MTA TURNSTILE TRAFFIC ANALYSIS

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

Introduction

Background and Motivation:

Bagels & Brew, a restaurant that focuses on breakfast offerings, with considerations to open a new location in New York City, planned for spring of 2022. The owners are seeking to make an informed decision about which location(s) are ideal for their business, with particular focus on morning weekday hours as it closely aligns with breakfast offerings.

Objectives/Goals:

Which subway stations experience the most traffic?
Which days of the week experiences the most traffic?
Which stations experience the most relevant traffic to determine ideal locations for a potential opening of a new Bagels & Brew business?

Languages:
- Python
- SQL

Libraries:
- SQLITE and SQLAlchemy   (for data query and importing)
- Pandas and NumPy   (for data manipulation)
- Matplotlib and Seaborn   (for data visualization)

NYC MTA Turnstile Recordings
- 13 Weeks of data, JAN-MAR 2020
- Individual counts, recorded every 4 hours
- Counts are cumulative

- Rows: 2,677,639
- Stations: 378
- Turnstiles: 4,969

Sample Snapshot of Data

	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

|	
Station	Represents the station name the turnstile is located
Turnstile	Represents a individual, unique 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_Exits	Represents non-cumulative Exits recorded since last capture (occurs every 4 hours)
Total_Traffic	Represents sum of Daily_Entries and Daily_Exits at a specific date/time (occurs every 4 hours)

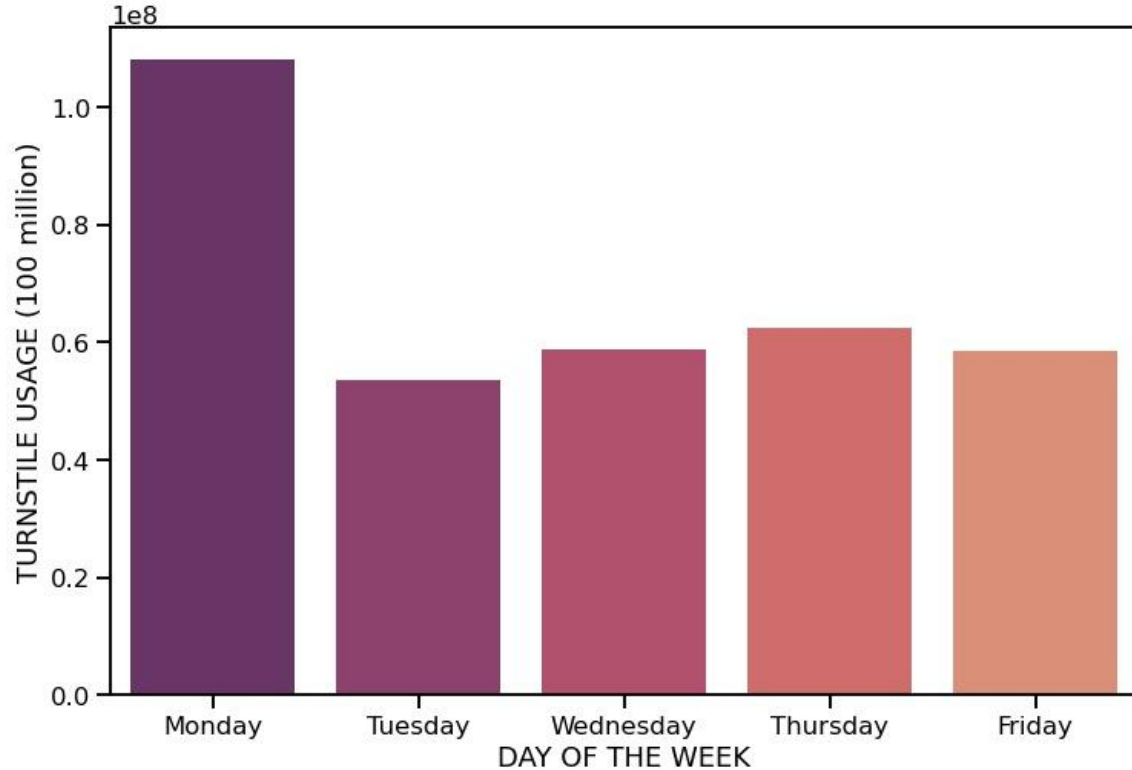# Results
Busiest Stations during the Weekday (4AM-12PM)
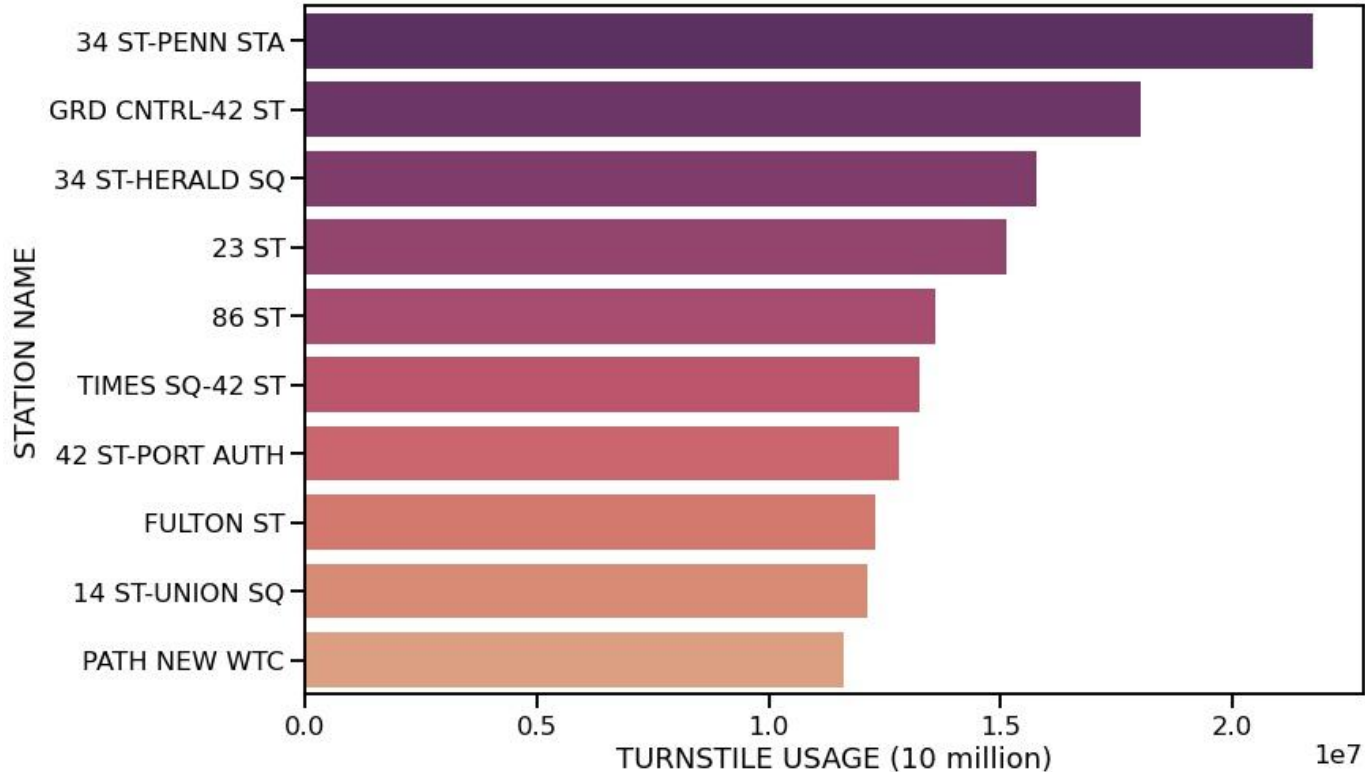
Results
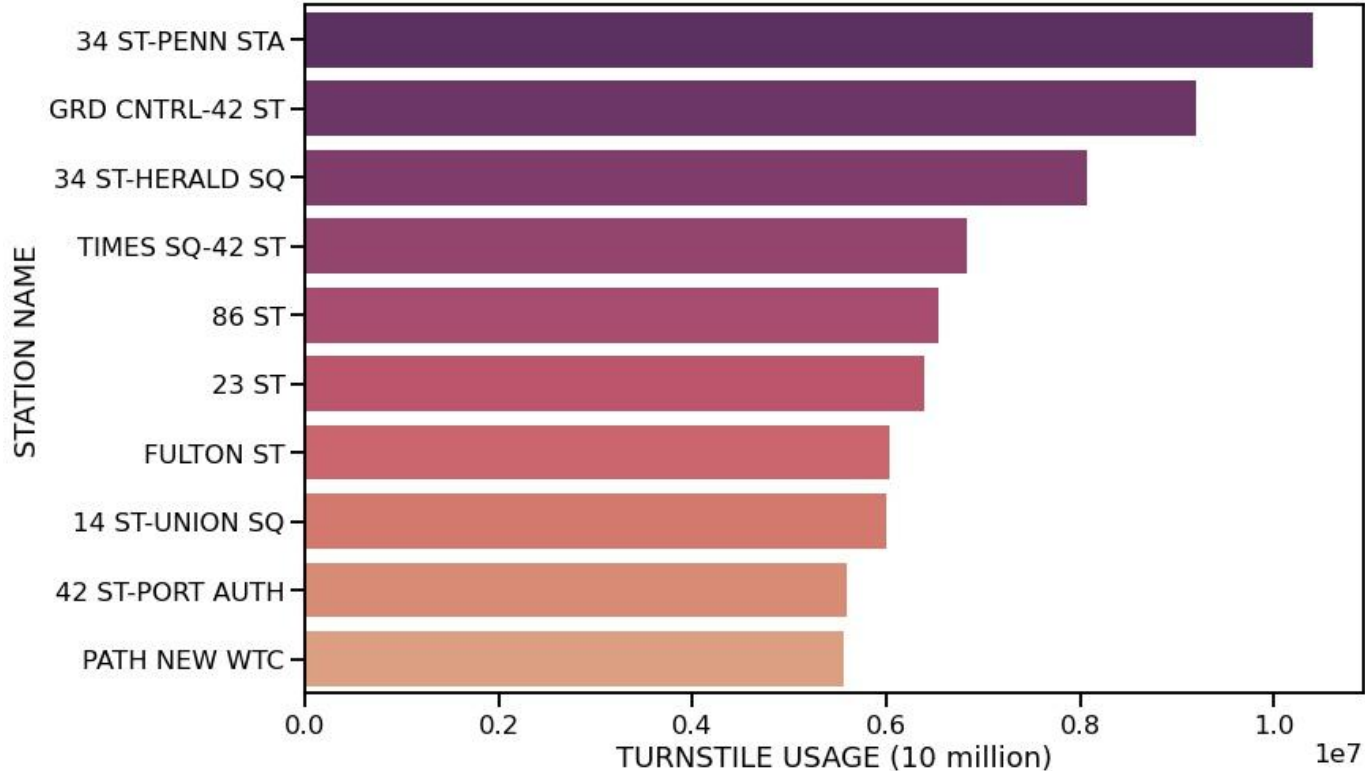## Turnstile Traffic (Total) by Weekday (4AM-12PM)
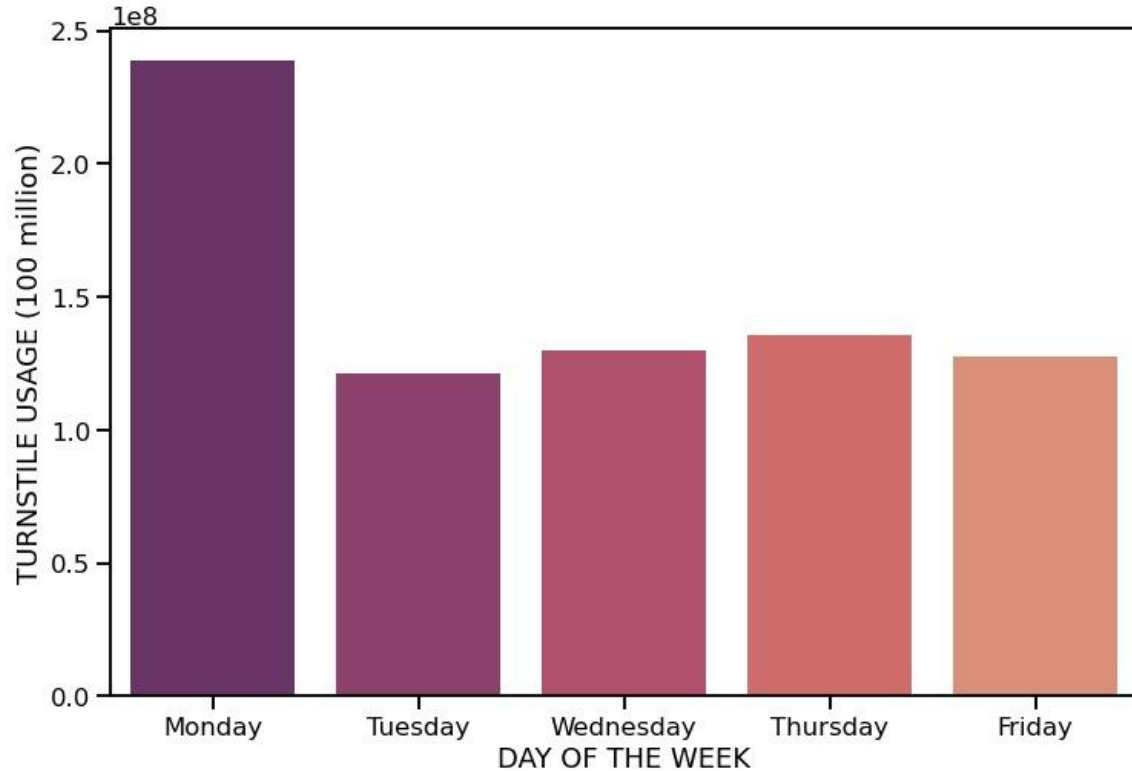
Results
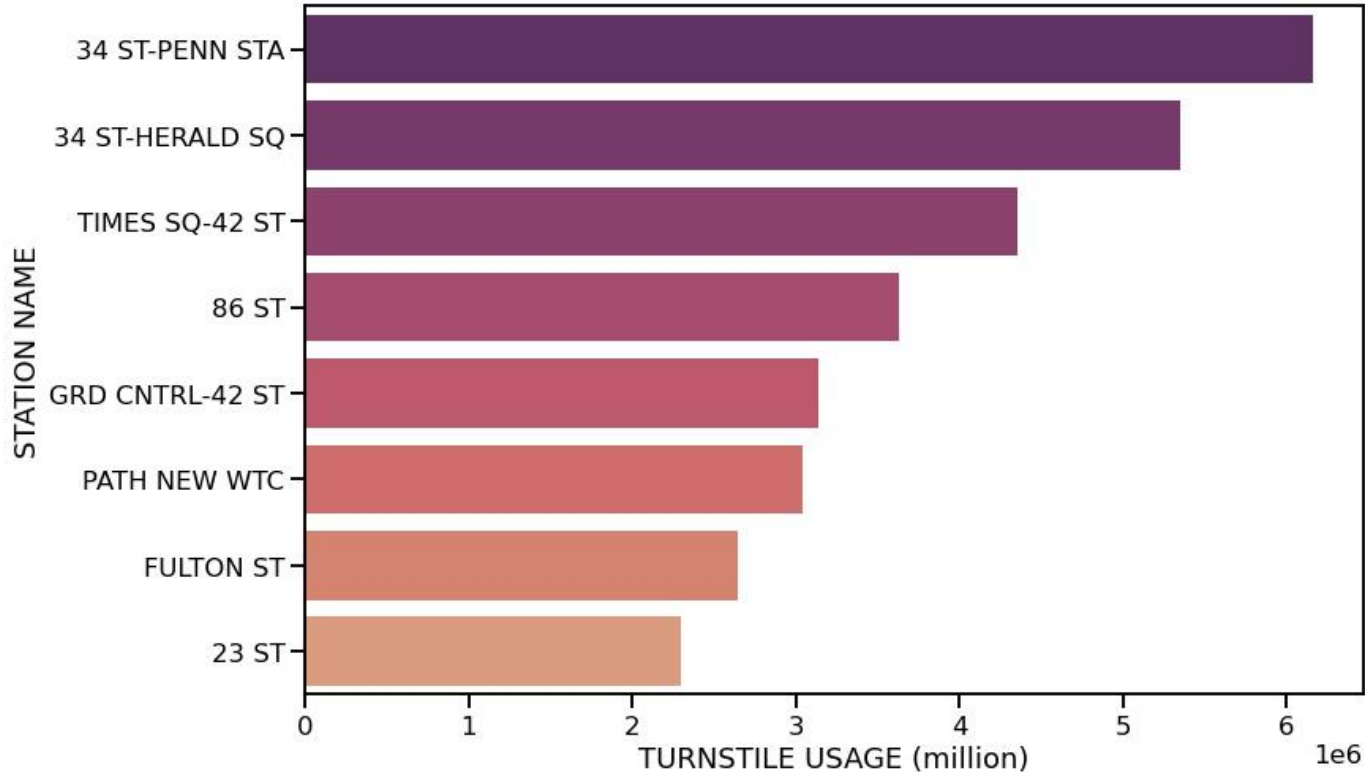## Turnstile Traffic (Exits Only) by Weekday (4AM-12PM)

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

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

placeholder

Future Considerations
placeholder

placeholder

Appendix

placeholder