# Increasing Attendance at the WTWY Summer Gala

Billy Cottrell, Joyce Lee, Alan Lin, Auste Mastaviciute, Chelan Patton, Xu Zhou

7/6/18

# Objectives

WomenTechWomenYes wants to optimize placement of street teams around subway stations, in order to:

- Maximize Signups
- Maximize Attendance
- Maximize Donations

# Target Audience Profile

← This is Becky.

- 20 to 30-something woman
- Works for a tech company
- Enjoys Starbucks coffee

# Approach

Two part approach:

1. Maximize total number of people seen by street teams

    - Busiest* subway stations (MTA Data)

2. Maximize the likelihood of signups & attendance

    - Station proximity to a Starbucks (Google Maps Data)
    - Station proximity to Tech companies  (Google Maps Data)
    - Percentage gender score for each station (US Census Data)

*Stations with the most exits

# MTA Data Exploration

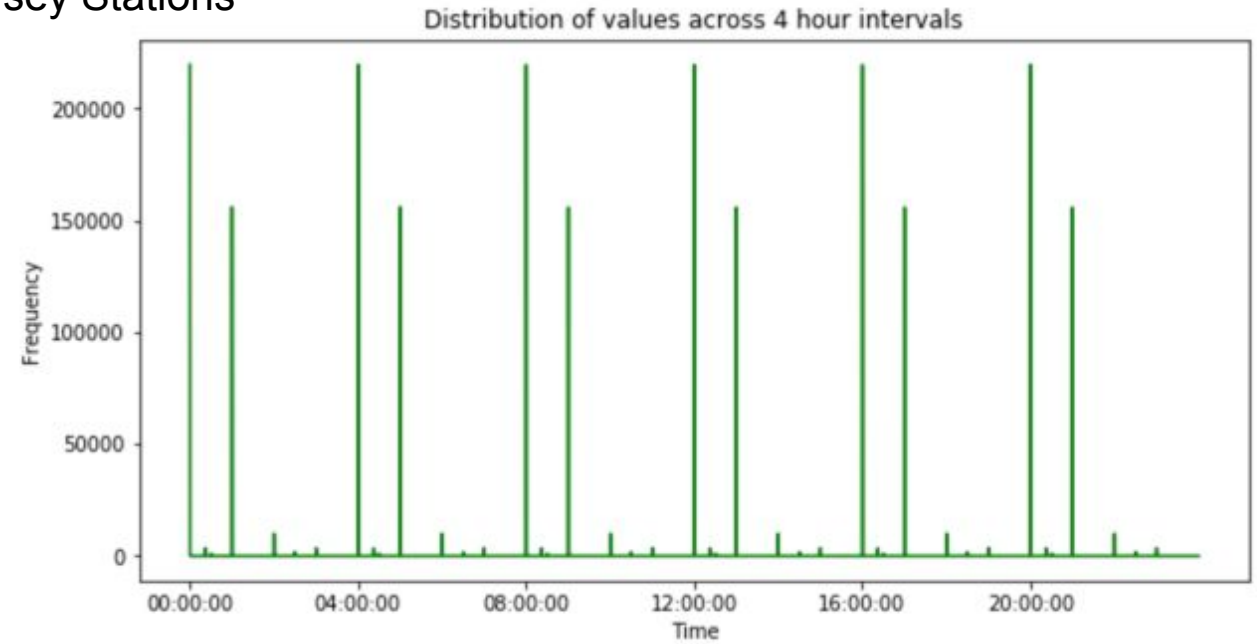
**We Identified:**

- Duplicated records
- Random Time Intervals
- Stations in New Jersey
- Extreme Number of Exits

# MTA Data Cleaning

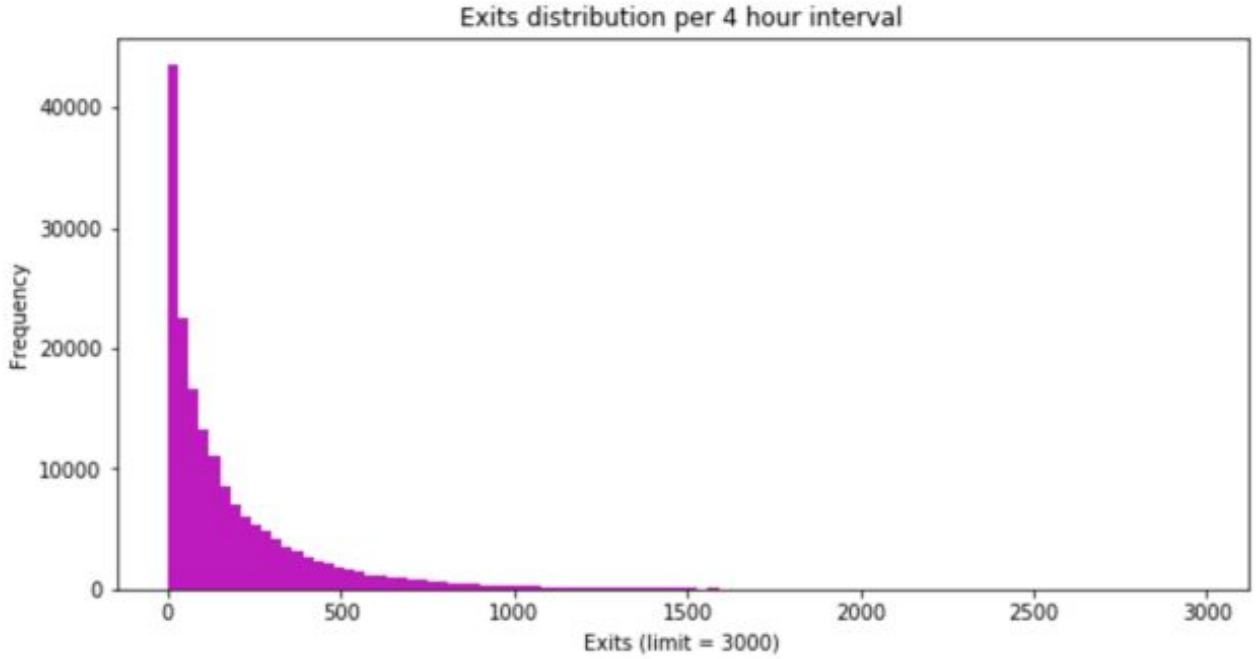
Records at random minutes:seconds

- 75% New Jersey Stations



# MTA Data Cleaning

Extreme values:



# MTA Analysis - Interactive Interface

## LET'S SEE IT IN ACTION!

```
In [ ]: activity, dct, df = main(month,day,hour,yrs_back)
```

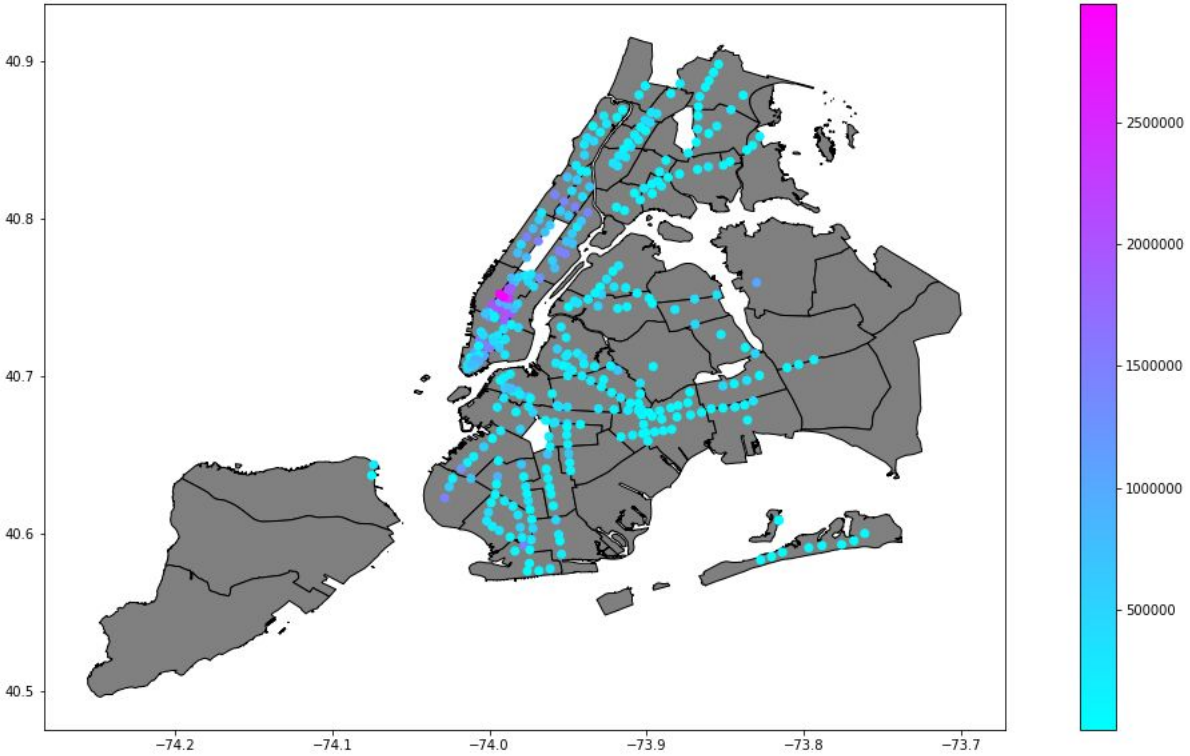
```
In [282]: activity_by_time(5,9,dct)[:10]
```

```
Out[282]: [['TIMES SQ-42 ST', 4257.166666666667],
           ['14 ST-UNION SQ', 3772.3333333333335],
           ['34 ST-HERALD SQ', 3469.1875],
           ['59 ST COLUMBUS', 1797.125],
           ['FLUSHING-MAIN', 1589.3125],
           ['W 4 ST-WASH SQ', 1574.25],
           ['59 ST', 1567.5625],
           ['8 AV', 1424.1875],
           ['BEDFORD AV', 1235.25],
           ["B'WAY-LAFAYETTE", 1229.9375]]
```

```
clean_df_cum(dct)
```

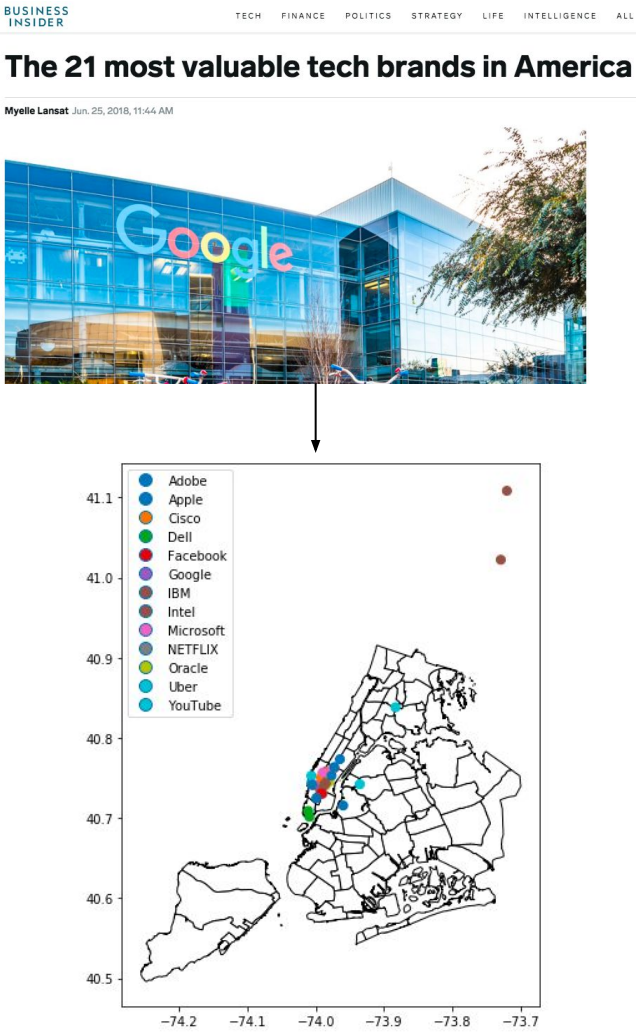
```
array([['TIMES SQ-42 ST', 3151.29935515873],
       ['14 ST-UNION SQ', 2986.3298611111113],
       ['34 ST-HERALD SQ', 2375.5850694444443],
       ['FLUSHING-MAIN', 2036.685267857143],
       ['ATL AV-BARCLAY', 1687.7100694444446],
       ['JKSN HT-ROOSVLT', 1686.421875],
       ['59 ST COLUMBUS', 1648.171875],
       ['BEDFORD AV', 1446.298859126984],
       ['59 ST', 1341.6450892857142],
       ['W 4 ST-WASH SQ', 1333.3482142857142],
```

# MTA - Heat Map


40.9
40.8
40.7
40.6
40.5
−74.2
−74.1
−74.0
−73.9
−73.8
−73.7
2500000
2000000
1500000
1000000
500000

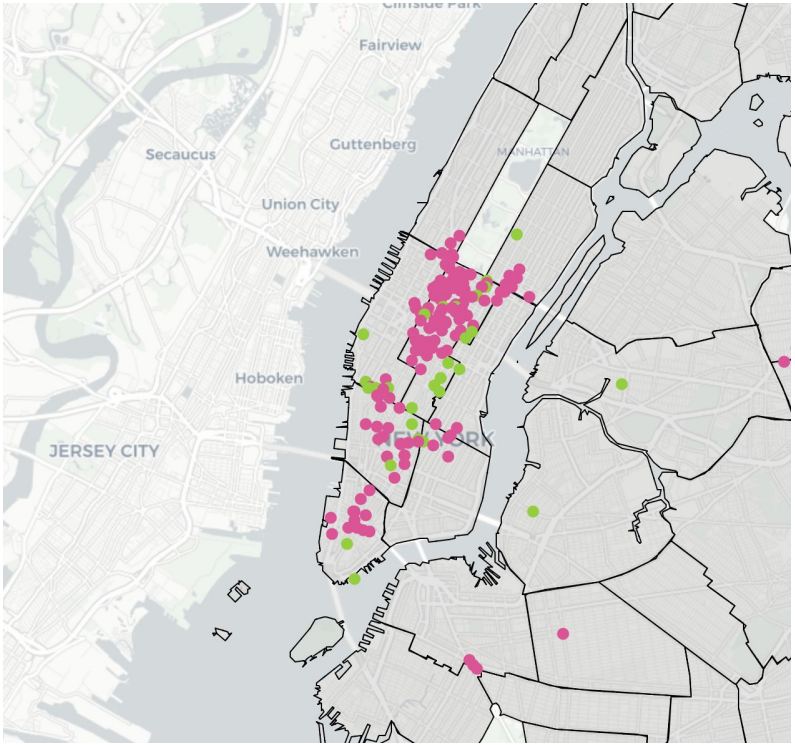
# Work - Tech Companies

- Top 21 most valuable American technology companies
- Use Google Places API to search for each company's location in New York City
- Calculate tech company proximity score for each station

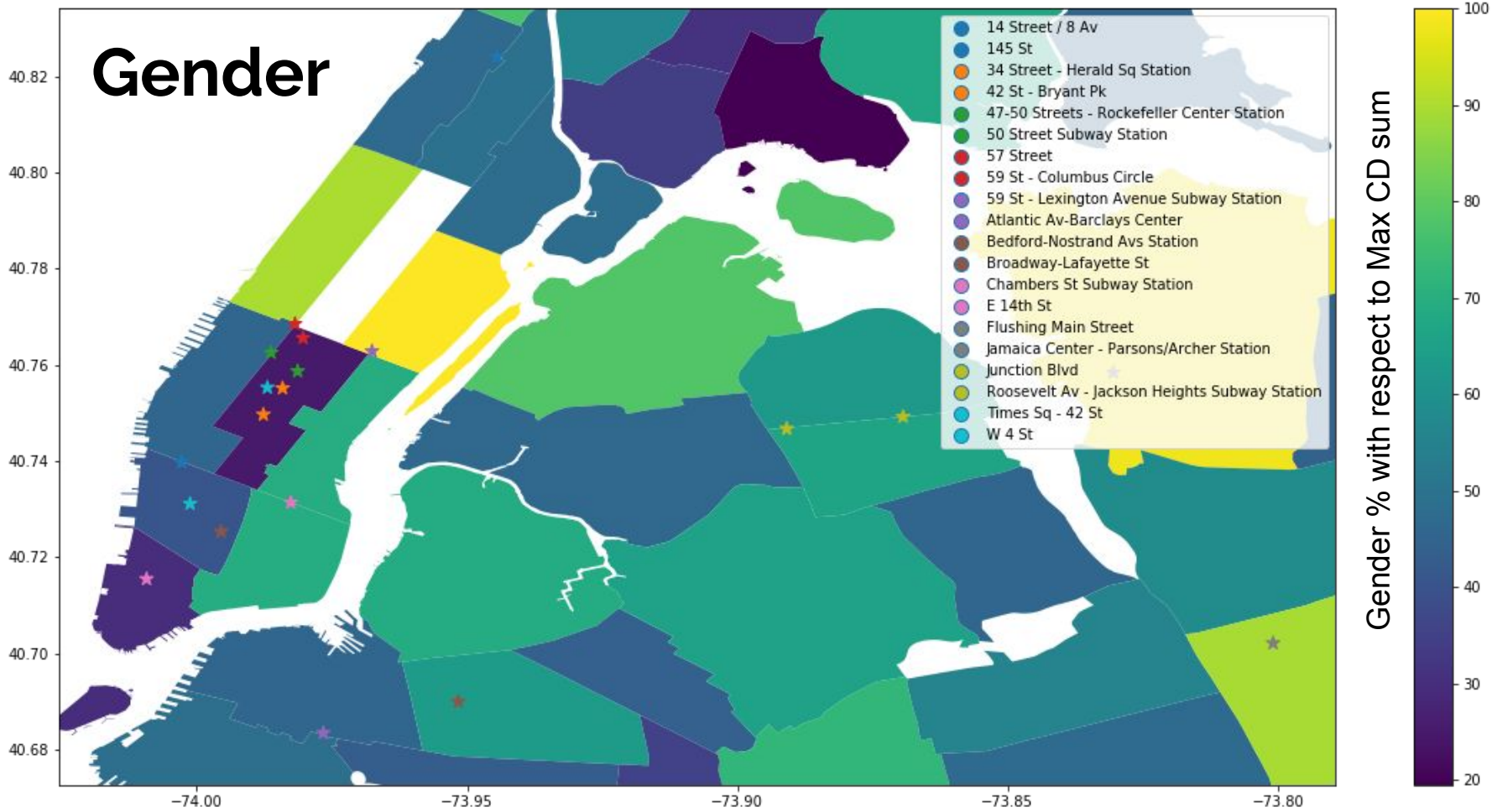


# Interests - Starbucks

- All Starbucks locations within 600 meters of the identified MTA stations were found
- Subway stations were scored based on the quantity and distance to Starbucks locations




- Green - tech companies
- Pink - Starbucks


Gender
14 Street / 8 Av
145 St
34 Street - Herald Sq Station
42 St - Bryant Pk
47-50 Streets - Rockefeller Center Station
50 Street Subway Station
57 Street
59 St - Columbus Circle
59 St - Lexington Avenue Subway Station
Atlantic Av-Barclays Center
Bedford-Nostrand Avs Station
Broadway-Lafayette St
Chambers St Subway Station
E 14th St
Flushing Main Street
Jamaica Center - Parsons/Archer Station
Junction Blvd
Roosevelt Av - Jackson Heights Subway Station
Times Sq - 42 St
W 4 St
40.82
40.80
40.78
40.76
40.74
40.72
40.70
40.68
−74.00
−73.95
−73.90
−73.85
−73.80
Gender % with respect to Max CD sum
100
90
80
70
60
50
40
30
20

# Final Scoring

Equation used for scoring:

$$FS = W_1 * Score_{MTA} + W_2 * Score_{Tech} + W_3 * Score_{Coffee} + W_4 * Score_{Gender}$$

- W1 = 0.5
- W2 , W3 = 0.15
- W4 = 0.2

# Overall Recommendations

Top 10 list of recommended stations:

1. Times Sq- 42 St
2. E 14th St
3. 34th St - Herald Sq Station
4. 59th St - Columbus Circle
5. Flushing Main Street
6. 59th St - Lexington Avenue Subway Station
7. 47 - 50 Rockefeller Center Station
8. 14th Street / 8 Av
9. W 4 St
10. Atlantic Av - Barclays Center

# Thank You!

# Assumptions

- Timeline of analysis is now, summer time
- In the summer season, we expected that a majority of volunteers would be available between 12pm and 12am. We examined data in June for the past three years as a subset

# MTA Analysis: Final Product

- User inputs month, day, hour and number of years to scrape.
- Program provides three main functions:

# Takeaway Points

- MTA analysis → ‘Busiest’ stations mostly located south of Central Park: Midtown, Flatiron, SoHo, Financial District
- Starbucks and tech company proximity scores, as well as demographic data can be used to further prioritize busiest stations

# Next Steps

- Explore non-linear regression models
- Short term vs. long term engagement
- Continue to collect feedback and learning
- Other factors to consider
  - Tech awareness
  - Education level
  - Income level

# Results Table

| | Subway_Station | MTA Score Percent (%) | Tech Company Percent (%) | Strbks Distance Score Percent (%) | Gender Percent (%) | Final_Score |
|---|---|---|---|---|---|---|
| 0 | Times Sq - 42 St | 100.000000 | 55.670240 | 78.772617 | 25.018381 | 75.170105 |
| 1 | E 14th St | 94.857815 | 38.314692 | 40.158419 | 69.226205 | 73.045115 |
| 2 | 34 Street - Herald Sq Station | 75.458110 | 54.950511 | 84.839765 | 25.018381 | 63.701273 |
| 3 | 59 St - Columbus Circle | 52.320407 | 30.560302 | 77.566706 | 89.901583 | 60.359572 |
| 4 | Flushing Main Street | 64.848899 | 5.554432 | 27.907102 | 98.332194 | 57.110118 |
| 5 | 59 St - Lexington Avenue Subway Station | 42.558717 | 37.314237 | 61.888235 | 100.000000 | 56.159729 |
| 6 | 50 Street Subway Station | 35.128152 | 40.312335 | 93.198653 | 45.682149 | 46.727154 |
| 7 | 47-50 Streets - Rockefeller Center Station | 18.413980 | 100.000000 | 94.633298 | 25.018381 | 43.405661 |
| 8 | 14 Street / 8 Av | 23.446992 | 79.833552 | 67.421549 | 41.068273 | 42.025416 |
| 9 | W 4 St | 42.352487 | 40.626471 | 38.419252 | 41.068273 | 41.246757 |
| 10 | Atlantic Av-Barclays Center | 54.102154 | 10.969329 | 21.809874 | 45.754382 | 41.118834 |
| 11 | 57 Street | 30.172809 | 36.908444 | 100.000000 | 25.018381 | 40.626347 |
| 12 | 42 St - Bryant Pk | 20.141195 | 56.020324 | 95.494891 | 25.018381 | 37.801556 |
| 13 | Roosevelt Av - Jackson Heights Subway Station | 53.567523 | 8.978976 | 3.751323 | 44.085286 | 37.510364 |
| 14 | Bedford-Nostrand Avs Station | 45.940253 | 10.830579 | 0.000000 | 63.784230 | 37.351559 |
| 15 | Jamaica Center - Parsons/Archer Station | 33.664250 | 4.453061 | 7.093317 | 89.389503 | 36.441983 |
| 16 | Chambers St Subway Station | 22.933191 | 23.127369 | 83.942019 | 29.619358 | 33.450875 |
| 17 | Broadway-Lafayette St | 23.711866 | 40.053577 | 24.795028 | 41.068273 | 29.796878 |
| 18 | 145 St | 33.592474 | 7.912415 | 9.170600 | 46.881732 | 28.735036 |
| 19 | Junction Blvd | 18.748353 | 7.346592 | 0.000000 | 63.784230 | 23.233011 |

# Introducing NYC's Top 100 tech companies

BY LIZ WARREN | NOVEMBER 7, 2017

# Built In NYC's 50 Startups to Watch in 2018

BY LIZ WARREN | JANUARY 16, 2018