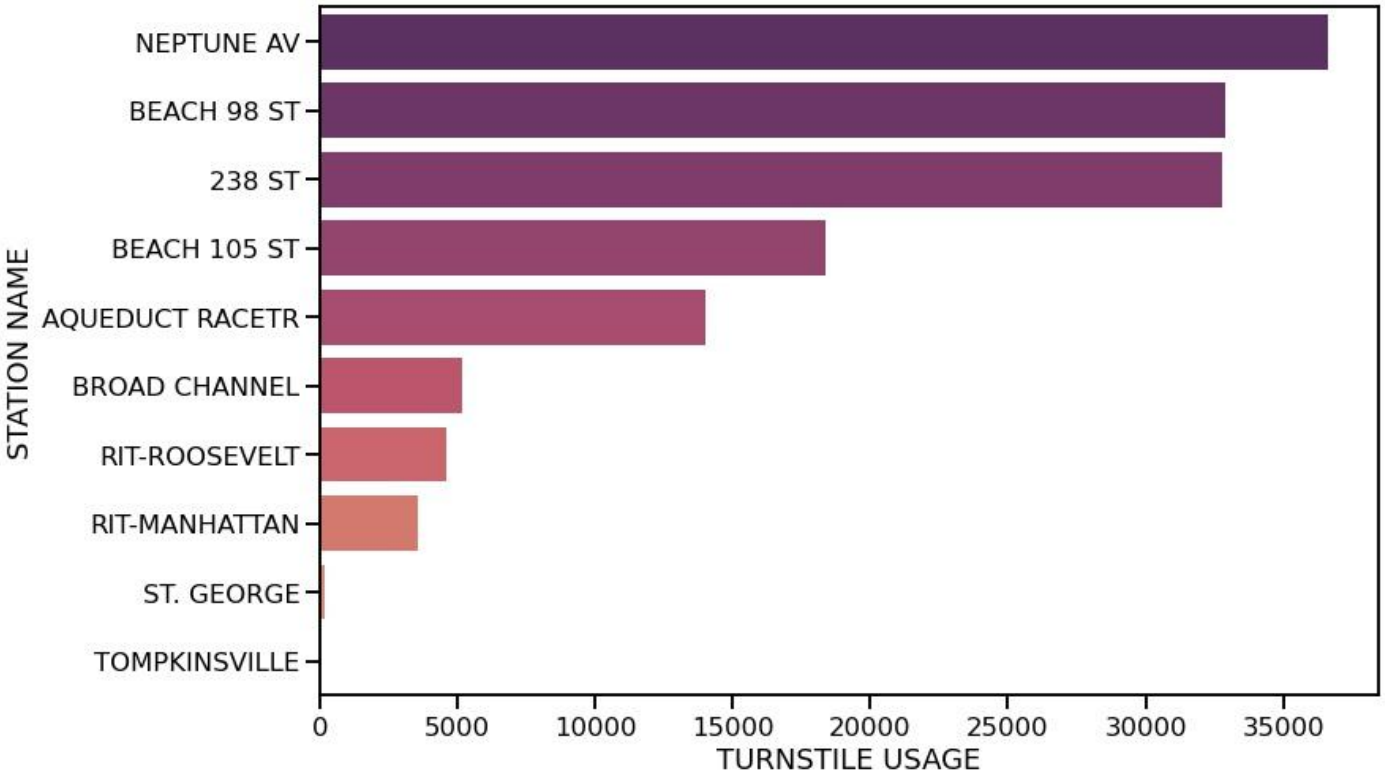
Appendix

Fewest total weekday morning traffic (4AM-12PM)



Appendix

Fewest exit weekday morning traffic (4AM-12PM)



Appendix

Maps of NYC Subway and Tech Companies