

The background image shows the interior of an MTA turnstile, featuring metal railings and a blue-tinted overlay. A piece of technical equipment, possibly a camera or sensor, is visible on a tripod in the center-left. The text "DAVID WISMER" is overlaid in white, and "MTA TURNSTILE ANALYSIS" is overlaid in large blue letters.

DAVID WISMER

---

# MTA TURNSTILE ANALYSIS

# THE GOAL

---

## ▶ **Client**

- ▶ Metropolitan Transportation Authority

## ▶ **Key Deliverables**

- ▶ Initial analysis of first 5 months of 2021, including total turnstile usage and top growing and declining stations
- ✱ Interactive dashboards for customers and employees to explore trends

## ▶ **Purpose**

- ▶ Facilitate resource management by providing improved access to data
- ▶ Empower customers to self-police overcrowding

# THE METHODS AND TOOLS

---

## ▶ **Data Ingestion and Storage**

- ▶ Pandas
- ▶ SQLite

## ▶ **Data Cleaning and Manipulation**

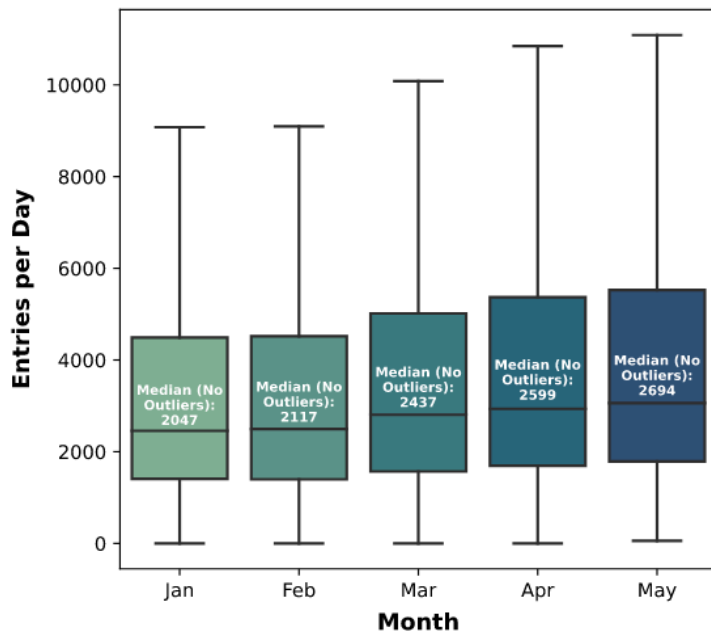
- ▶ Python Libraries: Pandas, Numpy, Itertools
- ▶ Reversed counters, counter resets, erroneous data, long gaps between audits
- ✱ Conversion to hourly data for analysis and presentation
- ✱ Removal of outliers

## ▶ **Presentation Tools**

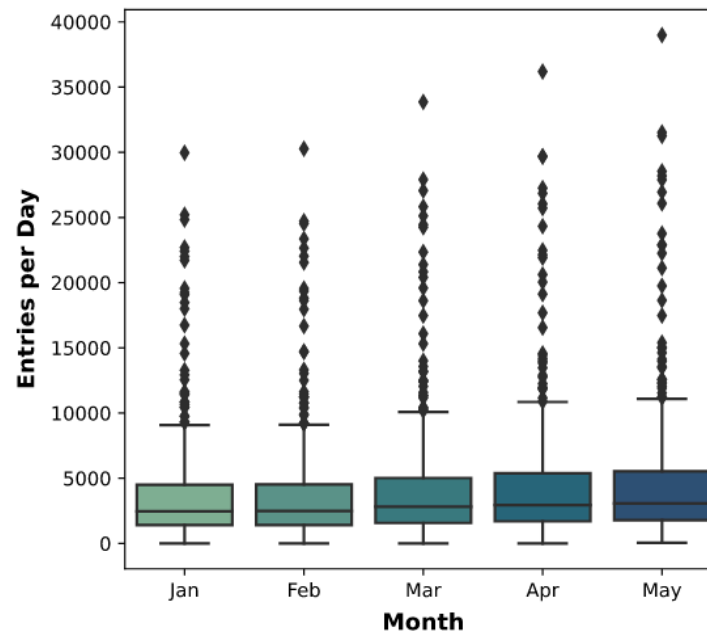
- ▶ Python Libraries: Matplotlib, Seaborn
- ▶ Tableau Public

# THE TRENDS

**Average Daily Entries Per Station  
(No Outliers)**



**Average Daily Entries Per Station  
(Outliers Included)**

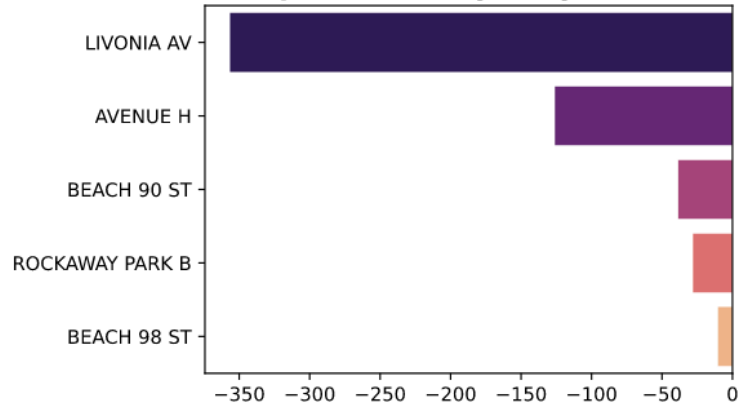


## Takeaways:

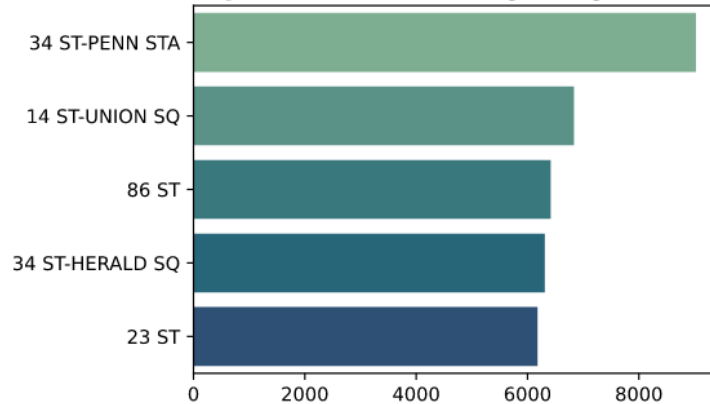
- Average entries are slowly returning to normal over the first 5 months of the year.
- There are long tails in entry data, and it will be important to understand patterns in the outliers.

# THE TRENDS – MAY VS JANUARY

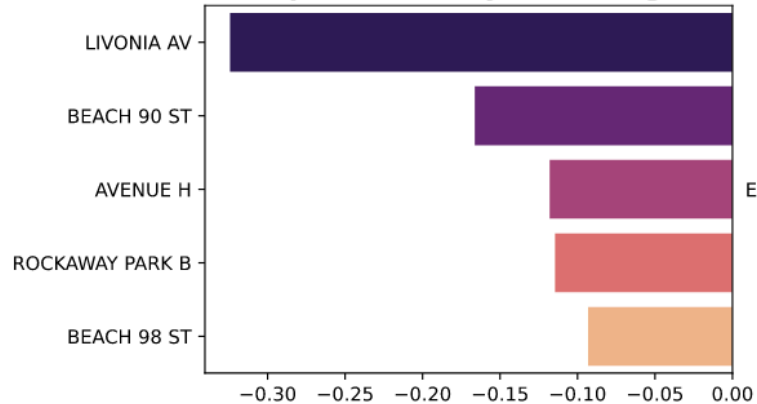
**Top Decliners by Daily Entries**



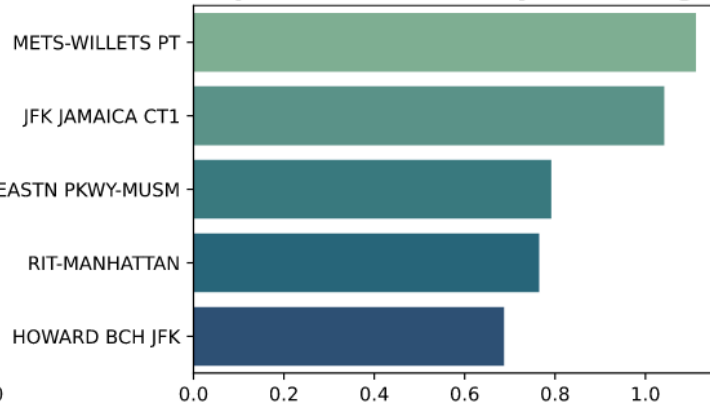
**Top Growth Drivers by Daily Entries**



**Top Decliners by Percentage**



**Top Growth Drivers by Percentage**



## Takeaways:

- The declining group is the same in absolute and percentage terms. These are worth looking into.
- The growth drivers in absolute terms are some of the largest stations. It will be more impactful to study the stations growing most in percentage terms.

# THE PRODUCT

### Average Hourly Activity by Station

Station	Avg. Entries Per Hour	Avg. Exits Per Hour	Avg. Total Traffic
34 ST-PENN STA	1,473	1,971	3,444
34 ST-HERALD SQ	1,228	1,597	2,825
86 ST	1,173	1,495	2,668
125 ST	1,073	1,324	2,398
GRD CNTRL-42 ST	1,064	1,454	2,518
23 ST	1,031	1,170	2,201
42 ST-PORT AUTH	1,028	1,214	2,242
14 ST-UNION SQ	1,023	1,399	2,422
FLUSHING-MAIN	882	966	1,848
96 ST	867	955	1,822
59 ST	856	1,023	1,879
JKSN HT-ROOSVLT	828	782	1,609
FULTON ST	807	1,044	1,851
TIMES SQ-42 ST	790	1,031	1,820
59 ST COLUMBUS	723	795	1,518
CANAL ST	721	877	1,598
14 ST	601	784	1,384

### Entries by Station

Station	Entries
34 ST-PENN STA	4,007,846
34 ST-HERALD SQ	3,380,870
86 ST	3,278,583
125 ST	3,114,835
GRD CNTRL-42 ST	2,875,797
42 ST-PORT AUTH	2,872,509
23 ST	2,861,034
14 ST-UNION SQ	2,820,675
FLUSHING-MAIN	2,587,596
59 ST	2,492,882
JKSN HT-ROOSVLT	2,424,794
96 ST	2,408,206
FULTON ST	2,221,747
TIMES SQ-42 ST	2,127,177
59 ST COLUMBUS	1,986,191
CANAL ST	1,946,908
CHURCH AV	1,702,564

### Average Hourly Activity by Day and Hour

Time	Weekday						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
00:00	25.3	18.2	21.0	23.7	24.2	25.3	28.9
01:00	13.6	10.1	11.3	12.5	12.9	13.4	15.3
02:00	12.5	9.8	10.9	12.1	12.3	12.8	14.1
03:00	14.2	35.6	38.6	41.3	40.4	36.9	18.5
04:00	29.8	116.8	125.1	129.1	126.9	116.1	44.4
05:00	40.3	164.6	175.4	178.9	175.9	162.5	60.2
06:00	41.5	169.4	181.3	184.8	181.0	167.6	62.0
07:00	62.5	216.1	233.3	239.4	234.2	216.0	92.9
08:00	92.1	237.1	255.3	262.6	257.3	241.0	130.6
09:00	103.2	224.1	241.4	249.5	245.2	232.6	143.9
10:00	104.6	222.2	238.8	247.1	243.2	231.3	145.6
11:00	118.1	206.5	218.9	226.5	223.3	219.8	162.8
12:00	141.8	217.2	230.2	237.7	234.5	241.7	193.0
13:00	151.0	239.2	254.0	261.1	257.3	266.9	205.0
14:00	151.8	242.9	258.3	265.4	261.2	271.3	206.3
15:00	155.2	274.4	297.7	307.3	299.7	304.9	212.5
16:00	146.0	291.8	319.9	329.2	320.5	317.9	202.9
17:00	137.5	278.1	306.8	315.5	308.0	303.1	193.0
18:00	136.5	274.8	303.3	311.8	304.6	299.4	191.6
19:00	113.7	213.3	234.1	239.1	235.6	235.4	156.9
20:00	79.5	118.2	131.4	133.8	132.9	139.9	107.9
21:00	65.4	84.7	94.9	98.6	96.9	105.4	89.0
22:00	63.9	81.4	91.6	95.1	93.8	101.6	86.9
23:00	45.7	55.5	62.1	63.7	63.8	70.4	61.7

### Activity By Station and Date

Station	Day of Date	Entries	Exits	Net Flow	Total Traffic
34 ST-PENN STA	May 13, 2021	41,437	55,387	-13,950	96,823
34 ST-PENN STA	May 7, 2021	40,227	55,774	-15,548	96,001
34 ST-PENN STA	May 12, 2021	40,519	52,005	-11,486	92,524
34 ST-PENN STA	April 30, 2021	38,512	53,430	-14,918	91,942
34 ST-PENN STA	May 6, 2021	39,398	51,907	-12,509	91,305
34 ST-PENN STA	May 11, 2021	39,945	50,751	-10,806	90,696
34 ST-PENN STA	April 23, 2021	37,922	52,199	-14,277	90,121
34 ST-PENN STA	May 4, 2021	39,129	50,217	-11,089	89,346
34 ST-PENN STA	May 5, 2021	38,867	49,831	-10,965	88,698
34 ST-PENN STA	April 28, 2021	38,684	49,991	-11,307	88,675
34 ST-PENN STA	May 14, 2021	36,256	52,022	-15,766	88,278
34 ST-PENN STA	April 16, 2021	36,889	50,508	-13,619	87,397
34 ST-PENN STA	April 22, 2021	37,623	49,418	-11,796	87,041
34 ST-PENN STA	May 10, 2021	39,104	47,728	-8,624	86,832

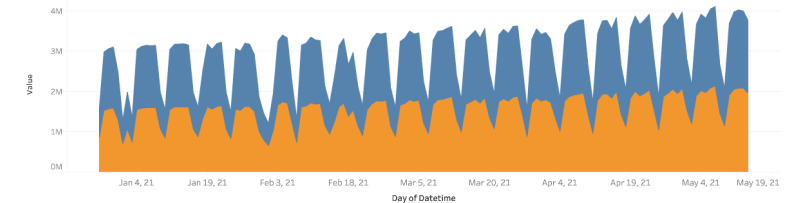
Date  
12/27/2020 5/14/2021

Weekday  
(All)

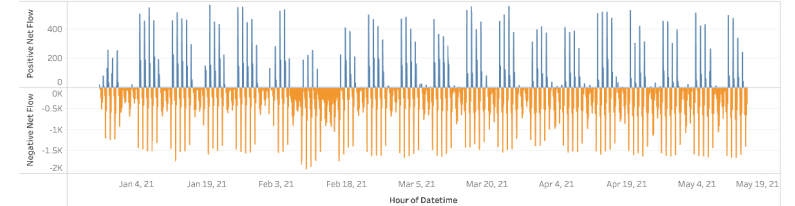
Time  
(All)

Station  
(All)

### Time Series Data



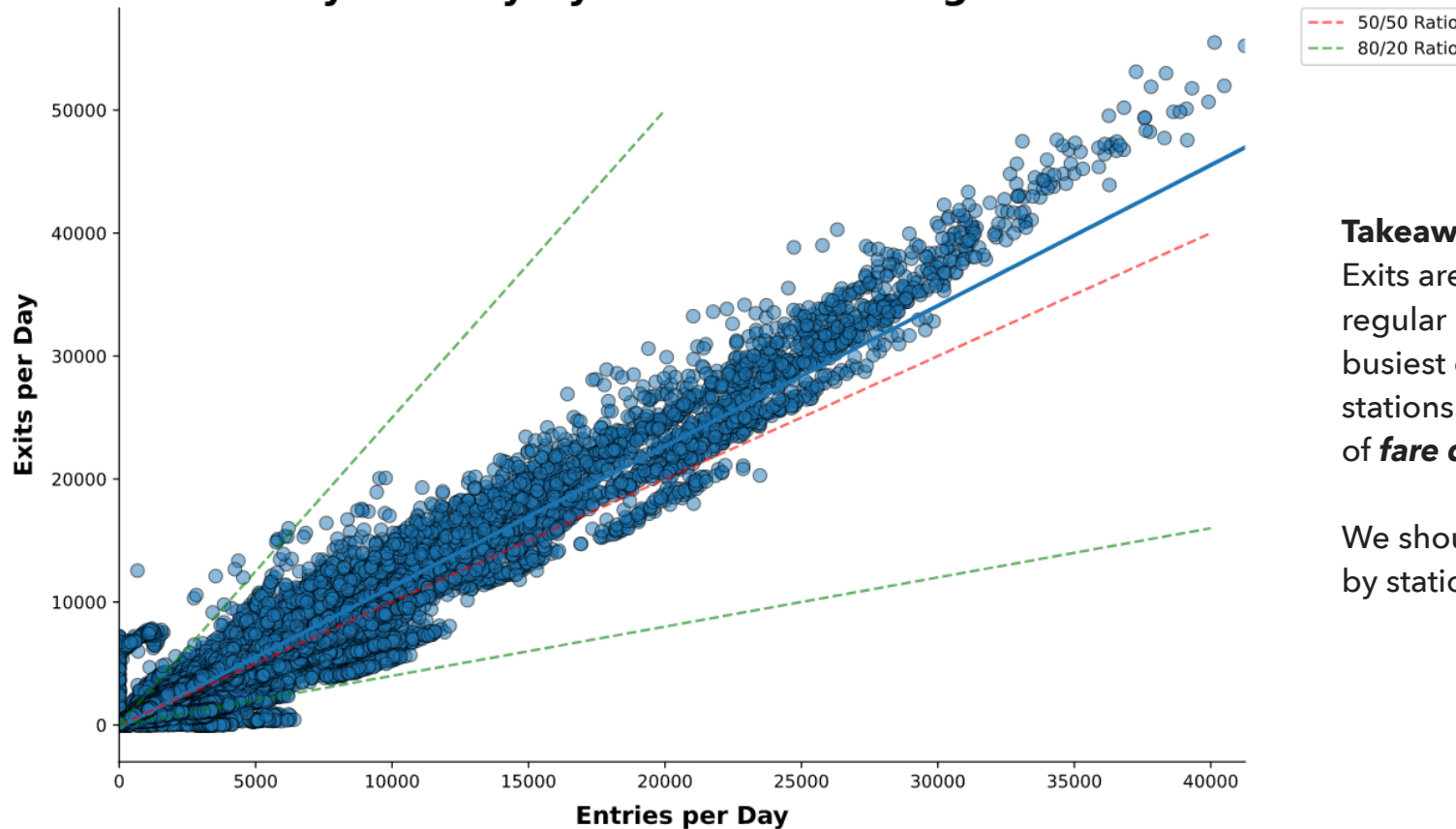
### Net Flow



► [Tableau Dashboards Link](#)

# DASHBOARD USE CASE – FARE DODGING

**Daily Activity by Station with Regression**



**Takeaway:**

Exits are exceeding Entries on a regular basis, particularly on the busiest days at the busiest stations. This is a potential sign of **fare dodging**.

We should explore the **net flow** by station.

# DASHBOARD USE CASE – FARE DODGING

## Net Flow by Station

Station	
34 ST-PENN STA	-1,371,276
GRD CNTRL-42 ST	-936,755
34 ST-HERALD SQ	-920,501
14 ST-UNION SQ	-840,945
NEWARK HW BMEBE	-793,751
86 ST	-780,203
125 ST	-766,533
NEWARK HM HE	-701,860
FULTON ST	-630,623
TIMES SQ-42 ST	-594,917
1 AV	-557,086
42 ST-PORT AUTH	-536,588
BEDFORD AV	-531,190
B'WAY-LAFAYETTE	-514,725
W 4 ST-WASH SQ	-489,563
28 ST	-461,543
72 ST	-442,109

Selected Measure



Measure Selector

Net Flow

Date

12/27/2020 5/14/2021

Weekday

(All)

Time

(All)

Station

(All)

## Average Hourly Activity by Day and Hour

Time	Weekday						
	Sunday	Monday	Tuesday	Wednesday	Thursday	Friday	Saturday
00:00	-127	-77	-82	-100	-101	-88	-130
01:00	-119	-77	-73	-87	-84	-76	-116
02:00	-114	-74	-61	-69	-72	-70	-110
03:00	-118	-70	-61	-50	-47	-65	-145
04:00	-157	-37	-12	-4	-21	-72	-222
05:00	-176	88	163	177	153	62	-243
06:00	-205	217	303	308	278	139	-285
07:00	-338	79	156	143	108	-58	-505
08:00	-465	-152	-107	-114	-136	-300	-643
09:00	-491	-346	-348	-366	-393	-516	-708
10:00	-527	-552	-581	-595	-616	-713	-759
11:00	-591	-626	-696	-723	-752	-823	-799
12:00	-627	-728	-812	-841	-892	-970	-810
13:00	-644	-864	-1,002	-1,053	-1,099	-1,148	-833
14:00	-635	-1,066	-1,225	-1,274	-1,324	-1,333	-843
15:00	-521	-1,059	-1,249	-1,362	-1,377	-1,412	-755
16:00	-373	-950	-1,147	-1,215	-1,289	-1,329	-686
17:00	-303	-821	-1,003	-1,051	-1,118	-1,182	-623
18:00	-207	-590	-748	-799	-871	-951	-559
19:00	-96	-344	-440	-443	-522	-611	-387
20:00	-21	-126	-193	-231	-235	-349	-292
21:00	-24	-62	-97	-131	-134	-253	-252
22:00	-50	-22	-53	-78	-79	-203	-218
23:00	-66	-59	-81	-109	-84	-177	-184

Avg. Selected Measure

-1,412 308

Measure Selector

Net Flow

Date

12/27/2020 5/14/2021

Weekday

(All)

Time

(All)

Station

34 ST-PENN STA

► [Tableau Dashboards Link](#)



# THE FUTURE

---

- ▶ **Standardization of Data Pipeline**

- ▶ Automated collection at consistent intervals
- ▶ Real-time reporting of turnstile activity

- ▶ **Targeted Studies**

- ▶ Fare dodging analysis by station
- ▶ Staffing planning by station and time of day
- ▶ Examination of outliers (holidays, events)

- ▶ **Implement Strategies and Test**

- ▶ Provide trip planner tools to the public and measure impact
- ▶ Revisit trending after implementing resource management changes