

Istanbul Data Science Academy Project

MTA Traffic Analysis for WTWY



Presented by

Pınar YILDIZ

Alican Kartal

Muammer Uğur

Özge Mutlu

Effective placement of street teams is critical for attracting more participants to the gala and raising awareness.

Problem



Solution



Analyze MTA subway data to strategically position street teams in areas with high concentrations of potential participants.

Utilizing additional data sources along with MTA subway data, identify the busiest stations, days, and time periods to determine the most suitable locations for maximizing the effectiveness of invitation distribution.

Target



Data Sources



New York City Subway

From January 1st to March 18th 2023

Tools



GeoPandas



pandas

matplotlib



NumPy



seaborn

NYC OpenData
Based on Census 2019

eventbrite

Women & Tech Events in NYC

About the Data

C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
A002	R051	02-00-00	59 ST	NQR456W	BMT	12/31/2022	03:00:00	REGULAR	7811594	2771224
A002	R051	02-00-00	59 ST	NQR456W	BMT	12/31/2022	07:00:00	REGULAR	7811598	2771239
A002	R051	02-00-00	59 ST	NQR456W	BMT	12/31/2022	11:00:00	REGULAR	7811625	2771306
A002	R051	02-00-00	59 ST	NQR456W	BMT	12/31/2022	15:00:00	REGULAR	7811696	2771355
A002	R051	02-00-00	59 ST	NQR456W	BMT	12/31/2022	19:00:00	REGULAR	7811815	2771390

MTA Turnstile Data

Metadata	Column Name	Description
	C/A	Control Area (e.g., A002)
	UNIT	Remote Unit for a station (e.g., R051)
	SCP	Subunit Channel Position represents a specific address for a device (e.g., 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	Represents the "REGULAR" scheduled audit event (Normally occurs every 4 hours). 1. Audits may occur more than 4 hours due to planning or troubleshooting activities. 2. Additionally, there may be a "RECOVR AUD" entry: This refers to a missed audit that was recovered.
	ENTRIES	The cumulative entry register value for a device
	EXITS	The cumulative exit register value for a device

* This data is provided by the Metropolitan Transportation Authority.

Data Preparation

Index	C/A	UNIT	SCP	STATION	LINENAME	DIVISION	DATE	TIME	DESC	ENTRIES	EXITS
1715591	D005	R398	00-06-BD	NEW UTRECHT AV	ND	BMT	02/26/2023	04:00:00	REGULAR	1743915	739152253
1715592	D005	R398	00-06-BD	NEW UTRECHT AV	ND	BMT	02/26/2023	04:00:00	RECOVR AUD	1733163	739152253

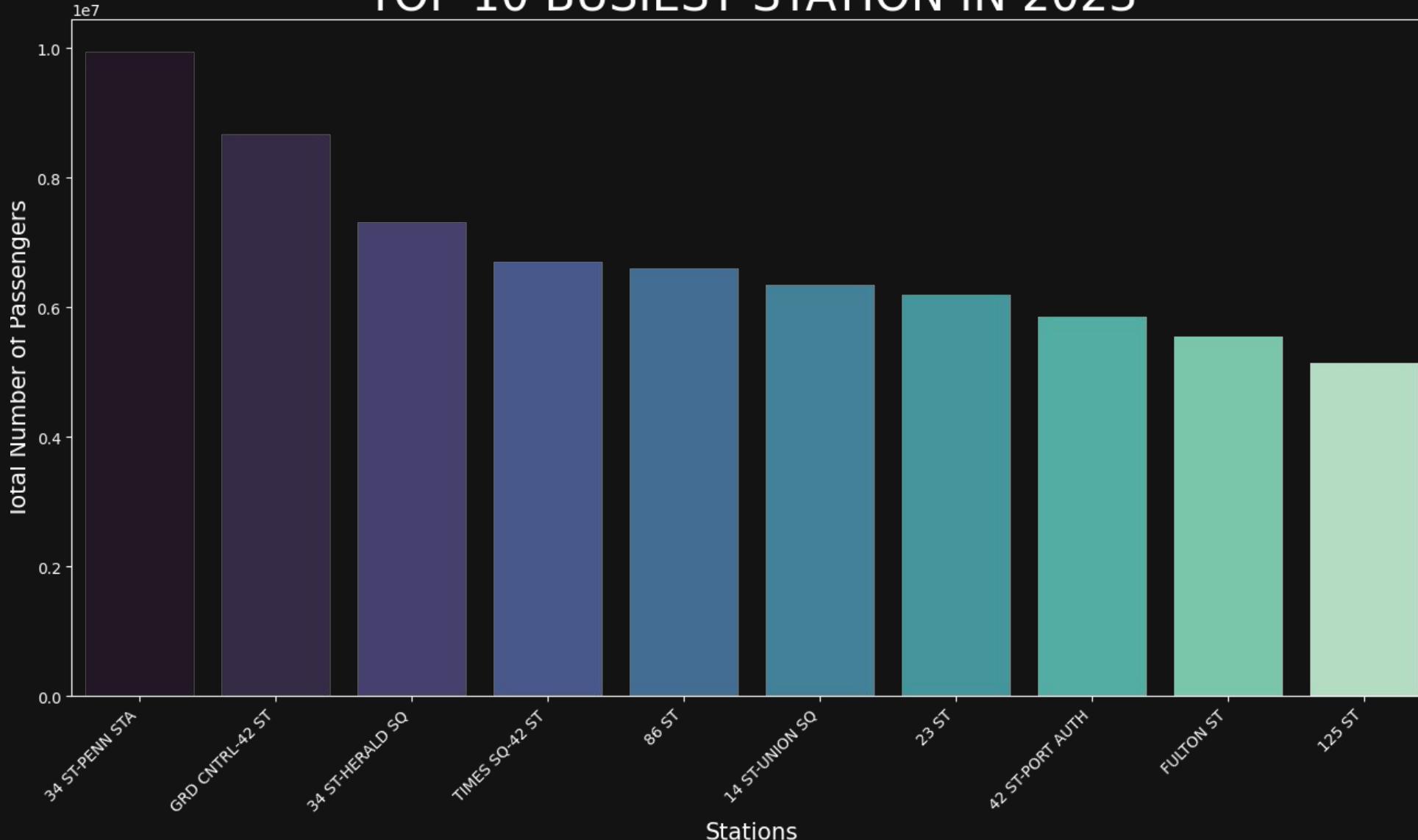
Upon examining the data, we noticed some inconsistencies and issues. We removed irrelevant columns and made necessary corrections on empty and erroneous data within the dataset.

Data Preparation

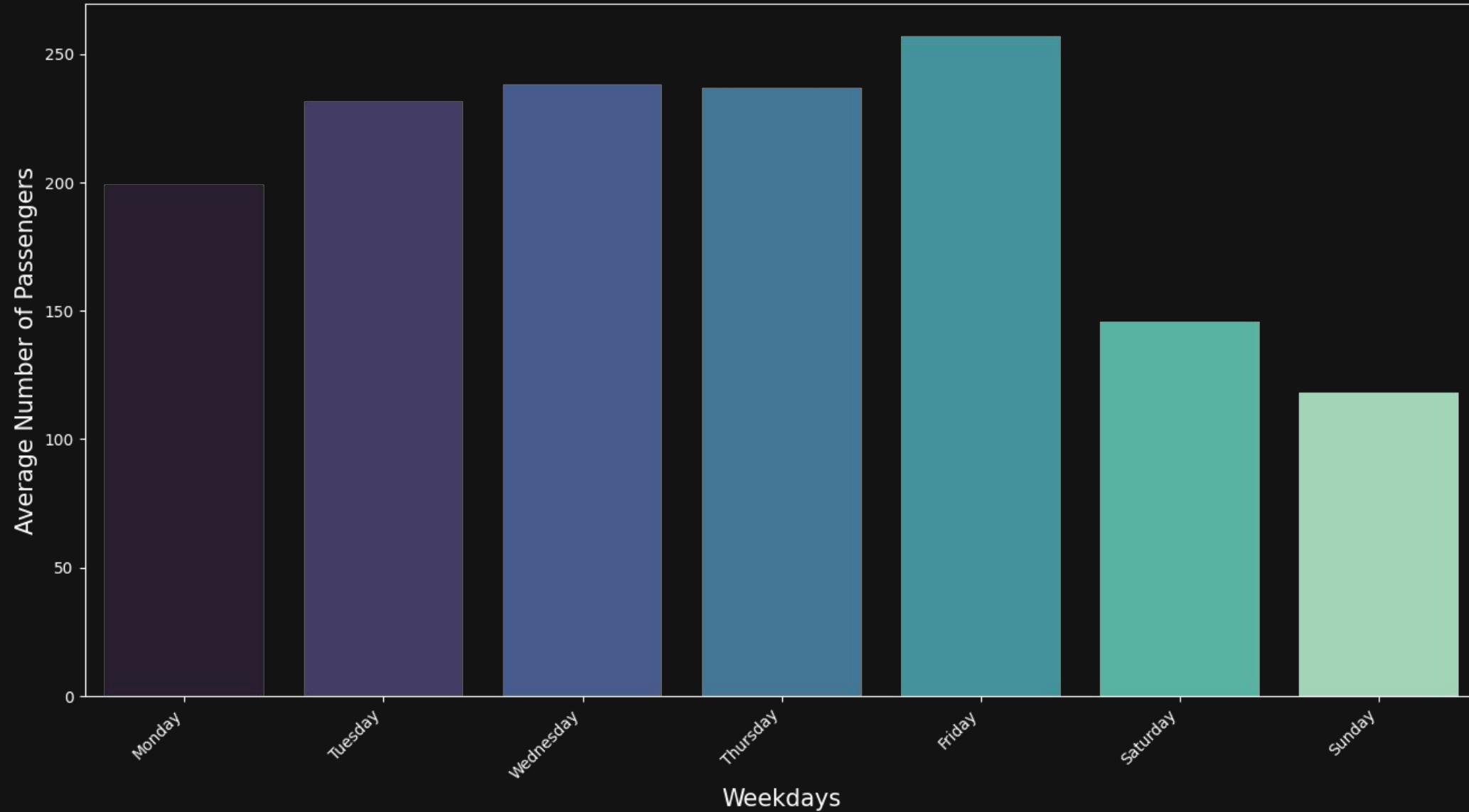
Index	STATION	TURNSTILE	DATE	TIME	TIME_STAMP	WEEKDAYS	ENTRIES	EXITS	NET_ENTRIES	NET_EXITS
38	59 ST	A002-R051-02-00-00	01/06/2023	11:00:00	2023-01-06 11:00:00	Friday	7814080	2773145	107	57
39	59 ST	A002-R051-02-00-00	01/06/2023	15:00:00	2023-01-06 15:00:00	Friday	7814187	2773202	294	54
40	59 ST	A002-R051-02-00-00	01/06/2023	19:00:00	2023-01-06 19:00:00	Friday	7814481	2773256	88	29
41	59 ST	A002-R051-02-00-00	01/06/2023	23:00:00	2023-01-06 23:00:00	Friday	7814569	2773285	-906529	-1140159
42	59 ST	A002-R051-02-00-01	12/31/2022	03:00:00	2022-12-31 03:00:00	Saturday	6908040	1633126	1	3
43	59 ST	A002-R051-02-00-01	12/31/2022	07:00:00	2022-12-31 07:00:00	Saturday	6908041	1633129	16	31
44	59 ST	A002-R051-02-00-01	12/31/2022	11:00:00	2022-12-31 11:00:00	Saturday	6908057	1633160	42	20

Considering the cumulative increase in entry and exit numbers, we calculated the net entry and exit counts by taking the difference between each data point and its predecessor. During this process, we encountered outliers. By examining the data in four-hour intervals, we cleaned these outliers from our dataset.

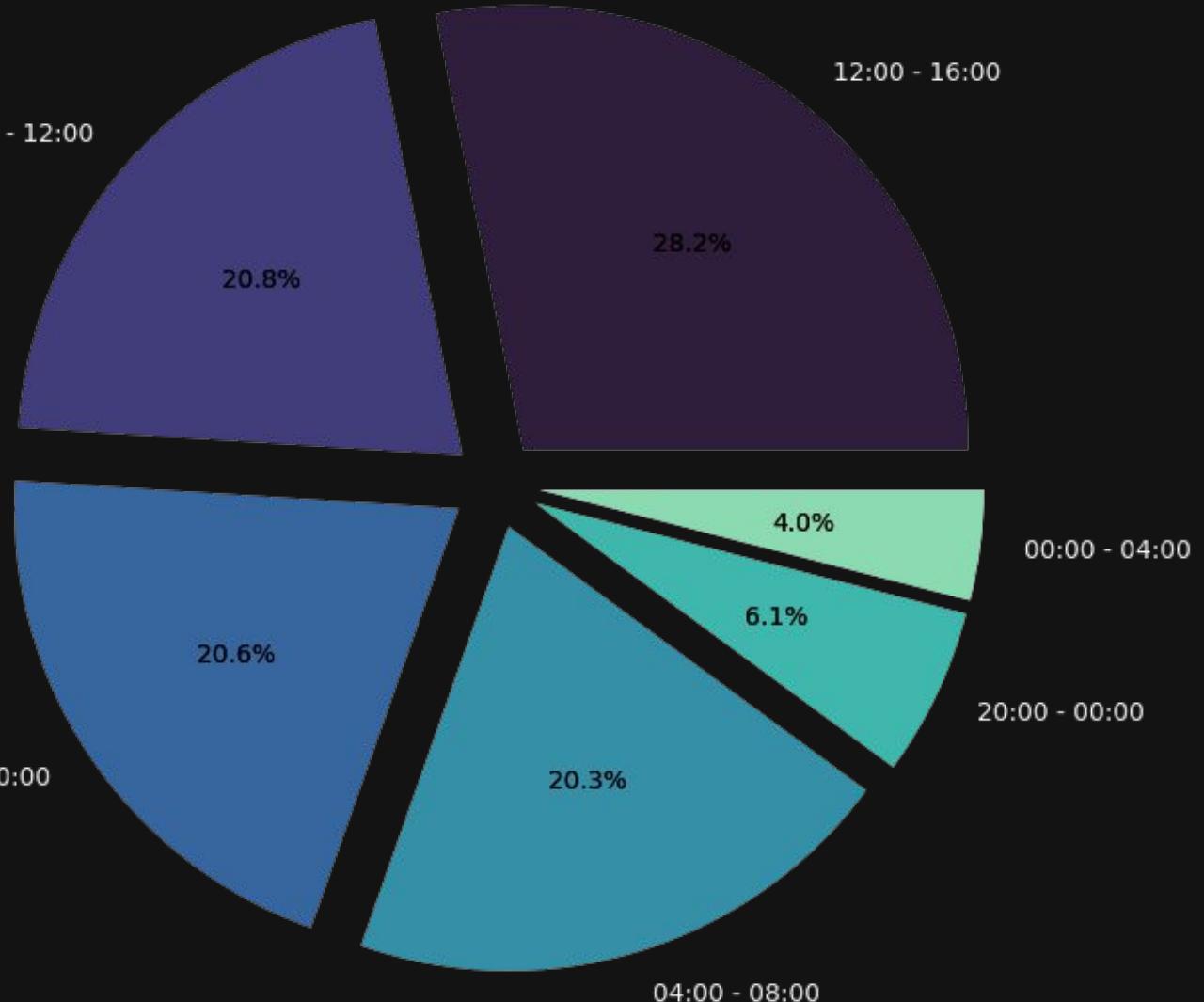
TOP 10 BUSIEST STATION IN 2023



TOTAL TRAFFIC ACCORDING TO WEEKDAYS IN 2023



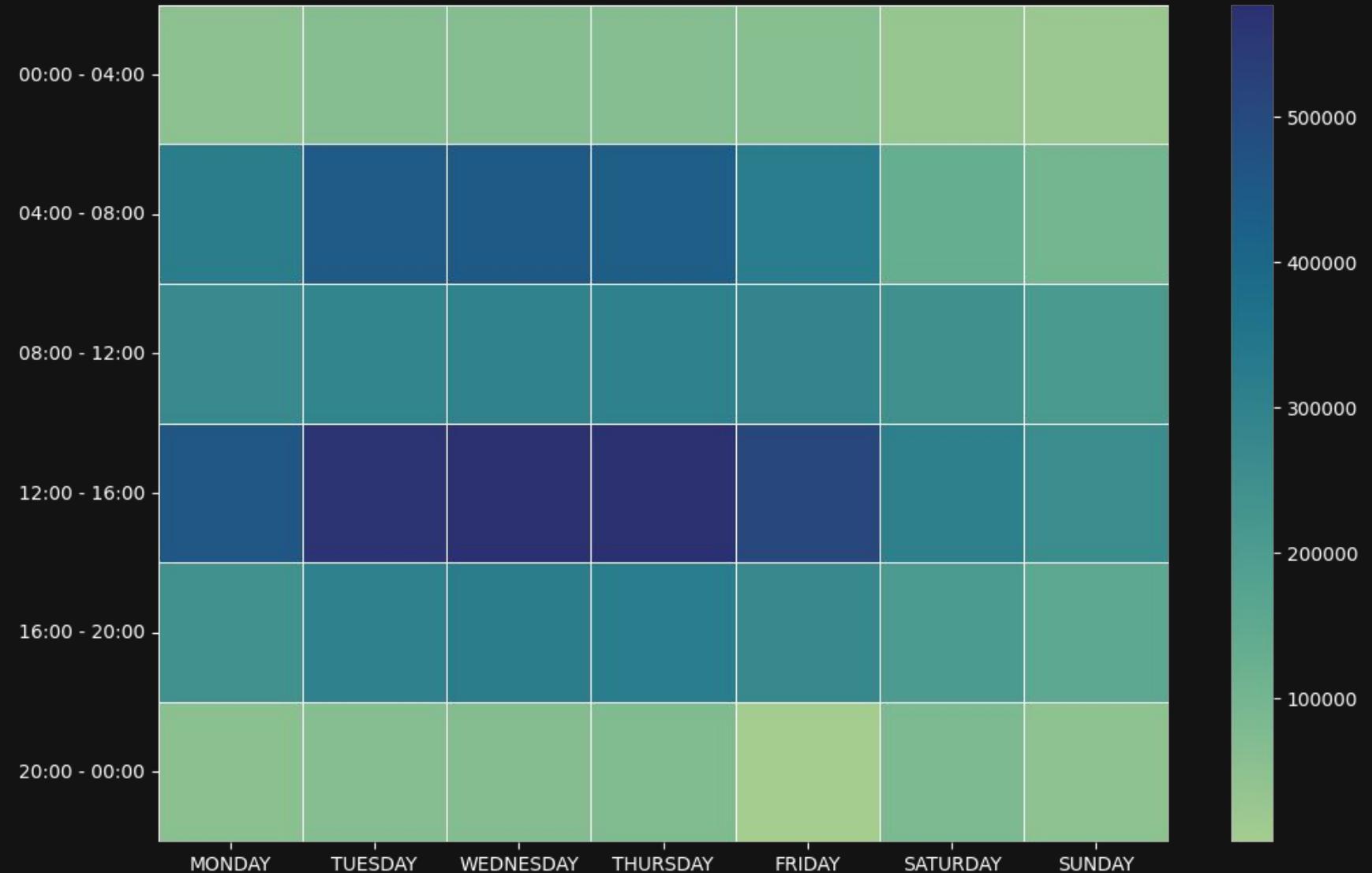
TOTAL TRAFFIC ACCORDING TO TIMES OF DAY 2023



This pie chart titled "Total traffic according to times of day 2023" illustrates the total subway traffic observed during different time periods throughout the day. The purpose of the chart is to assist in effectively positioning street teams during the busiest hours.

BUSIEST DAYS AND TIMES FOR 34 ST-PENN STATION

This heat map titled "Busiest days and times for 34 ST-PENN STATION" displays the congestion levels at the 34 ST-PENN Station according to days and hours. The purpose of the map is to assist in efficiently positioning street teams at this station during the peak times.



Optimizing Street Team Deployment: Leveraging Census Data in Metro Station Areas

By analyzing the female population in areas close to metro stations, we aim to effectively place our street teams. Using census data, we prioritize stations with a high concentration of women. With this strategic placement, our goal is to attract more interest in our event.



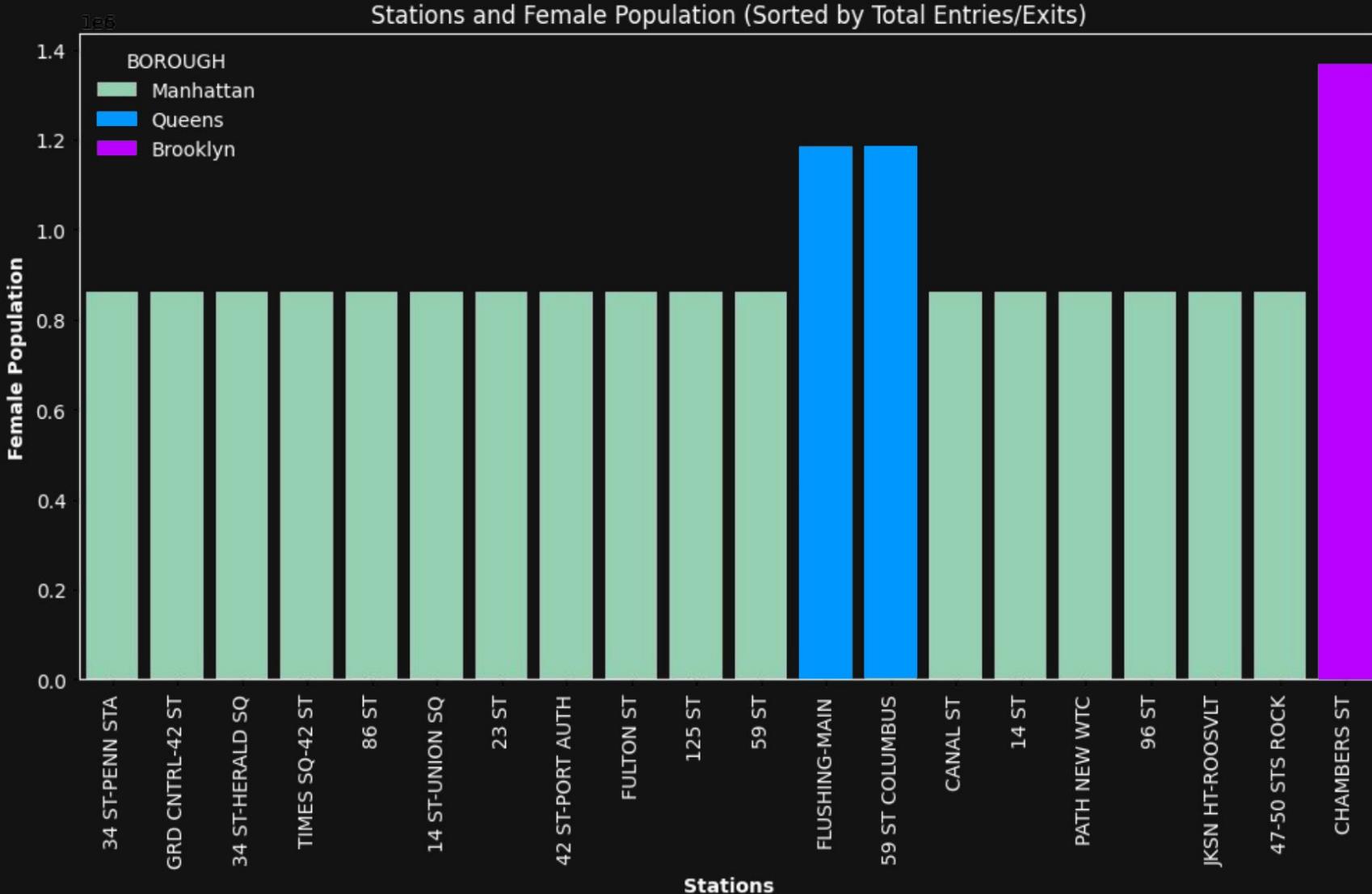
Note: The census data used in this analysis is from 2019 and should be considered valid until more recent data becomes available.

About the Data

Census Tract	County	Borough	Total Pop	Men	Women	Income	Income PerCap
36005000100	Bronx	Bronx	7703	7133	570	59282,86971	2440
36005000200	Bronx	Bronx	5403	2659	2744	72034	22180
36005000400	Bronx	Bronx	5915	2896	3019	74836	27700
36005001600	Bronx	Bronx	5879	2558	3321	32312	17526
36005001900	Bronx	Bronx	2591	1206	1385	37936	17986
36005002000	Bronx	Bronx	8516	3301	5215	18086	12023
36005002300	Bronx	Bronx	4774	2130	2644	14479	9781
36005002400	Bronx	Bronx	150	109	41	59282,86971	49365
36005002500	Bronx	Bronx	5355	2338	3017	17226	11493
36005002701	Bronx	Bronx	3016	1375	1641	20153	10317

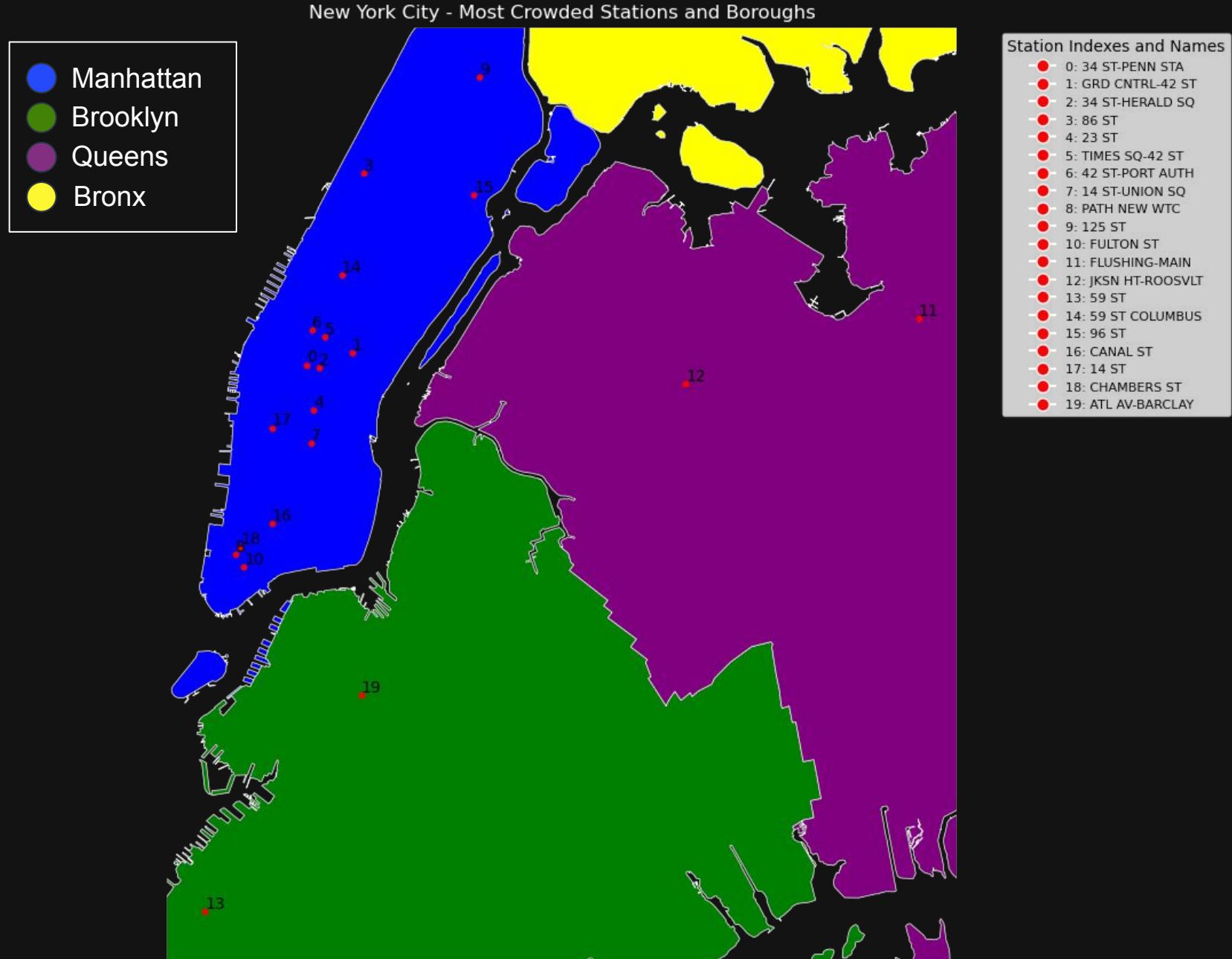
Metadata

Census Tract	A numerical code identifying census tracts within the United States.
County	The name of the county in which the census tract is located.
Borough	The name of the borough in which the census tract is located.
Total Pop	The total population of the census tract.
Men	The male population of the census tract.
Women	The female population of the census tract.
Income	The mean annual household income in the census tract, in US dollars.
IncomePerCap	The mean annual per capita income in the census tract, in US dollars.



Borough	Female Population
Brooklyn	1.366.258
Queens	1.185.680
Manhattan	860.073
Bronx	755.910
Staten Island	243.778

This map displays the busiest stations in New York City and their associated boroughs. The stations are marked with red dots, and the station indices and names are shown in a box in the upper left corner of the map. The boroughs are represented by different colors, with the name of each borough faintly written in white over the respective area. This visualization provides valuable insights into New York City's public transportation system and its geographical distribution.





- * When looking at the top ten busiest stations, 34th Street-Penn Station stands out as the busiest.
- * Evaluating the total traffic based on weekdays, Friday is identified as the busiest day.
- * Assessing the total traffic according to times of the day:
 - Peak hours: 12:00 PM - 4:00 PM
 - Least busy hours: 12:00 AM - 4:00 AM
- * For 34th Street-Penn Station specifically, the busiest hours are again between 12:00 PM and 4:00 PM.

Future Work

Social Media
Targeting

QR Code-Featured
Posters





How can we
further
delve into
this subject?



 April 8

 Central Park Central
Park New York, NY





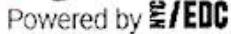
April 11



New York Genome
Center 101 6th Avenue
New York, NY



Spring St.
Canal St.

Presented by: **women.nyc**
Powered by 

NYC
Small Business Services

careers
businesses
neighborhoods





-  April 24
-  43 W 23rd St 43 West 23rd Street New York, NY
-  23 St



LIVE EVENT

Women Transforming Technology

Panel Discussion Networking

Presented by:

women.nyc
Powered by E/EDC



April 26



Alianza Dominicana
Cultural Center 530 W
166th St New York, NY

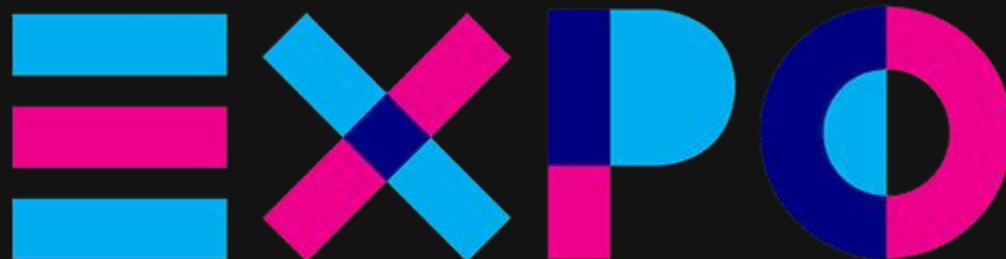


168 St - Washington Hts
168 St

TECHSPO

NEW YORK 2023

TECHNOLOGY



APRIL 27 - 28



April 27



New York Marriott at the
Brooklyn Bridge Hotel 333
Adams St New York, NY



Jay St - MetroTech
Borough Hall

STARTUP+ PRESENTS

STARTUP+ NETWORKING

An IN-PERSON Tech Startup Networking Mixer



FIRST THURSDAYS | 5 PM
NYC



May 4



Sour Mouse 110
Delancey Street New
York, NY



Delancey St · Essex St
Bowery

womentech
network

CODING
GIRLS

WOMEN IN TECH NEW YORK

GLOBAL IMPACT SUMMIT
MAY 12, 2023 IN-PERSON



CONNECT. GROW. LEAD.

womentech

GLOBAL CONFERENCE 2023

This event is part of the Women in Tech Global Conference 2023 local series.



May 12



Looking for Host





Thank you for your attention