

EDA Presentation Slides

Identifying & Ranking Post-Pandemic Declines in NYC Subway Ridership

George Pappy – 18 August 2021

Introduction

- Motivation:

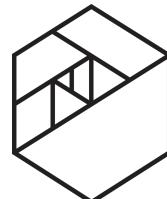
- NYC MTA under severe budget shortfall for years¹
- Projected \$15.9B deficit by late 2024¹ → Made far worse by COVID
- MTA relies on paid fares: **Lost Riders = Lost Revenue!**

- Objectives and Goals:

- Quantify ridership decline in 2021
- Pinpoint stations with the largest declines and identify median neighborhood incomes associated with each
- Provide MTA with information that will help target marketing/promotional outreach campaigns to win back as many lost riders as possible

¹ "Why has the New York City subway gone off the rails?" CBS 60 Minutes, 21 Oct. 2018.

² "Crowded Subways? Yes, in Neighborhoods Where People Have to Go to Work." The New York Times, 2 Dec. 2020.



Methodology

- **Primary Data:** MTA Turnstile Counts (recorded every 4 hrs.)

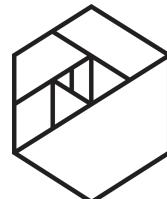
- May - June, all stations, 2019 & 2021 (no Staten Island)
- Same two months in both years avoids seasonality variations
- 2019 clearly captures “pre-pandemic” ridership levels
- May-June 2021 pre-dates recent public concern (Delta variant)

- **Supplementary Data:**

- MTA Station, Complex and Location (Lat/Long) datasets³
- US Census Bureau median income by zip code⁴

³ <https://qli.cloud/nyc-transit-data/>

⁴ <https://data.cccnewyork.org/data/table/66/median-incomes#66/107/62/a/a>



Methodology (con't.)

- Metrics

- Change in Total Riders per day per station

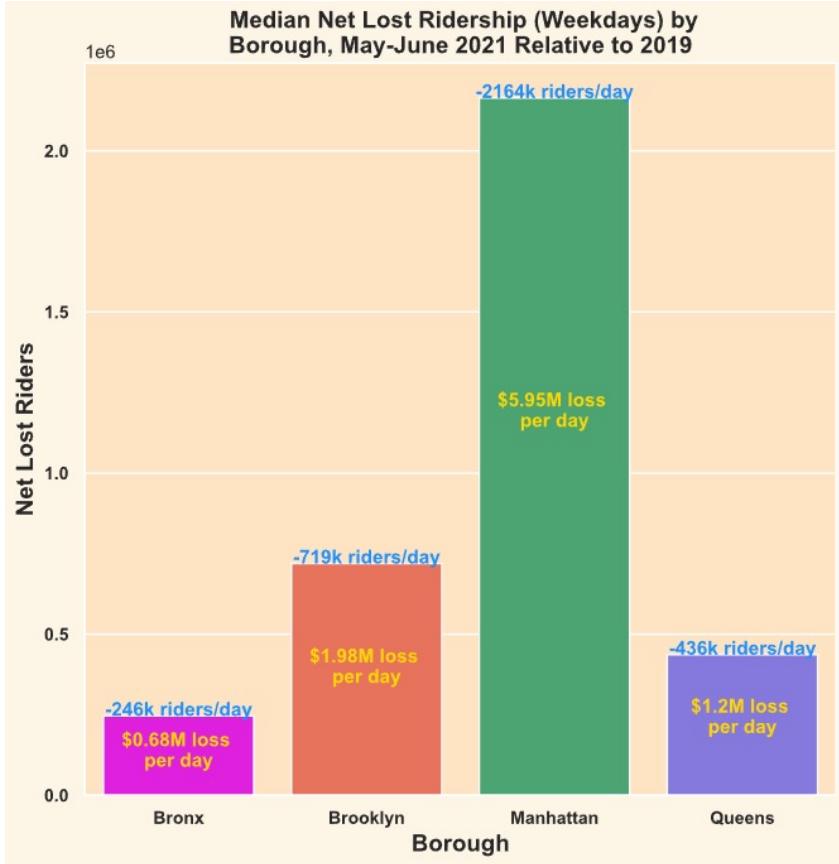
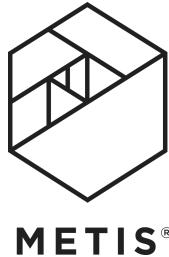
- Tools

- **SQLite/SQLAlchemy**: gather MTA data, check for NaNs, import to Pandas
 - **Python 3.8/Pandas**: clean, explore, aggregate & import additional data
 - **Matplotlib/Seaborn**: visualization during data exploration and for final results

- How the data is being used:

- Aggregations by Station, Borough, Median Income Level
 - Visualizations: identifying & ranking stations by degree of lost ridership

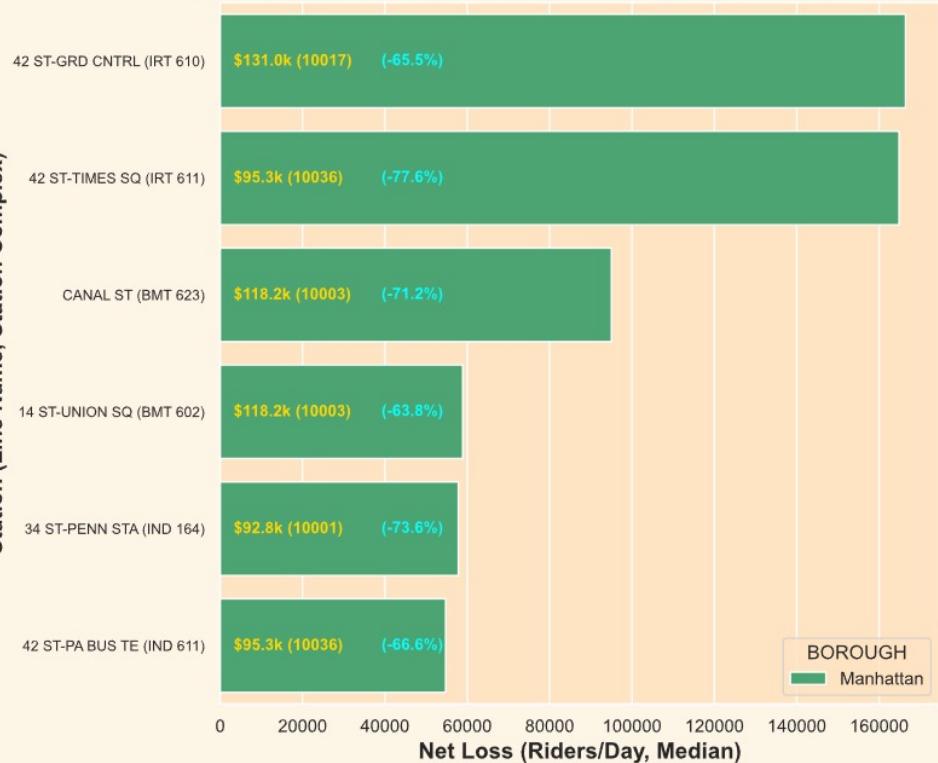
Results



- Daily revenue loss in Manhattan eclipses the other 3 boroughs combined (based on Basic Fare = \$2.75/rider)
- Should outreach campaign focus only on Manhattan?
 - But which Manhattan stations?
 - Are there some stations in the other boroughs also worth targeting?

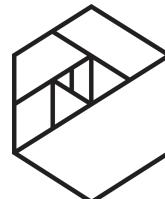
Results (con't.)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
 (Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



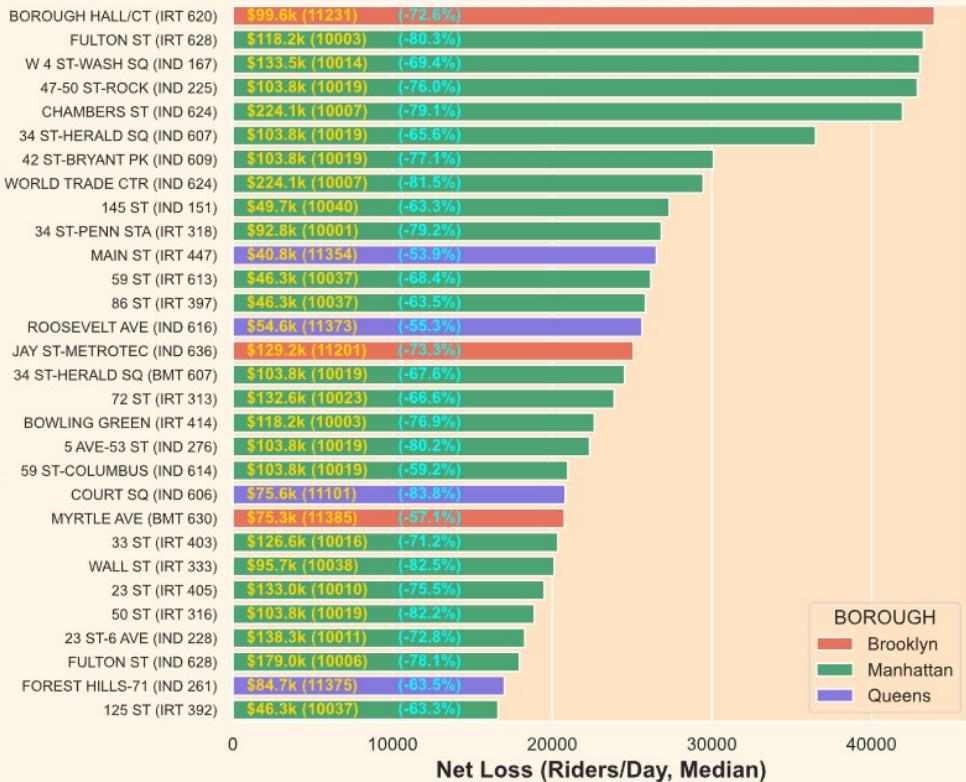
- The “Big 6” Stations in Manhattan:

- Lost > 50,000 riders/day
- Together, they account for \$1.64M/day in lost revenue
- These are unique stations
 - Major transport hubs
 - Tourist destinations



Results (con't.)

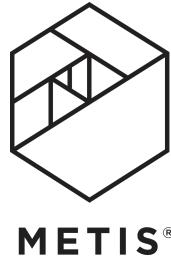
Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



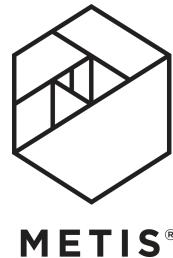
- The Next 30 Stations:

- Together, they account for \$2.20M/day in revenue loss
- It's not just Manhattan:
 - There are stations in Brooklyn and Queens worth targeting
 - No Bronx stations

Conclusions

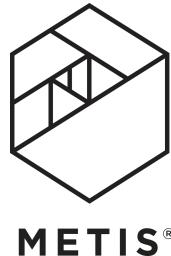


- Recommendations
 - Target outreach campaign(s) to stations with the biggest ridership losses
 - The “Big 6” may require unique targeting (i.e., tourists vs. commuters)
 - Possibly reach out to large employers near certain stations (e.g., Discount MTA passes)
 - Should also pursue general “untargeted” outreach (radio/TV/print ads)
- Interesting insights
 - Manhattan takes an outsized role in overall ridership loss
 - The Bronx plays a relatively minor role in ridership loss



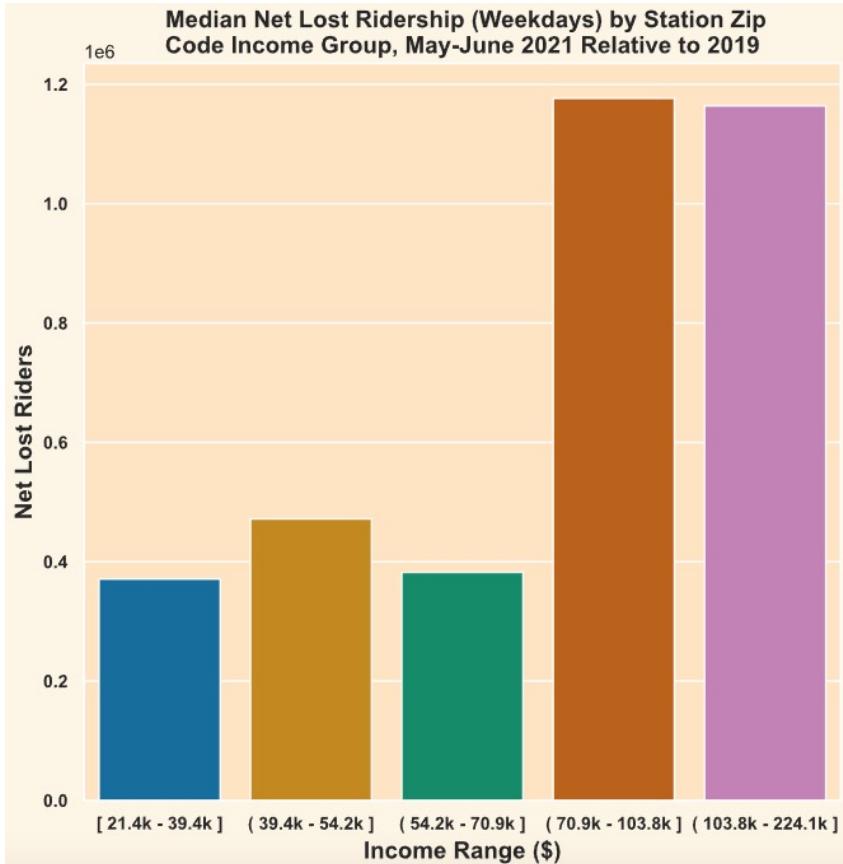
Appendix

Future Work

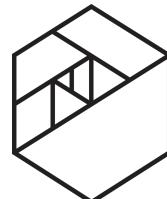


- Find improved methods of assuring highly accurate daily counts from the MTA Turnstile data
 - Dealing with counter resets/discontinuities (e.g., interpolation)
 - Handling Counter Audits, non-contiguous days of turnstile operation, etc.
- Associate more Census Bureau demographic data with the stations
 - Only the Median Income Data is provided by Zip Code
 - All other data comes at the “Community District” level (multiple Zip Codes); did not have time to map Zip Codes to Districts

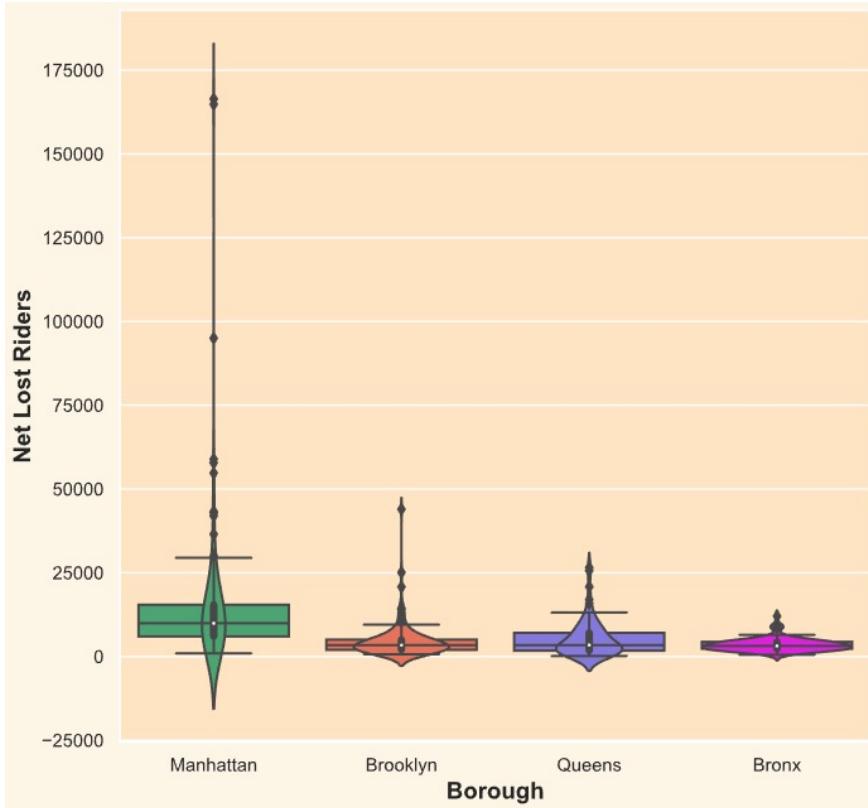
Early Exploratory Visualizations



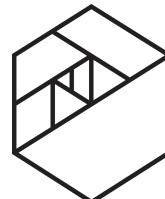
- No surprise (widely reported): Weekday ridership decline seems to correlate with income level
- **Caution:** This might reflect a connection between higher income level and ability to work from home (or use alternative transportation)
- But there's no way to be certain the people entering a given station actually live in that neighborhood



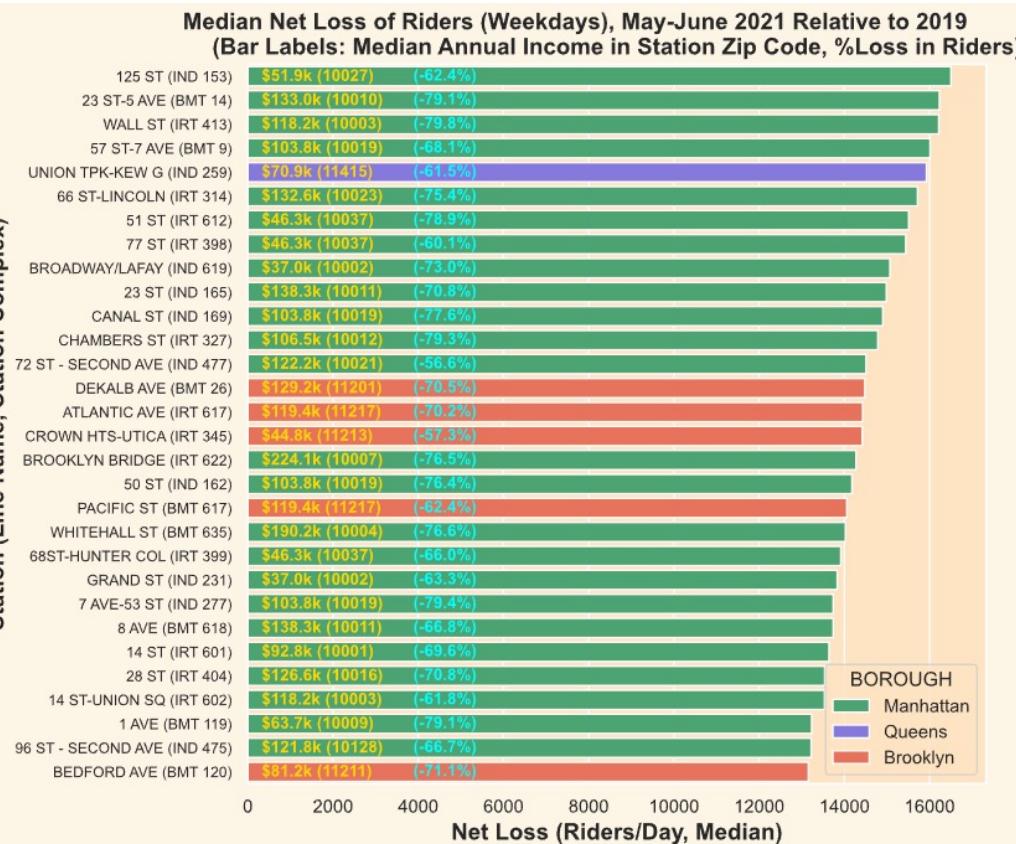
Early Exploratory Visualizations



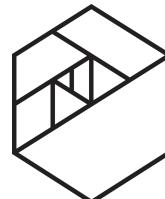
- Overlaying Box and Violin Plots by gives a nice insight into how the distributions of Net Lost Riders varies by borough
- Easy to see the “Big 6” in the Manhattan outliers
- The Bronx’s “tight” distribution reveals how relatively unchanged its ridership patterns were between 2019 and 2021 relative to the other boroughs



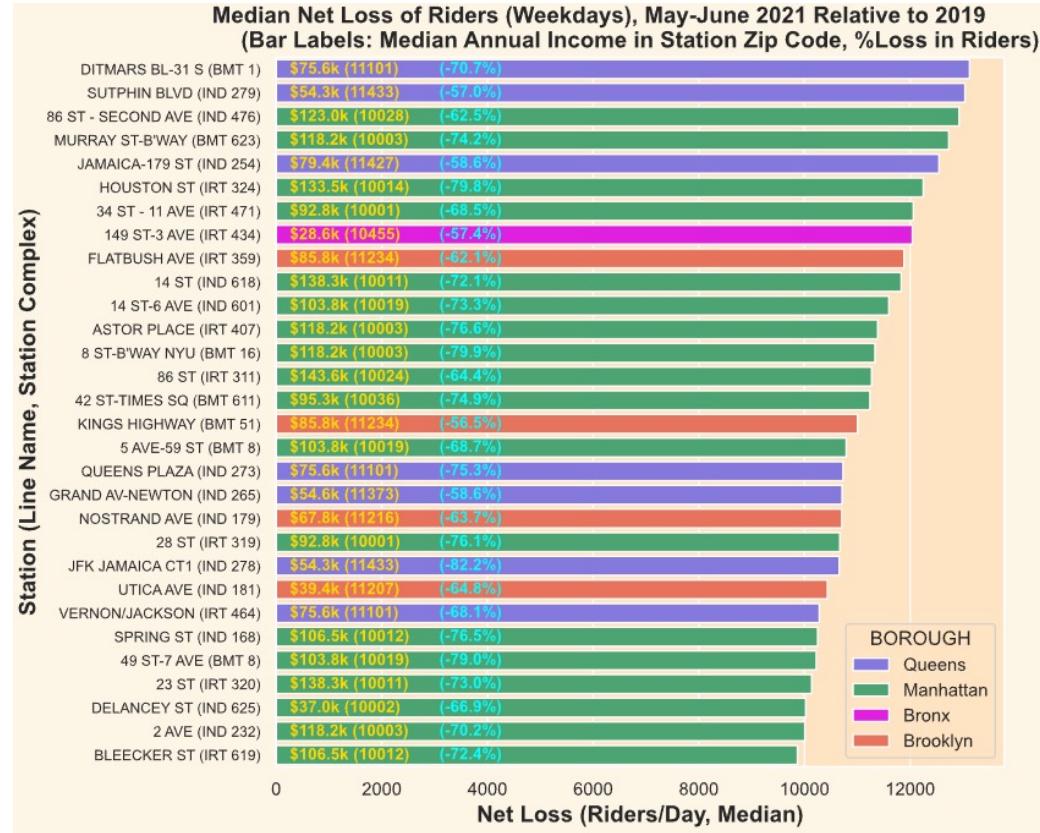
All Remaining Stations, 30 at a Time (Weekdays)



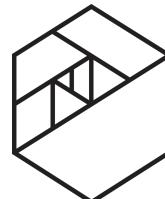
- \$1.20M/day in revenue loss
- Still mostly Manhattan stations



All Remaining Stations, 30 at a Time (Weekdays)



- \$0.93M/day in revenue loss
- Queens stations showing up a lot more, as well as (finally) a Bronx station



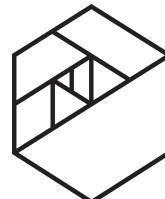
All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



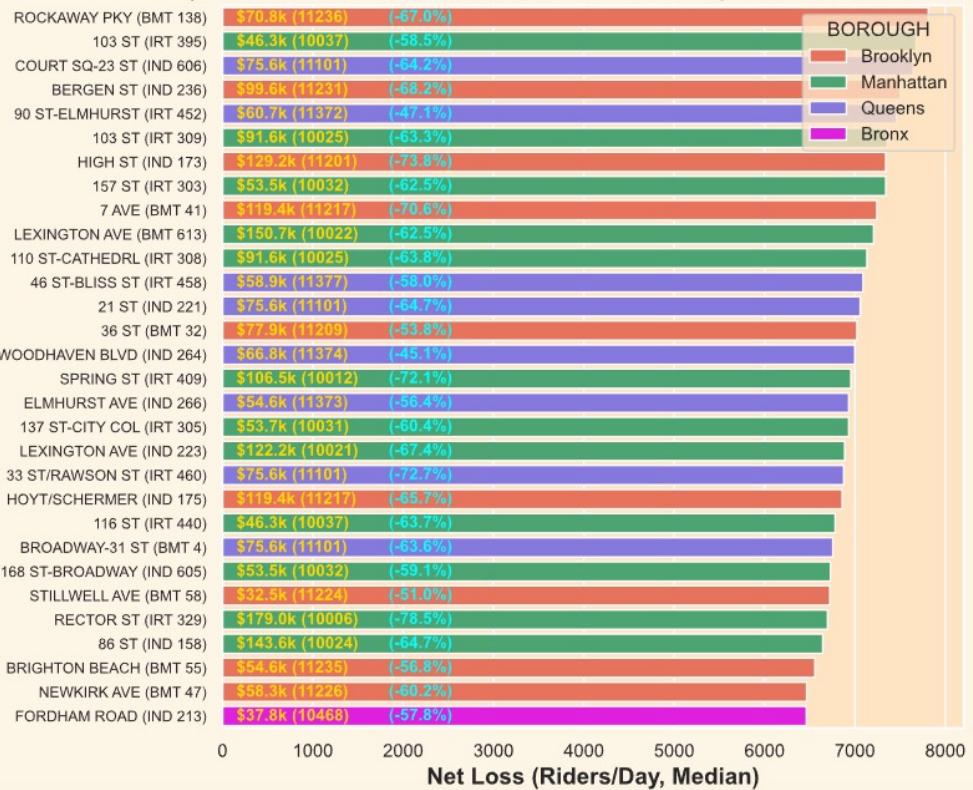
- \$0.73M/day in revenue loss
- Even more Bronx stations now

METIS®

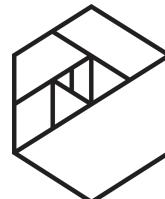


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



- \$0.58M/day in revenue loss

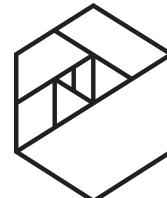


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

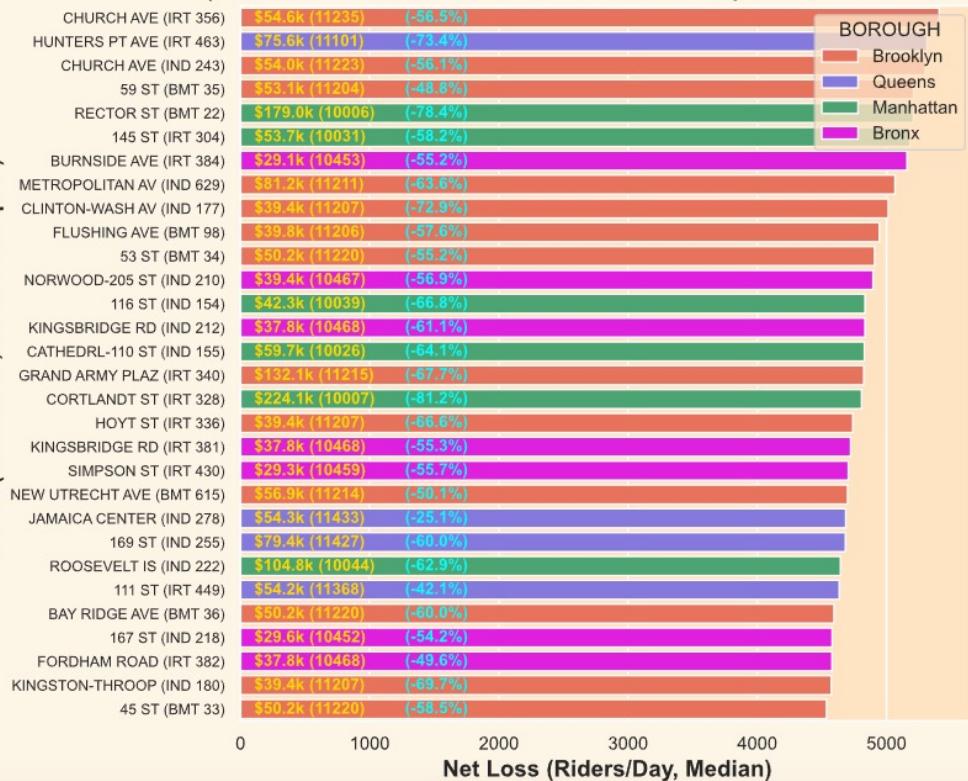


- \$0.49M/day in revenue loss
- Note that there are more lower-income Zip Code stations than before

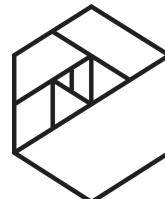


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

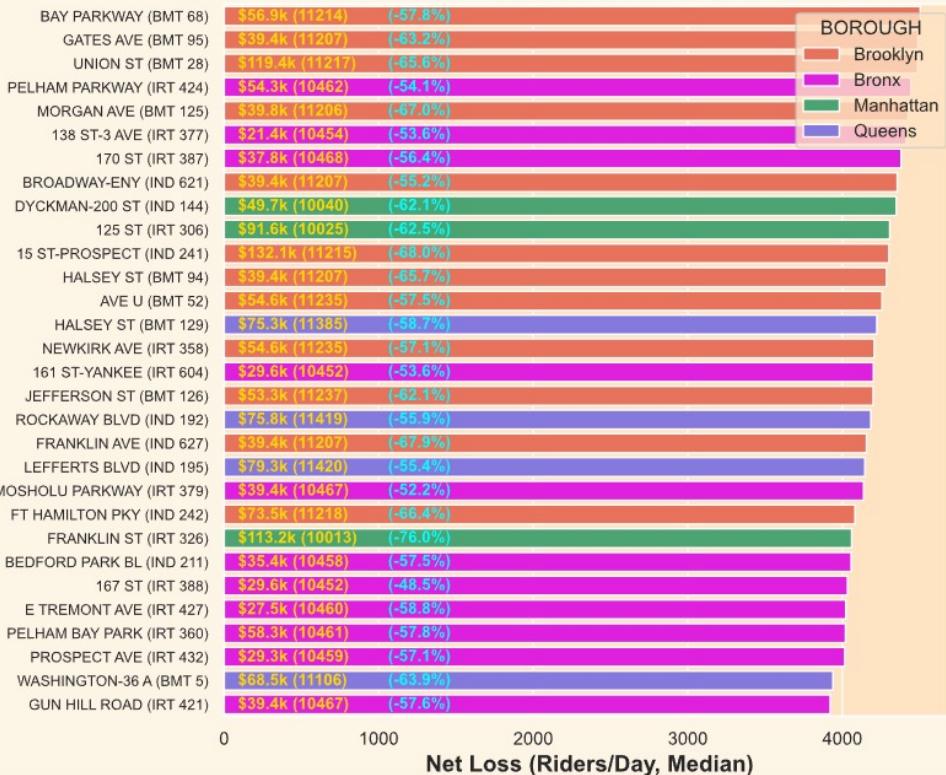


- \$0.40M/day in revenue loss
- A lot fewer Manhattan stations now

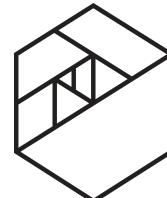


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

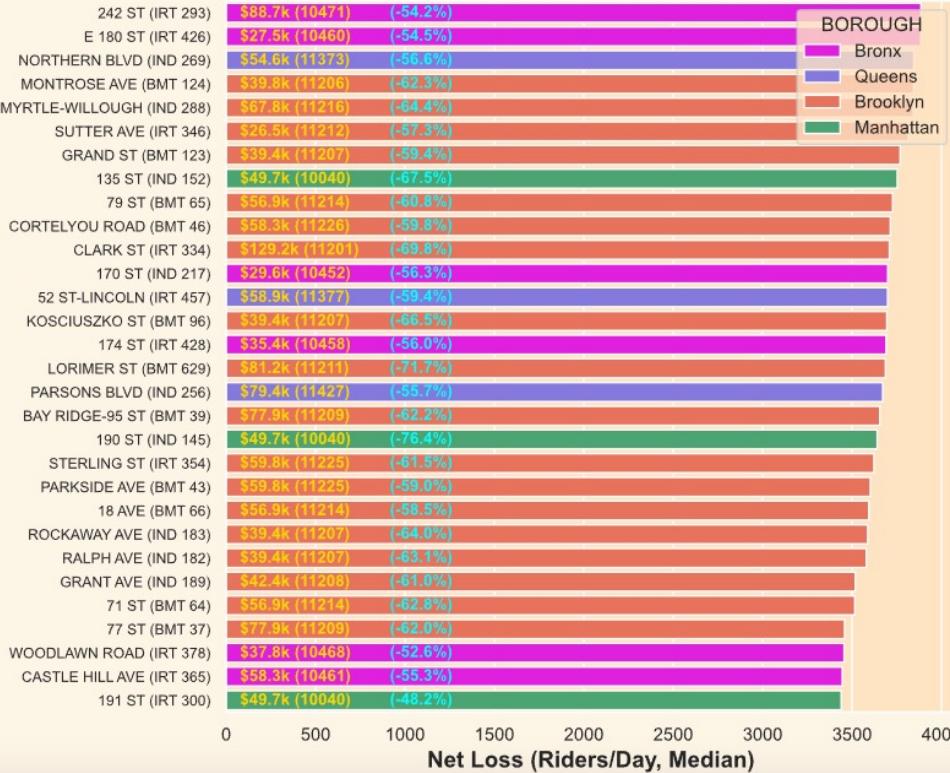


- \$0.35M/day in revenue loss

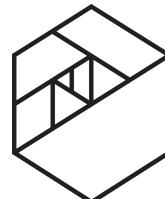


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



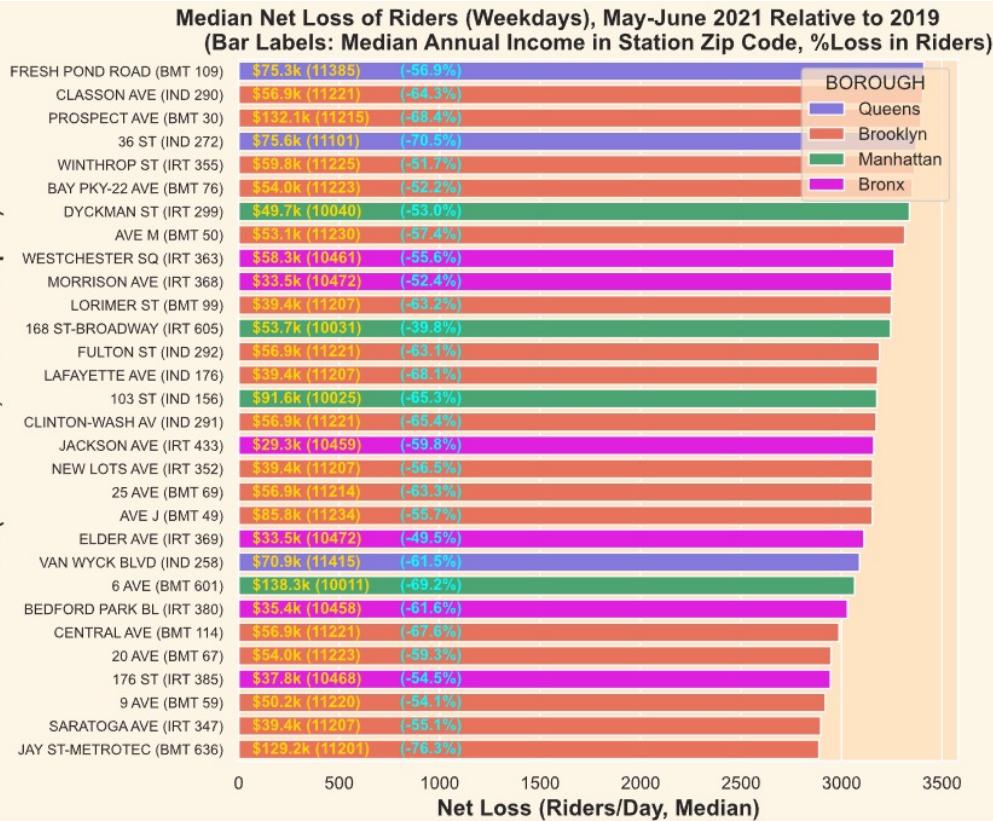
- \$0.30M/day in revenue loss
- Almost no high-income Zip Codes now



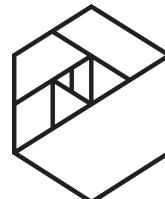
All Remaining Stations, 30 at a Time (Weekdays)

METIS®

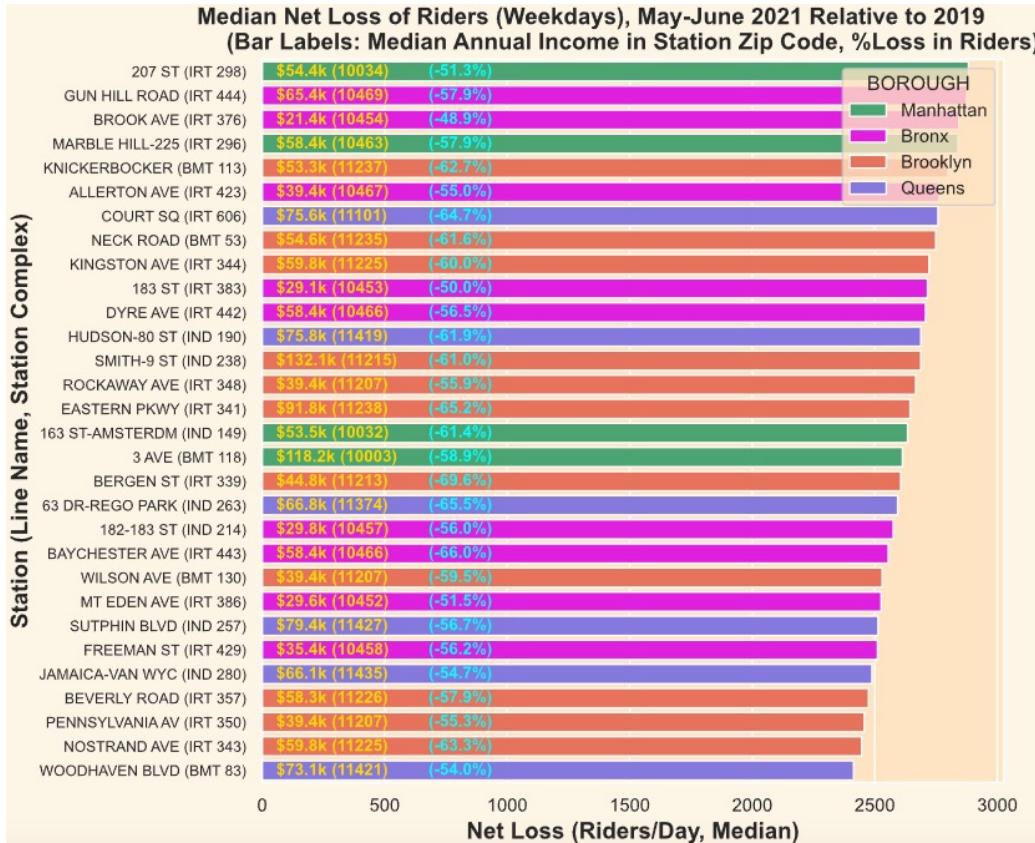
Station (Line Name, Station Complex)



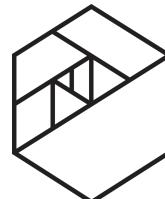
- \$0.26M/day in revenue loss



All Remaining Stations, 30 at a Time (Weekdays)



- \$0.22M/day in revenue loss

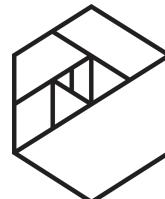


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

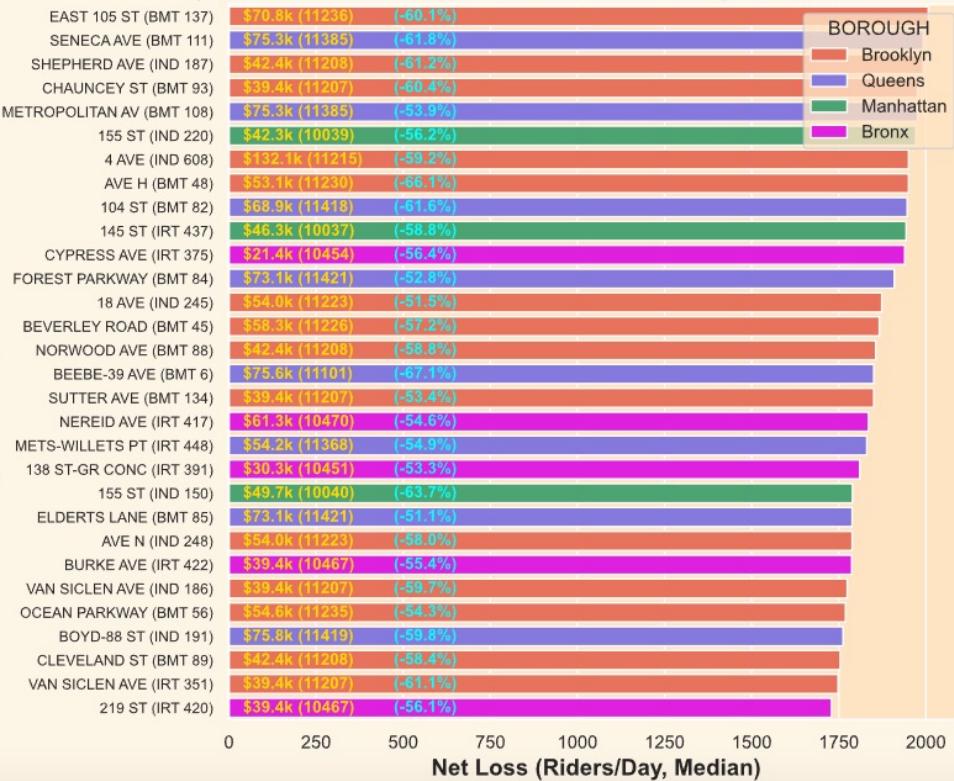


- \$0.18M/day in revenue loss

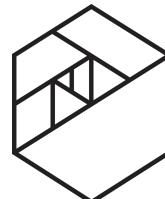


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

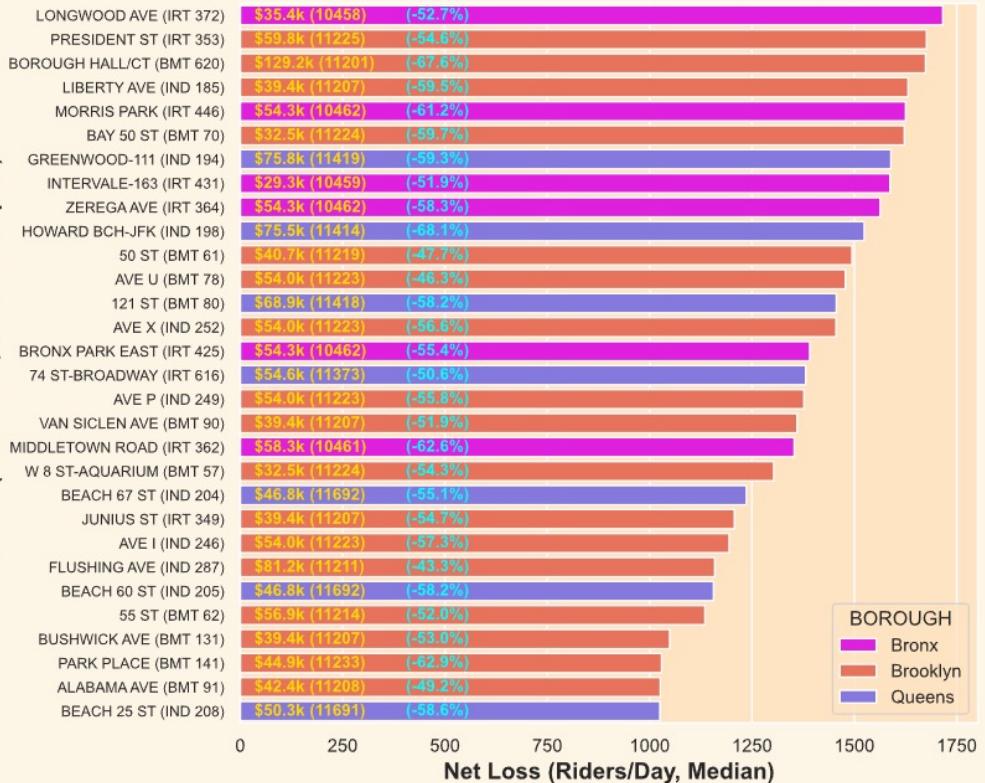


- \$0.15M/day in revenue loss
- The few remaining Manhattan stations are in lower-income neighborhoods now

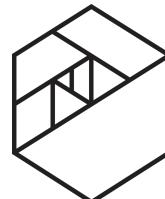


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

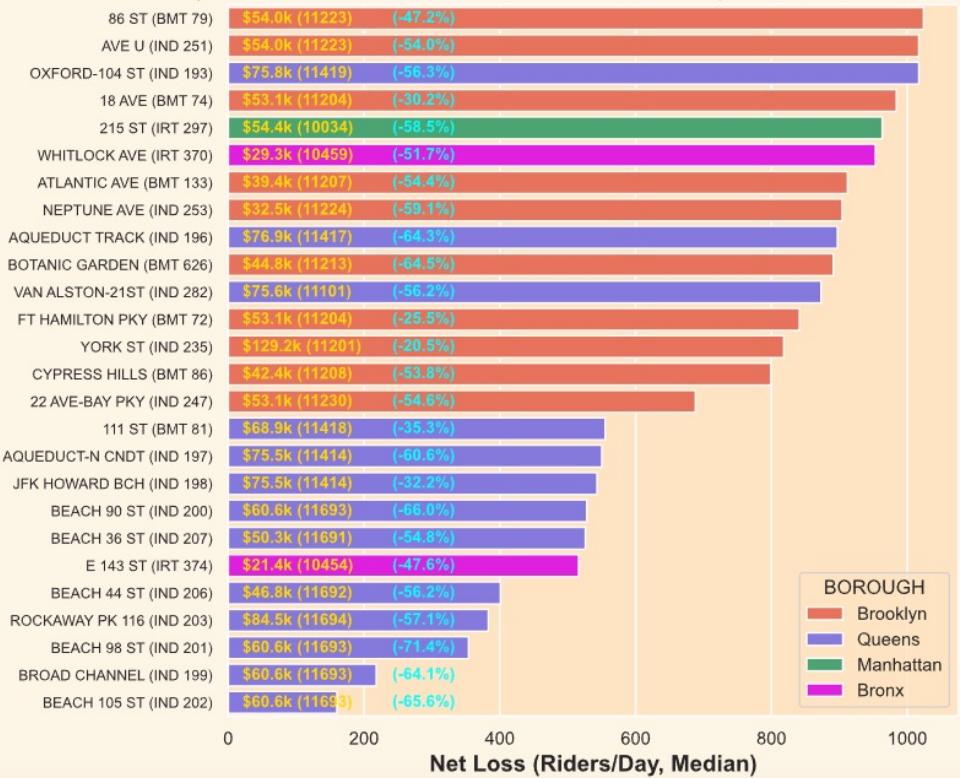


- \$0.11M/day in revenue loss
- No Manhattan stations at all

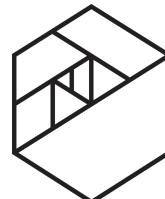


All Remaining Stations, 30 at a Time (Weekdays)

Median Net Loss of Riders (Weekdays), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

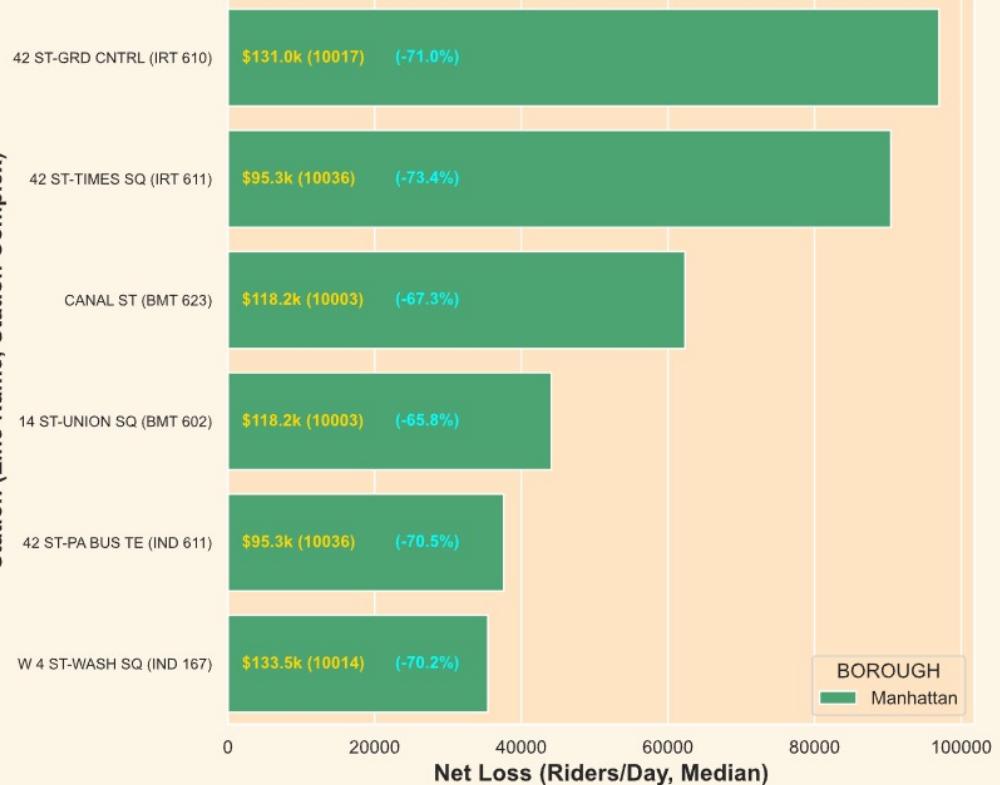


- \$0.05M/day in revenue loss

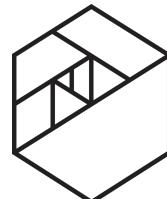


The “Big 6” Stations (Weekends)

Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

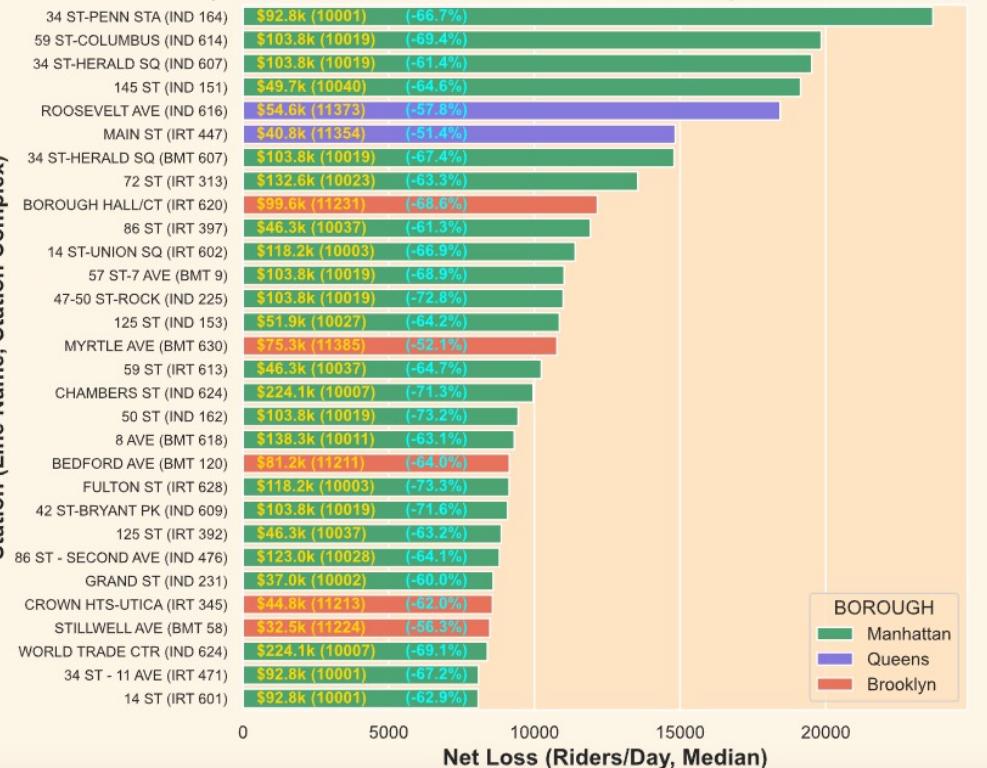


- \$1.01M/day in revenue loss
- Less ridership on Weekends, but for loss leaders, this much lost revenue is still non-trivial

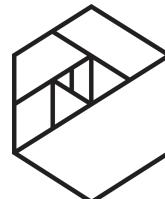


All Remaining Stations, 30 at a Time (Weekends)

Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

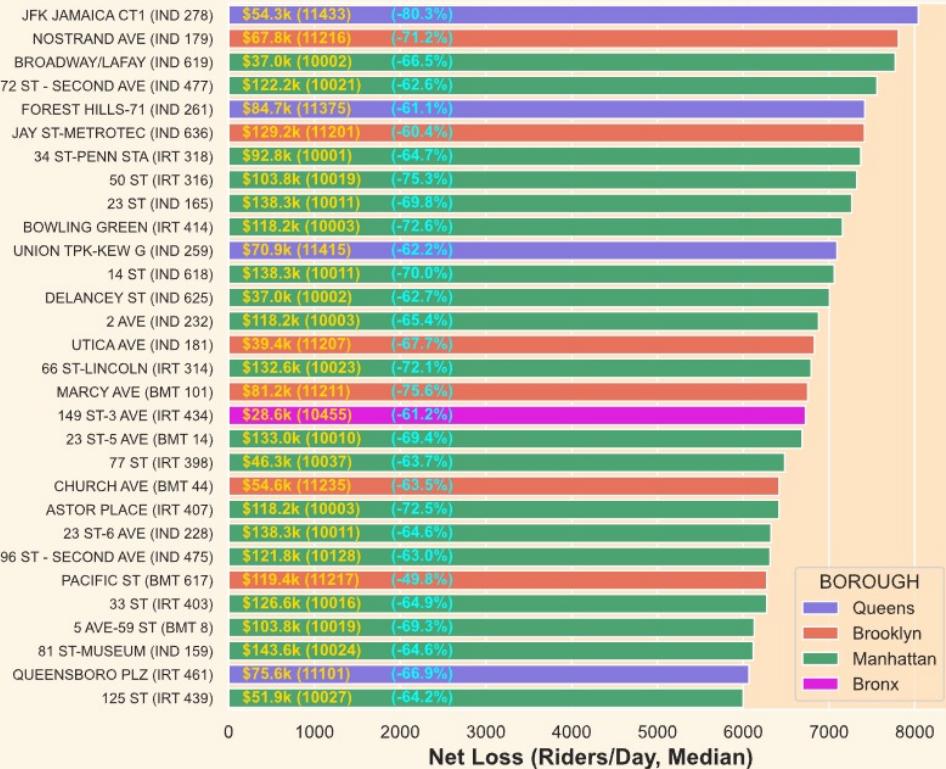


- \$0.98M/day in revenue loss

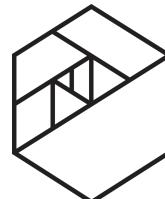


All Remaining Stations, 30 at a Time (Weekends)

Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

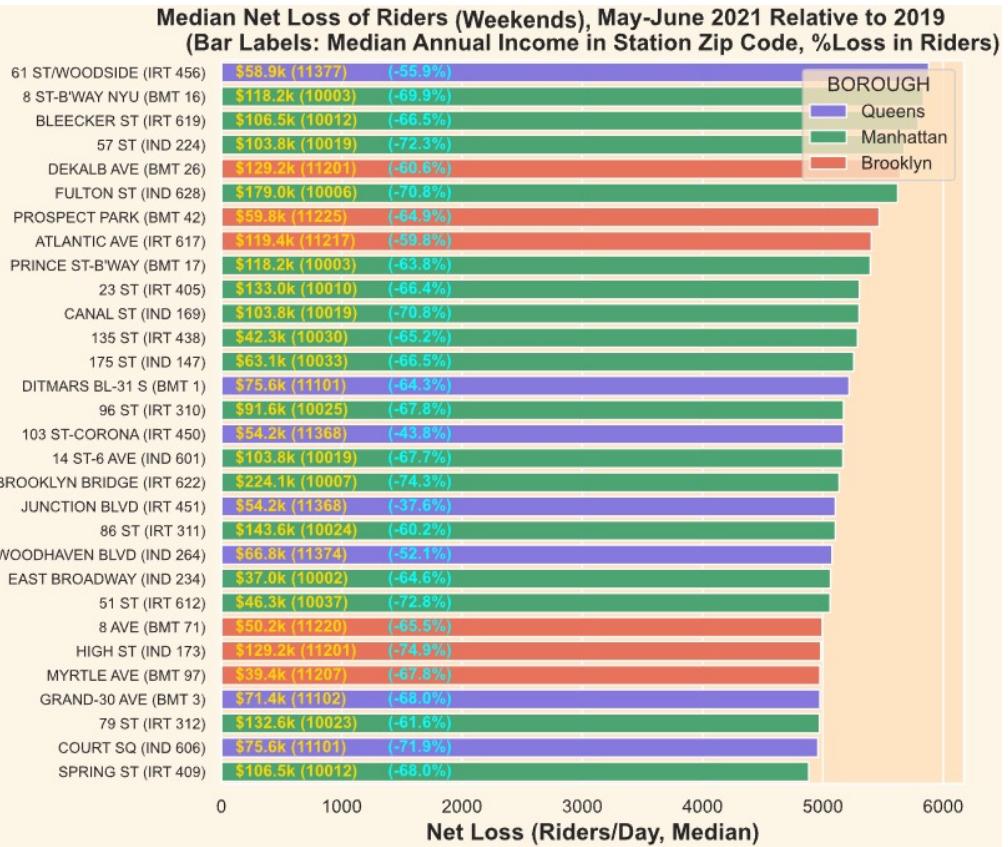


- \$0.57M/day in revenue loss

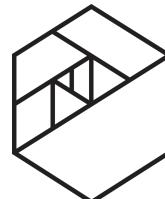


All Remaining Stations, 30 at a Time (Weekends)

Station (Line Name, Station Complex)



- \$0.43M/day in revenue loss

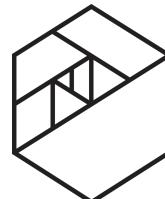


All Remaining Stations, 30 at a Time (Weekends)

Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

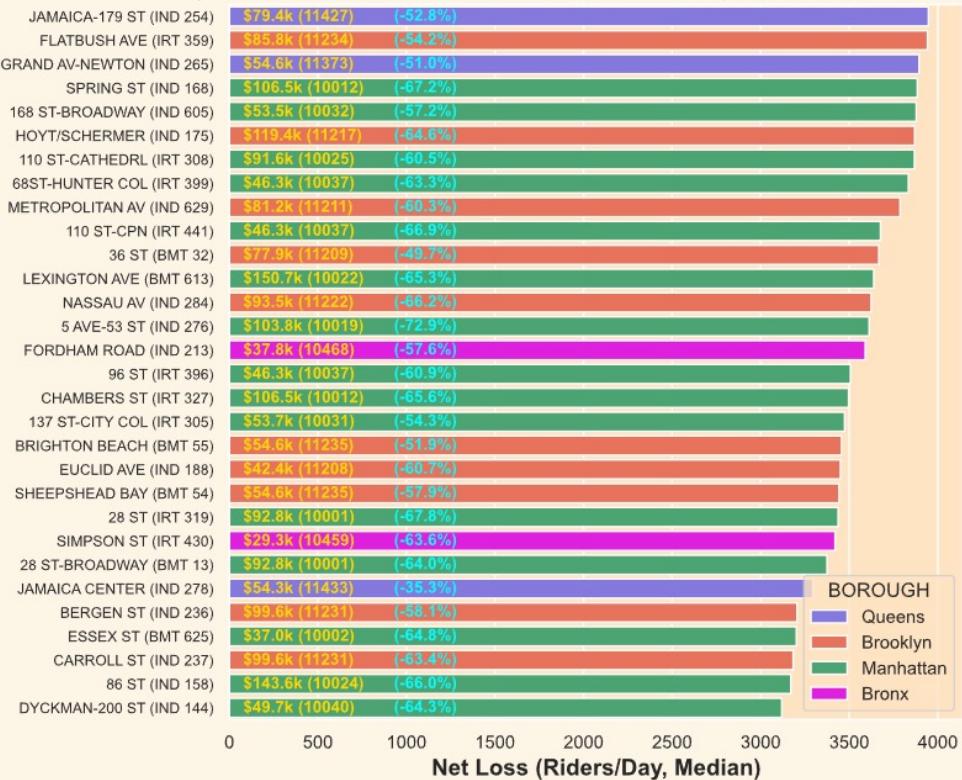


- \$0.35M/day in revenue loss

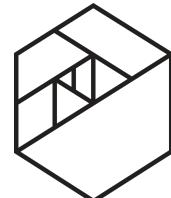


All Remaining Stations, 30 at a Time (Weekends)

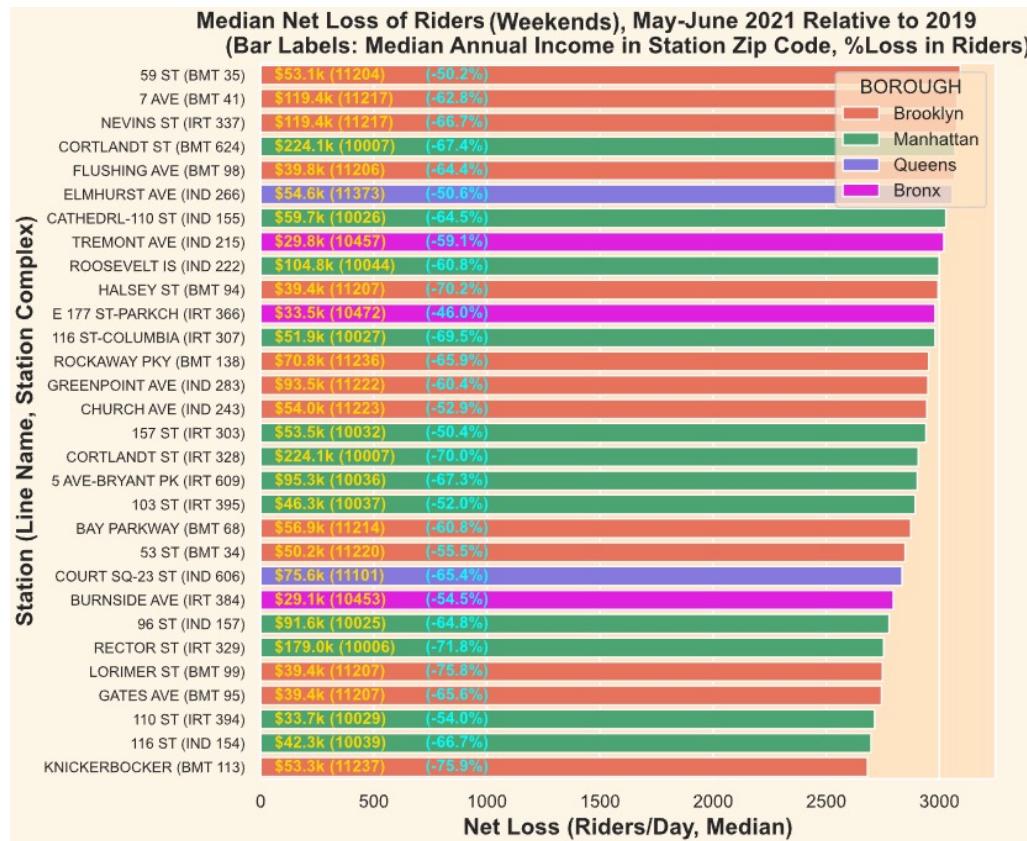
Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



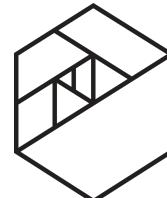
- \$0.29M/day in revenue loss



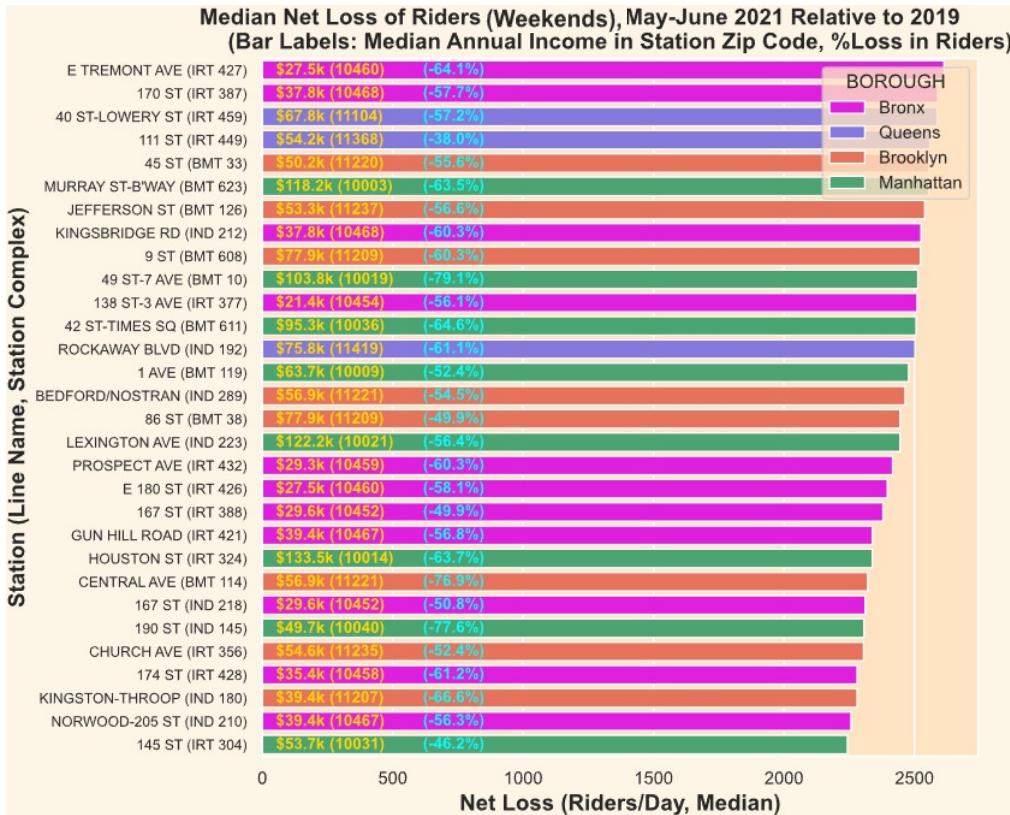
All Remaining Stations, 30 at a Time (Weekends)



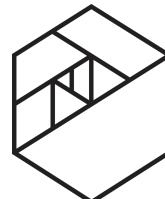
- \$0.24M/day in revenue loss



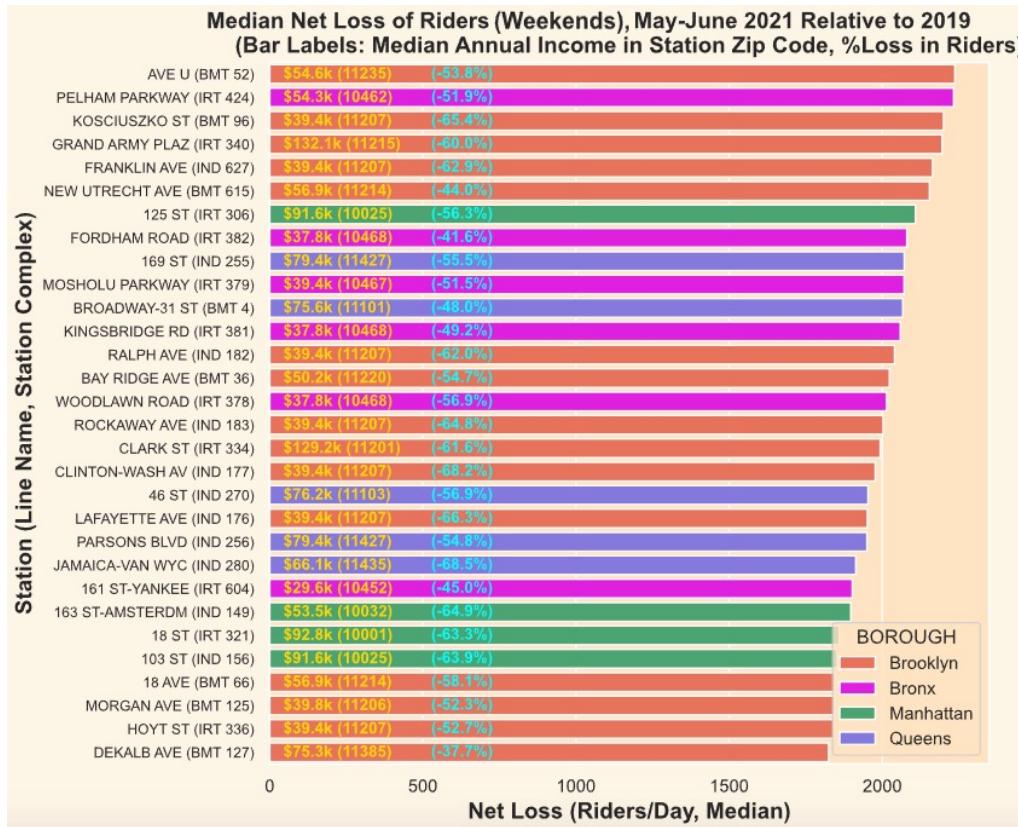
All Remaining Stations, 30 at a Time (Weekends)



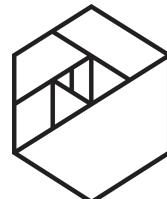
- \$0.20M/day in revenue loss



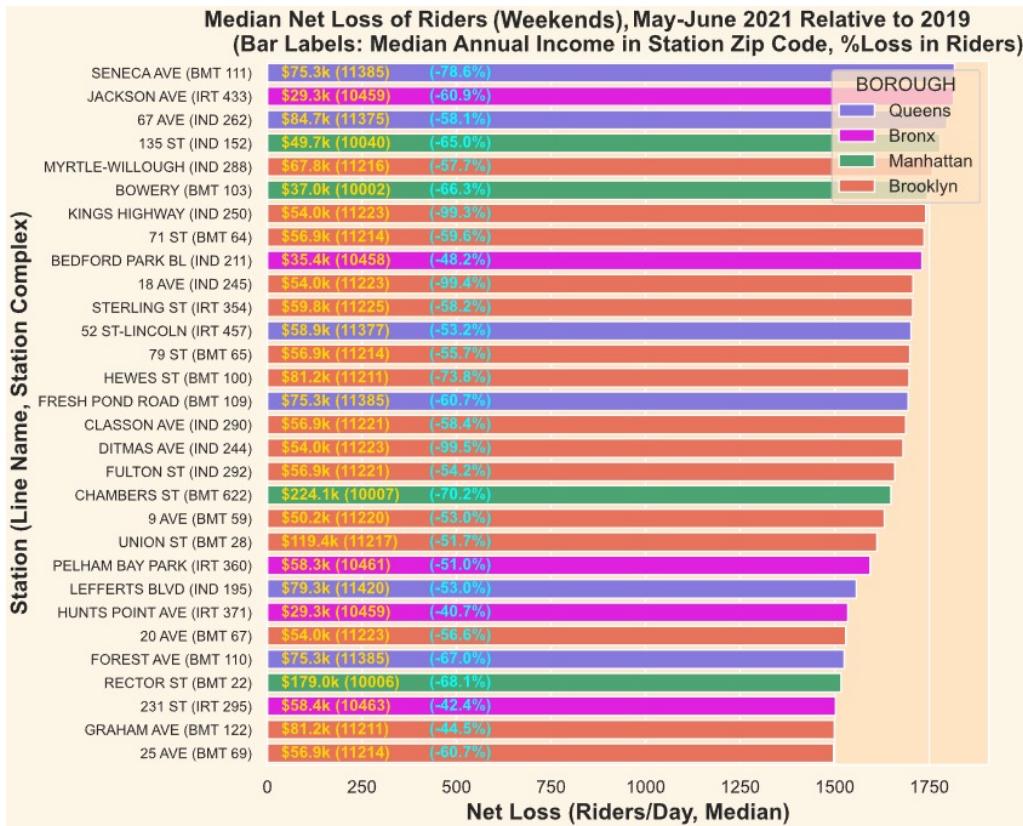
All Remaining Stations, 30 at a Time (Weekends)



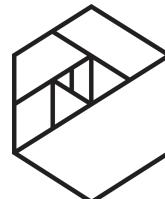
- \$0.17M/day in revenue loss



All Remaining Stations, 30 at a Time (Weekends)

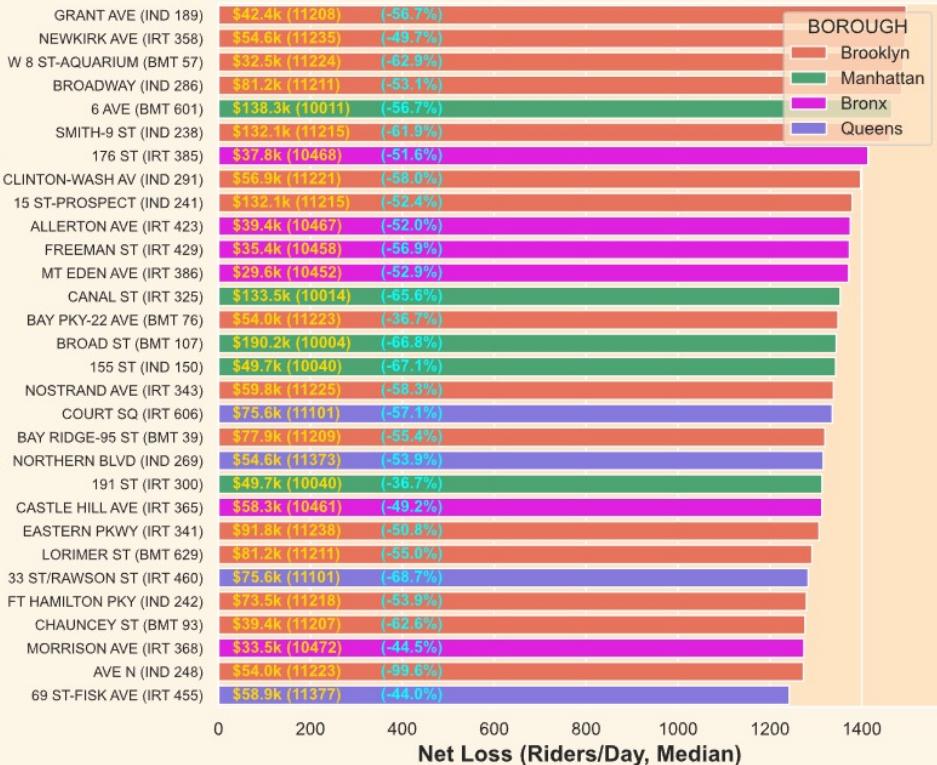


- \$0.14M/day in revenue loss

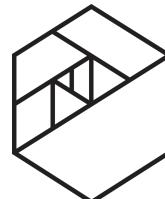


All Remaining Stations, 30 at a Time (Weekends)

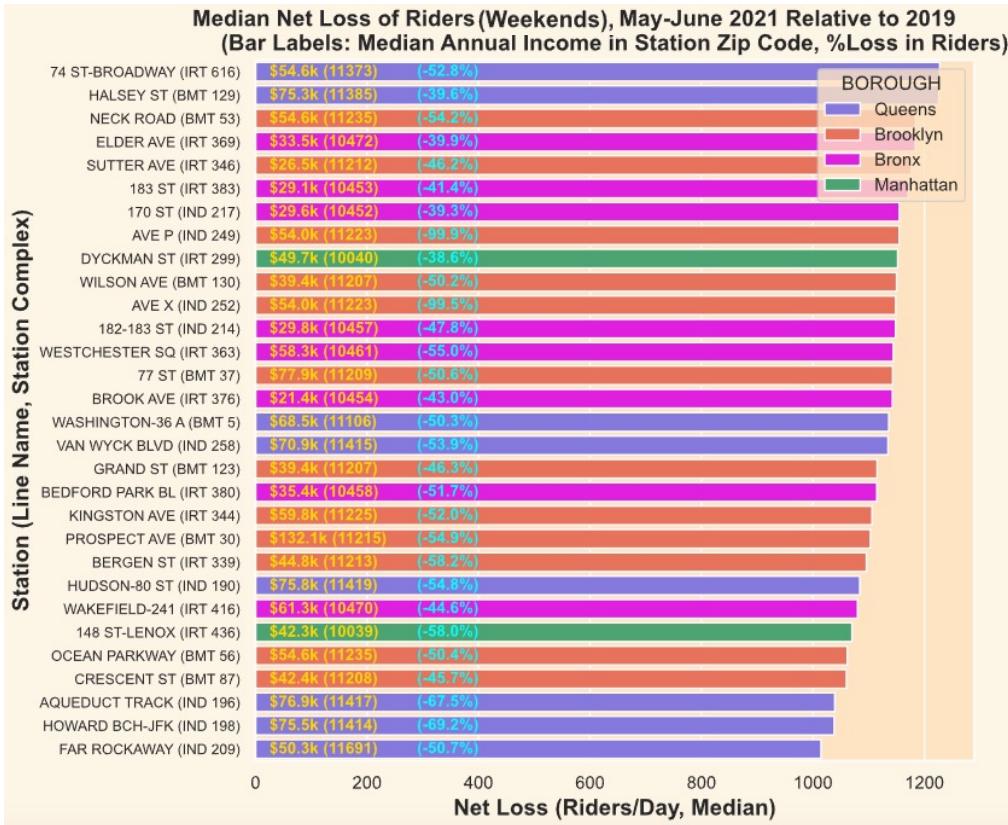
Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



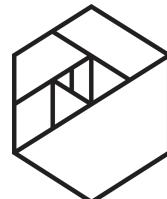
- \$0.11M/day in revenue loss



All Remaining Stations, 30 at a Time (Weekends)



- \$0.09M/day in revenue loss

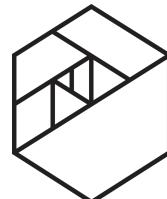


All Remaining Stations, 30 at a Time (Weekends)

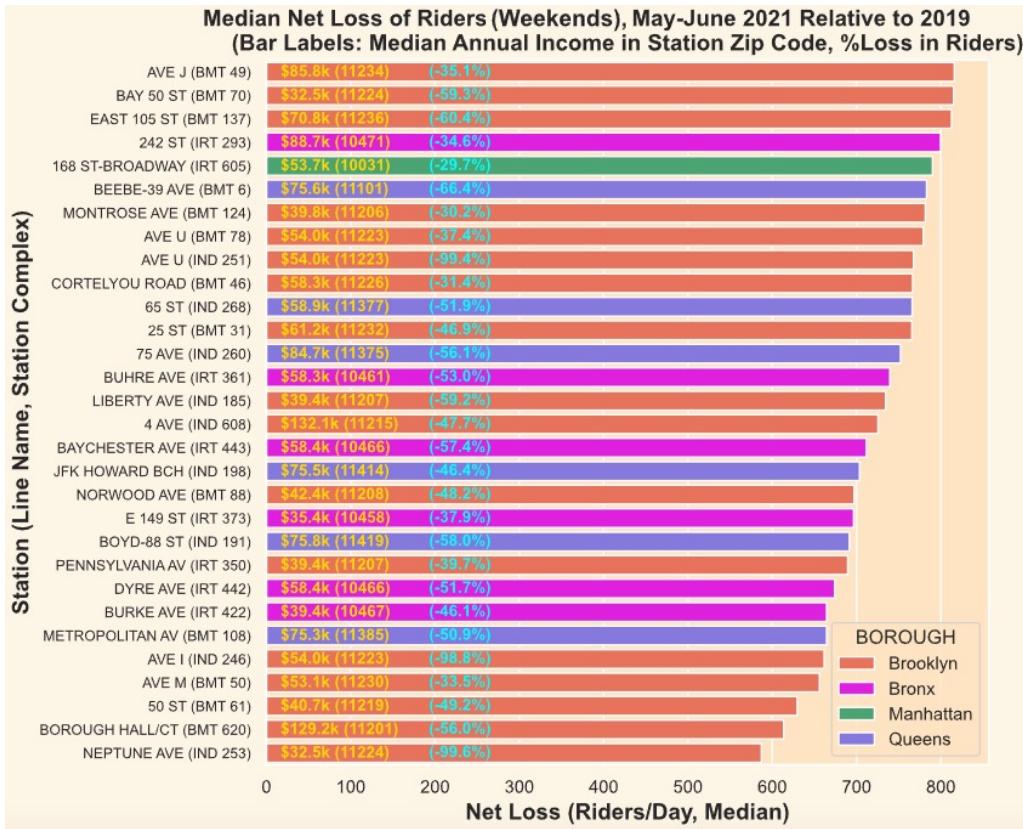
Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)



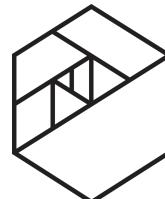
- \$0.07M/day in revenue loss



All Remaining Stations, 30 at a Time (Weekends)



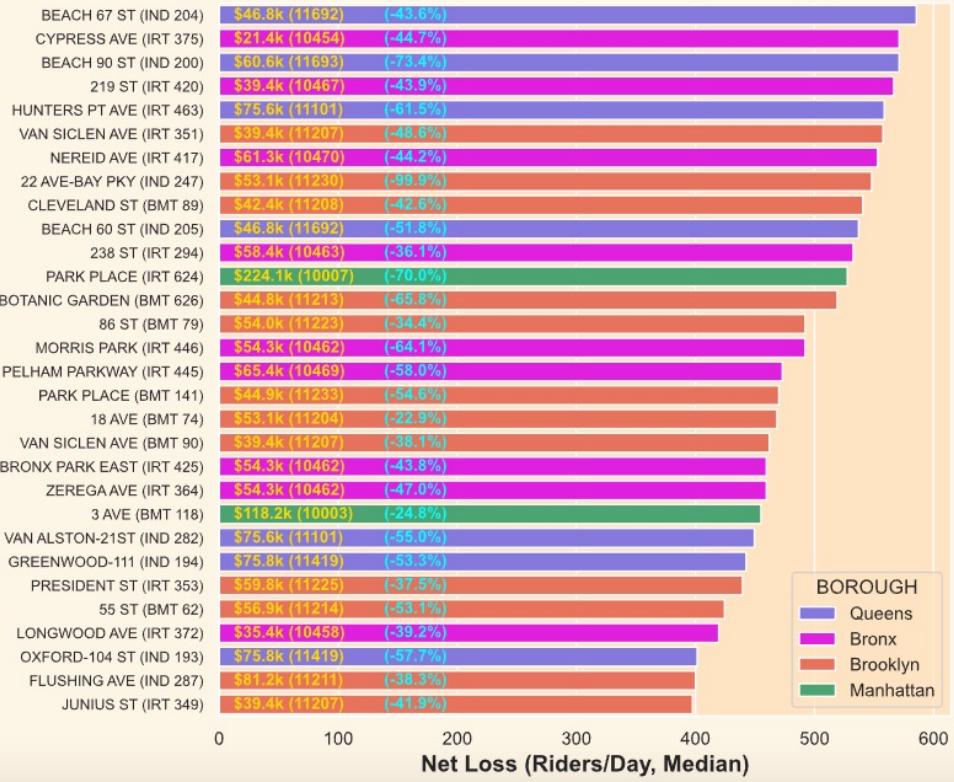
- \$0.06M/day in revenue loss



All Remaining Stations, 30 at a Time (Weekends)

Median Net Loss of Riders (Weekends), May-June 2021 Relative to 2019
(Bar Labels: Median Annual Income in Station Zip Code, %Loss in Riders)

Station (Line Name, Station Complex)



- \$0.04M/day in revenue loss