

The background image shows the interior of an MTA turnstile. A camera is mounted on a tripod in the center of the frame, pointing towards the turnstile. The turnstile has a metal frame and a glass door. The lighting is dim, and the overall color scheme is dark blue and black.

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**

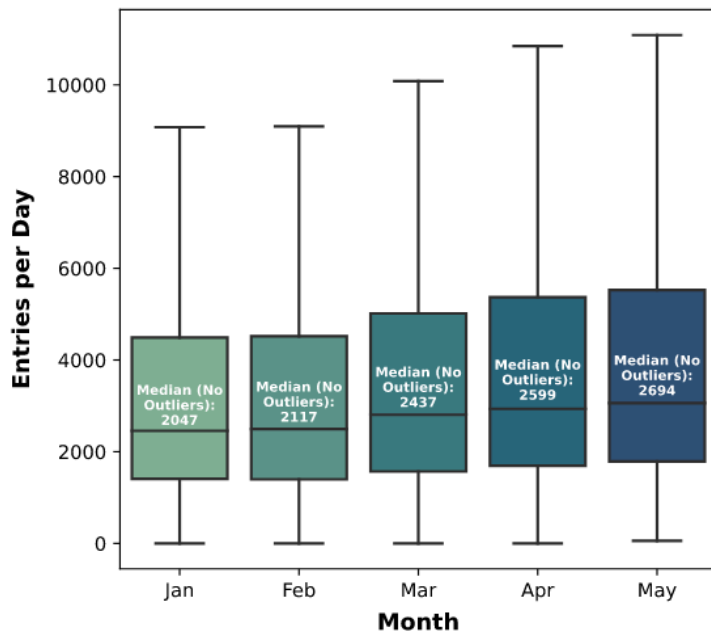
- ▶ Python Libraries: Pandas, Numpy, Itertools
- ▶ Reversed counters, counter resets, erroneous data, long gaps between audits
- ✱ Conversion to hourly data for analysis and presentation
- ✱ Removal of outliers

▶ **Presentation Tools**

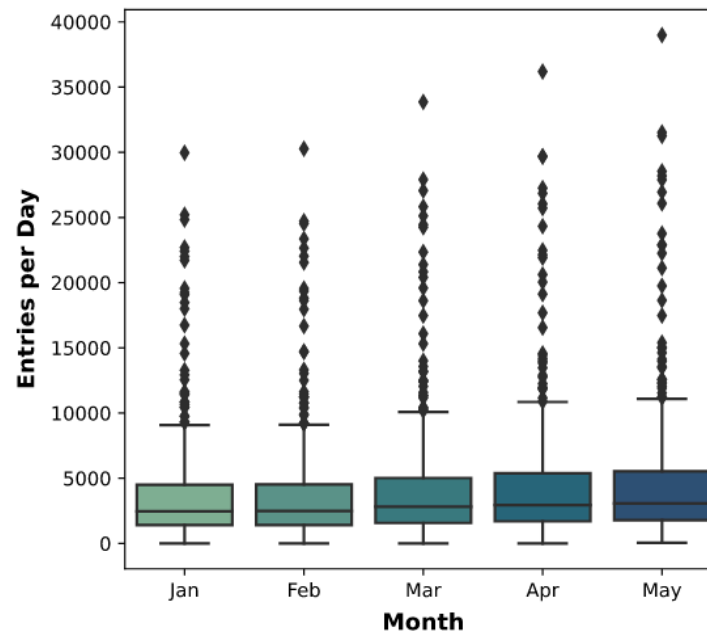
- ▶ Python Libraries: Matplotlib, Seaborn
- ▶ Tableau Public

THE RESULTS

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



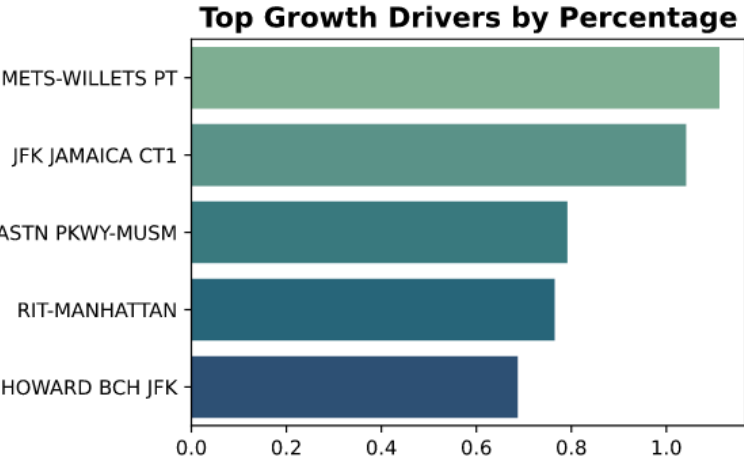
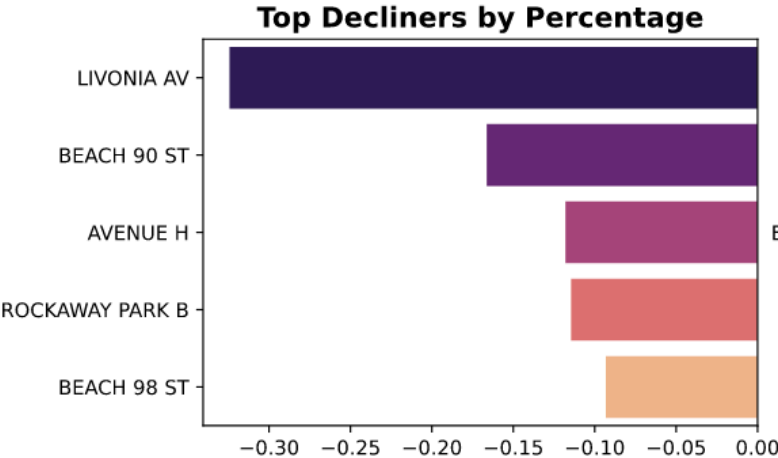
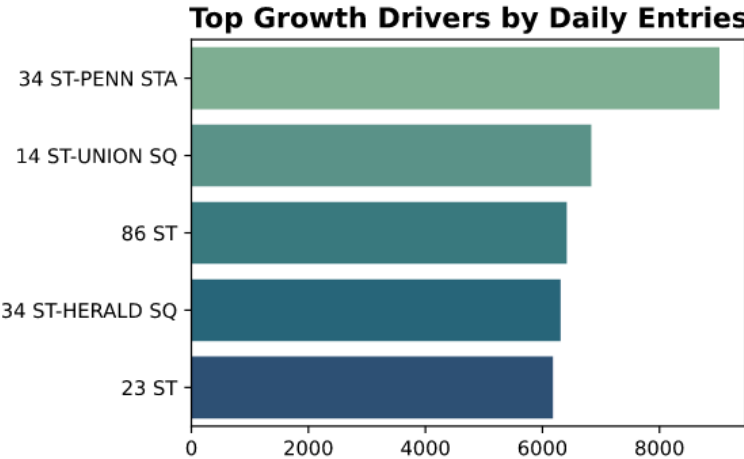
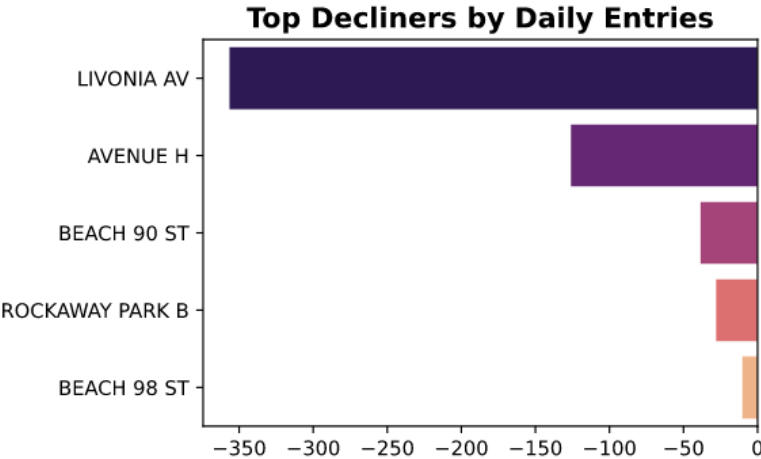
**Average Daily Entries Per Station
(Outliers Included)**



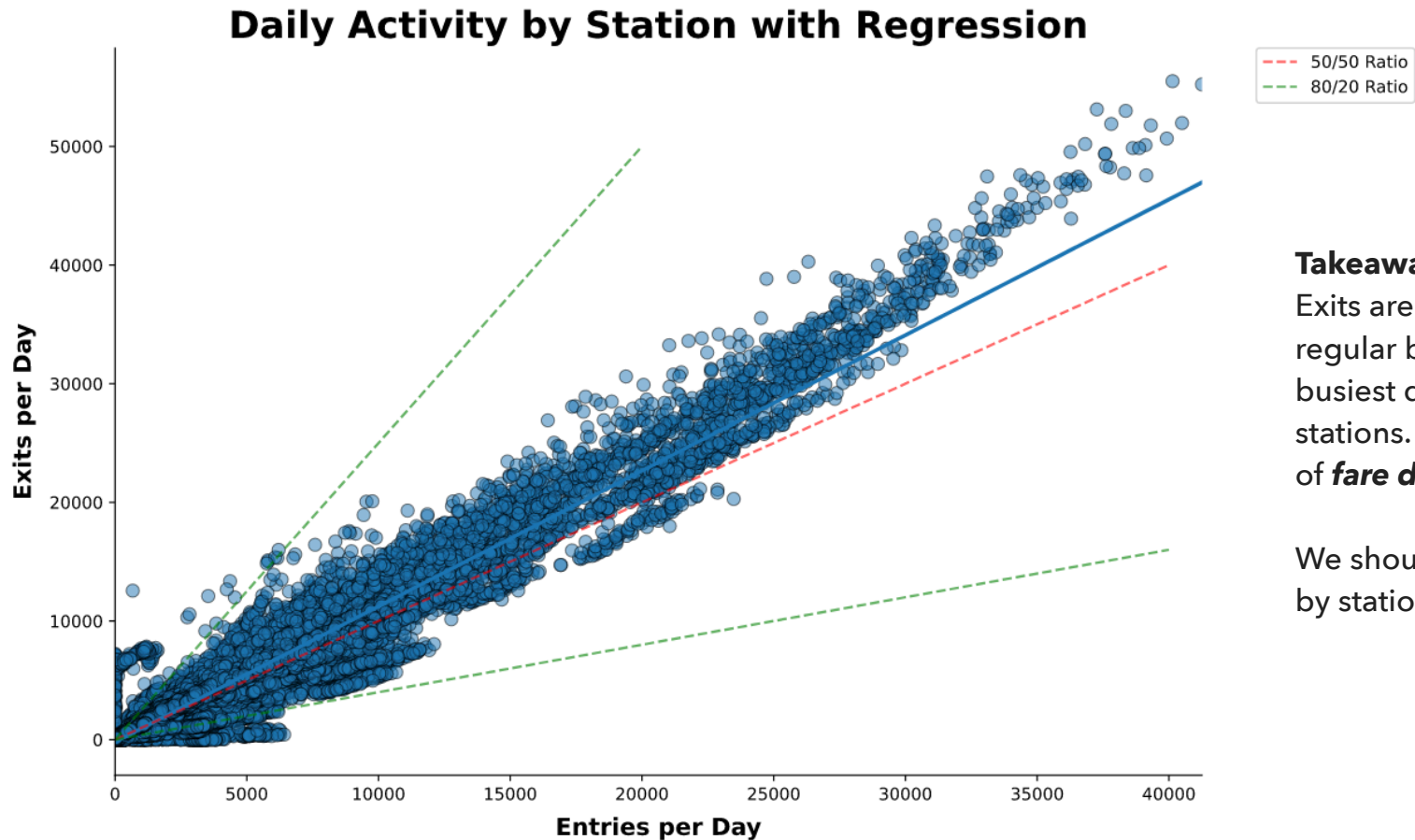
Takeaways:

- 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 RESULTS



DASHBOARD USE CASE – FARE DODGING



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

Interactive Tableau Exploration:

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