The launch of Spotify KIDS
A data-driven approach

By Valentina Rizzati

The Ask: Target list of MTA stations for Outdoor campaign

Spotify KIDS: Outdoor campaign 📩 Inbox ✕

Skylar Newman <sky.newman@spotify.com>
to me ▾

Mon, 19 Apr, 18:23 (6 days ago) ☆

Hey Vale,

As discussed in our meeting last week, we are planning an Out-Of-Home campaign as part of
marketing campaign for the launch of Spotify KIDS. I've already shared all the documentation
launch in a separate email. Lucy Zein is the main poc on the Product Marketing team and I a
launch.

As a reminder, the target for the campaign is:
● Geo: NYC (five boroughs)
● Demo: Families
  ○ Age > 30
  ○ Household Income > $90k
  ○ Persons Per Household > 2

We intend to run the campaign close to MTA stations in neighbourhoods that both fit into our
high foot traffic.

We need your support to identify these MTA stations. Can you please provide a target list of MTA stat
next week? We need to pass the information to the agency.

Thanks and let me know if you have any question.

Cheers,
Sky

As a reminder, the target for the campaign is:
● Geo: NYC (five boroughs)
● Demo: Families
  ○ Age > 30
  ○ Household Income > $90k
  ○ Persons Per Household > 2

The plan of action

Next Steps (4)

(1) Data Requirements

Final Recommendation (4)

(2) Data Model, Process and Tools

Identify high-traffic stations (4)

(3) Define Target PUMAs

(1) Data Requirements

Census Data
Source: [ACS 1-Year Estimates-Public Use Microdata Sample (2019)](#)
- Level of aggregation: Public Use Microdata Sample (PUMA)
  - 55 PUMAs in the five boroughs of NYC
- Period of Analysis: 2019
- Data retrieved:
  - Average Age by PUMA
  - Average Persons per Household by PUMA
  - Average Household Income by PUMA

MTA Data
Source: [Turnstile Data](#)
- Level of aggregation: MTA station
  - 378 unique stations
- Period of Analysis: 12 weeks (3/1/21 - 3/27/21)
- Data retrieved:
  - CA / UNIT / SCP / STATION = 1 turnstile
  - LINENAME: available train lines
  - DIVISION: line station belonged to
  - DATE (MM-DD-YY)
  - TIME (hh:mm:ss)
  - DESC: REGULAR or RECOVR AUD
  - ENTRIES: daily cumulative record
  - EXITS: daily cumulative record

2 Data Model, Process and Tools

1 Set up the data

2 Explore and clean the data

3 Model and visualize the data

TOOLS

OUTPUT

mta.db

mta_data table

census.db

census_data table

Renamed column HOUSEHOLD_INCOME and changed all data types to floats

CA	TEXT
UNIT	TEXT
SCP	TEXT
STATION	TEXT
LINENAME	TEXT
DIVISION	TEXT
DATE	TEXT
TIME	TEXT
DESC	TEXT
ENTRIES	INTEGER
EXITS	INTEGER

PUMA	TEXT
AVG_AGE	TEXT
AVG_PERSONS_HOUSEHOLD	TEXT
HOUSEHOLD_INCOME	TEXT

Entries and Exits by hour for an anomalous turnstile

Box Plot of Daily Entries

Top 30 MTA stations by unique traffic
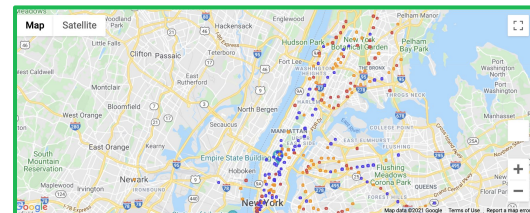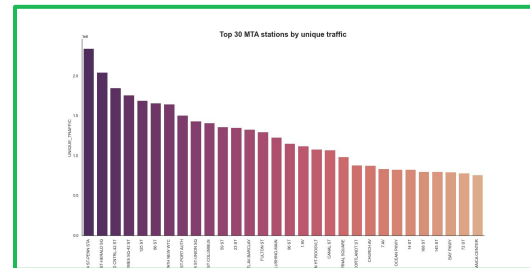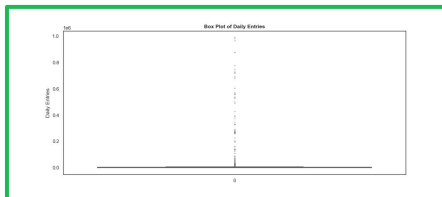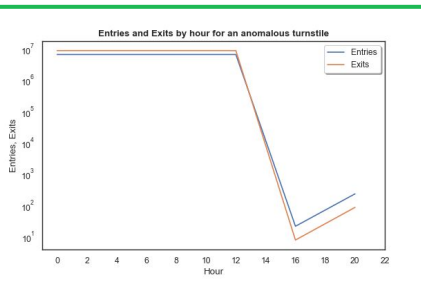
Define Target Public Use Microdata Areas

New York City Economic Development Region 2010 PUMAs
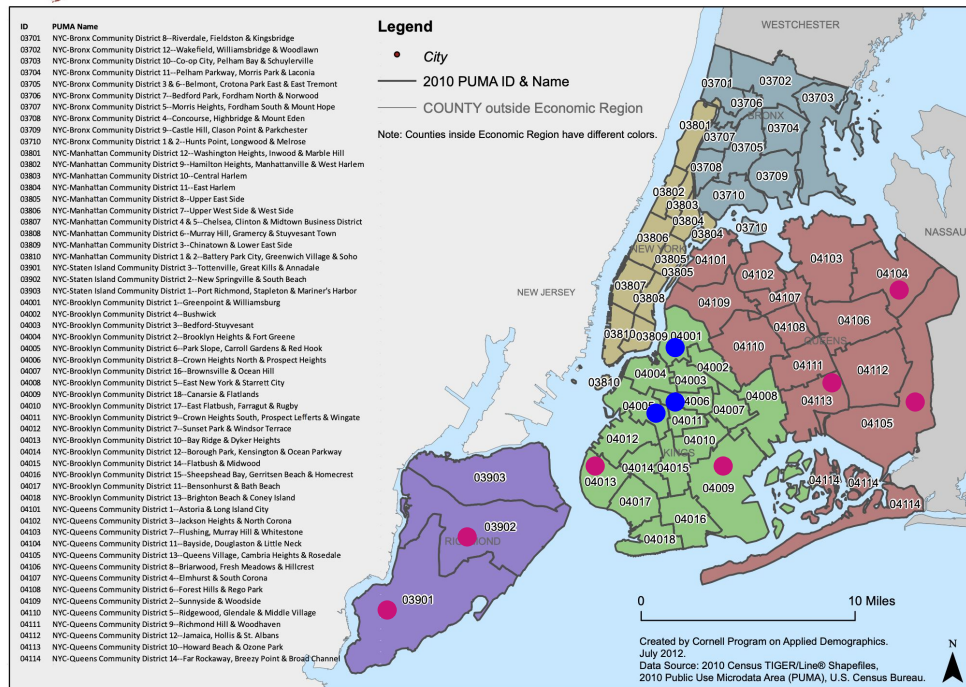
PUMA		AVG_AGE	AVG_PERSONS_HOUSEHOLD	AVG_HOUSEHOLD_INCOME	IN_TARGET
20	NYC-Staten Island Community District 3--Tottenville, Great Kills & Annadale PUMA; New York	40.199216	2.610094	114466.988500	YES
21	NYC-Staten Island Community District 2--New Springville & South Beach PUMA; New York	41.665055	2.629408	109321.873700	YES
23	NYC-Brooklyn Community District 1--Greenpoint & Williamsburg PUMA; New York	35.148875	2.231105	119603.896700	YES
27	NYC-Brooklyn Community District 6--Park Slope, Carroll Gardens & Red Hook PUMA; New York	37.341553	2.240369	190640.735300	YES
28	NYC-Brooklyn Community District 8--Crown Heights North & Prospect Heights PUMA, New York	37.556544	2.082794	91788.768770	YES
31	NYC-Brooklyn Community District 18--Canarsie & Flatlands PUMA, New York	40.056227	2.581528	90082.101790	YES
35	NYC-Brooklyn Community District 10--Bay Ridge & Dyker Heights PUMA, New York	41.135425	2.354281	90954.328810	YES
44	NYC-Queens Community District 11--Bayside, Douglaston & Little Neck PUMA; New York	43.431911	2.406867	99957.402150	YES
45	NYC-Queens Community District 13--Queens Village, Cambria Heights & Rosedale PUMA; New York	42.245283	2.864101	103646.073300	YES
53	NYC-Queens Community District 10--Howard Beach & Ozone Park PUMA, New York	41.656728	2.778062	98921.659610	YES

Public Use Microdata Areas (PUMA): non-overlapping, statistical geographic areas that partition each state or equivalent entity into geographic areas containing no fewer than 100,000 people each. Source: United States Census Bureau

PUMAs have firstly been targeted based on:
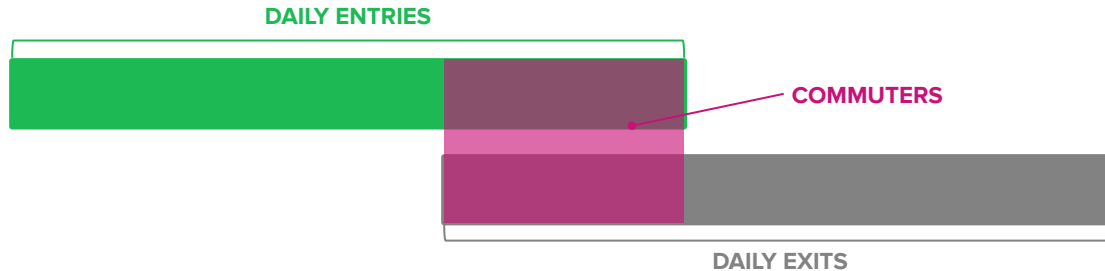- Average Age > 30
- Average Household Income > $90k
- Average Persons Per Household > 2

Because of considerations regarding the demographics and foot traffic, the resulting 10 PUMAs in target have been reduced to 3:
- **04001**: NYC Brooklyn District 1 - Greenpoint, Williamsburg
- **04005**: NYC Brooklyn District 6 - Park Slope, Carrolls Gardens
- **04006**: NYC Brooklyn District 8 - Crown Heights North, etc..

Source: Pad Human Cornell, 2010

Identify high-traffic stations (1/2)

4

1

DAILY ENTRIES

COMMUTERS

DAILY EXITS

2

UNIQUE ENTRIES

COMMUTERS

UNIQUE EXITS

UNIQUE TRAFFIC

KPI the algorithm optimizes for

3

NEW TRAFFIC

BASELINE TRAFFIC

NEW TRAFFIC

UNIQUE TRAFFIC
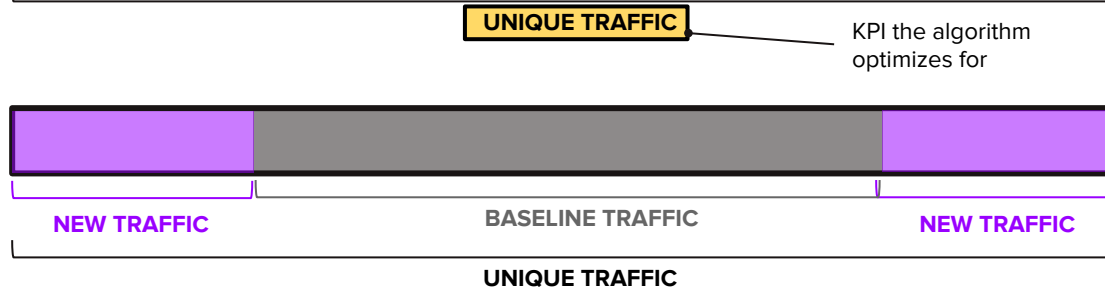
FORMULAS

TOTAL TRAFFIC = DAILY ENTRIES + DAILY EXITS

TRAFFIC LEVEL:
- HIGH - 3rd tertile of TOTAL TRAFFIC
- MED - 2nd tertile of TOTAL TRAFFIC
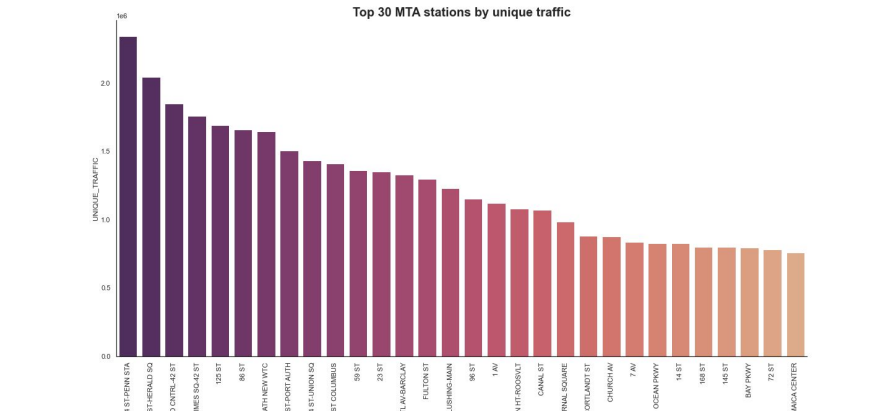- LOW - 1st tertile of TOTAL TRAFFIC

COMMUTERS[1] = x% * TOTAL TRAFFIC

UNIQUE TRAFFIC = DAILY ENTRIES + DAILY EXITS - COMMUTERS

UNIQUE BASELINE TRAFFIC[1] = y% * UNIQUE TRAFFIC

UNIQUE NEW TRAFFIC = UNIQUE TRAFFIC - UNIQUE BASELINE TRAFFIC

1. x% and y% determined in line with assumptions

Identify high-traffic stations (2/2)

Top 30 MTA stations by unique traffic

MTA stations are associated with different colors based on traffic level:
- **HIGH TRAFFIC**
- **MEDIUM TRAFFIC**
- **LOW TRAFFIC**

To finalize the MTA stations selection for the campaign:
1. Within the targeted PUMAs, these high traffic stations were selected:
- **04001**: NYC Brooklyn District 1 - Greenpoint & Williamsburg
  - Bedford Ave
  - Lorimer Street
- **04005**: NYC Brooklyn District 6 - Park Slope, Carrolls Gardens, Red Hook
  - No MTA stop selected
- **04006**: NYC Brooklyn District 8 - Crown Heights North, Prospect Heights
  - Franklin Av-Medgar Evers College
  - Crown Hts - Utica Av
2. Based on their proximity to PUMA 04005 these high traffic stations were selected:
  - DeKalb Av
  - Jay St - MetroTech
3. Based on the traffic level analysis these stations were selected:
  - 34th Street-Penn Station
  - Grand Central 42nd Street
  - Times Square 42nd Street
  - 14th Street Union Square

> Selection based on traffic level and location in order to ensure a greater geo coverage of the campaign

5 Final Recommendation

Station	PUMA	Traffic Level	Selection Rationale
Bedford Avenue	**04001**	NYC Brooklyn District 1 - Greenpoint & Williamsburg	HIGH	Demographic, Traffic
Lorimer Street	**04001**	NYC Brooklyn District 1 - Greenpoint & Williamsburg	HIGH	Demographic, Traffic
Franklin Av-Medgar Evers College	**04006**: NYC Brooklyn District 8 - Crown H.N., Prospect H.	HIGH	Demographic, Traffic
Crown Hts - Utica Av	**04006**: NYC Brooklyn District 8 - Crown H.N., Prospect H.	HIGH	Demographic, Traffic
DeKalb Av	**04004:** NYC Brooklyn District 2 - Brooklyn Heights, Fort G.	HIGH	Demographic, Traffic
Jay St - MetroTech	**04004:** NYC Brooklyn District 2 - Brooklyn Heights, Fort G.	HIGH	Demographic, Traffic
34th Street-Penn Station	**03807:** NYC Manhattan District 4&5 - Chelsea, Midtown, etc..	HIGH	Traffic, Geo Coverage
Grand Central 42nd Street	**03807:** NYC Manhattan District 4&5 - Chelsea, Midtown, etc..	HIGH	Traffic, Geo Coverage
Times Square 42nd Street	**03807:** NYC Manhattan District 4&5 - Chelsea, Midtown, etc..	HIGH	Traffic, Geo Coverage
14th Street Union Square	**03810:** NYC Manhattan District 1&2 - Soho, Battery Park, etc..	HIGH	Traffic, Geo Coverage

6   Next Steps

- Iterate on the current version of the Google map for MTA stations:
    - Add polygons to visualize PUMAs
    - Explore filtering options for the map
    - Explore color gradient instead of three different colors based on traffic level
    - Explore Tableau integration to make map more interactive
- Analyze unique traffic by day of the week and time of day so as to enhance the targeting algorithm's accuracy
- Define marketing spend allocation by MTA station or cluster of MTA stations
- Build a tool on Tableau that allows the marketing team to include their level of marketing spend, demographic target, time of campaign and outputs which MTA stations or other locations they should advertise at

Thank You!